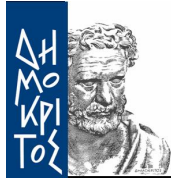




ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΕΡΕΥΝΑΣ ΦΥΣΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ «ΔΗΜΟΚΡΙΤΟΣ»
ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΑΝΑΛΥΣΗ ΚΑΙ ΣΥΓΚΡΙΣΗ ΑΛΓΟΡΙΘΜΩΝ ΕΥΚΑΙΡΙΑΚΟΥ
ΧΡΟΝΟΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΕ ΑΣΥΡΜΑΤΑ ΔΙΚΤΥΑ
ΕΠΙΚΟΙΝΩΝΙΩΝ: ΜΕΘΟΔΟΛΟΓΙΕΣ ΑΝΑΛΥΣΗΣ ΚΑΙ
ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΥΓΚΡΙΤΙΚΗΣ ΑΞΙΟΛΟΓΗΣΗΣ

ΙΩΑΝΝΗΣ Ν. ΓΙΑΝΝΟΥΛΑΚΗΣ

ΑΘΗΝΑ,
ΙΟΥΛΙΟΣ 2013



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

**ΑΝΑΛΥΣΗ ΚΑΙ ΣΥΓΚΡΙΣΗ ΑΛΓΟΡΙΘΜΩΝ ΕΥΚΑΙΡΙΑΚΟΥ
ΧΡΟΝΟΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΣΕ ΑΣΥΡΜΑΤΑ ΔΙΚΤΥΑ
ΕΠΙΚΟΙΝΩΝΙΩΝ: ΜΕΘΟΔΟΛΟΓΙΕΣ ΑΝΑΛΥΣΗΣ ΚΑΙ
ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΥΓΚΡΙΤΙΚΗΣ ΑΞΙΟΛΟΓΗΣΗΣ**

ΙΩΑΝΝΗΣ Ν. ΓΙΑΝΝΟΥΛΑΚΗΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Νικόλαος Μήτρου, Καθηγητής Ε.Μ.Π.

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Νικόλαος Μήτρου, Καθηγητής Ε.Μ.Π.

Μιχαήλ Θεολόγου, Καθηγητής Ε.Μ.Π.

Ευστάθιος Συκάς, Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 08/07/2013.

.....
Νικόλαος Μήτρου,
Καθηγητής Ε.Μ.Π.

.....
Μιχαήλ Θεολόγου,
Καθηγητής Ε.Μ.Π.

.....
Ευστάθιος Συκάς,
Καθηγητής Ε.Μ.Π.

.....
Μιλτιάδης Αναγνώστου,
Καθηγητής Ε.Μ.Π.

.....
Φίλιππος Κωνσταντίνου,
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Στασινόπουλος,
Καθηγητής Ε.Μ.Π.

.....
Κίμων Κοντοβασίλης,
Διευθυντής Ερευνών
Ε.Κ.Ε.Φ.Ε. «ΔΗΜΟΚΡΙΤΟΣ»

Αθήνα, Ιούλιος 2013

.....
ΙΩΑΝΝΗΣ Ν. ΓΙΑΝΝΟΥΛΑΚΗΣ

Φυσικός, Παν. Ιωαννίνων

Copyright © 2013 ΙΩΑΝΝΗΣ Ν. ΓΙΑΝΝΟΥΛΑΚΗΣ

Με επιφύλαξη παντός δικαιώματος.

All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου ή του Ε.Κ.Ε.Φ.Ε. «Δημόκριτος».

ΠΕΡΙΛΗΨΗ

Τα δίκτυα ασύρματων επικοινωνιών γνωρίζουν ιδιαίτερη ανάπτυξη και χαρακτηρίζονται από τη συνεχώς αυξανόμενη ανάγκη για βελτιωμένες επιδόσεις και υψηλούς ρυθμούς μετάδοσης. Η διατριβή ασχολείται με τη μελέτη της επίδοσης σε συστήματα τηλεπικοινωνιών όπως το HSDPA και το HDR που βασίζονται σε αλγορίθμους ευκαιριακού χρονοπρογραμματισμού για την επιλογή του τερματικού που θα εξυπηρετηθεί σε κάθε χρονοσχιμή, λαμβάνοντας υπόψη την κατάσταση της ασύρματης ζεύξης. Βασικό γνώρισμα αυτού του τύπου χρονοπρογραμματισμού αποτελεί το γεγονός ότι ο μέσος ρυθμός εξυπηρέτησης δεν παραμένει σταθερός αλλά μεταβάλλεται ανάλογα με το πλήθος των ενεργών τερματικών.

Ένα δίκτυο με τα παραπάνω χαρακτηριστικά αντιστοιχείται στο μοντέλο μερισμού του εξυπηρετητή με εξάρτηση από την κατάσταση, το οποίο στην πιο γενική μορφή που θεωρούμε εδώ, διαθέτει ξεχωριστές κλάσεις τερματικών ανάλογα με το μέσο επιτεύξιμο ρυθμό εξυπηρέτησης και γενική μορφή χώρου καταστάσεων ως αποτέλεσμα του ελέγχου αποδοχής κλήσεων. Στα πλαίσια της διδακτορικής διατριβής, για το συγκεκριμένο σύστημα υπολογίζεται η κατανομή του χρόνου απόκρισης κάποιου τερματικού δεδομένης κλάσης, η οποία εκφράζεται μέσω του εκθετικού ενός κατάλληλου πίνακα τύπου M. Η μορφή αυτή επιτρέπει τον χαρακτηρισμό του ασυμπτωτικού ρυθμού φθίσης ουράς της κατανομής με τη βοήθεια της ελάχιστης ιδιοτιμής του συγκεκριμένου πίνακα και ταυτόχρονα παρέχει εύκολα υπολογίσιμα άνω και κάτω φράγματα για την τιμή της.

Επίσης, η διατριβή ασχολείται με τα φαινόμενα που προκύπτουν σε περιβάλλον διαχωρισμού χρονικής κλίμακας, διαχωρίζοντας τα τερματικά σε αργές και γρήγορες κλάσεις σύμφωνα με το μέσο ρυθμό μετάδοσης. Με εκμετάλλευση της σχεδόν πλήρως αποσυζεύξιμης δομής του πλήρους συστήματος, λαμβάνουμε ικανοποιητικές προσεγγίσεις για την επίδοση των αργών και των γρήγορων κλάσεων, οι οποίες προσφέρουν μεγαλύτερη ευκολία στην κατανόηση και υπολογιστική απλότητα. Παράλληλα, θεωρώντας ειδικότερες περιπτώσεις για τους ρυθμούς μετάδοσης των κλάσεων ή/και της μορφής του ελέγχου αποδοχής κλήσεων, καταλήγουμε σε χρήσιμα συμπεράσματα σχετικά με την επίδραση των αργών τερματικών στην κατανομή του χρόνου απόκρισης των γρήγορων τερματικών.

Τα παραπάνω αποτελέσματα του γενικότερου προβλήματος εφαρμόζονται για την περίπτωση των υπό μελέτη συστημάτων τηλεπικοινωνιών και ακολουθεί η σύγκριση δύο βασικών αλγορίθμων ευκαιριακού χρονοπρογραμματισμού, όσον αφορά θέματα επίδοσης και δικαιοσύνης. Παρατίθενται χρήσιμα

συμπεράσματα σχετικά με την αξία και την καταλληλότητα του καθενός, την επίδραση της μορφής του ελέγχου αποδοχής κλήσεων και τον προσδιορισμό της χειρότερης δυνατής περίπτωσης για τα τερματικά των γρήγορων κλάσεων. Επιπρόσθετα, προτείνεται ένας νέος αλγόριθμος ευκαιριακού χρονοπρογραμματισμού που συνδυάζει τα ευνοϊκά χαρακτηριστικά από τα ήδη μελετηθέντα σχήματα. Η διατριβή ολοκληρώνεται με την αριθμητική σύγκριση των θεωρητικών ευρημάτων.

Λέξεις κλειδιά: ευκαιριακός χρονοπρογραμματισμός, μερισμός εξυπηρετητή με εξάρτηση από την κατάσταση, κατανομή χρόνου απόκρισης, διαχωρισμός χρονικής κλίμακας, κέρδη διαφορισμού πολλαπλών χρηστών, ασύρματο δίκτυο με παρακολούθηση καναλιού, αποσύνθεση, συνάθροιση

ABSTRACT

Wireless communication networks develop rapidly, featuring excessive demand for performance improvements. This thesis studies performance and fairness issues of opportunistic schedulers, widely used in modern wireless standards like HSDPA and HDR. Opportunistic schedulers take into account the condition of the wireless link in order to select a terminal for service during a timeslot, by allocating to it all the available network resources. The main attribute of this type of scheduling is that the average transmission rate does depend on the number of active terminals, in the form of a multiuser diversity gain which acts as a multiplicative factor on the average rate.

The above characteristics yield to a State Dependent Processor Sharing (SDPS) model, which in its general form considered here, features separate classes of terminals and a generic form of the state space, as a result of the admission control scheme. From a mathematical point of view, we manage to provide the distribution of the sojourn time for an SDPS system, which is expressed through the exponential of an appropriate M-matrix. The form allows the characterization of the asymptotic decay rate of the sojourn time's distribution tail and provides easily computable upper and lower bounds.

Furthermore, the conducted research studies time-scale separation phenomena, by separating the classes of terminals into a slow and a fast group, according to their average transmission rate. By leveraging on the nearly completely decomposable structure of the full system, we arrive in accurate approximations that offer conceptual and computational simplicity. Additionally, we consider some special important cases of the service rate function and/or the admission control scheme, which allow to determine the effect of the presence of slow class terminals on the performance of the terminals belonging in the fast group of classes.

The results and conclusions of the generalized problem have been applied in the practical case of a channel-aware wireless network. A comparison of two basic opportunistic schedulers with respect to performance and fairness issues follows. We result in several conclusions regarding the value and the suitability of each scheduler, the appropriate form of the service pool and the specification of the worst case scenario related to the fast class terminals. Moreover, a new scheduling scheme is proposed, combining the nice properties of the previous two schedulers. The theoretical findings are compared and verified numerically.

Keywords: opportunistic scheduling, state-dependent processor sharing, sojourn time distribution, time scale separation, multiuser diversity gains, channel-aware wireless network, decomposition, aggregation

ΕΥΧΑΡΙΣΤΙΕΣ

Η διδακτορική διατριβή εκπονήθηκε στο εργαστήριο Δικτύων Τηλεπικοινωνιών του ινστιτούτου Πληροφορικής και Τηλεπικοινωνιών του Ε.Κ.Ε.Φ.Ε «Δημόκριτος», υποστηριζόμενη μέσω υποτροφίας τετραετούς συνεργασίας με τη μορφή μηνιαίου χρηματικού βοηθήματος και υπό την επίβλεψη του Διευθυντή Ερευνών Δρα. Κίμωνα Κοντοβασίλη.

Έχοντας δείξει απεριόριστη εμπιστοσύνη προς το πρόσωπό μου σε όλη τη μακρόχρονη περίοδο της διδακτορικής μου έρευνας, ο Δρ. Κοντοβασίλης υπήρξε άριστος καθοδηγητής, υπομονετικός σύμβουλος και ηθικός συμπαράστατης. Η συνεισφορά του υπήρξε καθοριστική, από την αρχική μου μελέτη και τον καθορισμό του προβλήματος ως και την υπεράσπιση της διατριβής. Για τους παραπάνω λόγους αλλά και για την κάθε είδους στήριξη που μου έχει δείξει όλο αυτόν τον καιρό θα ήθελα να τον ευχαριστήσω θερμά.

Θα ήθελα επίσης να εκφράσω θερμές ευχαριστίες στον Καθηγητή της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Ε.Μ.Π. Νικόλαο Μήτρου, για το ειλικρινές ενδιαφέρον του και την απεριόριστη διάθεση συνεργασίας που επέδειξε όλα αυτά τα χρόνια ως επιβλέπων μέλος της τριμελούς επιτροπής παρακολούθησης.

Από το περιβάλλον του εργαστηρίου στο Δημόκριτο θα ήθελα να ευχαριστήσω όλους τους συναδέλφους για το ενδιαφέρον τους αλλά και τις όμορφες στιγμές που περάσαμε. Θα αναφέρω ονομαστικά και θα ευχαριστήσω τον φίλο Δρα. Εμμανουήλ Καφετζάκη, με τον οποίο διατρέξαμε μία παράλληλη χρονικά ερευνητική διαδρομή, αντιμετωπίσαμε κοινές δυσκολίες και ανησυχίες και ανταλλάξαμε εποικοδομητικά απόψεις, αναζητώντας λύσεις σε ένα ευρύ φάσμα προβλημάτων.

Τέλος, νιώθω ιδιαίτερη υποχρέωση να ευχαριστήσω ένα μεγάλο αριθμό δικών μου ανθρώπων, τη σύζυγό μου Διονυσία, την οικογένεια και τους φίλους μου, για τη συμπαράσταση που επέδειξαν και την κατανόηση με την οποία με περιέβαλλαν κατά τα χρόνια διεξαγωγής της διδακτορικής έρευνας.

Αθήνα,
Ιούλιος 2013
Ιωάννης Γιαννουλάκης

Περιεχόμενα

Σελίδα Έγκρισης	iii
Περίληψη	vii
Ευχαριστίες	xi
Περιεχόμενα	xv
Κατάλογος Σχημάτων	xvii
Κατάλογος Πινάκων	xix
1 Εισαγωγή	1
1.1 Ευκαιριακός Χρονοπρογραμματισμός - Περιγραφή του Προβλήματος	2
1.2 Συνεισφορά της Διατριβής	4
1.3 Δομή της Διατριβής	6
2 Βασικά Χαρακτηριστικά Ευκαιριακού Χρονοπρογραμματισμού	9
2.1 Επισκόπηση αλγορίθμων ευκαιριακού χρονοπρογραμματισμού	9
2.2 Μοντελοποίηση ασύρματος περιβάλλοντος και σχετικές ιδιότητες	12
2.3 Υπολογισμός κέρδους διαφορισμού πολλαπλών χρηστών και σχε- τικές ιδιότητες	15
2.4 Μεγάλες διαφορές ισχύος σήματος και εφαρμογή στον BR	17
3 Επισκόπηση Μοντέλου PS και Ορισμοί Απαιτούμενων Μεγεθών	23
3.1 Βιβλιογραφική επισκόπηση μοντέλων PS	24
3.2 Μεθοδολογία επίλυσης	28
3.3 Ορισμοί και μαρκοβιανή δομή του μοντέλου	31
4 Κατανομή του Χρόνου Απόκρισης και Συναφή Αποτελέσματα	35

4.1	Φασματικές ιδιότητες του B_i	39
4.2	Δίκτυα με ιδιότητα ισοροπίας	43
5	Διαχωρισμός Χρονικής Κλίμακας	47
5.1	Η Σχεδόν Πλήρως Αποσυσζεύξιμη δομή του πίνακα ρυθμών μεταβάσεων	49
5.2	Κατανομή του χρόνου απόκρισης για πελάτες γρήγορων κλάσεων	53
5.3	Κατανομή του χρόνου απόκρισης για πελάτες αργών κλάσεων .	56
6	Επιπλέον Αποτελέσματα για Περιβάλλον Διαχωρισμού Χρονικής Κλίμακας	59
6.1	Διαχωρισμός χρονικής κλίμακας με ρυθμούς εξυπηρέτησης προνομιακούς για γρήγορες κλάσεις	59
6.2	Περαιτέρω χαρακτηρισμός της συμπεριφοράς ουράς για κλάσεις με γρήγορη δυναμική	64
7	Επέκταση των Αποτελεσμάτων για Γενικότερες Κατανομές Απαιτήσεων Εξυπηρέτησης	71
8	Μέτρα Επίδοσης και Διαχωρισμός Χρονικής Κλίμακας στον Ευκαιριακό Χρονοπρογραμματισμό	77
8.1	Συνολικά αποτελέσματα επίδοσης	78
8.2	Αποτελέσματα επίδοσης για κλάσεις με μεγάλες διαφορές ρυθμών εξυπηρέτησης	80
9	Σύγκριση των Αλγορίθμων BR και NBR	85
9.1	Αποτελέσματα επίδοσης του NBR	85
9.2	Αποτελέσματα επίδοσης του BR	89
10	Εισαγωγή του Αλγορίθμου SG-PF	95
11	Αριθμητικά Αποτελέσματα	101
11.1	Αριθμητικά αποτελέσματα για την επαλήθευση του μοντέλου SDPS101	
11.2	Αριθμητικά αποτελέσματα για τη σύγκριση του SG-PF με τους BR και NBR	113
12	Συμπεράσματα και Κατευθύνσεις Μελλοντικής Έρευνας	127
12.1	Συμπεράσματα	127
12.2	Κατευθύνσεις Μελλοντικής Έρευνας	129
	Παράρτημα	133

I	Απόδειξη ότι το γινόμενο $n_i \phi_i(n)$ είναι γνησίως αύξουσα συνάρτηση ως προς $n_i, i \in \mathcal{K}$	133
II	Απόδειξη του Θεωρήματος 1	133
III	Απόδειξη της Πρότασης 1	135
IV	Απόδειξη του Λήμματος 2	136
V	Απόδειξη του Λήμματος 3	137
	Βιβλιογραφία	141
	Γλωσσάρι	147
	Συμβολισμοί Μεγεθών	151
	Ευρετήριο	155

Κατάλογος σχημάτων

11.1 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, ρυθμοί εξυπηρέτησης DPS - Κλάση 1	103
11.2 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, ρυθμοί εξυπηρέτησης DPS - Κλάση 2	104
11.3 Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 1	105
11.4 Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 2	106
11.5 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης - Κλάση 1	108
11.6 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης - Κλάση 3	109
11.7 Κοινός θάλαμος εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπη- ρέτησης υπέρ της γρήγορης κλάσης - Κλάση 1	111
11.8 Κοινός θάλαμος εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπη- ρέτησης υπέρ της γρήγορης κλάσης - Κλάση 3	111
11.9 Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 1 - $L = 10$	112
11.10 Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 2 - $L = 10$	112
11.11 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7, L_f = 4$ - Κλάση 1	116
11.12 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7, L_f = 4$ - Κλάση 2	117
11.13 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7, L_f = 4$ - Κλάση 3	118
11.14 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7, L_f = 4$ - Κλάση 4	119

Κατάλογος πινάκων

11.1 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, ρυθμοί εξυπηρέτησης DPS102	
11.2 Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS	105
11.3 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης	107
11.4 Κοινός θάλαμος εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπη- ρέτησης υπέρ της γρήγορης κλάσης	110
11.5 Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS, $L = 10$	110
11.6 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7, L_f = 4$	114
11.7 Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 3, L_f = 4$	120
11.8 Κοινός θάλαμος εξυπηρέτησης, $L = 7$	121
11.9 Κοινός θάλαμος εξυπηρέτησης, $L = 7$ - Σκέδαση Nakagami με $m = 2$	122
11.10 Κέρδη διαφορισμού πολλαπλών χρηστών για σκέδαση Rayleigh ($m = 1$) και σκέδαση Nakagami με $m = 2$	122
11.11 Διαχωρισμένοι θάλαμοι εξυπηρέτησης - BR	123
11.12 Διαχωρισμένοι θάλαμοι εξυπηρέτησης - NBR	123
11.13 Διαχωρισμένοι θάλαμοι εξυπηρέτησης - SG-PF	124
11.14 Κοινός θάλαμος εξυπηρέτησης - BR	124
11.15 Κοινός θάλαμος εξυπηρέτησης - NBR	125
11.16 Κοινός θάλαμος εξυπηρέτησης - SG-PF	125

Κεφάλαιο 1

Εισαγωγή

ΤΑ ασύρματα δίκτυα τρίτης γενιάς δημιουργήθηκαν για να προσφέρουν ασύρματες ευρυζωνικές υπηρεσίες στους χρήστες. Η εμφάνισή τους έγινε με την εισαγωγή των προτύπων Universal Mobile Telecommunications System (UMTS) στην Ευρώπη και CDMA2000 στις Η.Π.Α. Η ανάγκη για υψηλότερους ρυθμούς επικοινωνίας και βέλτιστη χρήση του διατιθέμενου φάσματος οδήγησε στη δημιουργία ενός νέου προτύπου που αποτελεί προσθήκη στο UMTS και ονομάζεται High Speed Packet Access (HSPA) και αποτελεί την συνένωση δύο πρωτοκόλλων, του High Speed Packet Downlink Access (HSDPA) και του High Speed Packet Uplink Access (HSUPA). Ο στόχος κατά τη δημιουργία του ήταν να υποστηρίξει ρυθμούς δεδομένων της τάξης των 10 Mb/s. Παρόμοιο πρότυπο αποτελεί το High Data Rate (HDR) (μετονομασμένο σε CDMA2000 1xEVolution-Data Optimized (1xEV-DO)) που υλοποιήθηκε στις Η.Π.Α. κατά το αντίστοιχο χρονικό διάστημα.

Κατά τη λειτουργία τεχνολογιών αυτού του τύπου παρέχεται η δυνατότητα μεταξύ άλλων, να εξυπηρετούνται τα τερματικά κατά τη διάρκεια ενός μικρού χρονικού διαστήματος (χρονοσχισμή (αγ.: timeslot) τάξης 1-2 ms), δεσμεύοντας το σύνολο των διαθέσιμων πόρων του εξυπηρετητή που στη συγκεκριμένη περίπτωση είναι ο σταθμός βάσης. Παράλληλα, η απόφαση για την επιλογή του τερματικού που θα εξυπηρετηθεί σε κάθε χρονοσχισμή (όταν αυτά είναι περισσότερα του ενός), λαμβάνεται από το σταθμό βάσης μέσω ενός αλγορίθμου που αξιοποιεί την πληροφορία για την κατάσταση της ασύρματης ζεύξης κάθε τερματικού. Η απόφαση αυτή προφανώς επηρεάζει την επίδοση κάθε τερματικού, αλλά και του δικτύου συνολικά. Γι' αυτό και η επιλογή του αλγορίθμου που θα υλοποιηθεί είναι βαρύνουσας σημασίας.

Η παρούσα διατριβή θέτει το πρόβλημα του ευκαιριακού χρονοπρογραμματισμού (αγ.: opportunistic scheduling) σε ασύρματα δίκτυα τηλεπικοινωνώ-

νίων και αποσκοπεί στη σύγκριση και ποσοτικοποίηση τέτοιων αλγορίθμων. Η μελέτη επιδιώκει τον υπολογισμό μεγεθών επίδοσης, συμπεριλαμβανομένης της κατανομής του χρόνου απόκρισης, η οποία αποτελεί ένα σύνθετο μαθηματικό πρόβλημα, αλλά ταυτόχρονα και ένα σοβαρό μέτρο επίδοσης κατά την εξυπηρέτηση σε οποιοδήποτε σύστημα αναμονής. Ακολουθεί μία σύντομη περιγραφή και τοποθέτηση του προβλήματος. Στη συνέχεια παρουσιάζονται η συνεισφορά και τα σημαντικότερα αποτελέσματα που προέκυψαν κατά τη διάρκεια εκπόνησής της διατριβής και η εισαγωγή κλείνει παραθέτοντας τη δομή της διατριβής.

1.1 Ευκαιριακός Χρονοπρογραμματισμός - Περιγραφή του Προβλήματος

Ένα τυπικό ασύρματο δίκτυο τρίτης γενιάς το οποίο ενσωματώνει τεχνολογίες που επιτρέπουν τη γρήγορη μετάδοση δεδομένων, όπως π.χ., HSDPA, HDR κ.λπ., καταλήγει σε ένα σταθμό βάσης ανά κυψέλη, που αναλαμβάνει την επικοινωνία των κινητών τερματικών με το υπόλοιπο δίκτυο. Όταν περισσότερα του ενός τερματικά ζητούν να επικοινωνήσουν με το σταθμό βάσης, εκείνος επιλέγει ένα και μόνο τερματικό, στο οποίο διαθέτει το σύνολο των διαθέσιμων πόρων του δικτύου (π.χ., ορθογώνιοι κώδικες για την περίπτωση που χρησιμοποιείται ευρυζωνική πολλαπλή πρόσβαση με διαίρεση κώδικα). Η ανάθεση διαρκεί για το διάστημα μίας μόνο χρονοσχισμής, η οποία έχει μικρή διάρκεια (π.χ., 2 ms στο HSDPA). Η επιλογή του τερματικού που τελικά θα λάβει τους πόρους του δικτύου προς εξυπηρέτηση, μπορεί να στηρίζεται μέσω κατάλληλων αλγορίθμων και στην πληροφορία που αποστέλλει κάθε τερματικό στο σταθμό βάσης σχετικά με την κατάσταση της ασύρματης ζεύξης του. Η πληροφορία αυτή αντικατοπτρίζεται στο σηματοθορυβικό λόγο (αγγ.: Signal to Noise Ratio (SNR)) που μετρά το τερματικό σε κάθε χρονοσχισμή και γνωστοποιείται στο σταθμό βάσης μέσω ενός ανεξάρτητου ασύρματου καναλιού, αποκλειστικά δεσμευμένου για το σκοπό αυτό.

Ο σηματοθορυβικός λόγος της ασύρματης ζεύξης σχετίζεται άμεσα με τον υποστηριζόμενο ρυθμό μετάδοσης προς ή από κάθε τερματικό και κατά συνέπεια, η συγκεκριμένη πληροφορία αποτελεί ένα από τα σημαντικότερα χαρακτηριστικά μιας ζεύξης. Η κατάλληλη αξιοποίηση της συγκεκριμένης πληροφορίας μπορεί να παρουσιάσει σημαντικά οφέλη για τις επιδόσεις κάθε τερματικού. Σε αντίθεση, αν η πληροφορία δεν αξιοποιηθεί σωστά μπορεί να οδηγήσει σε κακή διαχείριση των πόρων και σε χαμηλές επιδόσεις. Για το λόγο αυτό, έχουν προταθεί αρκετοί αλγόριθμοι ευκαιριακού χρονοπρογραμματι-

σμού. Δύο από αυτούς, καθώς και κάποιες τροποποιήσεις τους, αποτελούν τα βασικότερα τέτοια σχήματα, επειδή είναι αρκετά αποτελεσματικοί αλλά και αρκετά απλοί ως προς την τεχνική υλοποίησή τους.

Ο πρώτος αλγόριθμος βασίζεται αποκλειστικά στην πληροφορία για το σηματοθορυβικό λόγο της ασύρματης ζεύξης του τερματικού και ονομάζεται αλγόριθμος μέγιστου SNR (αγ.: maxSNR). Ο συγκεκριμένος αλγόριθμος επιλέγει για εξυπηρέτηση πάντοτε το τερματικό με το μέγιστο SNR και προφανώς αποτελεί ένα απλό σχήμα που παρέχει τη δυνατότητα να αυξηθεί η συνολική ρυθμαπόδοση (αγ.: throughput) του δικτύου. Ωστόσο τέτοιου είδους προβλήματα δεν επιδέχονται πάντα μία απλή λύση. Όταν υπάρχουν μεγάλες διαφορές μεταξύ τερματικών όσον αφορά τη μέση τιμή του SNR, ο maxSNR μοιραία αδικεί τα τερματικά με πιο ασθενές σηματοθορυβικό λόγο. Για να καταπολεμηθεί το φαινόμενο αυτό, έχει προταθεί μία κανονικοποιημένη έκδοση του maxSNR, που ονομάζεται αλγόριθμος «Αναλογικά Δίκαιος» (αγ.: Proportional Fair (PF)). Ο PF αντιμετωπίζει το πρόβλημα της δικαιοσύνης επιλέγοντας το τερματικό που διαθέτει τη μέγιστη τιμή στο κλάσμα σηματοθορυβικού λόγου προς τη ρυθμαπόδοση που έχει επιτύχει το τερματικό σε κάποιο δεδομένο και σταθερό χρονικό διάστημα. Παρόλο που αποτελεί καλή λύση για θέματα δικαιοσύνης, ο PF υστερεί σε κάποια μεγέθη επίδοσης, όπως για παράδειγμα στο χρόνο απόκρισης των τερματικών με υψηλότερο μέσο SNR.

Για την ορθή επιλογή του κατάλληλου αλγορίθμου, ένα σημαντικό πρόβλημα καθώς και αντικείμενο έρευνας για τη θεμελίωση χρήσιμων συμπερασμάτων, αποτελεί η κατανομή του χρόνου απόκρισης (αγ.: sojourn time) μίας νεοαφηχθείσας αίτησης προς εξυπηρέτηση (π.χ., μεταφορά αρχείου κ.λπ.). Έχοντας υπόψη προηγούμενες εργασίες, οι οποίες έχουν συνδέσει επιτυχώς τη λειτουργία του υπό μελέτη δικτύου με το μαθηματικό μοντέλο εξυπηρέτησης Μερισμού του Εξυπηρετητή (αγ.: Processor Sharing (PS)), στα πλαίσια της παρούσας διατριβής κρίθηκε σκόπιμο να μελετηθεί το μαθηματικό πρόβλημα της κατανομής του χρόνου απόκρισης σε συστήματα PS. Η μελέτη του συγκεκριμένου προβλήματος χρειάστηκε να γενικευθεί για την περίπτωση μεγάλων διαφορών του μέσου SNR των τερματικών ώστε να περιγραφεί με ακρίβεια η επίδραση στη δικαιοσύνη των παραπάνω αλγορίθμων. Τα αποτελέσματα της μελέτης επέτρεψαν τη λεπτομερή περιγραφή και σύγκριση των αλγορίθμων, τόσο σε πλαίσια μέσων τιμών, όσο και ως προς την ασυμπτωτική συμπεριφορά του χρόνου απόκρισης.

1.2 Συνεισφορά της Διατριβής

Προκειμένου να συγκριθούν λεπτομερώς οι βασικοί αλγόριθμοι ευκαιριακού χρονοπρογραμματισμού, κρίθηκε απαραίτητη η αναγωγή του προβλήματος στη μελέτη ενός θεωρητικού γενικότερου συστήματος PS με εξάρτηση του ρυθμού εξυπηρέτησης από την κατάσταση του συστήματος (αγ.: State Dependent Processor Sharing (SDPS)). Τα αποτελέσματα της διαδικασίας αυτής επέτρεψαν να προσδιοριστούν τα θετικά γνωρίσματα του κάθε αλγορίθμου, να αναγνωριστούν οι αδυναμίες του καθενός και να προβούμε σε μία λεπτομερή σύγκριση όσον αφορά θέματα επίδοσης και δικαιοσύνης κατά την εφαρμογή τους.

Διαθέτοντας την περιγραφή του προβλήματος ως σύστημα SDPS με πολλαπλές κλάσεις πελατών, υπολογίστηκε στη γενική της μορφή η κατανομή του χρόνου απόκρισης κάποιου «πελάτη» του συστήματος με τη μορφή του εκθετικού ενός κατάλληλου πίνακα τύπου M. Κατά τον υπολογισμό χρησιμοποιήθηκε η έννοια ενός «μόνιμου πελάτη» στο σύστημα, δίνοντας έτσι βάση στη μαρκοβιανή δομή του προβλήματος και επιτρέποντας μία ουσιώδη περιγραφή. Επιπλέον, η μορφή της συνάρτησης κατανομής μέσω του εκθετικού ενός πίνακα επιτρέπει την άμεση εξαγωγή συμπερασμάτων για την ασυμπτωτική της συμπεριφορά καθώς και τον υπολογισμό άνω και κάτω φραγμάτων όσον αφορά το ρυθμό φθίσης ουράς. Παράλληλα, παρέχονται αποτελέσματα σε κλειστή μορφή για συστήματα που διαθέτουν την επιπλέον ιδιότητα «ισορροπίας» των συναρτήσεων ρυθμού εξυπηρέτησης, αποτελέσματα που είναι ιδιαίτερα χρήσιμα για την (πιο σύνθετη) περιγραφή του PF.

Μία άλλη σημαντική συνιστώσα του προβλήματος αποτελεί η περίπτωση διαχωρισμού χρονικής κλίμακας (αγ.: time-scale separation) η οποία προκύπτει όταν συνυπάρχουν στο δίκτυο τερματικά με πολύ μεγάλες διαφορές ως προς το μέσο SNR (και κατ' επέκταση στο μέσο ρυθμό μετάδοσης). Η παρούσα διατριβή ασχολείται λεπτομερώς με το συγκεκριμένο πρόβλημα, διαχωρίζοντας τις κλάσεις των τερματικών σε ομάδες με πολύ ασθενές και πολύ ισχυρό SNR (πολύ αργούς και πολύ γρήγορους μέσους ρυθμούς εξυπηρέτησης αντίστοιχα). Στη συνέχεια με εκμετάλλευση της σχεδόν πλήρως αποσυνζεύξιμης (αγ.: Nearly Completely Decomposable (NCD)) δομής του συστήματος, αποδεικνύεται ότι το πλήρες σύστημα μπορεί να μελετηθεί μέσω μιας διαδικασίας αποσύνθεσης και συνάθροισης. Η αποσύνθεση (αγ.: decomposition) του μοντέλου γίνεται προς μειωμένα μοντέλα, καθένα από τα οποία διαθέτει ένα μόνιμο αριθμό αργών τερματικών. Εντός καθενός τέτοιου υποσυστήματος, ο αριθμός των γρήγορων τερματικών εξελίσσεται κανονικά και ανεξάρτητα από τα υπόλοιπα επί μέρους υποσυστήματα. Αντίστοιχα, η διαδικασία συνάθροισης (αγ.:

aggregation) δημιουργεί ένα άλλο μειωμένο σύστημα που αναπαριστά τη δυναμική μόνο των αργών τερματικών, καταλήγοντας έτσι σε μία ικανοποιητική προσέγγιση. Η διαδικασία αυτή παρέχει τον τρόπο να μελετηθεί το πρόβλημα με όρους μειωμένων μοντέλων, παρέχοντας υπολογιστική και νοητική απλότητα. Όσον αφορά την επίδοση των γρήγορων τερματικών, η παραπάνω διαδικασία επιτρέπει τον προσδιορισμό του χειρότερου δυνατού σεναρίου από πλευράς ύπαρξης συνδυασμού αργών τερματικών, ώστε να μπορεί αυτό να αποφευχθεί με κατάλληλες επιλογές του ελέγχου αποδοχής κλήσεων.

Τα αποτελέσματα της μελέτης μέσω μοντέλων SDPS, χρησιμοποιούνται για να συγκριθούν οι αλγόριθμοι maxSNR και PF ως προς το χρόνο απόκρισης και το βαθμό δικαιοσύνης που παρέχουν στα ενεργά τερματικά. Η σύγκριση επεκτείνεται για διάφορες μορφές του θαλάμου εξυπηρέτησης (αγ.: service pool) ως προς την επίδραση που έχει στις επιδόσεις και στη δικαιοσύνη μεταξύ των τερματικών. Έτσι μελετάται ένας πλήρως διαμοιραζόμενος χώρος μεταξύ ασθενών (αργών) και ισχυρών (γρήγορων) κλάσεων και συγκρίνεται με τον αντίστοιχο χώρο καταστάσεων που διαθέτει ξεχωριστά όρια στον αριθμό των αργών και γρήγορων τερματικών που μπορεί να δεχτεί. Αυτό καθιστά δυνατό τον προσδιορισμό του βέλτιστου σχήματος αποδοχής νέων ενεργών τερματικών.

Παράλληλα, κάνοντας χρήση των συμπερασμάτων που προκύπτουν από τη σύγκριση, καθώς και των χαρακτηριστικών γνωρισμάτων κάθε αλγορίθμου, προτείνουμε ένα νέο αλγόριθμο που συνδυάζει τις ευνοϊκές ιδιότητες των προηγούμενων δύο σχημάτων. Ο συγκεκριμένος αλγόριθμος με την ονομασία «Αναλογικά Δίκαιος - Κατηγορίας Ισχύος (αγ.: Strength Group - Proportional Fair (SG-PF)) ευνοεί καταρχήν τα τερματικά των κλάσεων με υψηλό SNR, εισάγει όμως και την έννοια της δικαιοσύνης μεταξύ τους. Ταυτόχρονα, δικαιοσύνη προσδίδεται και στα τερματικά κλάσεων με ασθενές SNR, στο μέτρο που δεν επηρεάζεται η επίδοση των γρήγορων. Αφού δοθεί η μαθηματική περιγραφή του SG-PF, αριθμητικά αποτελέσματα επιβεβαιώνουν τους ανωτέρω ισχυρισμούς.

Έχοντας υπόψη τα παραπάνω, οι κυριότερες συνεισφορές της διατριβής συνοψίζονται στα ακόλουθα:

- Τη μοντελοποίηση του δικτύου και των βασικών αλγορίθμων ευκαιριακού χρονοπρογραμματισμού ως ένα θεωρητικό σύστημα SDPS.
- Τον υπολογισμό μέσων ποσοτήτων επίδοσης, της κατανομής του χρόνου απόκρισης και των ασυμπτωτικών της ιδιοτήτων για ένα σύστημα SDPS με πολλαπλές κλάσεις πελατών και εκθετικά κατανομημένες απαιτήσεις εξυπηρέτησης.
- Την απλοποίηση και επαλήθευση των εκφράσεων για δίκτυα που διαθέ-

των την ιδιότητα ισορροπίας με χρήση της θεωρίας που αναπτύχθηκε.

- Τη γενίκευση του προβλήματος της κατανομής του χρόνου απόκρισης για την περίπτωση διαχωρισμού χρονικής κλίμακας και την προσέγγιση της πλήρους λύσης μέσω μιας διαδικασίας αποσύνθεσης σε μειωμένα μοντέλα.
- Τη σύγκριση των αλγορίθμων maxSNR και PF πάνω σε θέματα δικαιοσύνης και κατανομής του χρόνου απόκρισης.
- Την πρόταση ενός καινούριου αλγορίθμου που συνδυάζει τα επιθυμητά χαρακτηριστικά των προηγούμενων δύο.

Επίσης, στα πλαίσια της ερευνητικής προσπάθειας που διενεργήθηκε, προέκυψαν οι παρακάτω δημοσιεύσεις:

- I. Giannoulakis, K. Kontovasilis, N. Mitrou, **Sojourn time distributions and time scale separation results for multiclass state-dependent processor sharing systems with a finite service pool and exponential service requirements**, *Performance Evaluation*, 68 (7) (2011): 586-613, doi:10.1016/j.peva.2011.03.008
- I. Giannoulakis, K. Kontovasilis, N. Mitrou, **Balancing the performance-fairness tradeoff for wireless network access through channel-aware opportunistic schedulers under heterogeneous radio conditions**, *Tech. Report*, 2012, submitted for publication.
- I. Giannoulakis, K. Kontovasilis, N. Mitrou, **Performance, Dimensioning and Interference Tradeoffs for Two-tier Wireless Networks**, *accepted for publication in IEEE PIMRC 2013, London, UK.*

1.3 Δομή της Διατριβής

Η διατριβή αποτελείται από τρία κυρίως μέρη. Στο πρώτο μέρος περιγράφεται το πρόβλημα της επίδοσης αλγορίθμων ευκαιριακού χρονοπρογραμματισμού σε ασύρματα δίκτυα τηλεπικοινωνιών και αποδεικνύονται κάποιες βασικές ιδιότητες που θα χρησιμεύσουν στην ανάλυση που θα ακολουθήσει. Το δεύτερο μέρος αφιερώνεται στη μελέτη του γενικού μαθηματικού προβλήματος που αφορά συστήματα SDPS (ειδική περίπτωση των οποίων αποτελεί και το πρόβλημα του ευκαιριακού χρονοπρογραμματισμού που αποτέλεσε το έναυσμα για την ερευνητική δραστηριότητα της διατριβής), αλλά και προσεγγιστικά αποτελέσματα για την περίπτωση διαχωρισμού χρονικής κλίμακας. Το

τελευταίο μέρος επικεντρώνεται στην εφαρμογή των συμπερασμάτων και αποτελεσμάτων του παραπάνω γενικότερου προβλήματος σε συστήματα που επιτρέπουν ευκαιριακό χρονοπρογραμματισμό καθώς και στην πρόταση ενός νέου αλγορίθμου που διατηρεί τα ευνοϊκά χαρακτηριστικά των προηγούμενων.

Πιο συγκεκριμένα, στο Κεφάλαιο 2 παρουσιάζονται τα κύρια χαρακτηριστικά ενός ασύρματου δικτύου τηλεπικοινωνιών με ευκαιριακό χρονοπρογραμματισμό, οι βασικότεροι σχετικοί αλγόριθμοι καθώς και η αντίστοιχη βιβλιογραφική επισκόπηση. Επίσης αναφέρονται οι υποθέσεις που υιοθετούμε για το υπό μελέτη μοντέλο, σχολιάζεται η επίδραση της κατανομής σκέδασης της ασύρματης ζεύξης και γίνεται εισαγωγή του περιβάλλοντος διαχωρισμού χρονικής κλίμακας.

Στα Κεφάλαια 3-7 μεταφερόμαστε στο γενικότερο μαθηματικό πρόβλημα που αφορά συστήματα αναμονής τύπου SDPS. Η αρχή γίνεται στο Κεφάλαιο 3 με την αναφορά άλλων εργασιών που έχουν χρησιμοποιήσει αυτή την μοντελοποίηση για δίκτυα ευκαιριακού χρονοπρογραμματισμού. Έπειτα παραθέτουμε μία βιβλιογραφική επισκόπηση του θεωρητικού μοντέλου SDPS και περιγράφονται οι κύριοι άξονες της μεθοδολογίας που ακολουθήθηκε για την επίλυση του προβλήματος. Το κεφάλαιο κλείνει με ορισμό των μεγεθών που θα χρειαστούν στη συνέχεια.

Στο Κεφάλαιο 4 υπολογίζεται η κατανομή του χρόνου απόκρισης για ένα σύστημα SDPS με πολλαπλές κλάσεις πελατών και απαιτήσεις εξυπηρέτησης εκθετικά κατανεμημένες. (Οι κλάσεις διαχωρίζονται από το μέσο ρυθμό εξυπηρέτησης που λαμβάνουν οι πελάτες τους.) Επίσης, παρέχονται ιδιότητες για την ασυμπτωτική συμπεριφορά της κατανομής του χρόνου απόκρισης, υπολογίζονται άνω και κάτω φράγματα που σχετίζονται με αυτή, ενώ αναλύεται και η ειδική περίπτωση συστημάτων που διαθέτουν την ιδιότητα ισορροπίας, για τα οποία οι εκφράσεις των αποτελεσμάτων αποκτούν απλούστερη μορφή.

Το Κεφάλαιο 5 παρουσιάζει προσεγγιστικά αποτελέσματα για την περίπτωση που υφίσταται διαχωρισμός χρονικής κλίμακας μεταξύ ομάδων κλάσεων. Προς το σκοπό αυτό, γίνεται χρήση μίας διαδικασίας αποσύνθεσης και συνάθροισης μειωμένων μοντέλων, η οποία πηγάζει από τη δομή NCD του υποκείμενου πλήρους συστήματος.

Στο Κεφάλαιο 6 αναλύονται κάποιες ειδικές περιπτώσεις του διαχωρισμού χρονικής κλίμακας, οι οποίες θα χρησιμοποιηθούν σε επόμενα κεφάλαια κατά την εφαρμογή των αποτελεσμάτων της ανάλυσης στο πραγματικό σύστημα ασύρματων επικοινωνιών που μελετάμε. Παράλληλα, διερευνάται η επίδραση της ύπαρξης πελατών αργών κλάσεων αλλά και της μορφής του θαλάμου εξυπηρέτησης σε σχέση με την ασυμπτωτική συμπεριφορά ουράς της κατανομής του χρόνου απόκρισης των πελατών των γρήγορων κλάσεων.

Το δεύτερο κομμάτι της διατριβής που αναλύει το γενικευμένο θεωρητικό πρόβλημα, ολοκληρώνεται με το Κεφάλαιο 7, το οποίο δείχνει ότι η ουσία των αποτελεσμάτων διατηρείται ακόμα και αν οι απαιτήσεις εξυπηρέτησης δεν είναι εκθετικά κατανομημένες αλλά χαρακτηρίζονται από μία γενικότερη οικογένεια κατανομών.

Στο Κεφάλαιο 8 εφαρμόζονται τα βασικά αποτελέσματα ώστε να εξαχθούν μετρικές επίδοσης και μέσα ποσοτικά μεγέθη για συστήματα που εφαρμόζουν ευκαιριακό χρονοπρογραμματισμό, παρουσία ή μη φαινομένων διαχωρισμού χρονικής κλίμακας.

Στο Κεφάλαιο 9 αναλύονται αποτελέσματα σχετικά με την επίδοση των αλγορίθμων maxSNR και PF και εκτελείται η μεταξύ τους σύγκριση. Ακολουθούν βασικά συμπεράσματα που προκύπτουν από την ανάλυση του δεύτερου μέρους της διατριβής και σχετίζονται με την επίδραση της ταυτόχρονης παρουσίας τερματικών αργών και γρήγορων κλάσεων και της μορφής του θαλάμου εξυπηρέτησης στην ασυμπτωτική συμπεριφορά της κατανομής του χρόνου απόκρισης.

Η περιγραφή του νέου αλγορίθμου που προτείνεται στη διατριβή, ακολουθεί στο Κεφάλαιο 10, με ταυτόχρονη ανάλυση για τα θέματα επίδοσης και δικαιοσύνης που προκύπτουν κατά την υιοθέτηση του σχήματος αυτού.

Στο Κεφάλαιο 11 επαληθεύονται αριθμητικά τα προσεγγιστικά αποτελέσματα που αφορούν το διαχωρισμό χρονικής κλίμακας. Ταυτόχρονα, παρουσιάζονται αριθμητικά αποτελέσματα που συγκρίνουν τους τρεις αλγορίθμους και επαληθεύουν τα θεωρητικά ευρήματα.

Τέλος, τα συμπεράσματα της διατριβής και οι πιθανές μελλοντικές ερευνητικές προσπάθειες συνοψίζονται στο Κεφάλαιο 12.

Στο παράρτημα παραθέτουμε κάποιες τεχνικές αποδείξεις που κρίθηκε σκόπιμο να παραληφθούν από το κυρίως κείμενο της διατριβής.



Κεφάλαιο 2

Βασικά Χαρακτηριστικά Ευκαιριακού Χρονοπρογραμματισμού

2.1 Επισκόπηση αλγορίθμων ευκαιριακού χρονοπρογραμματισμού

Σ Ε ένα δίκτυο ευκαιριακού χρονοπρογραμματισμού τα τερματικά που μπορούν να υποστηρίξουν υψηλό ρυθμό μετάδοσης επιλέγονται για εξυπηρέτηση κατά βάση συχνότερα σε σχέση με εκείνα που διαθέτουν χαμηλότερο ρυθμό. Για το σκοπό αυτό γίνεται χρήση κατάλληλων αλγορίθμων που λαμβάνουν υπόψη το στιγμιαίο σηματοθορυβικό λόγο του τερματικού. Υπενθυμίζουμε εδώ ότι από τα δυο βασικά σχήματα ευκαιριακού χρονοπρογραμματισμού, το πρώτο στηρίζεται αποκλειστικά στην πληροφορία για την κατάσταση της ασύρματης ζεύξης και λειτουργεί με το να επιλέγει πάντοτε για εξυπηρέτηση το τερματικό με το μέγιστο SNR (maxSNR). Ωστόσο, προκειμένου να καταπολεμηθούν φαινόμενα αδικίας μεταξύ τερματικών με πολύ διαφορετικές συνθήκες μετάδοσης, έχει προταθεί ο αλγόριθμος PF. Ο PF εισάγει στα κριτήρια που χρησιμοποιεί την επιτευχθείσα ρυθμαπόδοση κάθε τερματικού για κάποιο σταθερό χρονικό διάστημα που προηγήθηκε, διαιρώντας με αυτήν τη στιγμιαία τιμή του SNR.

Εφόσον υψηλότερες τιμές του σηματοθορυβικού λόγου αντιστοιχούν σε μεγαλύτερες τιμές στο ρυθμό μετάδοσης, και οι δύο παραπάνω αλγόριθμοι μπορούν να εκφραστούν σε όρους του υποστηριζόμενου ρυθμού μετάδοσης αντί του ληφθέντος σηματοθορυβικού λόγου. Συγκεκριμένα, για τον maxSNR μπορούμε να δώσουμε μια ισοδύναμη περιγραφή που προσφέρει μεγαλύτερη

ευελιξία κατά τη μοντελοποίηση. Συμβολίζοντας με \tilde{R} , τον επιτεύξιμο ρυθμό μετάδοσης, προκύπτει ο αλγόριθμος μέγιστου ρυθμού (αγ.: Best Rate (BR)), που περιγράφεται από τη σχέση $u = \arg \max_v \tilde{R}_v$ όπου ο δείκτης v αναφέρεται σε όλα τα τερματικά που ζητούν εξυπηρέτηση. Σύμφωνα με το [37], ο BR αποτελεί την πολιτική που μεγιστοποιεί τη συνολική ρυθμαπόδοση και ως εκ τούτου, είναι ένα χρήσιμο σημείο αναφοράς. Όσον αφορά τον PF, η μέση ρυθμαπόδοση μπορεί να αντικατασταθεί από το μέσο επιτεύξιμο ρυθμό μετάδοσης κάθε τερματικού, \tilde{r} , καταλήγοντας έτσι στον αλγόριθμο κανονικοποιημένου μέγιστου ρυθμού (αγ.: Normalized Best Rate (NBR)). Σε αντιστοιχία με την περιγραφή του BR, ο NBR επιλέγει για εξυπηρέτηση ένα τερματικό σύμφωνα με τον κανόνα $u = \arg \max_v \tilde{R}_v / \tilde{r}_v$. Στο [32] έχει αποδειχθεί η ισοδυναμία μεταξύ PF και NBR δεδομένης της γραμμικής εξάρτησης του σηματοθορυβικού λόγου και του ρυθμού εξυπηρέτησης καθώς και κάποιων πιο τεχνικών προϋποθέσεων, που επίσης συζητούμε στην Ενότητα 2.2. Στην παρούσα διατριβή χρησιμοποιούνται οι παραλλαγές που περιγράφονται μέσω ρυθμών μετάδοσης (δηλαδή BR και NBR), επειδή επιτρέπουν πιο άνετη μαθηματική περιγραφή.

Παρόμοιοι αλγόριθμοι έχουν επίσης προταθεί στη βιβλιογραφία (βλ., π.χ. [4, 8, 33, 35, 40, 41, 46]). Το κοινό χαρακτηριστικό τους έγκειται στο ότι όλοι χρησιμοποιούν με κάποιο τρόπο τη στιγμιαία τιμή του σηματοθορυβικού λόγου κατά την απόφαση χρονοπρογραμματισμού των τερματικών. Για παράδειγμα, το [8] προτείνει ένα μηχανισμό που περιλαμβάνει στατιστικά στοιχεία του ρυθμού μετάδοσης σαν εναλλακτική πρόταση για τη μέση ρυθμαπόδοση, ώστε να υπερκεράσει το εμπόδιο της ασταθούς κατανομής της σκέδασης, με κόστος τη μεγαλύτερη πολυπλοκότητα κατά την υλοποίηση του. Στο [60] οι συγγραφείς δείχνουν ότι μπορεί να υπάρξει ένα είδος συμβιβασμού μεταξύ δικαιοσύνης και ρυθμού μετάδοσης, ομαδοποιώντας τα τερματικά σύμφωνα με τη στατιστική του ρυθμού τους και στη συνέχεια, εφαρμόζοντας τον BR και τον NBR σε ξεχωριστά βήματα. Εξάλλου, ένας παρόμοιος αλγόριθμος με τον NBR, ο οποίος λαμβάνει υπόψη το σχετικό σφάλμα αντί για το λόγο στιγμιαίου και μέσου ρυθμού, προτείνεται στο [6]. Το εκεί μοντέλο υιοθετεί σκέδαση Rayleigh, λογαριθμική σχέση μεταξύ SNR και στιγμιαίου ρυθμού μετάδοσης και καταλήγει στο ότι η επιλογή ταυτόχρονα περισσότερων του ενός τερματικών με ευνοϊκές συνθήκες βελτιώνει την εκμετάλλευση του καναλιού.

Αν και ο maxSNR αποτελεί την απλούστερη εκδοχή μεταξύ αυτών, στην πλειοψηφία τους, προσπαθούν να προσφέρουν μεγαλύτερη αίσθηση δικαιοσύνης μεταξύ των τερματικών αναιρώντας το «βίαιο» διαχωρισμό που στηρίζεται αποκλειστικά στο SNR. Η αναίρεση αυτή κατά κανόνα υλοποιείται με κάποιο κόστος στην επί μέρους και στη συνολική ρυθμαπόδοση. Παρά το γεγονός ότι μία πληθώρα αλγορίθμων έχει προταθεί προς τη συγκεκριμένη κατεύθυνση, οι

βασικές ιδιότητες ενός σχήματος που λαμβάνει υπόψη την ισχύ του σήματος αλλά και την επιτευχθείσα ρυθμαπόδοση, μπορούν να αποδοθούν ικανοποιητικά από τον PF.

Οι αναλυτικές μελέτες των αλγορίθμων ευκαιριακού χρονοπρογραμματισμού στοχεύουν κυρίως σε μέσα μεγέθη επίδοσης, όπως για παράδειγμα η μέση καθυστέρηση μίας δεδομένης απαίτησης εξυπηρέτησης. Η κυρία μέθοδος για να επιτευχθεί αυτό περιλαμβάνει τον υπολογισμό των πιθανοτήτων σταθερής κατάστασης του συστήματος. Συγκεκριμένα, η μέση καθυστέρηση υπολογίζεται από το άθροισμα όλων των πιθανών καθυστερήσεων που μπορούν να τύχουν για μία επιτυχημένη μετάδοση, πολλαπλασιασμένων με τις πιθανότητες που τους αντιστοιχούν. Ακολουθώντας τη συγκεκριμένη μεθοδολογία, θεωρώντας δίκτυο με ένα και μόνο κανάλι-εξυπηρετητή και σκέδαση Rayleigh, το [54] υπολογίζει τη μέση καθυστέρηση του BR (η οποία ορίζεται ως ο ελάχιστος αριθμός χρονοσχισμών που απαιτούνται για την επιτυχή μετάδοση ενός προκαθορισμένου αριθμού δεδομένων). Το μέγεθος κάθε χρονοσχισμής θεωρείται σταθερό, όπως άλλωστε και σε όλα τα μοντέρνα πρότυπα. Ένα παρόμοιο αποτέλεσμα παράγεται στο [59], αλλά σε αντίθεση με προηγουμένως, το μέγεθος της χρονοσχισμής θεωρείται μεταβλητό. Αν και η συγκεκριμένη μεθοδολογία είναι σχετικά απλή στη σύλληψή της, μπορεί να καταλήξει ιδιαίτερα περίπλοκη κατά τη θεωρητική ανάλυση της επίδοσης των αλγορίθμων, ιδίως όσον αφορά πολύπλοκες παραλλαγές τους.

Η σύγκριση μεταξύ BR και NBR δεν είναι απλή, πρωτίστως επειδή δεν υπάρχει μία πλήρης μαθηματική περιγραφή. Κάποιες προσπάθειες χρησιμοποιούν αποτελέσματα προσομοιώσεων, όπως για παράδειγμα στο [53] για ένα μείγμα κίνησης πραγματικού και μη πραγματικού χρόνου και στο [31], που στοχεύει στις περιόδους αποκλεισμού των τερματικών (αγ.: starvation period). Προσομοιώσεις συμπεριλαμβάνονται και σε μελέτες που έχουν αναφερθεί ήδη (π.χ. [4,6,8,9,41,46,60]), κυρίως για λόγους σύγκρισης των νεοεισαχθέντων σχημάτων με τα προϋπάρχοντα. Στις μελέτες αυτές αποτελούν κοινές παραδοχές η σκέδαση Rayleigh και ομοιογενές μοντέλο κίνησης (δηλαδή τερματικά χωρίς διαχωρισμό κλάσεων). Οι μετρικές επίδοσης που εξετάζονται σχετίζονται με τη μέση ρυθμαπόδοση και θέματα δικαιοσύνης. Επιπρόσθετα, στο [35] εμφανίζεται η έννοια πολλαπλών κλάσεων, όπου μία κυψέλη χωρίζεται σε τρεις τομείς, δίνοντας τρεις τύπους κίνησης. Η ανάλυση χρησιμοποιεί πιθανοτικά επιχειρήματα για να υπολογίσει τη μέση ρυθμαπόδοση και προτείνει έναν αλγόριθμο που αποτελείται από συνδυασμό του PF και μιας τροποποίησης του σχήματος κυκλικής εναλλαγής (αγ.: round robin).

2.2 Μοντελοποίηση ασύρματου περιβάλλοντος και σχετικές ιδιότητες

Η τυπική κυψέλη ενός ασύρματου δικτύου που είναι σε θέση να υποστηρίξει ευκαιριακό χρονοπρογραμματισμό αποτελείται από ένα σταθμό βάσης που βρίσκεται στο κέντρο της και ένα πεπερασμένο αριθμό τερματικών που μπορούν να επικοινωνήσουν αποκλειστικά με το σταθμό βάσης. Στην περίπτωση που κάποιο τερματικό επιζητεί εξυπηρέτηση (π.χ. λήψη - αποστολή δεδομένων κ.λπ.), θα αποκαλείται «ενεργό», δηλαδή σε κατάσταση που ανταγωνίζεται ώστε να λάβει τους πόρους του δικτύου. Ο σταθμός βάσης παρακολουθεί την κατάσταση της ασύρματης ζεύξης σε κάθε χρονοσχισμή για όλα τα ενεργά τερματικά (π.χ., μέσω ενός ανεξάρτητου καναλιού αποκλειστικά δεσμευμένο για το σκοπό αυτό) και θεωρούμε ότι η πληροφορία αυτή δεν υπόκειται σε σφάλματα.

Δεδομένων των ανωτέρω παραδοχών, ο σταθμός βάσης διαθέτει ακριβή πληροφορία για την τιμή του SNR κάθε ενεργού τερματικού κατά την έναρξη κάθε χρονοσχισμής. Το δίκτυο λειτουργεί με χρονική πολυπλεξία, δηλαδή σε κάθε χρονοσχισμή ο σταθμός βάσης αποφασίζει σύμφωνα με κάποιο αλγόριθμο που βασίζεται στην πληροφορία για την κατάσταση της ασύρματης ζεύξης για το ποιο τερματικό θα εξυπηρετήσει, παρέχοντας στο τερματικό το σύνολο των διαθέσιμων πόρων. Οι συγκεκριμένες χρονοσχισμές έχουν σταθερή διάρκεια αλλά αυτή μπορεί να διαφέρει από πρότυπο σε πρότυπο (π.χ., 1.67 ms στο HDR, 2 ms στο HSDPA κ.λπ.). Επίσης, η στιγμιαία τιμή της ισχύος του σήματος θεωρείται σταθερή για τη διάρκεια της χρονοσχισμής. Όταν περισσότερα του ενός τερματικά είναι ενεργά, η εναλλαγή τους στην εξυπηρέτηση μπορεί να λάβει χώρα ανά μικρά χρονικά διαστήματα, τα οποία είναι σαφώς μικρότερα από την τυπική απαίτηση εξυπηρέτησης.

Προκειμένου να μελετηθούν θέματα δικαιοσύνης υπό διαφορετικές συνθήκες μετάδοσης, στη διατριβή θα ασχοληθούμε με τη γενικότερη περίπτωση κατά την οποία τα τερματικά μπορούν εν γένει να κατηγοριοποιηθούν σε ένα πεπερασμένο αριθμό κλάσεων, k . Το σύνολο των κλάσεων θα συμβολίζεται με $\mathcal{K} \triangleq \{1, \dots, k\}$. Η κατηγοριοποίηση που χρησιμοποιείται λαμβάνει ως κριτήριο την απόσταση από το σταθμό βάσης, όπου τερματικά με διαφορετικές αποστάσεις από το σταθμό βάσης παρουσιάζουν επίσης διαφορετικό μέσο σηματοθορυβικό λόγο, $\tilde{\gamma}_i$, έτσι ώστε τερματικά της ίδιας κλάσης να υπόκεινται στο ίδιο μέσο SNR. Η κατάσταση του συστήματος θα αναπαρίσταται με ένα διάνυσμα της μορφής $\mathbf{n} = (n_1, \dots, n_k)$, όπου το n_i υποδηλώνει τον αριθμό των ενεργών τερματικών από την κλάση i .

Κατά τη μοντελοποίηση θεωρούμε μόνο τις μεταβολές του σηματοθορυ-

βικού λόγου του τερματικού που οφείλονται σε πολλαπλή σκέδαση (αγ.: multipath fading). Οι υπόλοιπες επιδράσεις (απώλειες διαδρομής, σκίασης κ.λπ.) συμπεριλαμβάνονται στη μέση τιμή του SNR, $\tilde{\gamma}_i$. Όλα τα τερματικά εντός της κυψέλης υφίστανται ομοιογενή σκέδαση, δηλαδή υπόκεινται σε μία κατανομή σκέδασης με τις ίδιες παραμέτρους, εκτός από το $\tilde{\gamma}_i$. Η μέση τιμή του SNR θεωρείται σταθερή για όλες τις κλάσεις. Εν αντιθέσει, η στιγμιαία τιμή του SNR εκλαμβάνεται σταθερή μόνο κατά τη διάρκεια μίας χρονοσχισμής. Λαμβάνοντας υπόψη τα παραπάνω, αν SNR_i είναι η τυχαία μεταβλητή της στιγμιαίας τιμής του SNR, τότε η τυχαία μεταβλητή του κανονικοποιημένου SNR_i , δηλαδή η $\text{SNR}_i/\tilde{\gamma}_i$, είναι ανεξάρτητη της κλάσης i , $i \in \mathcal{K}$ με συνάρτηση κατανομής πιθανότητας (αγ. Probability Distribution Function (PDF)) $\hat{F}(x)$.

Επιπρόσθετα, το μοντέλο μας βασίζεται στις ακόλουθες παραδοχές. Οι κανονικοποιημένες τιμές του στιγμιαίου SNR για την κλάση i θεωρούνται ανεξάρτητες τυχαίες μεταβλητές. Για όλες τις κλάσεις οι κανονικοποιημένες τιμές του στιγμιαίου SNR όλων των τερματικών της ίδιας κλάσης είναι ταυτοτικά καταμεμημένες με μέση τιμή ίση με μονάδα. Επιπρόσθετα, η $\text{SNR}_i(t)/\tilde{\gamma}_i$, $t \geq 0$ θεωρείται ότι είναι μία στάσιμη στοχαστική διαδικασία (αγ.: stationary stochastic process).

Στο σημείο αυτό είναι σκόπιμο να δειχτεί ότι η υπόθεση μιας κοινής PDF, $\hat{F}(x)$ όντως ισχύει για όλα τα ευρέως χρησιμοποιούμενα μοντέλα που χαρακτηρίζουν τα ασύρματα κανάλια. Ας θεωρήσουμε καταρχήν την κατανομή σκέδασης Rayleigh, ένα ευρέως χρησιμοποιούμενο μοντέλο όταν το ληφθέν σήμα δεν περιέχει συνιστώσα γραμμής οπτικής επαφής (αγ.: Line Of Sight (LOS)) (π.χ. εσωτερικοί χώροι, αστικό περιβάλλον κ.λπ.) και έστω ότι $\tilde{\gamma}$ είναι η μέση τιμή του SNR, χωρίς ακόμη διάκριση μεταξύ κλάσεων. Σε αυτή την περίπτωση, η συνάρτηση πυκνότητας πιθανότητας της στιγμιαίας τιμής του SNR είναι

$$f_{\text{Rayleigh}}(x) \triangleq \Pr\{x < \text{SNR} \leq x + dx\} = \frac{1}{\tilde{\gamma}} \exp\left(-\frac{x}{\tilde{\gamma}}\right), \quad x \geq 0.$$

Όταν αντίθετα, υπάρχει συνιστώσα LOS στο ληφθέν σήμα, συνήθως χρησιμοποιείται μία από τις κατανομές Rice ή Nakagami. Η συνάρτηση πυκνότητας πιθανότητας της πρώτης για το SNR δίνεται από τον τύπο

$$f_{\text{Rice}}(x) = \frac{(1+K)e^{-K}}{\tilde{\gamma}} \exp\left(-\frac{(1+K)x}{\tilde{\gamma}}\right) I_0\left(\sqrt{\frac{4K(1+K)x}{\tilde{\gamma}}}\right), \quad x \geq 0, K \geq 0. \quad (2.1)$$

όπου K είναι ο παράγοντας Rice και $I_0(\cdot)$ είναι η τροποποιημένη συνάρτηση Bessel, πρώτου είδους, μηδενικής τάξης. Το μοντέλο σκέδασης Nakagami m μπορεί να προσεγγίσει και την κατανομή Rayleigh και την κατανομή Rice,

μέσω κατάλληλης επιλογής της παραμέτρου m . Η συνάρτηση πυκνότητας πιθανότητας του SNR είναι

$$f_{Nakagami}(x) = \frac{m^m x^{m-1}}{\tilde{\gamma}^m \Gamma(m)} \exp\left(-\frac{mx}{\tilde{\gamma}}\right), \quad x \geq 0, m \geq 1/2, \quad (2.2)$$

όπου $\Gamma(\cdot)$ είναι η συνάρτηση γάμμα. Όταν $m = 1$, η (2.2) καταλήγει ακριβώς σε σκέδαση Rayleigh, ενώ υπάρχει και μία αμφιμονοσήμαντη απεικόνιση της παραμέτρου m προς τον παράγοντα Rice K , ώστε να προσεγγιστεί η κατανομή Rice, για $m > 1$. Και για τα δύο παραπάνω μοντέλα σκέδασης αποτελεί συνήθη πρακτική να παραμένουν οι παράμετροι K, m ανεξάρτητες της απόστασης του τερματικού από το σταθμό βάσης και να υποτίθεται ομοιογενές περιβάλλον σκέδασης. Τέτοια μοντελοποίηση έχει χρησιμοποιηθεί για παράδειγμα στο [21], όπου το περιβάλλον σκέδασης προσεγγίζεται από μία κοινή παράμετρο Nakagami, στα [3, 55], όπου όλα τα τερματικά θεωρούνται ότι υπόκεινται σε κοινή σκέδαση Nakagami, ακόμα και στην περίπτωση διαφορετικού μέσου SNR και στο [2] (βλ. π.χ., την Υπόθεση (39) εκεί), όπου μία κοινή παράμετρος m χρησιμοποιείται για λήψη από περισσότερους του ενός σταθμούς βάσης.

Για οποιοδήποτε από τα παραπάνω μοντέλα σκέδασης, έστω «C», η κανονικοποιημένη ποσότητα $\text{SNR}_i/\tilde{\gamma}_i$ διαθέτει κοινή PDF $\hat{F}(\cdot)$, αναλλοίωτη ως προς την κλάση, αν και μόνο αν οι αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας, δηλ., $f_{C,i}(\frac{x}{\tilde{\gamma}_i})$, συμπίπτουν με μία κοινή συνάρτηση $f_C(\cdot)$ ανά μοντέλο, για όλα τα $i \in \mathcal{K}$. Μπορούμε εύκολα να παρατηρήσουμε ότι

$$\hat{f}_C(x) = \exp(-x), \quad x \geq 0, \quad (\text{Rayleigh})$$

$$\hat{f}_C(x) = (1+K)e^{-K} \exp(-(1+K)x) I_0(\sqrt{4K(1+K)x}), \quad x \geq 0, K \geq 0, \quad (\text{Rice})$$

$$\hat{f}_C(x) = \frac{m^m x^{m-1}}{\Gamma(m)} \exp(-mx), \quad x \geq 0, m \geq 1/2, \quad (\text{Nakagami})$$

αποκαλύπτοντας ότι όλες παραμένουν αναλλοίωτες ως προς τη μέση τιμή του SNR, $\tilde{\gamma}$. Η παραπάνω ιδιότητα μπορεί επίσης να εκφραστεί ως αναλλοιωσιμότητα όλων των αντίστοιχων PDF, έστω $F_C(\cdot)$, της κανονικοποιημένης τιμής του σηματοθορυβικού λόγου, $\text{SNR}/\tilde{\gamma}$, καταλήγοντας σε μία κοινή PDF ανά μοντέλο, $\hat{F}_C(\cdot)$, ανεξάρτητη από τη μέση τιμή του SNR. Εφόσον στο μοντέλο μας το μέσο SNR είναι η ποσότητα που διαχωρίζει τα τερματικά σε κλάσεις, η $\hat{F}_C(\cdot)$ παραμένει ανεξάρτητη της κλάσης.

Επίσης σε ό,τι ακολουθεί, θα θεωρούμε ότι ο στιγμιαίος εφικτός ρυθμός μετάδοσης του τερματικού, \tilde{R} είναι συνδεδεμένος γραμμικά με τη στιγμιαία τιμή του SNR. Η υπόθεση της γραμμικότητας είναι λογική και χρησιμοποιείται ευρέως κατά τη μοντελοποίηση, κυρίως όταν τα τερματικά δε βρίσκονται

υπερβολικά κοντά στο σταθμό βάσης. Ως αποτέλεσμα της υπόθεσης αυτής οι κανονικοποιημένες στιγμιαίες τιμές \tilde{R}_i/\tilde{r}_i , όπου το \tilde{r}_i δηλώνει το μέσο ρυθμό μετάδοσης ενός τερματικού κλάσης i , διαθέτουν την ίδια PDF για όλες τις κλάσεις, η οποία επιπρόσθετα είναι ίση με την κοινή PDF της κανονικοποιημένης τιμής του SNR, έτσι ώστε

$$F_i(x) = \Pr\{\tilde{R}_i \leq x\} = \hat{F}\left(\frac{x}{\tilde{r}_i}\right), \quad i \in \mathcal{K}. \quad (2.3)$$

Δεδομένης της υπόθεσης γραμμικότητας και του παραπάνω σχόλιου, στη συνέχεια οι PDF $F_i(\cdot)$, $\hat{F}(\cdot)$ θα αναφέρονται στους ρυθμούς μετάδοσης και όχι σε στο SNR.

2.3 Υπολογισμός κέρδους διαφορισμού πολλαπλών χρηστών και σχετικές ιδιότητες

Οι δύο κύριοι αλγόριθμοι χρονοπρογραμματισμού, BR και NBR, μπορούν να περιγραφούν ως παραλλαγές του ίδιου κανόνα. Ειδικότερα, και οι δύο προκύπτουν με κατάλληλη επιλογή των βαρών a_v στον τύπο

$$u = \arg \max_v a_v \tilde{R}_v, \quad (2.4)$$

όπου ο δείκτης v αναφέρεται σε όλα τα ενεργά τερματικά. Για παράδειγμα, η επιλογή $a_v = 1$ οδηγεί άμεσα στον BR, το ανάλογο του maxSNR στην περίπτωση που επιλέγουμε να συγκρίνουμε τους ρυθμούς μετάδοσης. Από την άλλη πλευρά, όταν εφαρμόζεται ο NBR, η μετρική σύμφωνα με την οποία επιλέγεται ένα τερματικό προς εξυπηρέτηση είναι ο κανονικοποιημένος ρυθμός μετάδοσης $a_v = 1/\tilde{r}_v$.

Η υπόθεση της γραμμικότητας μεταξύ ρυθμού εξυπηρέτησης και SNR που εισήχθη προηγουμένως, επιφέρει μέσω της (2.3) κάποιες σημαντικές ιδιότητες με επιπλέον συνέπειες στην επίδοση των αλγορίθμων. Η βαθύτερη διαίσθηση που προκύπτει είναι χρήσιμη σε περιπτώσεις που η υπόθεση γραμμικότητας δεν ισχύει επακριβώς. Επιπλέον, ο χαρακτήρας ευκαιριακού χρονοπρογραμματισμού του συστήματος έχει σαν αποτέλεσμα την αύξηση του συνολικού μέσου ρυθμού μετάδοσης αναλόγως με το πλήθος (και τις κλάσεις στη γενική περίπτωση) των ενεργών τερματικών στο δίκτυο. Η μεταβολή αυτή περιγράφεται μέσω της έννοιας του «κέρδους διαφορισμού πολλαπλών χρηστών» (αγ.: multiuser diversity gain) κατά τη διάρκεια εξυπηρέτησης περισσότερων του ενός τερματικών. Ως κέρδος διαφορισμού πολλαπλών χρηστών, $\phi_i(\mathbf{n})$, ορίζεται ο λόγος του ρυθμού μετάδοσης που λαμβάνει κάποιο τερματικό, προς τον

αντίστοιχο ρυθμό που θα λάμβανε αν εξυπηρετούταν αποκλειστικά μόνο του στο σύστημα. Με αναφορά στην κλάση i , $i \in \mathcal{K}$, ορίζουμε

$$\phi_i(\mathbf{n}) \triangleq \tilde{r}_i^{-1} \mathbf{E} \left[\tilde{R}_i \mathbf{I}_{\text{επιλογή τερματικού κλάσης } i} \right],$$

όταν το σύστημα είναι στην κατάσταση \mathbf{n} , όπου $\mathbf{I}_{\mathcal{A}}$ είναι η δείκτρια συνάρτηση (αγ.: indicator function) του γεγονότος \mathcal{A} . Έτσι,

$$\phi_i(\mathbf{n}) = \tilde{r}_i^{-1} \int_0^\infty y (F_i(y))^{n_i-1} \prod_{\substack{j \in \mathcal{K} \\ j \neq i}} F_j\left(\frac{a_i y}{a_j}\right)^{n_j} dF_i(y), \quad i \in \mathcal{K}, n_i > 0$$

και με αλλαγή μεταβλητής στην έκφραση αυτή έχουμε ότι

$$\phi_i(\mathbf{n}) = \int_0^\infty x \hat{F}(x)^{n_i-1} \prod_{\substack{j \in \mathcal{K} \\ j \neq i}} \hat{F}\left(\frac{a_i \tilde{r}_i}{a_j \tilde{r}_j} x\right)^{n_j} d\hat{F}(x), \quad i \in \mathcal{K}, n_i > 0. \quad (2.5)$$

Από τα παραπάνω γίνεται φανερό πως το κέρδος διαφορισμού πολλαπλών χρηστών $\phi_i(\mathbf{n})$ είναι γνησίως φθίνουσα συνάρτηση ως προς τον αριθμό των ενεργών τερματικών n_i , $i \in \mathcal{K}$. Σημειώνουμε εδώ ότι το $\phi_i(\mathbf{n})$ αναφέρεται σε ένα τερματικό. Το συνολικό κέρδος που συνδέεται με την κλάση i ισούται με $n_i \phi_i(\mathbf{n})$, $i \in \mathcal{K}$. Στο παράρτημα αποδεικνύεται ότι το γινόμενο $n_i \phi_i(\mathbf{n})$ είναι γνησίως αύξουσα συνάρτηση ως προς n_i . Επίσης, σχετικά με την ανάλυση που προηγήθηκε και αφορούσε την επιλογή της κατανομής σκέδασης, σημειώνουμε ότι η (2.5) είναι πράγματι γενικής μορφής, με την έννοια ότι επιτρέπει διαφορετικές κατανομές σκέδασης, υπό τον όρο ότι η ανεξαρτησία της $\hat{F}(x)$ ως προς το \tilde{r}_i , $i \in \mathcal{K}$ παραμένει.

Ακολουθεί η εξειδίκευση της (2.5) για τους δύο βασικούς αλγορίθμους. Εφόσον ο BR προκύπτει με την επιλογή $a_i = 1$ στην 2.4, τα κέρδη διαφορισμού πολλαπλών χρηστών για τον BR μπορούν να γραφούν ως

$$\phi_i(\mathbf{n}) = \int_0^\infty x \hat{F}(x)^{n_i-1} \prod_{\substack{j \in \mathcal{K} \\ j \neq i}} \hat{F}\left(\frac{\tilde{r}_i}{\tilde{r}_j} x\right)^{n_j} d\hat{F}(x), \quad i \in \mathcal{K}, n_i > 0. \quad (2.6)$$

Όταν εφαρμόζεται ο NBR, από την (2.5) απορρέει μία επιπλέον ιδιότητα των κερδών διαφορισμού πολλαπλών χρηστών, χρήσιμη για την μετέπειτα ανάλυση. Έστω ότι $|\mathbf{n}|$ είναι το άθροισμα των στοιχείων του διανύσματος \mathbf{n} . Με αντικατάσταση των $a_i = \tilde{r}_i^{-1}$ και $a_j = \tilde{r}_j^{-1}$ στην (2.5), έχουμε ότι

$$\phi_i(\mathbf{n}) = g(|\mathbf{n}|), \quad n_i > 0, \quad (2.7)$$

όπου

$$g(k) \triangleq \int_0^\infty x \hat{F}(x)^{k-1} d\hat{F}(x). \quad (2.8)$$

Με άλλα λόγια τα κέρδη διαφορισμού πολλαπλών χρηστών για τον NBR εξαρτώνται μόνο από το άθροισμα των ενεργών τερματικών και παραμένουν ανεξάρτητα της κλάσης, συνεχίζοντας εκ των υστέρων τη γραμμική συσχέτιση μεταξύ στιγμιαίου SNR και στιγμιαίου ρυθμού εξυπηρέτησης. Σε περιβάλλον σκέδασης Rayleigh ισχύει ότι $\hat{F}(x) = 1 - e^{-x}$, και η (2.8) οδηγεί στο ότι $|m|g(m) = \sum_{k=1}^m 1/k$, όπως άλλωστε έχει δειχθεί στο [14]. Το αποτέλεσμα της (2.8) γενικεύεται για όλα τα σημαντικά μοντέλα σκέδασης.

Ο αλγόριθμος PF απαιτεί κατά την επιλογή των a_v στην (2.4) να πάρουν ως τιμή το αντίστροφο της επιτευχθείσας ρυθμαπόδοσης για κάποιο λογικά επιλεγμένο χρονικό παράθυρο, αντί του μέσου επιτεύξιμου ρυθμού. Το θέμα αντιμετωπίζεται επιτυχώς στο [32], όπου αποδεικνύεται η ισοδυναμία των δύο επιλογών, υπό των προϋποθέσεων γραμμικής σχέσης μεταξύ επιτεύξιμου ρυθμού μετάδοσης και SNR και ότι τα δείγματα του ρυθμού μετάδοσης είναι ασυσχέτιστα μεταξύ τους. Πράγματι, η απόφαση χρονοπρογραμματισμού για τον PF λαμβάνεται ως

$$u = \arg \max_v \frac{\tilde{R}_v}{T_v},$$

όπου T_v είναι η ρυθμαπόδοση του τερματικού v . Ωστόσο ο παράγοντας ρυθμαπόδοσης είναι ευθέως αναλόγως με το μέσο ρυθμό της κλάσης, δηλαδή $T_i = \tilde{r}_i \phi_i(\mathbf{n})$, ενώ όπως υποδεικνύει και η (2.8), σε αυτή την περίπτωση τα κέρδη διαφορισμού είναι ανεξάρτητα της κλάσης, $\phi_i(\mathbf{n}) = g(|\mathbf{n}|)$. Έτσι ο κανόνας χρονοπρογραμματισμού για τον PF μετατρέπεται σε

$$u = \arg \max_v \frac{\tilde{R}_v}{\tilde{r}_v g(|\mathbf{n}|)} = \arg \max_v \frac{\tilde{R}_v}{\tilde{r}_v}, \quad (2.9)$$

θεμελιώνοντας την ισοδυναμία με τον NBR.

2.4 Μεγάλες διαφορές ισχύος σήματος και εφαρμογή στον BR

Στην τελευταία ενότητα του παρόντος κεφαλαίου αναφερόμαστε στο διαχωρισμό των κλάσεων, δηλαδή την κατηγοριοποίησή τους σε ασθενή και ισχυρή ομάδα σύμφωνα με το μέσο σηματοθορυβικό λόγο. Η περίπτωση αυτή εμφανίζεται σε αρκετά πρακτικά προβλήματα και επιτρέπει τη λήψη χρήσιμων συμπερασμάτων για τη συνολική συμπεριφορά του συστήματος. Επίσης, προσφέρει βαθύτερη ανάλυση του προβλήματος επειδή σχετίζεται με ακραία σενάρια και ως εκ τούτου βοηθά στην κατανόηση του περιβάλλοντος που επιβάλλεται από το διαχωρισμό της κλίμακας.

Οι διαφορές στο μέσο SNR αντικατοπτρίζονται μέσω της υπόθεσης γραμμικότητας σε διαφορές του μέσου ρυθμού μετάδοσης. Για το λόγο αυτό, υποθέτουμε το ακόλουθο σχήμα διαφοροποίησης των κλάσεων. Όλες οι κλάσεις του υποσυνόλου με ασθενές μέσο SNR διαθέτουν μέσο ρυθμό μετάδοσης που είναι πολύ μικρότερος από οποιοδήποτε μέσο ρυθμό κλάσης που ανήκει στην ομάδα με δυνατό μέσο SNR. Πιο λεπτομερώς, έστω ότι \mathcal{K}_s είναι η ομάδα κλάσεων με ασθενές σήμα και \mathcal{K}_f είναι η ομάδα κλάσεων με ισχυρό σήμα. Το γεγονός $i \in \mathcal{K}_s$ θα δηλώνει ευθέως ότι $\tilde{r}_i = \tilde{r}_i \epsilon$, όπου το \tilde{r}_i είναι μία σταθερά ανεξάρτητη του ϵ , μίας παραμέτρου που χρησιμοποιείται για να ποσοτικοποιήσει τη διαφοροποίηση. Για όλα τα $i \in \mathcal{K}_f$ θεωρούμε ότι ο μέσος ρυθμός μετάδοσης \tilde{r}_i είναι τάξης $O(1)$, καθώς $\epsilon \rightarrow 0$. Σε αντίθεση με την συνήθη πρακτική στη σχετική βιβλιογραφία (βλ. π.χ., [19, 56]) της θεώρησης μίας ξεχωριστής κλίμακας για κάθε κλάση (έτσι ώστε να διαφέρει κάθε κλάση με την επόμενη της), η οποία δεν αποτελεί ρεαλιστική μοντελοποίηση, εδώ θεωρούμε δύο συνολικά κλίμακες.

Όπως δηλώνει η (2.7), σε περιβάλλον διαχωρισμού κλάσεων όπως αυτό που μόλις εισήχθη, τα κέρδη διαφορισμού πολλαπλών χρηστών για τον NBR παραμένουν ανεπιπρόσπαστα, επειδή μόνο η κανονικοποιημένη τιμή του επιτευξιμού ρυθμού μετάδοσης λαμβάνεται υπόψη κατά το χρονοπρογραμματισμό. Από την άλλη πλευρά, σε αυτήν την περίπτωση τα αντίστοιχα κέρδη για τον BR επηρεάζονται σοβαρά. Κατά την εφαρμογή του BR περιμένει κανείς διαπισθητικά ότι όποτε συνυπάρχουν ενεργά τερματικά και από τις δυο ομάδες κλάσεων, τα κέρδη διαφορισμού πολλαπλών χρηστών των ασθενών κλάσεων μειώνονται καθώς μειώνεται το ϵ . Αντίθετα, τα αντίστοιχα κέρδη των ισχυρών κλάσεων αναμένεται να μην είναι ευαίσθητα στην παρουσία τερματικών από την ασθενή ομάδα. Τα συγκεκριμένα διαπισθητικά αποτελέσματα αποδεικνύονται αυστηρά στο Θεώρημα 1 που ακολουθεί. Πριν από την απόδειξη όμως, γίνεται η εισαγωγή ενός βοηθητικού τροποποιημένου συμβολισμού και κάποιων πρόσθετων παραμέτρων.

Το διάνυσμα κατάστασης \mathbf{n} διαχωρίζεται σύμμορφα με το διαχωρισμό των κλάσεων, έτσι ώστε $\mathbf{n} = (\mathbf{n}_s, \mathbf{n}_f)$ και θεωρούμε την κοινή PDF $\hat{F}(x)$ του κανονικοποιημένου SNR. Επιπρόσθετα, ορίζουμε τις ακόλουθες παραμέτρους: $\delta_1(\epsilon) \triangleq (1 - \hat{F}(\frac{d}{\sqrt{\epsilon}}))$, $\delta_2(\epsilon) \triangleq \hat{F}(\bar{D}\sqrt{\epsilon})$, $\delta_3(\epsilon) \triangleq \frac{1}{\sqrt{\epsilon}}(1 - \hat{F}(\frac{1}{\sqrt{\epsilon}})) + \int_{\frac{1}{\sqrt{\epsilon}}}^{\infty} (1 - \hat{F}(u))du$, όπου $d \triangleq \min_{\substack{i \in \mathcal{K}_f \\ m \in \mathcal{K}_s}} \tilde{r}_i / \tilde{r}_m$, $D \triangleq \max_{i, m \in \mathcal{K}_f} \tilde{r}_i / \tilde{r}_m$,

$\bar{D} \triangleq \max\{D, d^{-1}\}$. Είμαστε τώρα σε θέση να αποδείξουμε το κύριο αποτέλεσμα αυτής της ενότητας, η απόδειξη του οποίου βρίσκεται στο παράρτημα.

Θεώρημα 1. Όταν η κοινή PDF της κανονικοποιημένης τιμής του SNR $\hat{F}(x)$ δεν περιέχει μάζα πιθανότητας στο μηδέν και το άπειρο, δηλαδή $\lim_{x \rightarrow 0^+} \hat{F}(x) = 0$ και

$\lim_{x \rightarrow \infty} \hat{F}(x) = 1$, υπό τον αλγόριθμο BR ισχύει ότι:

1. Για όλες τις ισχυρές κλάσεις $i \in \mathcal{K}_f$, $\phi_i(\mathbf{n}_s, \mathbf{n}_f) = \phi_i(\mathbf{0}, \mathbf{n}_f) + h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$, καθώς $\epsilon \rightarrow 0$, όπου $\lim_{\epsilon \rightarrow 0} h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f) = 0$ με $h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f) \leq O(\max\{\delta_1(\epsilon)|\mathbf{n}_s| \phi_i(\mathbf{0}, \mathbf{n}_f), \sqrt{\epsilon} \delta_2(\epsilon)^{|\mathbf{n}_f|}\})$.
2. Για όλες τις ασθενείς κλάσεις $i \in \mathcal{K}_s$:
 - $\phi_i(\mathbf{n}_s, \mathbf{0}) = O(1)$, $\forall \epsilon > 0$.
 - Όταν $\mathbf{n}_f > \mathbf{0}$, $\phi_i(\mathbf{n}_s, \mathbf{n}_f) = h^{(s)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ καθώς $\epsilon \rightarrow 0$, με $h^{(s)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f) \leq O(\max\{\delta_2(\epsilon)^{|\mathbf{n}_f|} \phi_i(\mathbf{n}_s, \mathbf{0}), \delta_3(\epsilon)\})$.

Με τον τρόπο αυτό διαθέτουμε μία απλή και διαισθητικά ελκυστική, προσεγγιστική έκφραση για τα κέρδη διαφορισμού με την εισαγωγή συναρτήσεων διαταραχών που τείνουν στο μηδέν καθώς $\epsilon \rightarrow 0$. Σύμφωνα με το Θεώρημα 1, τα κέρδη διαφορισμού στον BR καθορίζονται απόλυτα από την ομάδα των κλάσεων στην οποία αναφέρονται. Συγκεκριμένα, ο BR συμπεριφέρεται ως μία πολιτική που σχεδόν παραμελεί τις ασθενείς κλάσεις όταν υπάρχουν ενεργά τερματικά των ισχυρών κλάσεων και το γεγονός αυτό εισάγεται όχι μόνο από το μέσο ρυθμό των αργών κλάσεων που παραμένει μικρός, αλλά επίσης αντικατοπτρίζεται και στα κέρδη διαφορισμού που μειώνονται μέσω μίας συνάρτησης του ϵ . Επιπρόσθετα το θεώρημα δηλώνει ευθέως ότι τα κέρδη διαφορισμού των ασθενών κλάσεων είναι άξια αναφοράς μόνο κατά την απουσία τερματικών από τις ισχυρές κλάσεις.

Η ακόλουθη πρόταση που αποδεικνύεται στο παράρτημα, μας επιτρέπει τον περαιτέρω χαρακτηρισμό της φθίσης των συναρτήσεων σφάλματος $h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ και $h^{(s)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ της εκφώνησης του Θεωρήματος 1, καθώς $\epsilon \rightarrow 0$.

Πρόταση 1. Αν υπάρχουν θετικές σταθερές a, θ_0 τέτοιες ώστε $\hat{F}(x) = O(x^a)$ καθώς το $x \rightarrow 0$ και $s(\theta) \triangleq \int_0^\infty e^{s\theta} d\hat{F}(x) < +\infty$, τότε:

- Η $h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ είναι το πολύ τάξης $O(\epsilon^{(a|\mathbf{n}_f|+1)/2})$, καθώς το $\epsilon \rightarrow 0$,
- Η $h^{(s)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ είναι το πολύ τάξης $O(\epsilon^{a|\mathbf{n}_f|/2})$, καθώς το $\epsilon \rightarrow 0$.

Όλα τα μοντέλα σκέδασης που έχουμε ήδη αναφέρει διαθέτουν εκθετική φθίση της $1 - \hat{F}(x)$ και έτσι μπορούμε να βρούμε μία θετική σταθερά θ_0 . Για το μοντέλο σκέδασης Rayleigh, η εκθετική μορφή της κατανομής της στιγμιαίας τιμής του SNR, εγγυάται ότι η παράμετρος a που εισήχθη στην Πρόταση 1 είναι ίση με μονάδα, δηλαδή $\hat{F}(x) = O(x)$ καθώς το $x \rightarrow 0$. Σχετικά με το μοντέλο σκέδασης Rice, βλέπουμε από την (2.1) ότι υπεισέρχεται στον υπολογισμό της πυκνότητας πιθανότητας η τροποποιημένη συνάρτηση Bessel πρώτου είδους,

$I_0(\cdot)$. Για σχετικά μικρά x , $I_0(x) \approx 1$ και έτσι η κατανομή της συμπεριφέρεται παρόμοια με αυτή της εκθετικής, καταλήγοντας στο ότι $a \geq 1$. Το ίδιο απλή είναι η περίπτωση για το μοντέλο σκέδασης Nakagami, εφόσον η κατανομή πιθανότητας δίνεται από την ατελή συνάρτηση γάμμα, η οποία διαθέτει άμεσο ανάπτυγμα σε σειρά. Είναι εύκολο να δειχτεί ότι $\hat{F}(x) = O(x^m)$ καθώς το $x \rightarrow 0$, όπου $m \geq 1/2$ είναι η παράμετρος Nakagami. Τέλος, η υπόθεση της μη ύπαρξης μοναδικότητας του θεωρήματος ικανοποιείται επίσης από όλα τα μοντέλα σκέδασης.

Είμαστε πλέον σε θέση να καθορίσουμε τη χειριστη περίπτωση για το ρυθμό φθίσης των συναρτήσεων σφάλματος $h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ και $h^{(s)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$, καθώς $\epsilon \rightarrow 0$. Συγκεκριμένα, εφόσον $|\mathbf{n}_f| \geq 1$, η χειριστη περίπτωση προκύπτει για $|\mathbf{n}_f| = 1$. Έχοντας υπόψη την Πρόταση 1, η $h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ είναι το πολύ τάξης $O(\epsilon^{(a+1)/2})$ και η $h^{(s)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ είναι το πολύ τάξης $O(\epsilon^{a/2})$, καθώς το $\epsilon \rightarrow 0$. Σε συνέχεια της παραπάνω συζήτησης για τα συνηθισμένα μοντέλα σκέδασης, η χειριστη περίπτωση προκύπτει για σκέδαση Nakagami με $m = 1/2$, που απλώς σημαίνει ότι η $h^{(f)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ είναι το πολύ τάξης $O(\epsilon^{3/4})$ και η $h^{(s)}(\epsilon; \mathbf{n}_s, \mathbf{n}_f)$ είναι το πολύ τάξης $O(\epsilon^{1/4})$, καθώς $\epsilon \rightarrow 0$. Οι ιδιότητες των κερδών διαφορισμού θα χρησιμοποιηθούν σε επόμενα κεφάλαια για την ποσοτική αξιολόγηση των αλγορίθμων.

Κλείνοντας το κεφάλαιο αυτό, σημειώνουμε ότι τα έως τώρα χαρακτηριστικά του μοντέλου, σε συνδυασμό με τις υποθέσεις για τη λειτουργία του συστήματος που εισήχθησαν στην Ενότητα 2.2 (για παράδειγμα χρονοσχιστές πολύ μικρότερης διάρκειας σε σχέση με τη συνολική απαίτηση εξυπηρέτησης, χρονοπρογραμματισμός που επαναλαμβάνεται σε κάθε νέα χρονοσχιστική, περιγραφή των ρυθμών μετάδοσης μέσω κερδών διαφορισμού πολλαπλών χρηστών κ.λπ.), θέτουν το πλαίσιο για τη μελέτη του συστήματος ως ένα σύστημα SDPS πολλαπλών κλάσεων. Άλλωστε, αυτού του είδους η μοντελοποίηση έχει εφαρμοστεί για συστήματα ευκαιριακού χρονοπρογραμματισμού στο παρελθόν, όπως για παράδειγμα στα [10, 12, 14, 17, 38]. Το συγκεκριμένο πλαίσιο επιτρέπει την εξαγωγή αποτελεσμάτων που αφορούν μέσες μετρικές επίδοσης (π.χ., μέσο χρόνο απόκρισης, πιθανότητα αποκλεισμού κ.λπ.) αλλά και αποτελεσμάτων για την κατανομή του χρόνου απόκρισης ενός τερματικού. Επιπλέον, η μοντελοποίηση του ασύρματου δικτύου ως ένα σύστημα SDPS προσφέρει τη δυνατότητα να μελετηθούν σε μεγαλύτερο βάθος φαινόμενα διαχωρισμού χρονικής κλίμακας, υποβοηθούμενη από τη χρήση των ειδικότερων αποτελεσμάτων της παρούσας ενότητας.

Λόγω της περιπλοκότητας που παρουσιάζει στη γενική του μορφή το πρόβλημα ενός συστήματος αναμονής με πολιτική διαμοιρασμού πόρων SDPS, απ' όσο είμαστε σε θέση να γνωρίζουμε δεν έχει μελετηθεί πλήρως στη σχετική βι-

βλιογραφία, παρά μόνο για εξειδικευμένες περιπτώσεις που αναφέρονται στη σχετική επισκόπηση που δίνεται στο επόμενο κεφάλαιο. Στην παρούσα διατριβή θεωρούμε ένα μοντέλο SDPS το οποίο περιγράφεται και επιλύεται με γνώμονα την εξαγωγή αποτελεσμάτων εφαρμόσιμων σε ασύρματα δίκτυα που εφαρμόζουν ευκαιριακό χρονοπρογραμματισμό. Στα Κεφάλαια 3-7 που ακολουθούν αναλύεται σε μαθηματικό επίπεδο ένα σύστημα SDPS και στο Κεφάλαιο 8 επανερχόμαστε στο παρόν πρόβλημα για την κατάλληλη εφαρμογή των αποτελεσμάτων.



Κεφάλαιο 3

Επισκόπηση Μοντέλου PS και Ορισμοί Απαιτούμενων Μεγεθών

ΤΟ όφελος που προκύπτει από τη χρήση ευκαιριακού χρονοπρογραμματισμού αντικατοπτρίζεται στο κέρδος διαφορισμού πολλαπλών χρηστών, δηλαδή σε ένα πολλαπλασιαστικό παράγοντα του μέσου ρυθμού μετάδοσης επειδή ο αλγόριθμος εξυπηρέτησης ευνοεί τα τερματικά με ισχυρότερο σήμα (βλ. π.χ., [52]). Η συγκεκριμένη ιδιότητα, όπως άλλωστε ποσοτικοποιήθηκε στην Ενότητα 2.3, σχετίζεται άμεσα με τον αριθμό ενεργών τερματικών κάθε κλάσης που υπάρχουν στο δίκτυο. Η επίδοση συστημάτων των οποίων ο ρυθμός εξυπηρέτησης είναι εξαρτημένος από την κατάσταση (δηλ. τον αριθμό παρόντων «πελατών») μπορεί να περιγραφεί με σαφήνεια και με μαθηματικά κομψό τρόπο μέσω ενός κατάλληλου μοντέλου PS. Ακολουθώντας τη συγκεκριμένη μέθοδο, το [14] μελετά την επίδοση σε επίπεδο χρήστη του αλγορίθμου NBR χρησιμοποιώντας τη μοντελοποίηση SDPS, με την εξάρτηση στην κατάσταση να οφείλεται ακριβώς στα κέρδη διαφορισμού πολλαπλών χρηστών. Αρχικά παραθέτει τη σύνδεση του προβλήματος με μία ουρά SDPS και στη συνέχεια χρησιμοποιεί έτοιμα αποτελέσματα από τη θεωρία αναμονής για να ληφθούν μέσα μεγέθη επίδοσης.

Αυτός ο τύπος μοντελοποίησης έχει εφαρμοστεί επίσης και για τη μελέτη του αλγορίθμου PF στο 1xEV-DO (βλ. π.χ., στο [17] για αποτελέσματα που αφορούν τη μέση ρυθμαπόδοση, τη μέση καθυστέρηση και την πιθανότητα αποκλεισμού). Με παρόμοια ανάλυση, το [12] στοχεύει παρόμοια αποτελέσματα για το HDR που σχετίζονται με τις διαστάσεις της κυψέλης και την αποτίμηση της κίνησης. Μία λίγο διαφορετική μέθοδος προτείνεται στο [38], στο οποίο γίνεται χρήση του μοντέλου PS ώστε να μελετηθεί το πρόβλημα του ευκαιριακού χρονοπρογραμματισμού του HDR. Το πρόβλημα εκεί συνδέεται με τον πε-

ριορισμό δικαιοσύνης μικρού χρόνου (αγ.: short term fairness), δηλαδή κάθε τερματικό πρέπει να λαμβάνει τουλάχιστον ένα προκαθορισμένο αριθμό χρόνων.

Από τα παραπάνω γίνεται φανερό ότι το μοντέλο PS και οι εξειδικεύσεις του αποτελούν ένα χρήσιμο εργαλείο προκειμένου να δοθεί μία πλήρης και λεπτομερής μαθηματική περιγραφή που αφορά συστήματα ευκαιριακού χρονοπρογραμματισμού. Στα πλαίσια της παρούσας διατριβής κρίθηκε σκόπιμη η γενικότερη μελέτη και η επίλυση του συγκεκριμένου μαθηματικού προβλήματος και κατόπιν η εφαρμογή των αποτελεσμάτων για την περίπτωση του τηλεπικοινωνιακού δικτύου, που αναφέρθηκε στο Κεφάλαιο 2. Με το σκεπτικό αυτό, το παρόν και τα επόμενα τέσσερα κεφάλαια αφιερώνονται ακριβώς στην τοποθέτηση και επίλυση του γενικότερου προβλήματος. Χάριν γενικότητας, στο μαθηματικό μοντέλο που θα αναπτυχθεί στη συνέχεια δανειζόμαστε έννοιες της θεωρίας αναμονής (π.χ., πελάτης αντί κινητού τερματικού, εξυπηρετητής αντί σταθμού βάσης κ.ο.κ.).

3.1 Βιβλιογραφική επισκόπηση μοντέλων PS

Το μοντέλο PS είναι χρήσιμο κατά τη μοντελοποίηση δομών που επιβάλλουν το διαμοιρασμό πόρων. Αρχικά ξεκίνησε σαν μία ειδική περίπτωση του αλγορίθμου κυκλικής εναλλαγής σε υπολογιστικά συστήματα διαμοιρασμού χρόνου [36]. Πιο πρόσφατα το μοντέλο και οι προεκτάσεις του προσέλκυσαν το ενδιαφέρον σε διάφορους χώρους, όπως για παράδειγμα στην επίδοση ροών σε πρωτόκολλα διαμοιρασμού εύρους ζώνης σε δίκτυα μεταγωγής πακέτων (βλ. π.χ., [48] και τις εκεί αναφορές) καθώς και στην επίδοση αλγορίθμων χρονοπρογραμματισμού που στηρίζονται στα χαρακτηριστικά του καναλιού, σε ασύρματα δίκτυα δεδομένων (βλ. π.χ., [14]).

Στην αρχική του μορφή, η οποία αναφέρεται και ισόνομο PS (αγ.: Egalitarian Processor Sharing (EPS)), το μοντέλο δίνει ίση πρόσβαση στον εξυπηρετητή σε όλους τους χρήστες, έτσι όταν υπάρχουν n πελάτες στο σύστημα, ο καθένας εξυπηρετείται με ρυθμό $1/n$. Η έννοια αυτή μπορεί να γενικευθεί ώστε να συμπεριλάβει πολλαπλές κλάσεις πελατών και αυθαίρετους ρυθμούς εξυπηρέτησης που εξαρτώνται από την κατάσταση, πιθανά διαφορετικούς για κάθε κλάση. Το γενικευμένο μοντέλο που αποκαλείται SDPS, θεωρεί καταστάσεις του συστήματος της μορφής $\mathbf{n} = (n_1, \dots, n_k)$, συμβολίζοντας την παρουσία n_i πελατών από την κλάση i και συναρτήσεις ρυθμού εξυπηρέτησης $\phi_i(\cdot)$, όπου $i = 1, \dots, k$. Όταν το σύστημα είναι στην κατάσταση \mathbf{n} , κάθε πελάτης της κλάσης i λαμβάνει εξυπηρέτηση με ρυθμό $\phi_i(\mathbf{n})$ και ο συνολικός ρυθμός εξυπηρέτησης ισούται με $\sum_{i=1}^k n_i \phi_i(\mathbf{n})$. Έτσι, εκτός από ασύμμετρο διαμοιρασμό του

εξυπηρετητή, το μοντέλο SDPS μπορεί να εκφράσει μεταβολές που εξαρτώνται από την κατάσταση του συνολικού ρυθμού εξυπηρέτησης.

Μία ειδικότερη μορφή του SDPS στην οποία οι πελάτες λαμβάνουν ίσο μερίδιο στον εξυπηρετητή και οι συνολικοί ρυθμοί εξυπηρέτησης εξαρτώνται μόνο από το συνολικό αριθμό παρόντων πελατών στο σύστημα, δηλαδή όταν οι ρυθμοί εξυπηρέτησης έχουν τη μορφή $\phi_i(\mathbf{n}) = g(\sum_{j=1}^k n_j)$ για όλες τις κλάσεις i και κάποια κατάλληλη συνάρτηση $g(\cdot)$, μελετήθηκε στο [25] (αποκαλούμενη εκεί «Γενικευμένος Μερεισμός Εξυπηρετητή» (αγ.: Generalized Processor Sharing (GPS))). Η συγκεκριμένη παραλλαγή διαθέτει ιδιότητες και ως ένα σημείο δυνατότητες εύρεσης αποτελεσμάτων συγκρίσιμες με το EPS. Μάλιστα, το μοντέλο αντιστοιχεί σε EPS όταν $g(x) = 1/x$. Μία άλλη ειδική περίπτωση του SDPS είναι ο Διακρίνων Μερεισμός Εξυπηρετητή (Discriminatory Processor Sharing (DPS)), σύμφωνα με τον οποίο ο (σταθερός) ρυθμός εξυπηρέτησης διαμοιράζεται στους πελάτες ανάλογα με κάποια προκαθορισμένα βάρη που εξαρτώνται από την κλάση $w_i, i = 1, \dots, k$. Σε αυτή την περίπτωση, οι συναρτήσεις του ρυθμού εξυπηρέτησης παίρνουν τη μορφή $\phi_i(\mathbf{n}) = w_i / \sum_{j=1}^k n_j w_j$. Όταν όλα τα βάρη είναι ίσα, το DPS μετατρέπεται σε EPS. Για ανασκόπηση μοντέλων DPS και σύγκριση με άλλα παρόμοια μοντέλα διαμοιρασμού πόρων αναφέρουμε το [1].

Σε συστήματα SDPS με αφίξεις Poisson, οι πιθανότητες σταθερής κατάστασης (και μέσω του νόμου του Little ο μέσος χρόνος παραμονής) εξαρτώνται από την κατανομή των απαιτήσεων εξυπηρέτησης μόνο μέσω της αντίστοιχης μέσης τιμής, αν και μόνο αν οι συναρτήσεις ρυθμών εξυπηρέτησης $\phi_i(\cdot)$ διαθέτουν μία, όπως αποκαλείται, ιδιότητα «ισορροπίας» (βλ. π.χ., [11]). Ωστόσο ακόμα και σε αυτή την περίπτωση, η αποκλειστική εξάρτηση από τη μέση τιμή δε διαδίδεται στη συνολική κατανομή του χρόνου εξυπηρέτησης, η οποία εξαρτάται από ιδιότητες ανώτερης τάξης των απαιτήσεων εξυπηρέτησης. Εν γένει, ο υπολογισμός της κατανομής του χρόνου απόκρισης είναι ένα πολύπλοκο πρόβλημα με αρκετά σημεία προς διερεύνηση.

Για αφίξεις Poisson και απαιτήσεις εξυπηρέτησης εκθετικά κατανομημένες (δηλ. M/M/1-EPS), ο μετασχηματισμός Laplace-Stieltjes της κατανομής του χρόνου απόκρισης (αγ.: Laplace-Stieltjes Transform (LST)) για ένα σύστημα EPS με μία κλάση, έχει παραχθεί στο [24]. Αντίστοιχα αποτελέσματα που εκτείνονται σε γενικά κατανομημένες απαιτήσεις εξυπηρέτησης παρήχθησαν αργότερα στα [45, 50, 61]. Η περίπτωση του M/G/1-EPS, μπορεί να αντιμετωπιστεί εναλλακτικά ως η οριακή περίπτωση ενός συστήματος M/M/1-FCFS με ανάδραση, όταν οι πιθανότητες ανάδρασης τείνουν στο μηδέν [27]. Για μία ανασκόπηση των μεθοδολογιών σχετικά με αυτά τα αποτελέσματα, αναφερόμαστε στο [62].

Η μορφή των μετασχηματισμών LST που μόλις αναφέρθηκε είναι αρκετά περίπλοκη και κατά συνέπεια, η απευθείας αντιστροφή τους ώστε να βρεθεί η πλήρης κατανομή του χρόνου απόκρισης είναι δύσκολη. Αυτός είναι ένας από του λόγους για τους οποίους αρκετές μελέτες εστιάζουν στον ασυμπτωτικό χαρακτηρισμό ουράς (αγ.: tail) της κατανομής του χρόνου απόκρισης. Τέτοιου είδους χαρακτηρισμοί παίρνουν απλούστερη μορφή όταν οι απαιτήσεις εξυπηρέτησης είναι βαριάς ουράς (αγ.: heavy tailed). Για αρκετές τέτοιες κατανομές (συμπεριλαμβανομένων των ομαλά μεταβαλλόμενων (αγ.: regularly varying) και των τετραγωνικής-ρίζας-αναλλοίωτων υποεκθετικών-κοίλων κατανομών (αγ.: square-root-invariant subexponential-concave distributions), έχει αποδειχθεί ότι η ουρά της κατανομής του χρόνου απόκρισης είναι ασυμπτωτικά ισοδύναμη προς μία κατάλληλα τροποποιημένη έκδοση της ουράς της κατανομής των απαιτήσεων εξυπηρέτησης. Αυτή η ισοδυναμία, που συχνά αποκαλείται προσέγγιση μειωμένου ρυθμού εξυπηρέτησης (αγ.: reduced service rate approximation), μπορεί να εφαρμοστεί επίσης σε περιπτώσεις πιο γενικές από το απλό PS, και κυρίως σε συστήματα DPS με πολλαπλές κλάσεις. Στο [18] υπάρχει μία ανασκόπηση ασυμπτωτικών αποτελεσμάτων σχετικά με την ουρά της κατανομής του χρόνου απόκρισης, για απαιτήσεις εξυπηρέτησης βαριάς ουράς.

Η κατάσταση είναι διαφορετική όταν η κατανομή των απαιτήσεων εξυπηρέτησης είναι ελαφριάς ουράς, (αγ.: light-tailed). Ο χαρακτηρισμός ουράς για το χρόνο απόκρισης του M/M/1-EPS είναι ασυμπτωτικά ισοδύναμος με ένα γινόμενο τριών παραγόντων με αντίστροφη δύναμη, Weibull και εκθετική μείωση αντίστοιχα [16], ενώ το σύστημα M/D/1-PS χαρακτηρίζεται από κατανομή του χρόνου απόκρισης με καθαρά εκθετικά μειούμενη κατανομή ουράς [29]. Μία άλλη προσπάθεια στοχεύει σε λογαριθμικά ασυμπτωτικά αποτελέσματα. Δεδομένων των απαιτήσεων εξυπηρέτησης ελαφριάς ουράς, (εκτός από «πολύ ελαφριές» ουρές, όπως π.χ., απαιτήσεις εξυπηρέτησης με φραγμένη υποστήριξη (αγ.: supported bound) και ντετερμινιστικές) είναι δυνατό να υπολογιστεί ο ρυθμός εκθετικής μείωσης των πιθανοτήτων ουράς του χρόνου απόκρισης για συστήματα GI/G/1-ME μέσω των ροπογεννητριών συναρτήσεων του χρόνου διαδοχικών αφίξεων και των απαιτήσεων εξυπηρέτησης [42]. Η συγκεκριμένη μεθοδολογία μπορεί να επεκταθεί ώστε να συμπεριλάβει εξυπηρετητές με στοχαστικά μεταβλητό ρυθμό (ανεξάρτητα από το φορτίο που εισέρχεται στο σύστημα) [28].

Οι προαναφερθείσες εργασίες για την κατανομή του χρόνου απόκρισης σχετίζονται με μοντέλα PS που περιλαμβάνουν μία κλάση και απεριόριστο θάλαμο εξυπηρέτησης. Μοντέλα SDPS με μία μόνο κλάση πελατών, απαιτήσεις εξυπηρέτησης εκθετικά κατανομημένες και πεπερασμένο θάλαμο εξυπηρέτησης μελετώνται στο [15]. Η ανάλυση αποσκοπεί στο μετασχηματισμό LST του

χρόνου απόκρισης, δεδομένου του αριθμού πελατών που βρίσκονται στο σύστημα κατά την άφιξη ενός πελάτη. Οι συγκεκριμένοι μετασχηματισμοί LST μπορούν να βρεθούν επιλύοντας ένα κατάλληλο γραμμικό σύστημα και έπειτα, προκειμένου να βρεθεί ο μετασχηματισμός που δεν αντιστοιχεί στον δεσμευμένο χρόνο, πολλαπλασιάζονται με κατάλληλα βάρη οπότε και λαμβάνει χώρα η τελική αντιστροφή. Η συγκεκριμένη μεθοδολογία γενικεύεται στο [19] για περισσότερες κλάσεις. (Αν και το [19] εστιάζει σε συστήματα DPS, η μεθοδολογία εκεί μπορεί να εφαρμοστεί και σε συστήματα SDPS.) Ο κύριος περιορισμός της μεθόδου των [15] και [19] βρίσκεται στον αριθμητικό τους χαρακτήρα. Ειδικότερα, η εύρεση των μετασχηματισμών LST και η ανάγκη για αντιστροφή τους επισκιάζει την πιθανοθεωρητική δομή του προβλήματος.

Κατά την ανάλυση που θα ακολουθήσουμε στη διατριβή, η κατανομή του χρόνου απόκρισης κάποιου πελάτη οποιασδήποτε κλάσης προσδιορίζεται θεωρώντας ένα τροποποιημένο σύστημα που περιλαμβάνει ένα μόνιμο πελάτη της ίδιας κλάσης και υπολογίζοντας την πιθανότητα με την οποία η συνολική εξυπηρέτηση που λαμβάνεται από το μόνιμο πελάτη μέσα σε κάποιο χρόνο, υπερβαίνει την απαίτηση εξυπηρέτησης ενός αρχικού μη μόνιμου πελάτη. Η έννοια των μόνιμων πελατών έχει χρησιμοποιηθεί επίσης και από άλλες δουλειές που μελετούν διάφορες ιδιότητες των μοντέλων PS, όπως για παράδειγμα στα [18, 22, 23, 27, 44], από τα οποία το [18] είναι το κοντινότερο στον τρόπο με τον οποίο χρησιμοποιείται εδώ. Επίσης, στο [44] χρησιμοποιείται η διαδικασία ανταμοιβής Markov (αγγ.: Markov Reward process) αλλά η έννοια δίνεται με διαφορετικό τρόπο. Συγκεκριμένα, μελετώνται δεσμευμένες μέσες τιμές του χρόνου απόκρισης μοντέλων PS με μία κλάση των οποίων ο εξυπηρετητής σχετίζεται με ένα τυχαίο μαρκοβιανό περιβάλλον.

Η έκφραση της κατανομής του χρόνου απόκρισης σε όρους του εκθετικού ενός πίνακα προδίδει ξεκάθαρα τη μαρκοβιανή δομή του μοντέλου και ως εκ τούτου, είναι ιδιαίτερα κατάλληλη για να μελετηθούν φαινόμενα που προκύπτουν κατά το διαχωρισμό χρονικής κλίμακας. Αυτή η κατάσταση προκύπτει όταν κάποιες κλάσεις πελατών έχουν χαρακτηριστικά αφίξεων και απαιτήσεων εξυπηρέτησης με δυναμική πολύ γρηγορότερη από εκείνη άλλων κλάσεων, όπως το περιβάλλον διαχωρισμού κλάσεων που εισήχθη στην Ενότητα 2.4. Λόγω της σπουδαιότητας της για τη μελέτη της εσωτερικής δομής του συστήματος, αλλά και επειδή οδηγεί σε απλουστευμένες εκφράσεις, η συγκεκριμένη κατάσταση έχει ξαναμελετηθεί σε προηγούμενες εργασίες που εξετάζουν παραλλαγές του μοντέλου PS, συμπεριλαμβανομένης της [13] για γενικά μοντέλα SDPS και της [56] για την εξειδικευμένη περίπτωση του DPS. Οι παραπάνω μελέτες αναφέρονται μόνο σε πιθανότητες σταθερής κατάστασης. Επιπρόσθετα, εφαρμόζουν την απλουστευτική και κάπως μη ρεαλιστική παραδοχή ότι κάθε

κλάση λειτουργεί σε διαφορετική χρονική κλίμακα, ώστε οι κλάσεις να κατατάσσονται σύμφωνα με την ταχύτητα εξυπηρέτησής τους. Κάθε κλάση εμφανίζεται ως «ρευστό» προς όλες τις αργότερες και ως «οιονεί στάσιμη» προς όλες τις γρηγορότερες κλάσεις.

Σχετικά με την κατανομή του χρόνου απόκρισης, η μόνη σχετική εργασία (καθόσον είμαστε σε θέση να γνωρίζουμε) είναι η [19], η οποία εφαρμόζει τα αποτελέσματα της [56] ώστε να παράγει απλούστερες εκφράσεις για το δεσμευμένο μετασχηματισμό LST της κατανομής του χρόνου απόκρισης, στοχεύοντας στο να μειώσει την υπολογιστική πολυπλοκότητα. Η μεθοδολογία της [19] είναι δύσκολο να γενικευθεί πέρα από την ειδική περίπτωση του DPS η οποία διερευνάται εκεί, επειδή βασίζεται σε κρίσιμο βαθμό στην ύπαρξη διαφορετικής χρονικής κλίμακας για κάθε κλάση και στην εκ των προτέρων υπόθεση (που χρησιμοποιείται επίσης στο [56]) ότι, παρουσία του συγκεκριμένου διαχωρισμού, τα συστήματα DPS αποσυντίθενται σε συστήματα PS μίας κλάσης (με ένα αριθμό πρόσθετων μόνιμων πελατών).

3.2 Μεθοδολογία επίλυσης

Το πρόβλημα που θεωρούμε εδώ και θα αναλύσουμε περαιτέρω στα Κεφάλαια 4-7, αποτελείται από ένα σύστημα SDPS με πελάτες χωρισμένους σε κλάσεις ανάλογα με το μέσο ρυθμό εξυπηρέτησης, πεπερασμένο θάλαμο εξυπηρέτησης τυχαίου «σχήματος» (το σχήμα του θαλάμου εξυπηρέτησης αναφέρεται στην συνθήκη αποδοχής νέου πελάτη κάθε κλάσης) και εκθετικά κατανεμημένες απαιτήσεις εξυπηρέτησης με μέση τιμή που εξαρτάται από την κλάση. Οι πελάτες εισέρχονται στο σύστημα σύμφωνα με μία διαδικασία Poisson που εξαρτάται από την κλάση αλλά πιθανά και από την κατάσταση του συστήματος, γενικεύοντας έτσι το ομοιογενές χρονικά πρότυπο αφίξεων Poisson. Η σύνδεση της συνολικής εξυπηρέτησης ενός μόνιμου πελάτη με μία διαδικασία ανταμοιβής Markov επιτρέπει την απευθείας έκφραση της κατανομής του χρόνου απόκρισης με τη βοήθεια του εκθετικού ενός πίνακα, αποφεύγοντας την αντιστροφή των μετασχηματισμών LST.

Παράλληλα, στο πλαίσιο της διατριβής αναλύουμε την ασυμπτωτική συμπεριφορά της κατανομής του χρόνου απόκρισης, συνδέοντάς τη με την επικρατούσα ιδιοτιμή του πίνακα που εμπεριέχεται στην αντίστοιχη έκφραση. Η αντιστοίχιση με τη διαδικασία ανταμοιβής Markov βοηθά στην εξαγωγή περαιτέρω αποτελεσμάτων, συμπεριλαμβανομένων ευκόλως υπολογίσιμων άνω και κάτω φραγμάτων για την ασυμπτωτική εκθετική φθίση της ουράς της κατανομής. Επιπρόσθετα, όπως θα αποδειχθεί, όταν οι ρυθμοί εξυπηρέτησης χαρακτηρίζονται από την ιδιότητα ισορροπίας, το εκθετικό του πίνακα μπορεί

να διαγωνιοποιηθεί, παρέχοντας περαιτέρω αποτελέσματα σε κλειστή μορφή.

Μετά τον υπολογισμό της κατανομής του χρόνου απόκρισης και των ιδιοτήτων που σχετίζονται με αυτή, στο Κεφάλαιο 5 αναλύουμε φαινόμενα διαχωρισμού της χρονικής κλίμακας με γενικότερους όρους. Συγκεκριμένα, αντί να θεωρήσουμε αυστηρό διαχωρισμό κλίμακας μεταξύ δύο οποιονδήποτε κλάσεων, θεωρούμε την πιο ρεαλιστική προσέγγιση του διαχωρισμού των κλάσεων σε δύο ξεχωριστά υποσύνολα, τέτοια ώστε όλες οι κλάσεις ενός υποσυνόλου να έχουν δυναμική (χαρακτηριστικά κίνησης) ίδιας τάξης μεγέθους και έτσι ώστε, η δυναμική των κλάσεων της γρήγορης ομάδας να είναι πολύ μεγαλύτερη αυτής της αργής ομάδας. (Προφανώς η κατασκευή αυτή μπορεί να γενικευθεί εισάγοντας περισσότερα από δύο υποσύνολα, με αυστηρή διαφοροποίηση χρονικής κλίμακας μεταξύ τους και δυναμική ίδιας τάξης μεγέθους για όλες τις κλάσεις που ανήκουν στην ίδια ομάδα.)

Λόγω του διαχωρισμού χρονικής κλίμακας, το σύστημα διαθέτει δομή NCD [26], η οποία εμφανίζεται στην αλυσίδα Markov που υπόκειται της έκφρασης για την κατανομή του χρόνου απόκρισης. Έτσι γίνεται δυνατό να εκφραστεί η συγκεκριμένη κατανομή μέσω απλουστευμένων μοντέλων μειωμένης πολυπλοκότητας. Ειδικότερα, αποδεικνύεται αυστηρά ότι η κατανομή του χρόνου απόκρισης για πελάτες οποιασδήποτε γρήγορης κλάσης είναι περίπου ίση προς ένα ζυγισμένο άθροισμα των κατανομών των χρόνων απόκρισης (για πελάτες της ίδιας κλάσης) που αντιστοιχούν σε μειωμένα συστήματα. Καθένα από αυτά τα συστήματα εξυπηρετεί πελάτες μόνο από τις γρήγορες κλάσεις, με ρυθμούς αφίξεων και εξυπηρέτησης προσαρμοσμένα ώστε να αντιστοιχούν στην μόνιμη παραμονή ενός σταθερού συνδυασμού πελατών από τις αργές κλάσεις. Με άλλα λόγια, οι κλάσεις του αργού συνόλου εμφανίζονται από κοινού ως οιονεί στάσιμες προς εκείνες του γρήγορου συνόλου.

Σε αναλογία, η κατανομή του χρόνου απόκρισης για πελάτες οποιασδήποτε αργής κλάσης αποδεικνύεται ότι είναι περίπου ίση με την κατανομή του χρόνου απόκρισης ενός μικρότερου, μειωμένου συστήματος. Το σύστημα αυτό φαίνεται να εξυπηρετεί μόνο πελάτες των αργών κλάσεων και διαθέτει τροποποιημένους ρυθμούς αφίξεων και εξυπηρέτησης που λαμβάνονται από τους αντίστοιχους του αρχικού μοντέλου αφού σταθμιστούν ως προς τους πελάτες των γρήγορων κλάσεων. Οι κλάσεις του γρήγορου συνόλου εμφανίζονται συνολικά ως ρευστό στις κλάσεις του αργού συνόλου. Η ανάλυση υπολογίζει το βαθμό της προσέγγισης, δείχνοντας ότι είναι κάποια συνάρτηση τάξης $O(\epsilon)$ (όπου ο παράγων σύζευξης ϵ είναι η μικρή παράμετρος που ποσοτικοποιεί το διαχωρισμό κλίμακας, όπως ακριβώς εισήχθη στην Ενότητα 2.4).

Εφόσον η κατανομές του χρόνου απόκρισης των πελατών των γρήγορων κλάσεων είναι προσεγγιστικά ίσες με ένα ζυγισμένο άθροισμα των κατανομών

που αντιστοιχούν στα μειωμένα υποσυστήματα, το συγκεκριμένο υποσύστημα με τη βαρύτερη ουρά κυριαρχεί στην ασυμπτωτική συμπεριφορά της συνολικής κατανομής. Εξερευνώντας την πιθανοθεωρητική δομή του συστήματος, αποδεικνύεται ότι όταν οι συναρτήσεις του ρυθμού εξυπηρέτησης υπακούουν σε μία ιδιότητα μονοτονίας (που ικανοποιείται πρακτικά σε όλα τα ρεαλιστικά συστήματα), το υποσύστημα με τη βαρύτερη ουρά συνδέεται με το μέγιστο δυνατό πληθυσμό πελατών των αργών κλάσεων που επιτρέπεται από τη μορφή του χώρου καταστάσεων. Το συγκεκριμένο αποτέλεσμα είναι χρήσιμο για να μειωθεί η υπολογιστική πολυπλοκότητα του καθορισμού της συμπεριφοράς της ουράς της κατανομής του χρόνου απόκρισης (επειδή μόνο ένα υποσύστημα πρέπει να μελετηθεί, γνωστό εκ των προτέρων). Το αποτέλεσμα είναι επίσης χρήσιμο επειδή καταλήγει στο ότι όταν υπάρχει ενδιαφέρον για την ουρά της κατανομής του χρόνου απόκρισης, ο χώρος καταστάσεων (θάλαμος εξυπηρέτησης) θα πρέπει να διαθέτει άνω φράγμα στον πληθυσμό των αργών πελατών, σε αντίθεση με τη χρήση ενός κοινού χώρου για όλες τις κλάσεις.

Επίσης, στο Κεφάλαιο 6 θεωρούμε μία ειδική παραλλαγή του περιβάλλοντος διαφοροποίησης της χρονική κλίμακας, κατά την οποία οι συναρτήσεις του ρυθμού εξυπηρέτησης διαθέτουν και μία «τροποποίηση υπέρ των γρήγορων κλάσεων» (αγ.: fast dynamics bias). Πιο συγκεκριμένα σε αυτή την περίπτωση οι αργές κλάσεις διαθέτουν όχι μόνο αργά χαρακτηριστικά κίνησης αλλά και αμελητέα μικρούς ρυθμούς εξυπηρέτησης όταν υπάρχουν πελάτες των γρήγορων κλάσεων. Εν αντιθέσει, οι ρυθμοί εξυπηρέτησης των γρήγορων κλάσεων παραμένουν σχεδόν ανεπηρέαστοι από την παρουσία αργών πελατών. Ο φορμαλισμός αυτός, πέρα από την άμεση σχέση με τις εκφράσεις των κερδών διαφορισμού του Θεωρήματος 1, είναι χρήσιμος για τη μελέτη αρκετών πρακτικών προβλημάτων, για παράδειγμα συστήματα DPS με βάρη κατάλληλα ώστε να εξυπηρετούν μείγμα αργών και γρήγορων ροών στο διαδικτυο [56].

Η ανάλυση αποδεικνύει ότι η κατανομή του χρόνου απόκρισης των γρήγορων πελατών παραμένει σχεδόν ανεπηρέαστη από την παρουσία αργών πελατών (εκτός από επιδράσεις λόγω της μορφής του χώρου καταστάσεων). Κατά συνέπεια, η κατανομή είναι προσεγγιστικά ίση με την κατανομή που αντιστοιχεί σε ένα σύστημα που περιέχει μόνο τους γρήγορους πελάτες. Εφόσον οι αργοί πελάτες έχουν πρόσβαση στον εξυπηρετητή μόνο όταν απουσιάζουν οι γρήγοροι πελάτες, η κατανομή του χρόνου απόκρισής τους είναι προσεγγιστικά ίση με την κατανομή ενός συστήματος που περιλαμβάνει μόνο τους αργούς πελάτες και ρυθμό εξυπηρέτησης ελαττωμένο κατά ένα παράγοντα ίσο με την πιθανότητα κενής κατάστασης του συστήματος που περιέχει τους γρήγορους χρήστες. Και σε αυτή την περίπτωση, το συγκεκριμένο γεγονός μπορεί να χρη-

οιμοποιηθεί για να ληφθούν αποτελέσματα μειωμένης πολυπλοκότητας.

3.3 Ορισμοί και μαρκοβιανή δομή του μοντέλου

Όπως έχει ήδη αναφερθεί, θεωρούμε ένα σύστημα PS με διαφοροποίηση κλάσεων, αφίξεις Poisson που εξαρτώνται από την κλάση και την κατάσταση του συστήματος, ρυθμούς εξυπηρέτησης που εξαρτώνται επίσης από την κλάση και την κατάσταση του συστήματος καθώς και πεπερασμένο μέγεθος του θαλάμου εξυπηρέτησης. Όταν υπάρχει ελεύθερη θέση στο θάλαμο εξυπηρέτησης, οι καινούριοι πελάτες γίνονται αποδεκτοί προς εξυπηρέτηση, διαφορετικά απορρίπτονται. Σε αντιστοιχία με τους ορισμούς της Ενότητας 2.2, οι πελάτες χωρίζονται σε ένα πεπερασμένο αριθμό κλάσεων, k . Το σύνολο των κλάσεων θα συμβολίζεται με $\mathcal{K} = \{1, \dots, k\}$ και η κατάσταση του συστήματος με ένα διάνυσμα της μορφής $\mathbf{n} = (n_1, \dots, n_k)$, όπου $n_i, i = 1, \dots, k$, είναι ο αριθμός των πελατών της κλάσης i που είναι παρόντες στο σύστημα.

Όταν το σύστημα είναι στην κατάσταση \mathbf{n} , οι πελάτες της κλάσης i αφιχνούνται σύμφωνα με μία διαδικασία Poisson με ρυθμό $\lambda_i(\mathbf{n}), i = 1, \dots, k$. Το μοντέλο της διαδικασίας αφίξεων που θεωρούμε εδώ, αποτελεί μία γενίκευση των απλών αφίξεων Poisson (που εξαρτώνται ωστόσο από την κλάση). Αν κάποιο αποτέλεσμα ισχύει αποκλειστικά για ομοιογενείς αφίξεις Poisson (δηλαδή με ρυθμό αφίξεων ο οποίος δεν εξαρτάται από την κατάσταση του συστήματος), το γεγονός αυτό θα αναφέρεται ρητά. Οι πελάτες που εισέρχονται στο σύστημα διαθέτουν απαιτήσεις εξυπηρέτησης εκθετικά κατανομημένες (εκπεφρασμένες σε μονάδες χρόνου), με μέση τιμή $1/\mu_i, i = 1, \dots, k$.

Ο εξυπηρετητής μοιράζεται μεταξύ των πελατών σύμφωνα με τον ακόλουθο τρόπο. Για κάθε κλάση $i, i = 1, \dots, k$, ορίζουμε τη συνάρτηση ρυθμών εξυπηρέτησης $\phi_i(\cdot), \mathbb{N}^k \rightarrow \mathbb{R}^+$ (σε επόμενο κεφάλαιο η $\phi_i(\cdot)$ θα ταυτοποιηθεί με το κέρδος διαφορισμού πολλαπλών χρηστών της (2.5)). Έτσι, όποτε το σύστημα βρίσκεται στην κατάσταση \mathbf{n} (με $n_i > 0$), κάθε πελάτης κλάσης i που είναι παρών εξυπηρετείται με ρυθμό $\phi_i(\mathbf{n})$. Στο «σύννηθες» EPS οι ρυθμοί εξυπηρέτησης έχουν τη μορφή $\phi_i(\mathbf{n}) = 1/|\mathbf{n}|$, για όλα τα $i \in \mathcal{K}$. Δεδομένων των συναρτήσεων $\phi_i(\cdot)$, ο συνολικός ρυθμός εξυπηρέτησης στην κατάσταση \mathbf{n} είναι ίσος με $\sum_{i \in \mathcal{K}} n_i \phi_i(\mathbf{n})$. Οι συναρτήσεις $\phi_i(\cdot)$ μπορούν να είναι γενικής μορφής, με μόνο περιορισμό (και για κάποια από τα αποτελέσματα) να έχουν την ακόλουθη ιδιότητα «μερικής» μονοτονίας:

Ιδιότητα 1. Για κάθε $i = 1, \dots, k$ και όλα τα $\mathbf{n}, \mathbf{n}' \in \mathbb{N}^k$ τέτοια ώστε $\mathbf{n} \leq \mathbf{n}', n_i > 0$, ισχύει ότι $\phi_i(\mathbf{n}) \geq \phi_i(\mathbf{n}')$.

(Η ανισότητα μεταξύ διανυσμάτων ίδιου μεγέθους ορίζεται ως ανισότητα με-

ταξύ όλων των αντίστοιχων στοιχείων.)

Η παραπάνω είναι μία φυσική ιδιότητα που ισχύει σχεδόν για όλα τα συστήματα πρακτικού ενδιαφέροντος και περιγράφει απλώς το γεγονός πως όσο μεγαλύτερος είναι ο αριθμός των πελατών στο σύστημα (ακόμα και από διαφορετικές κλάσεις), τόσο μειώνεται ο ρυθμός εξυπηρέτησης κάθε πελάτη. Ας σημειωθεί ότι η Ιδιότητα 1 δεν αποκλείει την πιθανότητα ο συνολικός ρυθμός εξυπηρέτησης $\sum_{i \in \mathcal{K}} n_i \phi_i(\mathbf{n})$ να παραμένει σταθερός ή και να αυξάνεται, όταν αυξάνεται το \mathbf{n} . Η «ελευθερία» αυτή συνάδει με την αντίστοιχη ιδιότητα των κερδών διαφορισμού πολλαπλών χρηστών η οποία αναφέρθηκε στην Ενότητα 2.3 και αποδεικνύεται στο παράρτημα. Όταν οι συναρτήσεις ρυθμών εξυπηρέτησης διαθέτουν την Ιδιότητα 1, θα ονομάζονται *μονότονες*. Η εργασία [13] θεωρεί συστήματα SDPS με μονότονους ρυθμούς εξυπηρέτησης και δείχνει ότι μέσω της μελέτης πιο απλών συστημάτων που διαθέτουν την ιδιότητα ισορροπίας μπορούν να παραχθούν φράγματα απόδοσης σχετικά με το μήκος της ουράς. Για τα αποτελέσματα που ακολουθούν και απαιτείται η εγκυρότητα της Ιδιότητας 1, το γεγονός αυτό επίσης θα αναφέρεται ρητά.

Εφόσον ο θάλαμος εξυπηρέτησης είναι πεπερασμένος σε μέγεθος, ο χώρος καταστάσεων \mathcal{S} έχει ένα σύνολο το οποίο στην περίπτωση πολλαπλών κλάσεων που μελετάμε εδώ, μπορεί να πάρει διάφορες μορφές. Κανένας περιορισμός δεν τίθεται εκ των προτέρων για το σχήμα του συνόρου, εκτός από την ακόλουθη ιδιότητα η οποία είναι φυσική όταν το πεπερασμένο μέγεθος του χώρου καταστάσεων προκύπτει ως αποτέλεσμα ελέγχου αποδοχής (αγ.: admission control).

Ιδιότητα 2. Για κάθε $\mathbf{n} \in \mathcal{S}$, αν $\mathbf{n}' \leq \mathbf{n}$, τότε $\mathbf{n}' \in \mathcal{S}$.

Η Ιδιότητα 2 δηλώνει ότι όποτε μία διάταξη πελατών είναι αποδεκτή από το σύστημα, όλες οι διατάξεις με λιγότερους πελάτες είναι επίσης αποδεκτές. Δεδομένης της Ιδιότητας 2, το σύνολο των συνοριακών καταστάσεων με αναφορά στην κλάση i , που συμβολίζεται ως \mathcal{C}_i , είναι

$$\mathcal{C}_i \triangleq \{\mathbf{n} \in \mathcal{S} \mid \mathbf{n} + \mathbf{e}_i \notin \mathcal{S}\}, \quad (3.1)$$

όπου το \mathbf{e}_i είναι ένα διάνυσμα κατάλληλου μεγέθους που έχει το i -στό του στοιχείο ίσο με μονάδα και όλα τα υπόλοιπα ίσα με μηδέν. Τα σύνολα συνοριακών καταστάσεων \mathcal{C}_i συνδέονται άμεσα με τον υπολογισμό της πιθανότητας αποκλεισμού ενός νέου πελάτη κλάσης i . Αυτό το γενικό σύνολο του χώρου καταστάσεων μπορεί να εφαρμοστεί σε πολλές ειδικές περιπτώσεις με μεγαλύτερο ενδιαφέρον. Έτσι, μία τέτοια περίπτωση περιλαμβάνει εντελώς διαχωρισμένους χώρους καταστάσεων αναφορικά με την κλάση, της μορφής καρτεσιανού γινομένου $\mathcal{S} = \mathcal{S}_1 \times \dots \times \mathcal{S}_k$, όπου

$$\mathcal{S}_i = \{\mathbf{n} \mid 0 \leq n_i \leq L_i\}, \quad i = 1, \dots, k \quad (3.2)$$

δηλαδή, κάθε κλάση έχει το δικό της ξεχωριστό θάλαμο εξυπηρέτησης μεγέθους L_i και δεν επηρεάζεται από την παρουσία πελατών άλλων κλάσεων. Σε αυτή την περίπτωση,

$$\mathcal{C}_i = \{\mathbf{n} \in \mathcal{S} \mid n_i = L_i\}.$$

Από την άλλη πλευρά, είναι επίσης αρκετά συνηθισμένο να έχουμε ένα κοινό θάλαμο εξυπηρέτησης για όλες τις κλάσεις. Αν θεωρήσουμε ότι το μέγεθος αυτού είναι L , τότε:

$$\mathcal{S} = \{\mathbf{n} \geq 0 \mid |\mathbf{n}| \leq L\}, \quad (3.3)$$

περίπτωση κατά την οποία το σύνολο

$$\mathcal{C}_i = \{\mathbf{n} \geq \mathbf{0} \mid |\mathbf{n}| = L\} \quad (3.4)$$

για όλα τα $i \in \mathcal{K}$ αποτελεί επίσης ένα κοινό σύνολο. Ενδιάμεσα σχήματα είναι επίσης επιτρεπτά, για παράδειγμα όταν οι κλάσεις των πελατών χωρίζονται σε ομάδες (όπως στη μελέτη του διαχωρισμού χρονικής κλίμακας στα Κεφάλαια 5-6). Στην περίπτωση αυτή ο χώρος καταστάσεων μπορεί να διαχωριστεί σε δύο μέρη της μορφής

$$\mathcal{S} = \mathcal{S}_s \times \mathcal{S}_f. \quad (3.5)$$

Οι θάλαμοι εξυπηρέτησης μπορεί να είναι κοινοί για όλες τις κλάσεις μίας ομάδας ή μπορεί να μη συμβαίνει αυτό. Για παράδειγμα, κάποιος ή όλοι οι υποχώροι στο καρτεσιανό γινόμενο μπορεί να είναι της μορφής (3.3).

Έστω ότι $N(t)$ είναι η τυχαία μεταβλητή που αντιστοιχεί στην κατάσταση του συστήματος τη χρονική στιγμή t , $t \geq 0$. Λόγω του μοντέλου αφίξεων που έχει υποθεθεί και των εκθετικά κατανεμημένων απαιτήσεων εξυπηρέτησης, η διαδικασία αυτή εξελίσσεται σύμφωνα με μία αλυσίδα Markov συνεχούς χρόνου με πίνακα ρυθμών μεταβάσεων καταστάσεων $M = [M(\mathbf{n}, \mathbf{n}')]$, του οποίου τα μη διαγώνια, μη μηδενικά στοιχεία δίνονται από τις ακόλουθες σχέσεις:

$$M(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_j(\mathbf{n}), & \mathbf{n}' = \mathbf{n} + \mathbf{e}_j, \mathbf{n} \notin \mathcal{C}_j, \quad j \in \mathcal{K} \\ n_j \mu_j \phi_j(\mathbf{n}), & \mathbf{n}' = \mathbf{n} - \mathbf{e}_j, n_j > 0, \quad j \in \mathcal{K} \end{cases} \quad (3.6)$$

Τα διαγώνια στοιχεία του M καθορίζονται από την απαίτηση ότι τα αθροίσματα των γραμμών του πρέπει να ισούνται με μηδέν. Είναι φανερό ότι ο M είναι μη αναγώγιμος (αγ.: irreducible) και ως εκ τούτου διαθέτει ένα διάνυσμα πιθανοτήτων σταθερής κατάστασης (αγ.: steady state probability vector) π , τέτοιο ώστε $\pi M = 0$ (δηλ. $\sum_{\mathbf{n} \in \mathcal{S}} \pi(\mathbf{n}) M(\mathbf{n}, \mathbf{n}') = 0, \forall \mathbf{n}' \in \mathcal{S}$).

Η παραπάνω μαρκοβιανή δομή μπορεί να μεταβληθεί ώστε να περιλαμβάνει $l = (l_1, \dots, l_k)$ μόνιμους πελάτες επιπλέον των πελατών που προσμετρώνται στην κατάσταση του συστήματος. Οι μόνιμοι πελάτες θεωρούνται ότι βρίσκονται στο σύστημα, εξυπηρετούνται και καταλαμβάνουν θέσεις στο θάλαμο

εξυπηρέτησης αλλά δεν ολοκληρώνουν ποτέ την εξυπηρέτησή τους ώστε να αναχωρήσουν. Ο χώρος καταστάσεων της τροποποιημένης αλυσίδας Markov συμβολίζεται ως

$$\mathcal{S}^{(l)} = \{\mathbf{n} \geq \mathbf{0} \mid \mathbf{n} + \mathbf{l} \in \mathcal{S}\}, \quad (3.7)$$

ενώ το σύνορο που σχετίζεται με την κλάση i είναι

$$\mathcal{C}_i^{(l)} = \{\mathbf{n} \in \mathcal{S}^{(l)} \mid \mathbf{n} + \mathbf{e}_i \notin \mathcal{S}^{(l)}\}.$$

Επεκτείνοντας το συμβολισμό που μόλις εισήχθη, για οποιαδήποτε μεταβλητή στο υπόλοιπο της διατριβής (έστω z) σχετιζόμενη με το αρχικό σύστημα χωρίς μόνιμους πελάτες, ο συμβολισμός με εκθέτη εντός παρενθέσεως (δηλ. $z^{(l)}$) θα αναφέρεται στην αντίστοιχη μεταβλητή που σχετίζεται με το τροποποιημένο σύστημα που περιέχει το συνδυασμό μόνιμων πελατών που υποδηλώνεται στον εκθέτη (δηλ. l). Η περίπτωση ενός και μόνο μόνιμου πελάτη της κλάσης i (δηλ. η περίπτωση $l = \mathbf{e}_i$), είναι κεντρικής σημασίας για τη μετέπειτα ανάλυση. Ως εκ τούτου, θα χρησιμοποιείται ο επιπλέον απλοποιημένος συμβολισμός που δηλώνει μόνο την κλάση στον εκθέτη (δηλ. $z^{(i)}$) αντί του $z^{(\mathbf{e}_i)}$, δίνοντας περισσότερη έμφαση στην κλάση του μόνιμου πελάτη.

Τα μη μηδενικά, μη διαγώνια στοιχεία του πίνακα ρυθμών μεταβάσεων καταστάσεων για την τροποποιημένη αλυσίδα $M^{(l)}$ μπορούν να εκφραστούν σε σχέση με τα στοιχεία του M του αρχικού συστήματος ως εξής:

$$\begin{aligned} M^{(l)}(\mathbf{n}, \mathbf{n} + \mathbf{e}_j) &= \lambda_j^{(l)}(\mathbf{n}) \mathbf{I}_{\mathbf{n} \notin \mathcal{C}_j^{(l)}} = \lambda_j(\mathbf{n} + \mathbf{l}) \mathbf{I}_{\mathbf{n} + \mathbf{l} \notin \mathcal{C}_j} = M(\mathbf{n} + \mathbf{l}, \mathbf{n} + \mathbf{l} + \mathbf{e}_j), \\ M^{(l)}(\mathbf{n}, \mathbf{n} - \mathbf{e}_j) &= n_j \mu_j \phi_j^{(l)}(\mathbf{n}) = n_j \mu_j \phi_j(\mathbf{n} + \mathbf{l}) = \frac{n_j}{n_j + l_j} M(\mathbf{n} + \mathbf{l}, \mathbf{n} + \mathbf{l} - \mathbf{e}_j) \end{aligned} \quad (3.8)$$

Προφανώς, ο $M^{(l)}$ είναι επίσης μη αναγώγιμος και διαθέτει ένα διάνυσμα πιθανοτήτων σταθερής κατάστασης, $\boldsymbol{\pi}^{(l)}$.



Κεφάλαιο 4

Κατανομή του Χρόνου Απόκρισης και Συναφή Αποτελέσματα

ΚΑΝΟΝΤΑΣ χρήση της μοντελοποίησης που εισήχθη στο προηγούμενο κεφάλαιο, συνεχίζουμε εδώ με τον υπολογισμό της συμπληρωματικής συνάρτησης κατανομής πιθανότητας (αγ.: Complementary Probability Distribution Function (CPDF)) του χρόνου απόκρισης που αντιστοιχεί σε ένα πελάτη κλάσης i , $i \in \mathcal{K}$, για τη γενική περίπτωση ενός συστήματος SDPS. Προς το σκοπό αυτό, ας θεωρήσουμε ότι ο πελάτης του οποίου μελετάμε το χρόνο απόκρισης εισέρχεται στο σύστημα τη χρονική στιγμή $t_0 = 0$. Για όλες τις χρονικές στιγμές t κατά τη διάρκεια της παραμονής του συγκεκριμένου πελάτη, με $N^{(i)}(t)$ θα συμβολίζεται το διάνυσμα των υπολοίπων πελατών που είναι παρόντες στο σύστημα. Είναι προφανές ότι η διαδικασία $\{N^{(i)}(t), t \geq 0\}$ εξελίσσεται σύμφωνα με μία αλυσίδα Markov που προκύπτει από μία τροποποίηση της αλυσίδας του αρχικού συστήματος, θεωρώντας ένα επιπλέον πελάτη κλάσης i μόνιμο (ως εκ τούτου και ο συμβολισμός $N^{(i)}(t)$). Τα μη μηδενικά, μη διαγώνια στοιχεία του αντίστοιχου απειροστικού γεννήτορα $M^{(i)}$ της τροποποιημένης αλυσίδας δίνονται από τη σχέση (3.8) μέσω της αντικατάστασης $\mathbf{l} = \mathbf{e}_i$.

Με την ίδια επιχειρηματολογία, το ποσό της εξυπηρέτησης $Q_i(t)$ που έχει ληφθεί από τον παρατηρούμενο πελάτη στο χρονικό διάστημα $(0, t]$ είναι ίσο με

$$Q_i(t) = \int_0^t \phi_i^{(i)}(\mathbf{N}^{(i)}(u)) du = \int_0^t \phi_i(\mathbf{N}^{(i)}(u) + \mathbf{e}_i) du. \quad (4.1)$$

Σχολιάζοντας την (4.1), $Q_i(t)$ είναι η συνολική «ανταμοιβή» που συγκεντρώθηκε στο διάστημα $(0, t]$ μέσω ενός μοντέλου ανταμοιβής Markov με απειροστικό γεννήτορα $M^{(i)}$ και ρυθμούς ανταμοιβής $\phi_i^{(i)}(\mathbf{n})$. Αν συγκεντρώσουμε

τους ρυθμούς ανταμοιβής σε ένα διαγώνιο πίνακα, έχουμε

$$R^{(i)} \triangleq \text{diag}_{\mathbf{n} \in \mathcal{S}^{(i)}} \{\phi_i^{(i)}(\mathbf{n})\} = \text{diag}_{\mathbf{n} \in \mathcal{S}^{(i)}} \{\phi_i(\mathbf{n} + \mathbf{e}_i)\}. \quad (4.2)$$

Επίσης, θεωρούμε τους δεσμευμένους μετασχηματισμούς LST $x_{\mathbf{n}}(\theta) \triangleq \mathbb{E} [e^{-\theta Q_i(t)} | \mathbf{N}^{(i)}(0) = \mathbf{n}]$, $\mathbf{n} \in \mathcal{S}^{(i)}$ και τους συγκεντρώνουμε σε ένα διάνυσμα στήλη $\mathbf{x}(\theta) = [x_{\mathbf{n}}(\theta)]$. Έτσι, εφαρμόζοντας ένα καθιερωμένο αποτέλεσμα για μαρκοβιανά μοντέλα ανταμοιβής (βλ., π.χ. [34]¹),

$$\mathbf{x}(\theta) = \exp\{(-\theta R^{(i)} + M^{(i)})t\} \mathbf{1}, \quad (4.3)$$

όπου με $\mathbf{1}$ συμβολίζεται το διάνυσμα στήλη κατάλληλου μήκους με όλα τα στοιχεία ίσα με τη μονάδα.

Στο σημείο αυτό εισάγουμε το σύμβολο $\nu_i(\mathbf{n}) = \Pr\{\mathbf{N}^{(i)}(0) = \mathbf{n}\}$. Συγκεντρώνουμε αυτές τις πιθανότητες σε ένα διάνυσμα γραμμή $\boldsymbol{\nu}_i = [\nu_i(\mathbf{n})]$ και εφαρμόζοντας την (4.3), παίρνουμε εν τέλει

$$\mathbb{E} [e^{-\theta Q_i(t)}] = \sum_{\mathbf{n} \in \mathcal{S}^{(i)}} \nu_i(\mathbf{n}) x_{\mathbf{n}}(\theta) = \boldsymbol{\nu}_i \exp\{(-\theta R^{(i)} + M^{(i)})t\} \mathbf{1}. \quad (4.4)$$

Οι αρχικές πιθανότητες $\nu_i(\mathbf{n})$ μπορούν να καθοριστούν ως ακολούθως. Ας θεωρήσουμε μία ενσωμάτωση (αγ.: embedding) της αρχικής διαδικασίας Markov (χωρίς μόνιμους πελάτες) κατά τις στιγμές που συμβαίνει αλλαγή της κατάστασης. Όπως είναι γνωστό, οι πιθανότητες σταθερής κατάστασης αυτής της ενσωματωμένης αλυσίδας Markov διακριτού χρόνου είναι ίσες με $w(\mathbf{n}) = -cM(\mathbf{n}, \mathbf{n})\pi(\mathbf{n})$, $\mathbf{n} \in \mathcal{S}$ και $c = -(\sum_{\mathbf{n} \in \mathcal{S}} M(\mathbf{n}, \mathbf{n})\pi(\mathbf{n}) - 1)$, όπου η σταθερά αναλογίας c καθορίζεται μέσω της συνθήκης κανονικοποίησης. Όταν η ενσωματωμένη διαδικασία είναι σε κάποια κατάσταση \mathbf{n} , έστω ότι με \mathcal{A} συμβολίζεται το γεγονός ότι το επόμενο βήμα σχετίζεται με μία άφιξη κλάσης i που γίνεται δεκτή. Τότε η πιθανότητα $\Pr\{\mathbf{n} | \mathcal{A}\}$ είναι η πιθανότητα με την οποία, δεδομένης μίας άφιξης κλάσης i , η διαδικασία θα κινηθεί στην κατάσταση $\mathbf{n} + \mathbf{e}_i$, δηλαδή, η πιθανότητα με την οποία ο καινούριος πελάτης της κλάσης i που θα εισέλθει στο σύστημα θα βρει \mathbf{n} άλλους πελάτες παρόντες. Έτσι από τον νόμο του Bayes,

$$\nu_i(\mathbf{n}) = \Pr\{\mathbf{n} | \mathcal{A}\} = \frac{\Pr\{\mathcal{A} | \mathbf{n}\}w(\mathbf{n})}{\sum_{\mathbf{n}' \in \mathcal{S}^{(i)}} \Pr\{\mathcal{A} | \mathbf{n}'\}w(\mathbf{n}')} = \frac{\pi(\mathbf{n})\lambda_i(\mathbf{n})}{\sum_{\mathbf{n}' \in \mathcal{S}^{(i)}} \pi(\mathbf{n}')\lambda_i(\mathbf{n}')}, \quad (4.5)$$

¹Το [34] παράγει το γενικό ανάλογο της (4.3) εφαρμόζοντας ένα σχετικό προηγούμενο αποτέλεσμα που βρέθηκε ανεξάρτητα από τα [39] και [47]. Εναλλακτικά, η (4.3) μπορεί να ληφθεί άμεσα λύνοντας ένα σύστημα διαφορικών εξισώσεων πρώτης τάξης, του τύπου «εξισώσεις προς τα εμπρός» (αγ.: forward equations).

όπου η τελευταία ισότητα πηγάζει από την έκφραση για το $w(\mathbf{n})$ και το γεγονός ότι $\Pr\{\mathcal{A} \mid \mathbf{n}\} = \lambda_i(\mathbf{n})/(-M(\mathbf{n}, \mathbf{n}))$, $\mathbf{n} \in \mathcal{S}^{(i)}$. Το άθροισμα στον παρονομαστή της (4.5) εκτείνεται στις καταστάσεις $\mathcal{S}^{(i)} = \mathcal{S} - \mathcal{C}_i$ επειδή $\Pr\{\mathbf{n} \wedge \mathcal{A}\} = 0$, για όλα τα $\mathbf{n} \in \mathcal{C}_i$. Συγκεκριμένα, ο παρονομαστής στην (4.5)

$$\lambda_{av,i} \triangleq \sum_{\mathbf{n}' \in \mathcal{S} - \mathcal{C}_i} \pi(\mathbf{n}') \lambda_i(\mathbf{n}') \quad (4.6)$$

αναγνωρίζεται άμεσα ότι είναι ίσος με το μέσο ρυθμό αφίξεων της κλάσης i που εισέρχονται στο σύστημα (δηλ. δεν αποκλείονται). Ας σημειωθεί επίσης, ότι όταν οι πελάτες της κλάσης i αφικνούνται σύμφωνα με μία ομοιογενή διαδικασία Poisson (δηλ. $\lambda_i(\mathbf{n}) = \lambda_i, \forall \mathbf{n} \in \mathcal{S}$), η (4.5) εξειδικεύεται στην έκφραση

$$\nu_i(\mathbf{n}) = \pi(\mathbf{n})/(1 - P_{B,i}), \quad \mathbf{n} \in \mathcal{S} - \mathcal{C}_i, \quad (4.7)$$

όπου $P_{B,i} = \sum_{\mathbf{n} \in \mathcal{C}_i} \pi(\mathbf{n})$ είναι η πιθανότητα αποκλεισμού για τις αφίξεις της κλάσης i . Αυτό το ειδικό αποτέλεσμα δεν αποτελεί τίποτα άλλο παρά μία εκδήλωση της ιδιότητας PASTA (Poisson Arrivals See Time Averages - Οι αφίξεις Poisson βλέπουν χρονικούς μέσους).

Τώρα διαθέτουμε όλα τα απαραίτητα συστατικά για να αναπτύξουμε το κύριο αποτέλεσμα της παρούσας ενότητας:

Θεώρημα 2. Η CPDF του χρόνου απόκρισης V_i ενός πελάτη κλάσης i είναι

$$S_i(t) \triangleq \Pr\{V_i > t\} = \nu_i \exp\{-B_i t\} \mathbf{1}, \quad (4.8)$$

όπου

$$B_i \triangleq \mu_i R^{(i)} - M^{(i)} \quad (4.9)$$

και τα $R^{(i)}$ και ν_i ορίζονται στις (4.2) και (4.5) αντίστοιχα.

Απόδειξη. Έστω ότι T_i είναι η αρχική απαίτηση εξυπηρέτησης ενός πελάτη κλάσης i . Από τον ορισμό του $Q_i(t)$ στην (4.1),

$$S_i(t) = \Pr\{T_i > Q_i(t)\} = \int_0^\infty \Pr\{T_i > x\} d\Pr\{Q_i(t) \leq x\} = \mathbb{E} [e^{-\mu_i Q_i(t)}], \quad (4.10)$$

όπου η τελευταία εξίσωση πηγάζει από το γεγονός ότι οι απαιτήσεις εξυπηρέτησης των πελατών της κλάσης i είναι εκθετικά κατανομημένες με μέση τιμή $1/\mu_i$. Η εφαρμογή της (4.4) με $\theta = \mu_i$ ολοκληρώνει την απόδειξη. \square

Ας σημειώσουμε ότι παρά το γεγονός ότι το εκθετικό στην (4.8) εμπεριέχει τον πίνακα B_i που σχετίζεται με την τροποποιημένη αλυσίδα Markov με

ένα μόνιμο πελάτη κλάσης i , το διάνυσμα πιθανοτήτων ν_i συνδέεται με τις πιθανότητες σταθερής κατάστασης της αρχικής αλυσίδας. Εν γένει $\nu_i \neq \pi^{(i)}$ και έτσι, αν και θεωρείται ότι η αρχική αλυσίδα έχει φτάσει σε στάσιμη κατάσταση, η διαδικασία $N^{(i)}(t)$ που σχετίζεται με το $Q_i(t)$ δε βρίσκεται σε στάσιμη κατάσταση.

Δεν είναι δύσκολο να δει κανείς ότι το αποτέλεσμα του Θεωρήματος 2 είναι συμβατό και με το [15], (που διαπραγματεύεται σύστημα μίας κλάσης) αλλά και με το [19] (το οποίο αν και εστιάζει στο DPS, χρησιμοποιεί μία μοντελοποίηση το ίδιο γενική με εδώ). Καθεμία από αυτές τις μελέτες παράγει ένα σύστημα γραμμικών εξισώσεων που επιλύεται ως προς το διάνυσμα των μετασχηματισμών LST των δεσμευμένων χρόνων απόκρισης, με δεδομένο ότι ο παρατηρούμενος πελάτης βρίσκεται κατά την άφιξή του n άλλους πελάτες στο σύστημα. Χρησιμοποιώντας το συμβολισμό της παρούσας διατριβής, δημιουργούν το σύστημα γραμμικών εξισώσεων $\psi(\theta) = [\psi_n(\theta)]$, όπου $\psi_n(\theta) \triangleq E[e^{-\theta V_i} | N(0) = n]$.

Έστω το διάνυσμα $S(t) = [S_n(t)]$, όπου $S_n(t) \triangleq \Pr\{V_i > t | N^{(i)}(0) = n\}$ και $S^*(\theta)$ είναι ο μετασχηματισμός LST του $S(t)$. Λόγω των ιδιοτήτων του LST, ισχύει ότι $\psi(\theta) = \mathbf{1} - \theta S^*(\theta)$. Επίσης, από τα επιχειρήματα στην απόδειξη του Θεωρήματος 2, $S(t) = \mathbf{x}(\mu_i) = \exp\{-B_i t\}\mathbf{1}$, έτσι ώστε $S^*(\theta) = (\theta I + B_i)^{-1}\mathbf{1}$, όπου I είναι ο μοναδιαίος πίνακας. Αυτό οδηγεί στο ότι $\psi(\theta) = [I - \theta(\theta I + B_i)^{-1}]\mathbf{1} = (\theta I + B_i)^{-1}B_i\mathbf{1}$, ισοδύναμα με τη σχέση

$$(\theta I + B_i)\psi(\theta) = B_i\mathbf{1} = \mu_i R^{(i)}\mathbf{1}, \quad (4.11)$$

όπου η τελευταία εξίσωση ακολουθεί από την (4.9) και το γεγονός ότι εφόσον ο $M^{(i)}$ είναι απειροστικός γεννήτορας, ικανοποιεί τη σχέση $M^{(i)}\mathbf{1} = \mathbf{0}$. Είναι εύκολο να αναγνωρίσουμε ότι η (4.11) είναι ακριβώς το σύστημα εξισώσεων που προκύπτει από τα [15] και [19].

Λαμβάνοντας υπόψη τα παραπάνω σχόλια, το Θεώρημα 2 μπορεί να θεωρηθεί ως μία απευθείας αντιστροφή των λύσεων του μετασχηματισμού στα [15, 19]. Ωστόσο, το Θεώρημα 2 είναι βαθύτερο με την έννοια ότι τονίζει ευθέως μέσω της (4.9) την υποκείμενη πιθανοτική σύνδεση με την διαδικασία ανταμοιβής Markov που αντιστοιχεί στην τροποποιημένη αλυσίδα Markov με ένα μόνιμο πελάτη κλάσης i . Όπως θα δείξουμε σε επόμενα κεφάλαια, η υποκείμενη διαδικασία ανταμοιβής Markov μπορεί να εφαρμοστεί άμεσα για τη λήψη αποτελεσμάτων πέραν αυτών που επιτρέπει ο μετασχηματισμός LST [15,19]. Τέτοια αποτελέσματα περιλαμβάνουν το όριο για το ρυθμό εκθετικής φθίσης της ουράς του χρόνου απόκρισης και, πιο σημαντικό, την επίσημη σύσταση αποτελεσμάτων για τη γενική αποσύνθεση του μοντέλου όταν οι πελάτες μπορούν να χωριστούν σε ομάδες κλάσεων που λειτουργούν σε πολύ διαφορετικές χρονικές κλίμακες.

Παρακάτω παραθέτουμε κάποιες επιπλέον ιδιότητες της πιθανοτικής δομής του Θεωρήματος 2, παρέχοντας μία απευθείας επαλήθευση ότι η (4.8) είναι συνεπής με το νόμο του Little. Ξεκινώντας από την (4.8),

$$\mathbb{E}[V_i] = \int_0^\infty S_i(t)dt = \boldsymbol{\nu}_i \left(\int_0^\infty \exp\{-B_i t\} dt \right) \mathbf{1} = \boldsymbol{\nu}_i B_i^{-1} \mathbf{1}, \quad (4.12)$$

όπου η τελευταία εξίσωση ισχύει λόγω του γεγονότος (που θα αποδειχτεί στην επόμενη ενότητα) ότι όλες οι ιδιοτιμές του B_i έχουν αποκλειστικά θετικό πραγματικό μέρος, επιβεβαιώνοντας και ότι ο B_i^{-1} υπάρχει αλλά και ότι το ολοκλήρωμα του εκθετικού σε πίνακα συγκλίνει. Το επιχειρήμα συνεχίζεται ορίζοντας το διάνυσμα γραμμή $\mathbf{q}_i = [q_i(\mathbf{n})]$, με στοιχεία

$$q_i(\mathbf{n}) \triangleq (n_i + 1)\pi(\mathbf{n} + \mathbf{e}_i), \quad \mathbf{n} \in \mathcal{S}^{(i)}. \quad (4.13)$$

Ας σημειωθεί ότι

$$\mathbf{q}_i \mathbf{1} = \sum_{\mathbf{n} \in \mathcal{S}^{(i)}} (n_i + 1)\pi(\mathbf{n} + \mathbf{e}_i) = \sum_{\substack{\mathbf{n} \in \mathcal{S} \\ n_i > 0}} n_i \pi(\mathbf{n}) = \sum_{\mathbf{n} \in \mathcal{S}} n_i \pi(\mathbf{n}) = \mathbb{E}[N_i(t)], \quad (4.14)$$

που είναι ο μέσος αριθμός πελατών κλάσης i που είναι παρόντες στο σύστημα.

Εφαρμόζοντας τις (4.9) και (3.8), την τελευταία όπως εφαρμόζεται για $\mathbf{l} = \mathbf{e}_i$, είναι εύκολο να επαληθευτεί, αν και κάπως επίπονο, ότι το \mathbf{n} -στο στοιχείο του διανύσματος $\mathbf{q}_i B_i$ (δηλ. $(\mathbf{q}_i B_i)(\mathbf{n})$) ικανοποιεί τη σχέση $(\mathbf{q}_i B_i)(\mathbf{n}) = \pi(\mathbf{n})M(\mathbf{n}, \mathbf{n} + \mathbf{e}_i) = \pi(\mathbf{n})\lambda_i(\mathbf{n})$. Θεωρώντας τις (4.5) και (4.6) και συγκεντρώνοντας τις τελευταίες εξισώσεις για $\mathbf{n} \in \mathcal{S}^{(i)}$ σε μορφή διανύσματος, λαμβάνουμε τελικά ότι $\mathbf{q}_i B_i = \lambda_{\text{av},i} \boldsymbol{\nu}_i$ και ο συνδυασμός με την (4.12) οδηγεί τελικά στο ότι $\lambda_{\text{av},i} \mathbb{E}[T_i] = \mathbf{q}_i \mathbf{1}$. Με χρήση της (4.14) το επιχειρήμα ολοκληρώνεται.

4.1 Φασματικές ιδιότητες του B_i

Από τον ορισμό του στην (4.9), ο πίνακας $-B_i$ είναι ουσιωδώς μη αρνητικός και μη αναγώγιμος, κληρονομώντας αυτές τις ιδιότητες από τον $M^{(i)}$. Έτσι μία απλή εφαρμογή της θεωρίας Perron-Frobenius (βλ. π.χ., [7], σελ.27, Θεώρημα 1.4b) είναι αρκετή για να αποδειχτεί ότι η ιδιοτιμή του $-B_i$ με το μεγαλύτερο πραγματικό μέρος (και έτσι η ιδιοτιμή του B_i με το μικρότερο πραγματικό μέρος) είναι απλή, μεγαλύτερη από το πραγματικό μέρος οποιασδήποτε άλλης ιδιοτιμής και αντιστοιχεί σε αποκλειστικά θετικά αριστερό και δεξί ιδιοδιανύσματα. Επιπλέον, ο B_i είναι ένας μη αναγώγιμος, ομαλός πίνακας τύπου M.

Πράγματι, η μη αναγωγιμότητα έχει ήδη θεμελιωθεί από το προηγούμενο επιχειρήματα για τον $-B_i$. Επίσης, τα μη διαγώνια στοιχεία του B_i είναι μη θετικά. Σύμφωνα με την Πρόταση I₂₇, σελ.136 στο [7], κάθε πίνακας A αυτού του τύπου είναι ομαλός πίνακας τύπου M αν και μόνον αν, υπάρχει ένα αποκλειστικά θετικό διάνυσμα $\mathbf{x} \gg \mathbf{0}$ τέτοιο ώστε $A\mathbf{x} \gg \mathbf{0}$. Εφόσον $\mathbf{1} \gg \mathbf{0}$ και $M^{(i)}\mathbf{1} = \mathbf{0}$, θεωρώντας τη (4.9) έχουμε $B_i\mathbf{1} = \mu_i R^{(i)}\mathbf{1} \gg \mathbf{0}$, θεμελιώνοντας την ιδιότητα του πίνακα τύπου M. Έτσι, από ένα καθιερωμένο αποτέλεσμα της θεωρίας των πινάκων αυτού του τύπου (βλ. π.χ., Ιδιότητα E₁₁, σελ.150 στο [7]), το πραγματικό μέρος όλων των ιδιοτιμών του B_i είναι θετικό. Τα συγκεκριμένα επιχειρήματα αποδεικνύουν το ακόλουθο:

Λήμμα 1. *Ο πίνακας B_i στην (4.9) είναι ομαλός, μη αναγωγίμος πίνακας τύπου M. Το πραγματικό μέρος όλων των ιδιοτιμών του είναι θετικό. Η ιδιοτιμή με το μικρότερο πραγματικό μέρος, συμβολιζόμενη ως $\sigma_1(B_i)$ είναι πραγματική, απλή, αποκλειστικά μικρότερη από το πραγματικό μέρος οποιασδήποτε άλλης ιδιοτιμής και συνδέεται με αποκλειστικά θετικά αριστερό και δεξί ιδιοδιανύσματα.*

Το Λήμμα 1 διασφαλίζει ότι ο B_i είναι θετικά σταθερός (δηλαδή το πραγματικό μέρος των ιδιοτιμών του είναι θετικό) και άρα $\exp\{-B_i t\} \rightarrow 0$ όταν $t \rightarrow \infty$.

Ας υποθέσουμε τώρα ότι ο B_i είναι διαγωνιοποιήσιμος. Τότε για καθεμία από τις ιδιοτιμές $\sigma_j(B_i)$, $j = 1, \dots, \text{card}(\mathcal{S}^{(i)})$, ακόμα και για εκείνες που η αλγεβρική τους πολλαπλότητα είναι μεγαλύτερη της μονάδας (αν υπάρχει κάποια τέτοια), μπορεί κανείς να αντιστοιχίσει ένα αριστερό και ένα δεξί ιδιοδιάνυσμα $\mathbf{y}_j(B_i)$ και $\mathbf{x}_j(B_i)$, τα οποία μπορούν να κανονικοποιηθούν ώστε $\mathbf{y}_j(B_i)\mathbf{x}_j(B_i) = 1$. Ορίζοντας τον $X(B_i) \triangleq [\mathbf{x}_j(B_i)]$ ($Y(B_i) \triangleq [\mathbf{y}_j(B_i)]$) ως τον πίνακα με στήλες $\mathbf{x}_j(B_i)$ (γραμμές $\mathbf{y}_j(B_i)$) έχουμε $X(B_i) = Y(B_i)^{-1}$, $\exp\{-B_i t\} = \sum_{j=1}^m \mathbf{x}_j(B_i)\mathbf{y}_j(B_i)e^{-\sigma_j(B_i)t}$ και δεδομένης της (4.8),

$$S_i(t) = \sum_j a_j(B_i)e^{-\sigma_j(B_i)t}, \quad (4.15)$$

$$a_j(B_i) \triangleq (\boldsymbol{\nu}_i \mathbf{x}_j(B_i))(\mathbf{y}_j(B_i) \mathbf{1}). \quad (4.16)$$

Ας σημειώσουμε ότι $\sum_j a_j(B_i) = \boldsymbol{\nu}_i X(B_i) Y(B_i) \mathbf{1} = \boldsymbol{\nu}_i \mathbf{1} = 1$, έτσι ώστε οι συντελεστές να αθροίζονται στη μονάδα. Το Λήμμα 1 εξασφαλίζει ότι $\text{Re}\{\sigma_j(B_i)\} > \sigma_1(B_i)$ και έτσι

$$S_i(t) \sim a_1(B_i)e^{-\sigma_1(B_i)t}, \quad t \rightarrow \infty, \quad (4.17)$$

όπου η ισοδυναμία $f(t) \sim g(t)$, $t \rightarrow \infty$ έχει το καθιερωμένο νόημα $\lim_{t \rightarrow \infty} f(t)/g(t) = 1$. Το Λήμμα 1 και η θετικότητα των $\boldsymbol{\nu}_i$ και $\mathbf{1}$ θεμελιώνουν ότι σε κάθε περίπτωση $a_1(B_i) > 0$.

Στη γενική περίπτωση ο B_i δεν είναι απαραίτητα διαγωνιοποιήσιμος. Σε μία τέτοια περίπτωση, η θεώρηση της γενικής κανονικής μορφής Jordan του B_i δείχνει ότι όροι της μορφής $a_{jh}(B_i)t^h e^{-\sigma_j(B_i)t}$ μπορεί να εμφανιστούν στο άθροισμα για την $S_i(t)$ στην (4.15). Ωστόσο, δεδομένου του Λήμματος 1 (πολλαπλότητα της $\sigma_1(B_i)$), αυτό μπορεί να συμβεί μόνο για $j \neq 1$, ενώ ο όρος για $j = 1$ είναι όπως στην (4.16). Με αυτή την παρατήρηση, η κυριαρχία της $\sigma_1(B_i)$ θεμελιώνει ότι η ασυμπτωτικά ισοδύναμη (4.17) ισχύει εν γένει.

Λαμβάνοντας υπόψη την (4.17), η $\sigma_1(B_i)$ παρέχει τον ασυμπτωτικό ρυθμό εκθετικής φθίσης της ουράς του χρόνου απόκρισης. Για την ακρίβεια, η (4.17) οδηγεί αμέσως στη

$$\lim_{t \rightarrow \infty} \frac{\log S_i(t)}{t} = -\sigma_1(B_i). \quad (4.18)$$

Το ακόλουθο αποτέλεσμα παρέχει εύκολα υπολογίσιμα φράγματα για αυτήν τη σημαντική ποσότητα.

Πρόταση 2. Η ποσότητα $\sigma_1(B_i)$ είναι φραγμένη μεταξύ του ελάχιστου και του μέσου ρυθμού εξυπηρέτησης του μόνιμου πελάτη της κλάσης i , δηλαδή

$$\mu_i \underline{\phi}_i \leq \sigma_1(B_i) \leq \mu_i \bar{\phi}_i, \quad (4.19)$$

όπου $\underline{\phi}_i \triangleq \min_{\mathbf{n} \in \mathcal{S}^{(i)}} \phi_i^{(i)}(\mathbf{n}) = \min_{\substack{\mathbf{n} \in \mathcal{S} \\ n_i > 0}} \phi_i(\mathbf{n})$ και $\bar{\phi}_i \triangleq \sum_{\mathbf{n} \in \mathcal{S}^{(i)}} \pi^{(i)}(\mathbf{n}) \phi_i^{(i)}(\mathbf{n}) = \sum_{\mathbf{n} \in \mathcal{S}^{(i)}} \pi^{(i)}(\mathbf{n}) \phi_i(\mathbf{n} + \mathbf{e}_i)$. Αν η συνάρτηση ρυθμών εξυπηρέτησης $\phi(\cdot)$ ικανοποιεί την

Ιδιότητα 1, τότε αναγκαστικά $\underline{\phi}_i = \min_{\mathbf{n} \in \cap_{j=1}^k \mathcal{C}_j} \phi_i(\mathbf{n})$.

Απόδειξη. Το κάτω φράγμα στη (4.19) προκύπτει άμεσα από τις (4.1), (4.10) και (4.18). Η ισότητα $\min_{\mathbf{n} \in \mathcal{S}^{(i)}} \phi_i^{(i)}(\mathbf{n}) = \min_{\mathbf{n} \in \mathcal{P}_i} \phi_i(\mathbf{n})$, όπου $\mathcal{P}_i \triangleq \{\mathbf{n} \in \mathcal{S} \mid n_i > 0\}$, είναι απλή συνέπεια του ορισμού των $\mathcal{S}^{(i)}$ και $\phi_i^{(i)}(\cdot)$. Ισχύει πάντα ότι $\cap_{j=1}^k \mathcal{C}_j \subseteq \mathcal{C}_i \subseteq \mathcal{P}_i$, ώστε $\min_{\mathbf{n} \in \cap_{j=1}^k \mathcal{C}_j} \phi_i(\mathbf{n}) \geq \underline{\phi}_i$. Θα κατασκευάσουμε αναδρομικά μία πεπερασμένη ακολουθία \mathbf{n}_m , $m = 0, \dots, k$, τέτοια ώστε το \mathbf{n}_m να φτάνει στο $\underline{\phi}_i$, δηλαδή $\phi_i(\mathbf{n}_m) = \underline{\phi}_i$ και $\mathbf{n}_m \in \cap_{j=1}^m \mathcal{C}_j$ (για $m = 0$ $\cap_{j=1}^m \mathcal{C}_j \triangleq \mathcal{P}_i$). Για $m = 0$, επιλέγουμε οποιοδήποτε $\mathbf{n}_0 \in \mathcal{P}_i$ που φτάνει στο ελάχιστο, δηλαδή $\phi_i(\mathbf{n}_0) = \underline{\phi}_i$. Τώρα θα δείξουμε πως κάποιος μπορεί να πάρει το \mathbf{n}_{m+1} , δεδομένου του \mathbf{n}_m , $m \geq 0$. Αν $\mathbf{n}_m \in \mathcal{C}_{m+1}$ τότε θέτουμε απλώς $\mathbf{n}_{m+1} = \mathbf{n}_m$ και το \mathbf{n}_m ικανοποιεί τετριμμένα τις ζητούμενες ιδιότητες. Σε διαφορετική περίπτωση, $\mathbf{n}_m \notin \mathcal{C}_{m+1}$, υπονοώντας ότι $\mathbf{n}_m + \mathbf{e}_{m+1} \in \mathcal{S}$, έτσι ώστε $1 \in \mathcal{D}_{m+1} \triangleq \{l \in \mathbb{N} \mid \mathbf{n}_m + l\mathbf{e}_{m+1} \in \mathcal{S}\}$. Συνεπώς το σύνολο \mathcal{D}_{m+1} είναι μη κενό και πεπερασμένο (επειδή το \mathcal{S} είναι πεπερασμένο). Ως εκ τούτου το \mathcal{D}_{m+1} διαθέτει κάποιο μεγιστοτικό (αγγ.: maximal)

στοιχείο, έστω l^* . Εκ κατασκευής $\mathbf{n}_{m+1} \triangleq \mathbf{n}_m + l^* \mathbf{e}_{m+1} \in \mathcal{C}_{m+1}$. Εφόσον $\mathbf{n}_m \in \mathcal{C}_j$ $\forall j = 1, \dots, m$, από την Ιδιότητα 2 έχουμε ότι επίσης $\mathbf{n}_{m+1} \in \mathcal{C}_j, \forall j = 1, \dots, m$. Έτσι, $\mathbf{n}_{m+1} \in \bigcap_{j=1}^{m+1} \mathcal{C}_j$, ικανοποιώντας τη δέσμευση τα μέλη της ακολουθίας να ανήκουν στο δεδομένο σύνολο. Η ύπαρξη του \mathbf{n}_k αποδεικνύει το ζητούμενο καθώς από την Ιδιότητα 1, $\phi_i(\mathbf{n}_k - \mathbf{e}_j) \geq \phi_i(\mathbf{n}_k), j = 1, \dots, k$ και $\phi_i(\mathbf{n}_k) \geq \min_{\mathbf{n} \in \bigcap_{j=1}^k \mathcal{C}_j} \phi_i(\mathbf{n})$.

Για να αποδείξουμε το άνω φράγμα στην (4.19), ας θεωρήσουμε τη συσσωρευτική αμοιβή $Q_i(t)$ στο διάστημα $(0, t]$, η οποία και πάλι ικανοποιεί την (4.1), αλλά τώρα εφαρμόζουμε τις αρχικές πιθανότητες $\Pr\{\mathbf{N}^{(i)}(0) = \mathbf{n}\} = \pi^{(i)}(\mathbf{n}), \forall \mathbf{n} \in \mathcal{S}^{(i)}$ (δηλ., τις πιθανότητες σταθερής κατάστασης της τροποποιημένης αλυσίδας Markov), αντί των πιθανοτήτων της (4.5). Δεδομένης της (4.3) όπως εφαρμόζεται για $\theta = \mu_i$, έχουμε ότι $\tilde{S}_i(t) \triangleq \mathbb{E} \left[e^{-\mu_i \tilde{Q}_i(t)} \right] = \boldsymbol{\pi}^{(i)} \exp\{-B_i t\} \mathbf{1}$. Πρέπει να τονιστεί εδώ ότι η $\tilde{S}_i(t)$ δε σχετίζεται με την CPDF της κατανομής του χρόνου απόκρισης κάποιου πελάτη της κλάσης i . Παρόλα αυτά, εμπλέκει το εκθετικό του ίδιου πίνακα, ώστε τα επιχειρήματα για τη φασματική επέκταση του που παρουσιάστηκαν παραπάνω σε αυτήν την ενότητα συνεχίζουν να είναι σε ισχύ. Συγκεκριμένα, υφίσταται μία ανάλογη σχέση με την (4.17) (με τους τροποποιημένους παράγοντες $\tilde{a}_1(B_i) \triangleq (\boldsymbol{\pi}^{(i)} \mathbf{x}_1(B_i))(\mathbf{y}_1(B_i) \mathbf{1})$), ώστε η $\tilde{S}_i(t)$ να ικανοποιεί επίσης την (4.18). Εφόσον η $\mathbf{N}^{(i)}(t)$ αρχικοποιήθηκε με το διάνυσμα των πιθανοτήτων $\boldsymbol{\pi}^{(i)}$ είναι στάσιμη και, θεωρώντας την (4.1), η $\tilde{Q}_i(t)$ έχει στάσιμες ελάχιστες μεταβολές (αγ.: stationary increments) με συνέπεια

$$\mathbb{E} \left[\tilde{Q}_i(t) \right] = t \mathbb{E} \left[\phi_i^{(i)}(\mathbf{N}^{(i)}(0)) \right] = t \bar{\phi}_i. \text{ Οπότε,}$$

$$\log \tilde{S}_i(t) = \log \mathbb{E} \left[e^{-\mu_i \tilde{Q}_i(t)} \right] \geq \mathbb{E} \left[\log e^{-\mu_i \tilde{Q}_i(t)} \right] = -\mu_i \mathbb{E} \left[\tilde{Q}_i(t) \right] = -\mu_i \bar{\phi}_i t,$$

όπου η ανισότητα ακολουθεί λόγω του ότι η λογαριθμική συνάρτηση είναι κοίλη. Διαιρώντας με t , παίρνοντας το όριο $t \rightarrow \infty$ και εφαρμόζοντας την (4.18), η απόδειξη ολοκληρώνεται. \square

Σημειώνουμε ότι το συνολικό συνοριακό σύνολο $\bigcap_{j=1}^k \mathcal{C}_j$ εξαρτάται από τη μορφή του θαλάμου εξυπηρέτησης. Για ξεχωριστούς θαλάμους εξυπηρέτησης της μορφής (3.2), η $\bigcap_{j=1}^k \mathcal{C}_j$ περιλαμβάνει το απλό διάνυσμα (L_1, \dots, L_k) , ενώ σε ένα κοινό θάλαμο της μορφής (3.3), έχουμε ότι $\bigcap_{j=1}^k \mathcal{C}_j = \{\mathbf{n} \mid |\mathbf{n}| = L\}$. Για τη δεύτερη περίπτωση, αν επιπρόσθετα η συνάρτηση ρυθμού εξυπηρέτησης εξαρτάται αποκλειστικά από το συνολικό αριθμό πελατών, δηλαδή $\phi_i(\mathbf{n}) = g_i(|\mathbf{n}|)$ για κάποια συνάρτηση $g_i(\cdot)$, τότε όλα τα μέλη της $\bigcap_{j=1}^k \mathcal{C}_j$ σημειώνουν το ελάχιστο $\underline{\phi}_i = g_i(L)$.

4.2 Δίκτυα με ιδιότητα ισορροπίας

Τώρα θα εφαρμόσουμε τα αποτελέσματα της προηγούμενης ενότητας στη σημαντική ειδική περίπτωση των δικτύων που διαθέτουν επιπλέον την ιδιότητα ισορροπίας (αγ.: balanced networks, Whittle networks) [48, 51], δηλαδή συστήματα των οποίων οι συναρτήσεις ρυθμού εξυπηρέτησης υπακούουν στην ακόλουθη ιδιότητα ισορροπίας

$$\frac{\phi_i(\mathbf{n} - \mathbf{e}_j)}{\phi_i(\mathbf{n})} = \frac{\phi_j(\mathbf{n} - \mathbf{e}_i)}{\phi_j(\mathbf{n})}, \quad \mathbf{n} \in \mathcal{S}, n_i, n_j > 0, i, j \in \mathcal{K}. \quad (4.20)$$

Σε ολόκληρη την παρούσα ενότητα θα περιοριστούμε σε ομοιογενείς αφίξεις Poisson, δηλαδή $\lambda_i(\mathbf{n}) = \lambda_i, i \in \mathcal{K}$, για όλα τα $\mathbf{n} \in \mathcal{S}$. Με δεδομένη την ιδιότητα ισορροπίας, η «συνάρτηση ισορροπίας» $\Phi(\cdot)$, οριζόμενη αναδρομικά ως

$$\Phi(\mathbf{n}) = \Phi(\mathbf{n} - \mathbf{e}_i)/\phi_i(\mathbf{n}), \quad \mathbf{n} \in \mathcal{S}, i \in \mathcal{K}, \Phi(\mathbf{0}) \triangleq 1,$$

καθορίζεται μοναδικά, ανεξάρτητα των δεικτών των κλάσεων που χρησιμοποιούνται στην αναδρομή. Επιπρόσθετα, με αναφορά στη μορφή των ρυθμών μεταβάσεων στην (3.6) και ορίζοντας τις εντάσεις κίνησης $\rho_i \triangleq \lambda_i/\mu_i$, είναι εύκολο να επαληθευτεί ότι οι πιθανότητες σταθερής κατάστασης μπορούν να δοθούν σε κλειστή μορφή ως

$$\pi(\mathbf{n}) = \pi(\mathbf{0})\Phi(\mathbf{n}) \prod_{l=1}^k \frac{\rho_l^{n_l}}{n_l!}, \quad \mathbf{n} \in \mathcal{S}, \quad (4.21)$$

όπου το $\pi(\mathbf{0})$ καθορίζεται από τη συνθήκη κανονικοποίησης. Οι πιθανότητες αυτές είναι αναλλοίωτες ως προς τις κατανομές των απαιτήσεων εξυπηρέτησης πέρα από τη μέση τιμή τους [11].

Όταν το αρχικό σύστημα διαθέτει την ιδιότητα ισορροπίας, αυτή συνοπάχει και στο τροποποιημένο σύστημα με ένα μόνιμο πελάτη κλάσης i . Πράγματι, δεδομένου ότι $\phi_j^{(i)}(\mathbf{n}) = \phi_j(\mathbf{n} + \mathbf{e}_i)$, είναι εύκολο να δούμε ότι οι συναρτήσεις ρυθμών εξυπηρέτησης $\phi_j^{(i)}(\cdot)$ διαθέτουν επίσης την ιδιότητα ισορροπίας, με συνάρτηση ισορροπίας που σχετίζεται με το αρχικό σύστημα μέσω της $\Phi^{(i)}(\mathbf{n}) = \Phi(\mathbf{n} + \mathbf{e}_i)/\phi_i(\mathbf{e}_i)$.

Εφαρμόζοντας το γεγονός αυτό και σε συνεργασία με τις (4.13) και (4.21), γίνεται εφικτό να εκφραστούν οι πιθανότητες σταθερής κατάστασης του τροποποιημένου συστήματος σε όρους των αντίστοιχων πιθανοτήτων του αρχικού συστήματος ως

$$\pi^{(i)}(\mathbf{n}) = \pi^{(i)}(\mathbf{0})\Phi^{(i)}(\mathbf{n}) \prod_{l=1}^k \frac{\rho_l^{n_l}}{n_l!} = \pi^{(i)}(\mathbf{0}) \frac{\Phi(\mathbf{n} + \mathbf{e}_i)}{\phi_i(\mathbf{e}_i)} \prod_{l=1}^k \frac{\rho_l^{n_l}}{n_l!} = c(n_i+1)\pi(\mathbf{n} + \mathbf{e}_i) = cq_i(\mathbf{n}),$$

όπου $c = \pi^{(i)}(\mathbf{0})/(\pi(\mathbf{0})\rho_i\phi_i(\mathbf{e}_i))$. Αντίστοιχα, σε διανυσματική μορφή έχουμε ότι $\pi^{(i)} = c\mathbf{q}_i$. Έχει ήδη αποδειχθεί ότι $\mathbf{q}_i\mathbf{1} = \mathbb{E}[N_i]$ (σχέση (4.14)), όπου $\mathbb{E}[N_i]$ είναι ο μέσος αριθμός πελατών κλάσης i που είναι παρόντες στο αρχικό σύστημα. Το γεγονός αυτό παρέχει μία εναλλακτική έκφραση για τη σταθερά c , οδηγώντας στο ότι $\pi^{(i)} = (\mathbb{E}[N_i])^{-1}\mathbf{q}_i$.

Με εφαρμογή του αποτελέσματος αυτού στην έκφραση για το άνω φράγμα της Πρότασης 2 μαζί με τις (4.7), (4.21) και το νόμο του Little, οδηγούμαστε στο εξής:

$$\begin{aligned}\mu_i\bar{\phi}_i &= \mu_i \sum_{\mathbf{n} \in \mathcal{S}^{(i)}} \pi^{(i)}(\mathbf{n})\phi_i(\mathbf{n} + \mathbf{e}_i) = \frac{\mu_i}{\mathbb{E}[N_i]} \sum_{\mathbf{n} \in \mathcal{S}^{(i)}} (n_i + 1)\pi(\mathbf{n} + \mathbf{e}_i)\phi_i(\mathbf{n} + \mathbf{e}_i) = \\ &= \frac{\mu_i\rho_i}{\mathbb{E}[N_i]} \sum_{\mathbf{n} \in \mathcal{S} - \mathcal{C}_i} \pi(\mathbf{n}) = \frac{\lambda_i(1 - P_{B,i})}{\mathbb{E}[N_i(t)]} \sum_{\mathbf{n} \in \mathcal{S} - \mathcal{C}_i} \nu_i(\mathbf{n}) = \frac{\lambda_i(1 - P_{B,i})}{\mathbb{E}[N_i(t)]} = \frac{1}{\mathbb{E}[V_i]}.\end{aligned}$$

Έτσι, παρατηρούμε ότι όταν το σύστημα διαθέτει την ιδιότητα ισορροπίας, ο ασυμπτωτικός εκθετικός ρυθμός φθίσης $\sigma_1(B_i)$ των πιθανοτήτων ουράς του χρόνου απόκρισης για πελάτες της κλάσης i φράσσεται άνω από το αντίστροφο του μέσου χρόνου απόκρισης (για πελάτες της ίδιας κλάσης).

Τα συστήματα με ιδιότητα ισορροπίας σχετίζονται με μία χρονικά αντιστρέψιμη αλυσίδα Markov. Κάνοντας χρήση των (3.6) και (4.21) μπορεί να δει κανείς ότι

$$\pi(\mathbf{n})M(\mathbf{n}, \mathbf{n} \pm \mathbf{e}_j) = \pi(\mathbf{n} \pm \mathbf{e}_j)M(\mathbf{n} \pm \mathbf{e}_j, \mathbf{n}), \quad \forall \mathbf{n} \in \mathcal{S}, j \in \mathcal{K}.$$

Εφόσον το τροποποιημένο σύστημα με ένα πελάτη κλάσης i διαθέτει επίσης την ιδιότητα ισορροπίας, η ανάλογη συνθήκη χρονικής αντιστρεψιμότητας ισχύει και για τον $M^{(i)}$, υπονοώντας ότι ο πίνακας $\Pi^{(i)1/2}M^{(i)}\Pi^{(i)-1/2}$, όπου $\Pi^{(i)} \triangleq \text{diag}_{\mathbf{n}}\{\pi^{(i)}(\mathbf{n})\}$, είναι συμμετρικός. Υπενθυμίζοντας ότι ο πίνακας $R^{(i)}$ είναι διαγώνιος, έχουμε ότι ο $\Pi^{(i)1/2}B_i\Pi^{(i)-1/2} = R^{(i)} - \Pi^{(i)1/2}M^{(i)}\Pi^{(i)-1/2}$ είναι επίσης συμμετρικός και ως εκ τούτου διαγωνιοποιήσιμος, έχοντας όλες τις ιδιοτιμές του πραγματικές και πλήρες σύνολο ιδιοδιανυσμάτων, με τα αριστερά ιδιοδιανύσματα να είναι ίσα με τα αντίστοιχα ανάστροφα δεξιά ιδιοδιανύσματα. Η εφαρμογή κάποιου κατάλληλου μετασχηματισμού ομοιότητας έχει ως αποτέλεσμα ότι ο B_i είναι επίσης διαγωνιοποιήσιμος με όλες τις ιδιοτιμές πραγματικές και αριστερά και δεξιά ιδιοδιανύσματα (κανονικοποιημένα κατάλληλα ώστε να έχουν εσωτερικό γινόμενο ίσο με 1) που συνδέονται σύμφωνα με τη σχέση $\mathbf{y}_j(B_i) = (\mathbf{x}_j(B_i)^T \Pi^{(i)} \mathbf{x}_j(B_i))^{-1} \mathbf{x}_j(B_i)^T \Pi^{(i)}$.

Έτσι, η ανάπτυξη της (4.15) σε συστήματα με την ιδιότητα ισορροπίας είναι πάντοτε εφικτή. Επίσης, οι παράγοντες $a_j(B_i)$ εκεί μπορούν να εκφραστούν

σε όρους μόνο των δεξιών ιδιοδιανυσμάτων ως

$$a_j(B_i) = \frac{(\boldsymbol{\nu}_i \mathbf{x}_j(B_i))(\mathbf{x}_j(B_i)^T \Pi^{(i)} \mathbf{1})}{\mathbf{x}_j(B_i)^T \Pi^{(i)} \mathbf{x}_j(B_i)}.$$

Η έκφραση αυτή μπορεί να απλοποιηθεί περαιτέρω, παρατηρώντας ότι $\mathbf{x}_j(B_i)^T \Pi^{(i)} \mathbf{1} = \mathbf{1}^T \Pi^{(i)} \mathbf{x}_j(B_i) = \boldsymbol{\pi}^{(i)} \mathbf{x}_j(B_i)$, πολλαπλασιάζοντας από αριστερά τη $B_i x_j(B_i) = \sigma_j(B_i) x_j(B_i)$ με $\boldsymbol{\pi}^{(i)}$ και εφαρμόζοντας την (4.9), ώστε να λάβουμε $\boldsymbol{\pi}^{(i)} \mathbf{x}_j(B_i) = (\mu_i / \sigma_j(B_i)) \boldsymbol{\pi}^{(i)} R^{(i)} \mathbf{x}_j(B_i)$. Ταυτόχρονα, η εφαρμογή της σχέσης $\boldsymbol{\pi}^{(i)} = (\mathbb{E}[N_i])^{-1} \mathbf{q}_i$, οι εξισώσεις (4.2), (4.7), (4.13), (4.21) και ο νόμος του Little δείχνουν ότι

$$\boldsymbol{\pi}^{(i)} R^{(i)} = \frac{1}{\mu_i \mathbb{E}[V_i]} \boldsymbol{\nu}_i.$$

Έτσι $\boldsymbol{\pi}^{(i)} \mathbf{x}_j(B_i) = (\sigma_j(B_i) \mathbb{E}[V_i])^{-1} \boldsymbol{\nu}_i \mathbf{x}_j(B_i)$ και με αντικατάσταση στην έκφραση για τους παράγοντες οδηγούμαστε τελικά στο

$$a_j(B_i) = \frac{1}{\sigma_j(B_i) \mathbb{E}[V_i]} \frac{(\boldsymbol{\nu}_i \mathbf{x}_j(B_i))^2}{\mathbf{x}_j(B_i)^T \Pi^{(i)} \mathbf{x}_j(B_i)}.$$

Η έκφραση αυτή, το γεγονός ότι $\boldsymbol{\pi}^{(i)} = (\mathbb{E}[N_i])^{-1} \mathbf{q}$ και η (4.13) καθιστούν δυνατό να εκφραστούν τα $a_j(B_i)$ σε όρους των πιθανοτήτων σταθερής κατάστασης του αρχικού συστήματος αντί αυτών του τροποποιημένου. Επιπρόσθετα, η παραπάνω έκφραση δείχνει ότι όλοι οι παράγοντες είναι θετικοί. Κάνοντας χρήση αυτού του γεγονότος, του ότι $\sum_j a_j(B_i) = 1$, των ανισοτήτων $\sigma_1(B_i) < \sigma_j(B_i)$, $j \neq 1$ και ολοκληρώνοντας το $S_i(t)$ στην (4.15) οδηγούμαστε σε ένα εναλλακτικό τρόπο για να φτάσουμε στο φράγμα $\sigma_1(B_i) < 1/\mathbb{E}[V_i]$ που συζητήθηκε νωρίτερα.



Κεφάλαιο 5

Διαχωρισμός Χρονικής Κλίμακας

Το παρόν κεφάλαιο αφιερώνεται στη μελέτη συστημάτων SDPS στα οποία οι κλάσεις των πελατών διαθέτουν δυναμική που μπορεί να διαχωριστεί ως προς διαφορετικές χρονικές κλίμακες. Όπως αναφέρθηκε και στο Κεφάλαιο 2, τέτοιου είδους φαινόμενα εμφανίζονται σε περιπτώσεις ασύρματων δικτύων που εξυπηρετούν τερματικά με μεγάλες διαφορές ρυθμών μετάδοσης. Ένα άλλο συγκεκριμένο παράδειγμα αποτελεί η μοντελοποίηση ροών δεδομένων στο διαδίκτυο με μεγάλες διαφορές μεγέθους. Όπως θα αποδειχθεί στη συνέχεια του κεφαλαίου, η δυναμική του συστήματος μπορεί να προσεγγιστεί από απλούστερα μοντέλα και η ακρίβεια της προσέγγισης εξαρτάται από το βαθμό του διαχωρισμού.

Η ανάλυση της επίδρασης του χρονικού διαχωρισμού καταλήγει σε μειωμένα συστήματα στα οποία οι γρήγοροι πελάτες εκλαμβάνουν την παρουσία των αργών πελατών ως αυτοί να ήταν μόνιμα παρόντες. Από την άλλη μεριά, η δυναμική των αργών πελατών κυβερνάται από ένα συναθροιστικό σύστημα SDPS με τροποποιημένα όμως χαρακτηριστικά κίνησης, τα οποία λαμβάνουν υπόψη τις διαταραχές λόγω των γρήγορων πελατών. Όπως αναφέρθηκε και κατά την περιγραφή της μεθοδολογίας στην Ενότητα 3.2, προς το σκοπό αυτό εφαρμόζονται τεχνικές NCD (βλ. π.χ., [26]) και αποδεικνύεται ότι το σύστημα μπορεί να διαχωριστεί πλήρως σε ομάδες κλάσεων διατηρώντας μία κομψή και εκμεταλλεύσιμη μορφή.

Η ανάλυση είναι σε θέση να δώσει προσεγγιστικά και ασυμπτωτικά αποτελέσματα καθώς και να υπολογίσει την τάξη της προσέγγισης. Για τις γρήγορες κλάσεις αποδεικνύεται ότι η κατανομή της καθυστέρησης υπολογίζεται από ένα σταθμισμένο άθροισμα των κατανομών όλων των πιθανών συνδυασμών αργών πελατών, οι οποίοι «φαίνονται» στους γρήγορους σαν μόνιμοι, συν ένα μόνιμο γρήγορο πελάτη. Ωστόσο, αντί να υποτεθεί *a priori* η μονιμότητα των

αργών πελατών, η παρούσα ανάλυση παρέχει μία επίσημη απόδειξη για το γεγονός αυτό και επίσης βοηθά στον υπολογισμό των κατάλληλων συντελεστών. Για τους πελάτες των αργών κλάσεων παρέχεται ένα ισοδύναμο συναθροιστικό σύστημα όπου συμμετέχουν μόνο οι αργές κλάσεις και οι παράμετροι συνοπολογίζονται μέσω διακυμάνσεων που οφείλονται στους γρήγορους χρήστες. Αξίζει να υπενθυμίσουμε στο σημείο αυτό ότι η πλειοψηφία σχετικών μελετών ακολουθεί την απλουστευτική αλλά μη ρεαλιστική παραδοχή ότι κάθε κλάση λειτουργεί σε διαφορετική κλίμακα οπότε οι κλάσεις κατατάσσονται σύμφωνα με την ταχύτητά τους (βλ. π.χ., [13] [56]). Η κάθε κλάση θεωρείται από τις βραδύτερες ως «ρευστό» και από τις ταχύτερες ως «οιονεί στάσιμη».

Σύμφωνα με το διαχωρισμό των κλάσεων που εισήχθη στην Ενότητα 2.4 και επανεισάγεται εδώ για τη γενικότερη μορφή του προβλήματος, το σύνολο των κλάσεων \mathcal{K} διαχωρίζεται σε δύο ξένα μεταξύ τους υποσύνολα \mathcal{K}_f και \mathcal{K}_s , με γρήγορη και αργή δυναμική αντίστοιχα. Εντός κάθε υποσυνόλου (ομάδας κλάσεων), οι ρυθμοί αφίξεων Poisson $\lambda_j(\cdot)$ και οι παράμετροι απαιτήσεων εξυπηρέτησης μ_j είναι της ίδιας τάξης μεγέθους για όλες τις κλάσεις. Η διαφοροποίηση μεταξύ υποσυνόλων λαμβάνει χώρα απαιτώντας ότι

$$\frac{\max_{i \in \mathcal{K}_s} \max_{\mathbf{n} \in \mathcal{S}} \lambda_i(\mathbf{n})}{\min_{i \in \mathcal{K}_f} \min_{\mathbf{n} \in \mathcal{S}: \lambda_i(\mathbf{n}) > 0} \lambda_i(\mathbf{n})} = O(\epsilon) \quad \text{και} \quad \frac{\max_{i \in \mathcal{K}_s} \mu_i}{\min_{i \in \mathcal{K}_f} \mu_i} = O(\epsilon), \quad (5.1)$$

όπου ο παράγοντας σύζευξης ϵ ποσοτικοποιεί το διαχωρισμό χρονικής κλίμακας καθώς τείνει στο μηδέν. Ας σημειωθεί ότι παρόλο που οι παράμετροι απαιτήσεων εξυπηρέτησης για όλες τις κλάσεις του \mathcal{K}_s είναι τάξης $O(\epsilon)$, οι αντίστοιχες εντάσεις κίνησης $E[\lambda_j(\mathbf{N}(t))]/\mu_j$ παραμένουν τάξης $O(1)$ καθώς $\epsilon \rightarrow 0$. Επίσης, όπου απαιτείται, επιστρέφουμε στο συμβολισμό του διανύσματος κατάστασης και των συναρτήσεων του που είναι σύμμορφος με το διαχωρισμό των κλάσεων, δηλαδή $\mathbf{n} = (\mathbf{n}_s, \mathbf{n}_f)$, $f(\mathbf{n}) = f(\mathbf{n}_s, \mathbf{n}_f)$ κ.ο.κ.

Παρακάτω εισάγουμε σύνολα που περιέχουν υποδιανύσματα, όπως αυτά απαιτούνται από τη δομή του χώρου καταστάσεων του συστήματος. Το σύνολο όλων των εφικτών διανυσμάτων των αργών κλάσεων είναι

$$\bar{\mathcal{S}} \triangleq \{ \mathbf{n}_s \geq \mathbf{0} \mid \exists \mathbf{n}_f \geq \mathbf{0} : (\mathbf{n}_s, \mathbf{n}_f) \in \mathcal{S} \} = \{ \mathbf{n}_s \geq \mathbf{0} \mid (\mathbf{n}_s, \mathbf{0}) \in \mathcal{S} \},$$

όπου η δεύτερη ισότητα ακολουθεί από την Ιδιότητα 2. Αργότερα στο κεφάλαιο αυτό θα συναντήσουμε ένα κατάλληλα ορισμένο σύστημα SDPS που εξυπηρετεί αποκλειστικά πελάτες των αργών κλάσεων και με $\bar{\mathcal{S}}$ θα συμβολίζεται ο χώρος καταστάσεων αυτού. Σε αυτό το πνεύμα, άλλα παραγόμενα σύνολα όπως το σύνορο $\bar{\mathcal{C}}_i$ ως προς την κλάση $i \in \mathcal{K}_s$ και το υποσύνολο $\bar{\mathcal{S}}^{(i)} = \bar{\mathcal{S}} - \bar{\mathcal{C}}_i$ που μπορεί να υποδεχτεί ένα ακόμα πελάτη κλάσης $i \in \mathcal{K}_s$, θα ορίζονται σε αναλογία με τους αντίστοιχους ορισμούς για τους χώρους καταστάσεων \mathcal{S} του

αρχικού συστήματος της Ενότητας 3.3. Παράλληλα, θα χρειαστούμε το σύνολο των υποδιανυσμάτων αργών κλάσεων που είναι εφικτά όταν υπάρχει στο σύστημα επιπρόσθετα ένας πελάτης από μία *γρήγορη* κλάση, δηλαδή

$$\bar{\mathcal{S}}^{(i)} = \{ \mathbf{n}_s \in \bar{\mathcal{S}} \mid (\mathbf{n}_s, \mathbf{e}_i) \in \mathcal{S} \}, \text{ για } i \in \mathcal{K}_f.$$

Ας θεωρήσουμε τώρα το σύνολο $\hat{\mathcal{S}}^{(n_s)}$ των υποδιανυσμάτων γρήγορων κλάσεων που είναι εφικτά όταν ένας πληθυσμός $\mathbf{n}_s \in \bar{\mathcal{S}}$ από πελάτες αργών κλάσεων είναι επίσης μόνιμα παρόντες στο σύστημα, δηλαδή

$$\hat{\mathcal{S}}^{(n_s)} \triangleq \{ \mathbf{n}_f \geq \mathbf{0} \mid (\mathbf{n}_s, \mathbf{n}_f) \in \mathcal{S} \}, \quad \mathbf{n}_s \in \bar{\mathcal{S}}.$$

Κάνοντας χρήση της Ιδιότητας 2, για κάθε $\mathbf{n}_1, \mathbf{n}_2 \in \bar{\mathcal{S}}$ με $\mathbf{n}_1 \leq \mathbf{n}_2$, έχουμε ότι $\hat{\mathcal{S}}^{(n_2)} \subseteq \hat{\mathcal{S}}^{(n_1)}$. Σύντομα θα γίνει χρήση «μειωμένων» υποσυστημάτων SDPS που εξυπηρετούν πελάτες μόνο από τη γρήγορη ομάδα κλάσεων \mathcal{K}_f . Τα συστήματα αυτά χαρακτηρίζονται από χώρο καταστάσεων, ρυθμούς αφίξεων Poisson και ρυθμούς εξυπηρέτησης που ορίζονται κατάλληλα ώστε να αντιστοιχούν στο αρχικό πλήρες σύστημα SDPS, όταν αυτό λειτουργεί υπό την μόνιμη παρουσία ενός αργού πληθυσμού \mathbf{n}_s . Με άλλα λόγια, το μειωμένο σύστημα με δείκτη \mathbf{n}_s θα χαρακτηρίζεται από χώρο καταστάσεων $\hat{\mathcal{S}}^{(n_s)}$ και σύμφωνα με την (3.8), ρυθμούς αφίξεων Poisson και ρυθμούς εξυπηρέτησης

$$\hat{\lambda}_i^{(n_s)}(\mathbf{n}_f) \triangleq \lambda_i(\mathbf{n}_s, \mathbf{n}_f) \quad \text{and} \quad \hat{\phi}_i^{(n_s)}(\mathbf{n}_f) \triangleq \phi_i(\mathbf{n}_s, \mathbf{n}_f) \quad (5.2)$$

αντίστοιχα, για όλα τα $i \in \mathcal{K}_f$ και όλα τα $\mathbf{n}_f \in \hat{\mathcal{S}}^{(n_s)}$. Ο συμβολισμός που εφαρμόστηκε παραπάνω επεκτείνεται και σε άλλες ποσότητες που εισήχθησαν στα Κεφάλαια 3 και 4. Πιο συγκεκριμένα, δεδομένης μίας ποσότητας x που αναφέρεται στο αρχικό σύστημα, ο συμβολισμός $\hat{x}^{(n_s)}$ θα αναφέρεται στην αντίστοιχη ποσότητα που συνδέεται με το μειωμένο σύστημα που διαθέτει \mathbf{n}_s επιπλέον μόνιμους αργούς πελάτες.

5.1 Η Σχεδόν Πλήρως Αποσυζεύξιμη δομή του πίνακα ρυθμών μεταβάσεων

Ο διαχωρισμός χρονικής κλίμακας έχει επίδραση στη δομή του απειροστικού γεννήτορα του συστήματος M , η οποία γίνεται περισσότερο αισθητή αν ο απειροστικός γεννήτορας γραφεί (πιθανά μετά από αναδιάταξη των στοιχείων του) σε μορφή υποπινάκων $M = [M_{\mathbf{n}_s, \mathbf{n}'_s}]$. Κάθε υποπίνακας $M_{\mathbf{n}_s, \mathbf{n}'_s}$ έχει μέγεθος $\text{card}(\hat{\mathcal{S}}^{(n_s)}) \times \text{card}(\hat{\mathcal{S}}^{(n'_s)})$ και τα στοιχεία του $M_{\mathbf{n}_s, \mathbf{n}'_s}(\mathbf{n}_f, \mathbf{n}'_f)$ αντιστοιχούν σε μεταβάσεις από το $(\mathbf{n}_s, \mathbf{n}_f)$ προς το $(\mathbf{n}'_s, \mathbf{n}'_f)$. Εφόσον οι μεταβάσεις της αλυσίδας Markov συνεχούς χρόνου σχετίζονται με μεταβολές ενός πελάτη και

μόνο, οι υποπίνακες των μη διαγώνιων θέσεων είναι μη μηδενικοί μόνο όταν $\mathbf{n}'_s = \mathbf{n}_s \pm \mathbf{e}_i$, $i \in \mathcal{K}_s$ και αυτοί οι μη μηδενικοί υποπίνακες χαρακτηρίζονται από μη μηδενικά στοιχεία μόνο κατά μήκος της κύριας διαγωνίου τους (δηλαδή μόνο τα $M_{\mathbf{n}_s, \mathbf{n}_s \pm \mathbf{e}_i}(\mathbf{n}_f, \mathbf{n}_f)$ είναι μη μηδενικά). Επιπρόσθετα, με χρήση της (5.1), έπεται ότι $M_{\mathbf{n}_s, \mathbf{n}_s \pm \mathbf{e}_i} = O(\epsilon)$. Οι υποπίνακες κατά μήκος της διαγωνίου, $M_{\mathbf{n}_s, \mathbf{n}_s}$ αντιστοιχούν σε μεταβάσεις λόγω μεταβολών στον πληθυσμό των πελατών μόνο των κλάσεων της γρήγορης ομάδας \mathcal{K}_f , ενώ ο πληθυσμός των πελατών των αργών κλάσεων παραμένει σταθερός και ίσος με \mathbf{n}_s . Ως αποτέλεσμα, τα μη διαγώνια στοιχεία του $M_{\mathbf{n}_s, \mathbf{n}_s}$ είναι ίσα με τα μη διαγώνια στοιχεία του αντίστοιχου $\hat{M}^{(\mathbf{n}_s)}$, του απειροστικού γεννήτορα που αντιστοιχεί στο μειωμένο σύστημα με δείκτη \mathbf{n}_s . Τα διαγώνια στοιχεία του $M_{\mathbf{n}_s, \mathbf{n}_s}$ διαφέρουν κατά τάξη $O(\epsilon)$ από τα αντίστοιχα στοιχεία του $\hat{M}^{(\mathbf{n}_s)}$, έτσι ώστε να λάβουν υπόψη τα θετικά αθροίσματα γραμμών των εκτός της διαγωνίου υποπινάκων $M_{\mathbf{n}_s, \mathbf{n}_s \pm \mathbf{e}_i}$, $i \in \mathcal{K}_s$, και να καταστήσουν το άθροισμα κάθε γραμμής του ίσο με μηδέν. Κατά συνέπεια ο M παίρνει την προσεγγιστικά μπλοκ διαγώνια μορφή

$$M = \text{diag}\{\hat{M}^{(\mathbf{n}_s)}\}_{\mathbf{n}_s \in \mathcal{S}} + O(\epsilon), \quad (5.3)$$

υποδεικνύοντας ότι το σύστημα είναι σχεδόν πλήρως αποσυσζεύξιμο (NCD) προς ένα σύνολο μειωμένων συστημάτων.

Γενικότερα, η ιδιότητα NCD εμφανίζεται σε μοντέλα που διαθέτουν ένα χώρο καταστάσεων που μπορεί να επιμερισθεί σε ένα αριθμό υποχώρων, έτσι ώστε οι ισχυρές αλληλεπιδράσεις (υψηλοί ρυθμοί μεταβάσεων) να περιορίζονται εντός των υποχώρων, ενώ οι ασθενείς αλληλεπιδράσεις (χαμηλοί ρυθμοί μεταβάσεων) να λαμβάνουν χώρα μεταξύ υποχώρων. Για μαρκοβιανά μοντέλα, ο πίνακας ρυθμών μεταβάσεων εμφανίζει μία ειδική δομή, έτσι ώστε οι υποπίνακες κατά μήκος της κύριας διαγωνίου εκφράζουν τις ισχυρές αλληλεπιδράσεις εντός υποχώρων και οι εκτός διαγωνίου υποπίνακες έχουν μικρά στοιχεία, εκφράζοντας τις ασθενείς αλληλεπιδράσεις μεταξύ υποχώρων. Κάθε υποχώρος κυβερνάται από μία μειωμένη αλυσίδα Markov που μπορεί να μελετηθεί σε απομόνωση. Η διαδικασία αυτή ονομάζεται αποσύνθεση. Η ασθενής σύνδεση των υποχώρων λαμβάνεται υπόψη μέσω μιας συμπληρωματικής διαδικασίας συνάθροισης [26]. Τα στοιχεία του συναθροιστικού πίνακα ρυθμών μεταβάσεων αναπαριστούν σύνθετες μεταβάσεις μεταξύ διαφορετικών υποχώρων και λαμβάνονται σταθμίζοντας τους ρυθμούς μεταβάσεων μεταξύ ξεχωριστών καταστάσεων των υποχώρων του πλήρους συστήματος. Εκ κατασκευής, το συναθροιστικό σύστημα χαρακτηρίζεται από μία κατάσταση ανά υποχώρο του αρχικού μοντέλου.

Εκφράζοντας τις έννοιες που μόλις συζητήθηκαν σε μαθηματικό συμβολισμό, έστω ότι $A = [A_{IJ}]$, $1 \leq I, J \leq K$ είναι ένας απειροστικός γεννή-

τορας με δομή NCD και διάνυσμα πιθανοτήτων σταθερής κατάστασης $\pi = (\pi_1, \dots, \pi_K)$, διαχωρισμένο σε υποδιανύσματα σύμμορφα με τη μπλοκ δομή του πίνακα. Από την ιδιότητα NCD έχουμε ότι $A_{II} = \hat{A}_I + O(\epsilon)$ και $A_{IJ} = O(\epsilon)$ για $I \neq J$, και έτσι $A = \text{diag}_I \{\hat{A}_I\} + O(\epsilon)$, όπου όλοι οι \hat{A}_I , $I = 1, \dots, K$, είναι μη αναγώγιμοι απειροστικοί γεννήτορες με μειωμένες διαστάσεις. Δεδομένης της δομής αυτής, τα υποδιανύσματα π_I του πλήρους διανύσματος πιθανοτήτων είναι προσεγγιστικά ανάλογα με τα διανύσματα σταθερής κατάστασης $\hat{\pi}_I$ των σχεδόν αποσυζεύξιμων μειωμένων αλυσίδων \hat{A}_I .

Οι σταθερές αναλογίας καθορίζονται από τον $K \times K$ συναθροιστικό πίνακα \bar{A} , με στοιχεία $\bar{A}(I, J) \triangleq \hat{\pi}_I A_{IJ} \mathbf{1}$, όλα από τα οποία είναι τάξης $O(\epsilon)$ και εκφράζουν την ασθενή σύζευξη στο σύστημα NCD. Εκ κατασκευής, ο \bar{A} είναι μη αναγώγιμος απειροστικός γεννήτορας. Δεδομένου του διανύσματος πιθανοτήτων σταθερής κατάστασης που του αντιστοιχεί, $\bar{\pi} = (\bar{\pi}(1), \dots, \bar{\pi}(K))$, μπορεί να αποδειχθεί ότι $\pi_I = \bar{\pi}(I) \hat{\pi}_I + O(\epsilon)$ καθώς $\epsilon \rightarrow 0$, για όλα τα $I = 1, \dots, K$ (βλ. π.χ., [26]).

Προκειμένου να εφαρμόσουμε τη γενική μεθοδολογία αποσύνθεσης και συνάθροισης στο παρόν πρόβλημα, υπενθυμίζουμε ότι τα σχεδόν αποσυζεύξιμα υποσυστήματα της συγκεκριμένης αλυσίδας στην (5.3) είναι τα μειωμένα υποσυστήματα SDPS με δείκτη \mathbf{n}_s , για $\mathbf{n}_s \in \bar{\mathcal{S}}$. Κάθε τέτοιο σύστημα αντιμετωπίζει κατά κάποιο τρόπο τις αργές κλάσεις συνολικά σαν οιονεί στάσιμες. Επίσης, με βάση τα σχόλια που προηγήθηκαν σχετικά με την μπλοκ δομή του M , καταλήγουμε στο ότι τα μόνα μη διαγώνια στοιχεία του συναθροιστικού απειροστικού γεννήτορα \bar{M} που είναι μη μηδενικά είναι τα $\bar{M}(\mathbf{n}_s, \mathbf{n}_s \pm \mathbf{e}_i)$, $i \in \mathcal{K}_s$ με μορφή

$$\bar{M}(\mathbf{n}_s, \mathbf{n}_s + \mathbf{e}_i) = \hat{\pi}^{(\mathbf{n}_s)} M_{\mathbf{n}_s, \mathbf{n}_s + \mathbf{e}_i} \mathbf{1} = \sum_{\mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s + \mathbf{e}_i)} \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f) \lambda_i(\mathbf{n}_s, \mathbf{n}_f), \quad \mathbf{n}_s \in \bar{\mathcal{S}} - \bar{\mathcal{C}}_i,$$

και

$$\bar{M}(\mathbf{n}_s, \mathbf{n}_s - \mathbf{e}_i) = \hat{\pi}^{(\mathbf{n}_s)} M_{\mathbf{n}_s, \mathbf{n}_s - \mathbf{e}_i} \mathbf{1} = n_i \mu_i \sum_{\mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s)} \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f) \phi_i(\mathbf{n}_s, \mathbf{n}_f), \quad \mathbf{n}_s \in \bar{\mathcal{S}}, \quad n_i > 0.$$

Κατά συνέπεια, ο συναθροιστικός πίνακας αντιστοιχεί σε ένα σύστημα SDPS που εξυπηρετεί πελάτες μόνο των αργών κλάσεων \mathcal{K}_s , χαρακτηρίζεται από χώρο καταστάσεων $\bar{\mathcal{S}}$, πίνακα ρυθμών μεταβάσεων της μορφής (3.6) και διαθέτει ρυθμούς αφίξεων Poisson και εξυπηρέτησης για $i \in \mathcal{K}_s$

$$\bar{\lambda}_i(\mathbf{n}_s) \triangleq \sum_{\mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s + \mathbf{e}_i)} \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f) \lambda_i(\mathbf{n}_s, \mathbf{n}_f) \quad \text{και} \quad \bar{\phi}_i(\mathbf{n}_s) \triangleq \sum_{\mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s)} \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f) \phi_i(\mathbf{n}_s, \mathbf{n}_f). \quad (5.4)$$

Μπορεί να δει κανείς ότι οι ρυθμοί αυτοί προκύπτουν σταθμίζοντας τις διακυμάνσεις λόγω των μεταβολών στους πελάτες των γρήγορων κλάσεων, σε σχέση

με το διάνυσμα πιθανοτήτων σταθερής κατάστασης $\hat{\pi}^{(n_s)}$ που αντιστοιχεί στο μειωμένο υποσύστημα με δείκτη n_s . Με άλλα λόγια, το συναθροιστικό σύστημα θεωρεί τις γρήγορες κλάσεις συνολικά σαν ρευστό.

Είναι φανερό ότι οποιοδήποτε από τα μειωμένα συστήματα SDPS που αντιστοιχεί στις αλυσίδες $\hat{M}^{(n_s)}$, όπως επίσης και το σύστημα που αντιστοιχεί στον συναθροιστικό πίνακα \bar{M} , μπορεί να θεωρηθεί και να εξεταστεί ανεξάρτητα. Τα αποτελέσματα του Κεφαλαίου 4 (και συγκεκριμένα το Θεώρημα 2) μπορούν να χρησιμοποιηθούν για τον υπολογισμό των αντίστοιχων CPDF των χρόνων απόκρισης, δηλαδή τις $\hat{S}_i^{(n_s)}(t)$ για τα μειωμένα συστήματα (με $i \in \mathcal{K}_f$) και $\bar{S}_i(t)$ για το συναθροιστικό σύστημα (με $i \in \mathcal{K}_s$). Όπως θα δούμε αργότερα, η κατανομή $S_i(t)$, δηλαδή η CPDF του χρόνου απόκρισης που αντιστοιχεί σε πελάτες της κλάσης i του αρχικού συστήματος, μπορεί να προσεγγιστεί με χρήση των απλούστερων CPDF που μόλις αναφέραμε, με τρόπο που εξαρτάται από τη δυναμική της κλάσης i .

Η εξίσωση (5.4) υποδηλώνει ότι οι ρυθμοί αφίξεων Poisson στο συναθροιστικό σύστημα μπορούν να εξαρτώνται από την κατάσταση του συστήματος, ακόμα και αν στο αρχικό σύστημα οι αφίξεις χαρακτηρίζονται από ομοιογένεια ως προς την κατάσταση. Πράγματι, για πλήρως διαμοιραζόμενο θάλαμο εξυπηρέτησης, δηλαδή χώρο καταστάσεων \mathcal{S} της μορφής (3.3), $\hat{\mathcal{S}}^{(n_s)} - \hat{\mathcal{S}}^{(n_s+e_i)} = \{\mathbf{n}_f \geq \mathbf{0} \mid |\mathbf{n}_f| = L - |\mathbf{n}_s|\} \neq \emptyset$ και έτσι, ακόμα και ομοιογενής ρυθμός αφίξεων λ_i του αρχικού συστήματος οδηγεί στους εξαρτώμενους από την κατάσταση ρυθμούς $\bar{\lambda}_i(\mathbf{n}_s) = \lambda_i(1 - \sum_{\mathbf{n}_f \in \hat{\mathcal{S}}^{(n_s)} - \hat{\mathcal{S}}^{(n_s+e_i)}} \hat{\pi}^{(n_s)}(\mathbf{n}_f))$ για το συναθροιστικό σύστημα. Το φαινόμενο αυτό δεν υφίσταται σε θαλάμους εξυπηρέτησης με διαχωρισμό των ομάδων των κλάσεων, δηλαδή για χώρους της μορφής (3.5). Στην περίπτωση αυτή $\bar{\mathcal{S}} = \mathcal{S}_s$ και $\hat{\mathcal{S}}^{(n_s)} = \mathcal{S}_f$, για όλα τα $\mathbf{n}_s \in \bar{\mathcal{S}}$. Η δυνατότητα του συναθροιστικού συστήματος να χαρακτηρίζεται από ρυθμούς αφίξεων Poisson που εξαρτώνται από την κατάσταση, ακόμα και αν αυτό το γνώρισμα δεν υφίσταται στο αρχικό σύστημα, ήταν το βασικό κίνητρο ώστε να υποθεθεί εξ αρχής το πιο γενικό μοντέλο με εξάρτηση του ρυθμού αφίξεων από την κατάσταση.

Δεδομένου του γενικού αποτελέσματος για συστήματα NCD, οι πιθανότητες σταθερής κατάστασης που αντιστοιχούν στον M μπορούν άμεσα να προσεγγιστούν ως $\pi(\mathbf{n}_s, \mathbf{n}_f) = \bar{\pi}(\mathbf{n}_s)\hat{\pi}^{(n_s)}(\mathbf{n}_f) + O(\epsilon)$ και σε συνδυασμό με την (4.5),

$$\nu_i(\mathbf{n}_s, \mathbf{n}_f) = \frac{\bar{\pi}(\mathbf{n}_s)\hat{\pi}^{(n_s)}(\mathbf{n}_f)\lambda_i(\mathbf{n}_s, \mathbf{n}_f)}{\sum_{(\mathbf{n}'_s, \mathbf{n}'_f) \in \mathcal{S}^{(i)}} \bar{\pi}(\mathbf{n}'_s)\hat{\pi}^{(n'_s)}(\mathbf{n}'_f)\lambda_i(\mathbf{n}'_s, \mathbf{n}'_f)} + O(\epsilon).$$

(Το κλάσμα σε αυτή την εξίσωση είναι πάντα τάξης $O(1)$, ακόμα και για $i \in \mathcal{K}_s$, επειδή σε αυτήν την περίπτωση οι παράγοντες τάξης $O(\epsilon)$ στα $\lambda_i(\cdot, \cdot)$ εμφανίζονται και στον αριθμητή και στον παρονομαστή και άρα απλοποιούνται.) Επι-

πρόσθετα, για δείκτες κλάσεων $i \in \mathcal{K}_f$ μπορεί κανείς να εφαρμόσει την (5.2) οπότε το άθροισμα στον παρονομαστή γίνεται ίσο με

$$\sum_{\mathbf{n}'_s \in \bar{\mathcal{S}}^{(i)}} \bar{\pi}(\mathbf{n}'_s) \sum_{\mathbf{n}'_f \in \hat{\mathcal{S}}(\mathbf{n}'_s) - \hat{\mathcal{C}}_i(\mathbf{n}'_s)} \hat{\pi}(\mathbf{n}'_s)(\mathbf{n}'_f) \hat{\lambda}_i(\mathbf{n}'_s)(\mathbf{n}'_f) = \sum_{\mathbf{n}'_s \in \bar{\mathcal{S}}^{(i)}} \bar{\pi}(\mathbf{n}'_s) \hat{\lambda}_{\text{av},i}(\mathbf{n}'_s),$$

όπου η εμφάνιση του $\hat{\lambda}_{\text{av},i}(\mathbf{n}'_s)$, του μέσου ρυθμού μη αποκλειόμενων αφίξεων κλάσης i στο υποσύστημα με δείκτη \mathbf{n}'_s , προκύπτει από την (4.6). Παρόμοια, εφαρμόζοντας τις (4.5) και (4.6) για το υποσύστημα με δείκτη \mathbf{n}_s , ο αριθμητής προκύπτει ίσος με $\bar{\pi}(\mathbf{n}_s) \hat{\lambda}_{\text{av},i}(\mathbf{n}_s) \hat{\nu}(\mathbf{n}_s)(\mathbf{n}_f)$ και έτσι τελικά,

$$\nu_i(\mathbf{n}_s, \mathbf{n}_f) = \bar{\gamma}_i(\mathbf{n}_s) \hat{\nu}_i(\mathbf{n}_s)(\mathbf{n}_f) + O(\epsilon), \quad i \in \mathcal{K}_f, \quad \mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}, \quad \mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s) - \hat{\mathcal{C}}_i(\mathbf{n}_s), \quad (5.5)$$

με τον ορισμό

$$\bar{\gamma}_i(\mathbf{n}_s) \triangleq \frac{\bar{\pi}(\mathbf{n}_s) \hat{\lambda}_{\text{av},i}(\mathbf{n}_s)}{\sum_{\mathbf{n}'_s \in \bar{\mathcal{S}}^{(i)}} \bar{\pi}(\mathbf{n}'_s) \hat{\lambda}_{\text{av},i}(\mathbf{n}'_s)}. \quad (5.6)$$

Συνεπώς το υποδιάνυσμα του ν_i που αντιστοιχεί στο υποσύστημα με δείκτη \mathbf{n}_s , είναι κατά προσέγγιση ίσο με $\bar{\gamma}_i(\mathbf{n}_s) \hat{\nu}_i(\mathbf{n}_s)$, δηλαδή ανάλογο του διανύσματος αρχικών πιθανοτήτων που αντιστοιχεί στο συγκεκριμένο υποσύστημα.

Παρόμοια διαδικασία μπορεί να ακολουθηθεί και για τις κλάσεις της αργής ομάδας. Ο παρονομαστής του κλάσματος για το $\nu_i(\mathbf{n}_s, \mathbf{n}_f)$ μπορεί και πάλι να γραφεί με όρους δύο αθροισμάτων και το εξωτερικό άθροισμα και πάλι αφορά το σύνολο $\bar{\mathcal{S}}^{(i)}$, όμως τώρα $i \in \mathcal{K}_s$ και συνεπώς $\bar{\mathcal{S}}^{(i)} = \bar{\mathcal{S}} - \bar{\mathcal{C}}_i$. Το εσωτερικό άθροισμα αφορά το σύνολο $\hat{\mathcal{S}}(\mathbf{n}'_s + \mathbf{e}_i)$. Έτσι, κάνοντας χρήση της (5.4), ο παρονομαστής είναι ίσος με $\sum_{\mathbf{n}'_s \in \bar{\mathcal{S}} - \bar{\mathcal{C}}_i} \bar{\pi}(\mathbf{n}'_s) \bar{\lambda}_i(\mathbf{n}'_s)$ και εφαρμογή της (4.5) στο συναθροιστικό σύστημα οδηγεί στο

$$\nu_i(\mathbf{n}_s, \mathbf{n}_f) = \bar{\nu}_i(\mathbf{n}_s) z_{\mathbf{n}_s, i}(\mathbf{n}_f) + O(\epsilon), \quad i \in \mathcal{K}_s, \quad \mathbf{n}_s \in \bar{\mathcal{S}} - \bar{\mathcal{C}}_i, \quad \mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s + \mathbf{e}_i), \quad (5.7)$$

όπου $z_{\mathbf{n}_s, i}(\mathbf{n}_f) \triangleq \hat{\pi}(\mathbf{n}_s)(\mathbf{n}_f) \lambda_i(\mathbf{n}_s, \mathbf{n}_f) / (\sum_{\mathbf{n}'_f \in \hat{\mathcal{S}}(\mathbf{n}_s + \mathbf{e}_i)} \hat{\pi}(\mathbf{n}_s)(\mathbf{n}'_f) \lambda_i(\mathbf{n}_s, \mathbf{n}'_f))$.

5.2 Κατανομή του χρόνου απόκρισης για πελάτες γρήγορων κλάσεων

Ας θεωρήσουμε τον πίνακα ρυθμών μεταβάσεων $M^{(i)}$ που αντιστοιχεί στην τροποποιημένη αλυσίδα Markov, η οποία περιλαμβάνει ένα μόνιμο πελάτη κλάσης i . Με επιχειρήματα ακριβώς ανάλογα εκείνων που χρησιμοποιήθηκαν στην αρχή της Ενότητας 5.1 σχετικά με τη δομή NCD του M , είναι φανερό ότι και ο $M^{(i)}$ έχει δομή NCD, με μορφή $M^{(i)} = \text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{ \hat{M}(\mathbf{n}_s)^{(i)} \} + O(\epsilon)$.

Επίσης, με εξέταση του διαγώνιου πίνακα $R^{(i)}$ που εμφανίζεται στην (4.9) καταλήγει κανείς στο ότι είναι της μορφής $\text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{\hat{R}^{(\mathbf{n}_s)^{(i)}}\}$. Συνεπώς, η (4.9) προτείνει ότι $B_i = \text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{\hat{B}_i^{(\mathbf{n}_s)}\} + O(\epsilon)$. Υπενθυμίζεται ότι οι $\hat{M}^{(\mathbf{n}_s)^{(i)}$, $\hat{R}^{(\mathbf{n}_s)^{(i)}$ και $\hat{B}_i^{(\mathbf{n}_s)}$ σχετίζονται με τα μειωμένα υποσυστήματα με δείκτες \mathbf{n}_s κατά τον ίδιο τρόπο που οι $M^{(i)}$, $R^{(i)}$ και B_i σχετίζονται με το πλήρες σύστημα.

Λόγω της συνέχειας είναι αναμενόμενο ότι καθώς $\epsilon \rightarrow 0$ ο πίνακας $\exp\{-B_i t\} \rightarrow \exp\{\text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{-\hat{B}_i^{(\mathbf{n}_s)} t\}\}$. Το παρακάτω λήμμα θεμελιώνει αυτόν τον ισχυρισμό.

Λήμμα 2. Έστω ότι ο A είναι ένας τετραγωνικός πίνακας με ιδιοτιμές $\sigma_j(A)$, τέτοιες ώστε σε κάθε ιδιοτιμή με το ελάχιστο πραγματικό μέρος, $\sigma_{\min}(A) \triangleq \min_j \text{Re}\{\sigma_j(A)\}$ αντιστοιχεί ένα διακριτό ιδιοδιάνυσμα. Έστω ότι ο E είναι ένας τετραγωνικός πίνακας, ίδιων διαστάσεων με τον A , που χαρακτηρίζεται από $\|E\| = O(\epsilon)$, καθώς $\epsilon \rightarrow 0$. Τότε, $\|e^{-(A+E)t} - e^{-At}\| \leq O(e^{-\sigma_{\min}(A)t}(e^{O(\epsilon)t} - 1)) \leq O(\epsilon t e^{-(\sigma_{\min}(A)+O(\epsilon))t})$, $\forall t \geq 0, \epsilon \rightarrow 0$.

Η απόδειξη βρίσκεται στο Παράρτημα IV. Ας σημειωθεί ότι, για οποιοδήποτε σταθερό $t \geq 0$, το δεξί μέρος της ανισότητας του λήμματος είναι τάξης $O(\epsilon)$.

Το Λήμμα 1 εξασφαλίζει ότι η μικρότερη ιδιοτιμή κάθε υποπίνακα $\hat{B}_i^{(\mathbf{n}_s)}$ είναι απλή και έτσι, ακόμα και αν οι μικρότερες ιδιοτιμές διαφορετικών υποπινάκων συμπίπτουν, η συνθήκη του Λήμματος 2 που σχετίζεται με τα ιδιοδιανύσματα ικανοποιείται από τον $\text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{\hat{B}_i^{(\mathbf{n}_s)}\}$. Αυτός ο απλούστερος πίνακας μπορεί να χρησιμοποιηθεί στη θέση του B_i , για μία προσέγγιση της κατανομής του χρόνου απόκρισης. Συγκεκριμένα, έχουμε το παρακάτω θεώρημα:

Θεώρημα 3. Για κάθε $i \in \mathcal{K}_f$ και κάθε $t \geq 0$,

$$S_i(t) = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i(\mathbf{n}_s) \hat{S}_i^{(\mathbf{n}_s)}(t) + O(\epsilon),$$

με το $\bar{\gamma}_i(\mathbf{n}_s)$ όπως ορίστηκε στην (5.6).

Απόδειξη. Συνδυάζοντας την έκφραση για το εκθετικό του πίνακα για την $S_i(t)$ στο Θεώρημα 2, την (5.5) και το Λήμμα 2, βλέπουμε ότι για κάθε $t \geq 0$,

$$\exp\{-B_i t\} = \exp\{\text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{-\hat{B}_i^{(\mathbf{n}_s)} t\}\} + O(\epsilon) = \text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{\exp\{-\hat{B}_i^{(\mathbf{n}_s)} t\}\} + O(\epsilon)$$

και ότι το ν_i μπορεί να διαχωριστεί σύμμορφα σε ιδιοδιανύσματα που είναι ίσα με $\bar{\gamma}_i(\mathbf{n}_s) \hat{\nu}_i^{(\mathbf{n}_s)} + O(\epsilon)$. Συνεπώς,

$$S_i(t) = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i(\mathbf{n}_s) \hat{\nu}_i^{(\mathbf{n}_s)} \exp\{-\hat{B}_i^{(\mathbf{n}_s)} t\} \mathbf{1} + O(\epsilon).$$

Η απόδειξη ολοκληρώνεται με εφαρμογή του Θεωρήματος 2 σε κάθε όρο του αθροίσματος. \square

Το θεώρημα αυτό αποκαλύπτει ότι η κατανομή του χρόνου απόκρισης για πελάτες οποιασδήποτε δεδομένης κλάσης της γρήγορης ομάδας είναι προσεγγιστικά ίση με ένα σταθμισμένο άθροισμα των κατανομών των χρόνων απόκρισης που αντιστοιχούν στα μειωμένα υποσυστήματα με δείκτες τα εφικτά υποδιανύσματα των αργών κλάσεων. Οι πιθανότητες $\bar{\gamma}_i(\mathbf{n}_s)$ που χρησιμοποιούνται σαν βάρη ελέγχονται από την αργή δυναμική των κλάσεων του \mathcal{K}_s και το γεγονός αυτό εκφράζεται μέσω της παρουσίας του $\bar{\pi}(\mathbf{n}_s)$ στην (5.6).

Ανεξάρτητα από το πόσο μικρός (αλλά πεπερασμένος) είναι ο παράγοντας σύζευξης ϵ , όταν το $t \rightarrow \infty$ όλες οι CPDF $S_i(t)$ και $\hat{S}_i^{(n_s)}(t)$ μειώνονται και είναι ενδιαφέρον να καθορίσουμε αν η προσέγγιση του Θεωρήματος 3 παραμένει χρήσιμη. Για το σκοπό αυτό παρατηρούμε ότι για μεγάλα t , η ασυμπτωτική ισοδυναμία της (4.17) εφαρμόζεται σε όλες τις $\hat{S}_i^{(n_s)}(t)$ της προσέγγισης και έτσι

$$\sum_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i(\mathbf{n}_s) \hat{S}_i^{(n_s)}(t) \sim \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i(\mathbf{n}_s) a_1(\hat{B}_i^{(n_s)}) e^{-\sigma_1(\hat{B}_i^{(n_s)})t} \sim a_i^\infty e^{-\sigma_i^\infty t},$$

όπου η δεύτερη ισοδυναμία εκφράζει την επικράτηση των κυρίαρχων όρων στο προηγούμενο άθροισμα καθώς το t μεγαλώνει. Κατά συνέπεια,

$$\sigma_i^\infty \triangleq \min_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{\sigma_1(\hat{B}_i^{(n_s)})\}, \quad a_i^\infty \triangleq \sum_{\mathbf{n}_s: \sigma_1(\hat{B}_i^{(n_s)}) = \sigma_i^\infty} \bar{\gamma}_i(\mathbf{n}_s) a_1(\hat{B}_i^{(n_s)}). \quad (5.8)$$

Από το Λήμμα 1, οι ιδιοτιμές $\sigma_1(\hat{B}_i^{(n_s)})$, $\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}$, είναι απλές. Εφόσον ο B_i είναι μια διαταραχή του μπλοκ διαγωνίου πίνακα τάξης $O(\epsilon)$, μία εφαρμογή της καθιερωμένης θεωρίας διαταραχών (βλ. π.χ., [58, Κεφάλαιο 2]) αποκαλύπτει ότι οι αντίστοιχες ιδιοτιμές του B_i είναι ίσες με $\sigma_1(\hat{B}_i^{(n_s)}) + O(\epsilon)$, $\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}$, όπου τα αριστερά και δεξιά ιδιοδιανύσματα αποτελούν επίσης διαταραχές τάξης $O(\epsilon)$ των αντίστοιχων του μπλοκ διαγωνίου πίνακα. Οι ιδιότητες αυτές, σε συνδυασμό με τη μορφή του $a_1(B_i)$ στην (4.16) και την (5.5), δείχνουν ότι

$$S_i(t) \sim a_1(B_i) e^{-\sigma_1(B_i)t} = (a_i^\infty + O(\epsilon)) e^{-(\sigma_i^\infty + O(\epsilon))t},$$

εξασφαλίζοντας ότι η προσέγγιση του Θεωρήματος 3 παραμένει σε ισχύ ακόμα και για μεγάλες τιμές του t , εισάγοντας μόνο διαταραχές τάξης $O(\epsilon)$ στον ακριβή ασυμπτωτικό εκθετικό ρυθμό φθίσης και στον αντίστοιχο παράγοντα.

Στο Κεφάλαιο 6 θα δούμε ότι για συγκεκριμένες διατάξεις του χώρου καταστάσεων είναι δυνατό να καθοριστούν εκ των προτέρων οι συνδυασμοί πελατών των αργών κλάσεων \mathbf{n}_s που αντιστοιχούν στον ελάχιστο ασυμπτωτικό εκθετικό ρυθμό φθίσης, απαλείφοντας έτσι την ανάγκη να αναλυθούν όλα τα μειωμένα υποσυστήματα.

5.3 Κατανομή του χρόνου απόκρισης για πελάτες αργών κλάσεων

Όταν $i \in \mathcal{K}_s$, ο απειροστικός γεννήτορας $M^{(i)}$ έχει και πάλι δομή NCD, ικανοποιώντας ότι $M^{(i)} = \text{diag}_{n_s \in \bar{s} - \bar{c}_i} \{\hat{M}^{(n_s + e_i)}\} + O(\epsilon)$. Αυτή τη φορά ωστόσο, ο όρος μ_i είναι τάξης $O(\epsilon)$, ώστε ο πρώτος όρος στο δεξί μέρος της (4.9) είναι επίσης μικρός και μπορεί να συνδυαστεί με τον όρο τάξης $O(\epsilon)$ του $M^{(i)}$ δίνοντας

$$B_i = -\text{diag}_{n_s \in \bar{s} - \bar{c}_i} \{\hat{M}^{(n_s + e_i)}\} + O(\epsilon). \quad (5.9)$$

Ας σημειωθεί ότι οι διαγώνιοι υποπίνακες $\hat{M}^{(n_s + e_i)}$ είναι μη αναγώγιμοι απειροστικοί γεννήτορες, με τον καθένα να διαθέτει μία απλή μηδενική ιδιοτιμή. Σε αυτή την περίπτωση δεν ενδείκνυται να ακολουθηθεί η ίδια διαδικασία με την περίπτωση όπου $i \in \mathcal{K}_f$ και απευθείας εφαρμογή του Λήμματος 2, επειδή η προκύπτουσα προσέγγιση θα αντικαταστήσει τις ιδιοτιμές του B_i με το ελάχιστο μέτρο (όλες έχουν θετικό πραγματικό μέρος λόγω του Λήμματος 1) με μηδέν και έτσι θα διαστρεβλωθούν οι ιδιότητες της CPDF $S_i(t)$, ιδίως για μεγάλες τιμές του t . Κατά συνέπεια στην τρέχουσα ενότητα αναπτύσσουμε μία ειδικότερη και διαφορετική μεθοδολογία, έτσι ώστε να υπερπηδήσουμε το εμπόδιο αυτό.

Θα διαπραγματευτούμε το παραπάνω ζήτημα θεωρώντας πίνακες που είναι διαταραχές τάξης $O(\epsilon)$ πλήρως αποσυνθέσιμων απειροστικών γεννητόρων, αλλά οι ίδιοι δεν είναι απειροστικοί γεννήτορες. Δε θα στοχεύσουμε σε προσέγγιση του αντίστοιχου πλήρους εκθετικού του πίνακα, αλλά σε συγκεκριμένες εκφράσεις που περιλαμβάνουν αυτό το εκθετικό και μας χρειάζονται εδώ. Πιο συγκεκριμένα, παραθέτουμε το ακόλουθο αποτέλεσμα:

Λήμμα 3. Έστω ότι ο A είναι ομαλός, μη αναγώγιμος πίνακας τύπου M με μπλοκ δομή $A = [A_{IJ}]$, $1 \leq I, J \leq K$, όπου οι διαγώνιοι υποπίνακες A_{II} είναι τετραγωνικοί. Υποθέτουμε ότι $A_{II} = -\hat{A}_I + O(\epsilon)$ και $A_{IJ} = O(\epsilon)$ για $I \neq J$, ώστε $A = \text{diag}_I \{-\hat{A}_I\} + O(\epsilon)$, με όλους τους \hat{A}_I , $I = 1, \dots, K$, να είναι μη αναγώγιμοι απειροστικοί γεννήτορες με αντίστοιχα διανύσματα πιθανοτήτων σταθερής κατάστασης $\hat{\pi}_I$. Έστω ότι ο \bar{A} είναι ο $K \times K$ πίνακας με στοιχεία $\bar{A}(I, J) \triangleq \hat{\pi}_I A_{IJ} \mathbf{1}$. Τέλος, για οποιοδήποτε διάνυσμα γραμμή $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_K)$, διαχωρισμένο σύμμορφα με τη μπλοκ δομή του A , ορίζουμε το αντίστοιχο διάνυσμα μεγέθους K , $\beta_{\mathbf{b}} \triangleq (\mathbf{b}_1 \mathbf{1}, \dots, \mathbf{b}_K \mathbf{1})$. Τότε έχουμε τα ακόλουθα:

1. Για κάθε $t \geq 0$, $\mathbf{b} \exp\{-At\} \mathbf{1} = \beta_{\mathbf{b}} \exp\{-\bar{A}t\} \mathbf{1} + O(\epsilon)$.
2. Όλες οι ιδιοτιμές των A και \bar{A} έχουν θετικό πραγματικό μέρος και οι ιδιοτιμές με το ελάχιστο πραγματικό μέρος, $\sigma_1(A)$ και $\sigma_1(\bar{A})$, αντίστοιχα, είναι και οι δύο

απλές και $\sigma_1(A) = \sigma_1(\bar{A}) + O(\epsilon^2)$. Τα αντίστοιχα αριστερά και δεξιά ιδιοδιανύσματα, $\mathbf{y}_1(A)$ και $\mathbf{x}_1(A)$ για την $\sigma_1(A)$ και τα $\mathbf{y}_1(\bar{A})$ και $\mathbf{x}_1(\bar{A})$ για την $\sigma_1(\bar{A})$, ικανοποιούν τη σχέση $(\mathbf{b}\mathbf{x}_1(A))(\mathbf{y}_1(A)\mathbf{1}) = (\boldsymbol{\beta}_b\mathbf{x}_1(\bar{A}))(\mathbf{y}_1(\bar{A})\mathbf{1}) + O(\epsilon)$.

Η απόδειξη παρατίθεται στο Παράρτημα V.

Το Λήμμα 1 και η (5.9) εξασφαλίζουν ότι το Λήμμα 3 μπορεί να εφαρμοστεί στον B_i , περίπτωση κατά την οποία τα διανύσματα πιθανοτήτων $\hat{\pi}^{(n_s+e_i)}$ παίζουν το ρόλο των $\hat{\pi}_I$ του λήμματος. Με χρήση του γεγονότος αυτού, της μπλοκ δομής του B_i και της (5.4), μπορούμε να δείξουμε ότι ο $\bar{B}_i = \mu_i \bar{R}^{(i)} - \bar{M}^{(i)}$ (δηλαδή ο πίνακας που εμφανίζεται στην έκφραση με το εκθετικό του πίνακα του Θεωρήματος 2, όπως αυτή εφαρμόζεται στο συναθροιστικό σύστημα SDPS) δρα ως ο \bar{A} του λήμματος. Επίσης, όταν το διάνυσμα ν_i χρησιμοποιείται στη θέση του \mathbf{b} στο λήμμα, η (5.7) υποδεικνύει ότι το $\bar{\nu}_i$ δρα σαν το $\boldsymbol{\beta}_b$. Για το λόγο αυτό το πρώτο μέρος του Λήμματος 3 μας δίνει $\nu_i \exp\{-B_i t\}\mathbf{1} = \bar{\nu}_i \exp\{-\bar{B}_i t\}\mathbf{1} + O(\epsilon)$ και εφαρμογή του Θεωρήματος 2 και στις δύο πλευρές της τελευταίας σχέσης, αποδεικνύει το ακόλουθο:

Θεώρημα 4. Για κάθε κλάση $i \in \mathcal{K}_s$ και κάθε $t \geq 0$, $S_i(t) = \bar{S}_i(t) + O(\epsilon)$.

Με άλλα λόγια, η κατανομή του χρόνου απόκρισης για πελάτες κλάσεων με αργή δυναμική ελέγχεται από το συναθροιστικό σύστημα, το οποίο εξυπηρετεί πελάτες μόνο από κλάσεις του \mathcal{K}_s και ρυθμούς αφίξεων και εξυπηρέτησης όπως αυτοί δίνονται στην (5.4).

Το δεύτερο μέρος του Λήμματος 3 μπορεί να χρησιμοποιηθεί για να εξασφαλιστεί ότι η προσέγγιση στο Θεώρημα 4 παραμένει σε ισχύ καθώς $t \rightarrow \infty$. Πράγματι, με τη συγκεκριμένη εφαρμογή του λήμματος στον B_i , τα αποτελέσματα παίρνουν τη μορφή

$$\sigma_1(B_i) = \sigma_1(\bar{B}_i) + O(\epsilon^2) \text{ και}$$

$$a_1(B_i) = (\nu_i \mathbf{x}_1(B_i))(\mathbf{y}_1(B_i)\mathbf{1}) = (\bar{\nu}_i \mathbf{x}_1(\bar{B}_i))(\mathbf{y}_1(\bar{B}_i)\mathbf{1}) + O(\epsilon) = a_1(\bar{B}_i) + O(\epsilon),$$

με χρήση επίσης της (4.16). Αυτές οι εκφράσεις και η (4.17) υποδεικνύουν ότι προσεγγίζοντας την ακριβή CPDF από αυτή που παίρνουμε από το συναθροιστικό σύστημα εισάγεται μια διαταραχή τάξης $O(\epsilon^2)$ στον ασυμπτωτικό εκθετικό ρυθμό φθίσης και μία διαταραχή τάξης $O(\epsilon)$ στον αντίστοιχο πολλαπλασιαστικό παράγοντα.

Παρατηρώντας ότι $\bar{B}_i = O(\epsilon)$, έχουμε ότι $\sigma_1(\bar{B}_i) = O(\epsilon)$ και παρόμοια, $\sigma_1(B_i) = O(\epsilon)$. (Μπορούμε να καταλήξουμε εκεί χρησιμοποιώντας την (5.9) σε συνεργασία με τη θεωρία διαταραχών ιδιοτιμών, αλλά μπορούμε επίσης να κάνουμε χρήση των φραγμάτων της Πρότασης 2, υπενθυμίζοντας ότι $\mu_i = O(\epsilon)$.) Έτσι, η σχέση $\sigma_1(B_i) = \sigma_1(\bar{B}_i) + O(\epsilon^2)$ εισάγει ένα σχετικό σφάλμα τάξης $O(\epsilon)$.



Κεφάλαιο 6

Επιπλέον Αποτελέσματα για Περιβάλλον Διαχωρισμού Χρονικής Κλίμακας

6.1 Διαχωρισμός χρονικής κλίμακας με ρυθμούς εξυπηρέτησης προνομιακούς για γρήγορες κλά- σεις

ΣΤΗ συνέχεια αναλύουμε το καθεστώς διαχωρισμού χρονικής κλίμακας σε συνδυασμό με ρυθμούς εξυπηρέτησης που φέρουν κάποια επιπλέον χαρακτηριστικά ώστε το σύστημα να προσφέρει ακόμα μεγαλύτερη εύνοια στους πελάτες των γρήγορων κλάσεων. Η ανάγκη για τη μελέτη της επιλογής αυτής προκύπτει από τη συσχέτιση των ρυθμών εξυπηρέτησης του παρόντος γενικού προβλήματος με τις αντίστοιχες εκφράσεις των κερδών διαφορισμού πολλαπλών χρηστών που προκύπτουν στο Θεώρημα 1, σχετικά με το ειδικότερο πρόβλημα εκεί. Πιο συγκεκριμένα, στο γενικότερο πλαίσιο μελέτης SDPS συστημάτων, θεωρούμε ρυθμούς εξυπηρέτησης της ακόλουθης μορφής:

$$\phi_i(\mathbf{n}_s, \mathbf{n}_f) = \begin{cases} \phi_i^*(\mathbf{n}_f), & i \in \mathcal{K}_f, \mathbf{n}_s = \mathbf{0}, \\ \phi_i^*(\mathbf{n}_f) + O(\delta), & i \in \mathcal{K}_f, \mathbf{n}_s \neq \mathbf{0}, \\ \phi_i^*(\mathbf{n}_s), & i \in \mathcal{K}_s, \mathbf{n}_f = \mathbf{0}, \\ O(\delta), & i \in \mathcal{K}_s, \mathbf{n}_f \neq \mathbf{0}, \end{cases} \quad \begin{aligned} \phi_i^*(\mathbf{n}_f) &\triangleq \phi_i(\mathbf{0}, \mathbf{n}_f) = O(1), i \in \mathcal{K}_f, \\ \phi_i^*(\mathbf{n}_s) &\triangleq \phi_i(\mathbf{n}_s, \mathbf{0}) = O(1), i \in \mathcal{K}_s. \end{aligned} \quad (6.1)$$

Με αυτήν την τροποποίηση, το ποσό εξυπηρέτησης που λαμβάνεται από τους πελάτες των γρήγορων κλάσεων παραμένει σχεδόν ανεπηρέαστο από την

παρουσία πελατών των αργών κλάσεων. Οι τελευταίοι λαμβάνουν εξυπηρέτηση με μη αμελητέο ρυθμό μόνο όταν δεν υπάρχουν πελάτες γρήγορων κλάσεων στο σύστημα. Η μικρή παράμετρος δ ποσοτικοποιεί την απόκλιση από την πλήρη κατάρρευση των μικρών ρυθμών, η οποία θεωρούμε ότι συμβαίνει καθώς $\delta \rightarrow 0$. Ας σημειωθεί ότι η παράμετρος δ μπορεί να είναι μικρότερη, ίση ή μεγαλύτερη από τον παράγοντα σύζευξης ϵ . Στην ανάλυση που ακολουθεί, συμπληρώνουμε την (6.1) με την απαίτηση ότι οι ρυθμοί αφίξεων των πελατών από τις γρήγορες κλάσεις παραμένουν επίσης ανεπηρέαστοι από την παρουσία αργών πελατών, δηλαδή

$$\lambda_i(\mathbf{n}_s, \mathbf{n}_f) = \lambda_i^*(\mathbf{n}_f) + O(\delta), \lambda_i^*(\mathbf{n}_f) \triangleq \lambda_i(\mathbf{0}, \mathbf{n}_f) = O(1), \forall \mathbf{n}_s \neq \mathbf{0}, i \in \mathcal{K}_f. \quad (6.2)$$

Για ομοιογενείς αφίξεις Poisson, η συνθήκη (6.2) ισχύει αυτόματα για μηδενικό όρο $O(\delta)$.

Εκτός από την άμεση συσχέτιση με τα κέρδη του Θεωρήματος 1 που προαναφέραμε, η δομή της (6.1) είναι πρόσφορη για τη μοντελοποίηση πολλών περιπτώσεων πρακτικού ενδιαφέροντος. Για παράδειγμα, ας θεωρήσουμε την επιλογή κατάλληλων ρυθμών μετάδοσης για την εξυπηρέτηση μεγάλων και μικρών ροών στο διαδίκτυο. Η εργασία [56] μελετά συστήματα με ρυθμούς εξυπηρέτησης κατάλληλα τροποποιημένους για το σκοπό αυτό και προτείνει ότι οι μεγάλες ροές θα πρέπει να λαμβάνουν μόνο ένα μικρό ποσοστό του εξυπηρέτητη. Σε συμφωνία με την παρατήρηση αυτή, οι ρυθμοί εξυπηρέτησης DPS $\phi_i(\mathbf{n}_s, \mathbf{n}_f) = w_i / (\sum_{j \in \mathcal{K}_f} n_j w_j + \sum_{j \in \mathcal{K}_s} n_j \bar{w}_j)$ μπορούν να κάνουν χρήση βαρών $w_i = O(1)$ για $i \in \mathcal{K}_f$ και $w_i = \delta \bar{w}_i$, με $\bar{w}_i = O(1)$, για $i \in \mathcal{K}_s$. Δεδομένου ότι ο θάλαμος εξυπηρέτησης είναι πεπερασμένος, και άρα όλα τα n_j είναι άνω φραγμένα, μία αντικατάσταση των βαρών αυτών στην έκφραση για τους ρυθμούς εξυπηρέτησης ενός συστήματος DPS και έπειτα μία ανάπτυξη κατά Taylor ως προς δ , θεμελιώνουν ότι οι ρυθμοί εξυπηρέτησης πράγματι ταιριάζουν στο πλαίσιο της (6.1), με $\phi_i^*(\mathbf{n}_f) = w_i / (\sum_{j \in \mathcal{K}_f} n_j w_j)$ για κλάσεις $i \in \mathcal{K}_f$ και $\phi_i^*(\mathbf{n}_s) = \bar{w}_i / (\sum_{j \in \mathcal{K}_s} n_j \bar{w}_j)$ για κλάσεις $i \in \mathcal{K}_s$.

Με συνδυασμό του διαχωρισμού χρονικής κλίμακας και της επιπλέον ευνοϊκής τροποποίησης των ρυθμών εξυπηρέτησης υπέρ των γρήγορων κλάσεων, γίνεται δυνατή η περαιτέρω προσέγγιση των κατανομών του χρόνου απόκρισης με μεγαλύτερη απλότητα τόσο υπολογιστικά αλλά και σε διαισθητικό και νοητικό επίπεδο. Η επιπλέον απλοποίηση εμφανίζεται με την αντικατάσταση των αρχικών ρυθμών εξυπηρέτησης από τους αντίστοιχους της (6.1), αφηφώντας τους όρους τάξης $O(\delta)$ που εμφανίζονται εκεί. Όπως προκύπτει από την (6.2), μία παρόμοια αντικατάσταση εφαρμόζεται και για τους ρυθμούς αφίξεων των κλάσεων $i \in \mathcal{K}_f$. Σε ό,τι ακολουθεί, οι ποσότητες που σχετίζονται με το τροποποιημένο σύστημα SDPS που προκύπτει μετά τις αλλαγές στους

ρυθμούς αφίξεων και εξυπηρέτησης που μόλις αναφέρθηκαν, θα εμφανίζονται με την προσθήκη ενός «*» στο συμβολισμό που χρησιμοποιείται για τις αντίστοιχες ποσότητες του αρχικού συστήματος. Αξίζει να σημειωθεί ωστόσο ότι οι χώροι καταστάσεων πριν και μετά τις αντικαταστάσεις συμπίπτουν πάντα (π.χ., $\mathcal{S} = \mathcal{S}^*$, $\hat{S}^{*(n_s)} = \hat{S}^{(n_s)}$ κ.ο.κ.).

Οι προσεγγίσεις των κατανομών του χρόνου απόκρισης λαμβάνονται με εφαρμογή των Θεωρημάτων 3 και 4 στο τροποποιημένο σύστημα SDPS. Έτσι για κάποια κλάση $i \in \mathcal{K}_f$,

$$S_i(t) \approx \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}^{*(i)}} \bar{\gamma}_i^{*(\mathbf{n}_s)} \hat{S}_i^{*(\mathbf{n}_s)}(t). \quad (6.3)$$

Το νέο στοιχείο εδώ είναι ότι λόγω των αντικαταστάσεων που βασίζονται στις (6.1) και (6.2), οι κατανομές $\hat{S}_i^{*(\mathbf{n}_s)}(t)$ αντιστοιχούν σε συστήματα SDPS που χαρακτηρίζονται από τους ίδιους ρυθμούς αφίξεων $\lambda_j^*(\cdot) = \lambda_j(\mathbf{0}, \cdot)$ και ρυθμούς εξυπηρέτησης $\phi_j^*(\cdot) = \phi_j(\mathbf{0}, \cdot)$, $j \in \mathcal{K}_f$, ίσους με τους ρυθμούς που θα είχε το αρχικό σύστημα αν λειτουργούσε χωρίς πελάτες αργών κλάσεων. Τα συστήματα αυτά διαφέρουν μόνο ως προς τους χώρους καταστάσεων $\hat{S}^{*(\mathbf{n}_s)}$. Συνεπώς, όταν υπάρχει η τροποποίηση υπέρ των γρήγορων κλάσεων, η επίδοση αυτών μπορεί να επηρεαστεί από πελάτες αργών κλάσεων μόνο όταν η παρουσία των τελευταίων μεταβάλλει το θάλαμο εξυπηρέτησης που διατίθεται στους πελάτες γρήγορων κλάσεων, όπως για παράδειγμα σε θαλάμους εξυπηρέτησης της μορφής (3.3) αλλά όχι της μορφής (3.2) ή (3.5).

Για κλάσεις αργής δυναμικής το Θεώρημα 4 εξασφαλίζει ότι $S_i(t) \approx \bar{S}_i^*(t)$, η οποία είναι η κατανομή του χρόνου απόκρισης που αντιστοιχεί στο συναθροιστικό σύστημα. Οι αντικαταστάσεις των ρυθμών εξυπηρέτησης και η (5.4) οδηγούν στο ότι $\bar{\phi}_i^*(\mathbf{n}_s) = \hat{\pi}^{*(\mathbf{n}_s)}(\mathbf{0}) \phi_i^*(\mathbf{n}_s)$, $i \in \mathcal{K}_s$. Με άλλα λόγια, ο ρυθμός εξυπηρέτησης στο συναθροιστικό σύστημα στην κατάσταση \mathbf{n}_s είναι ίσος με το ρυθμό που επιτυγχάνεται στο αρχικό σύστημα, όταν αυτό εξυπηρετεί αποκλειστικά τον πληθυσμό \mathbf{n}_s και χρησιμοποιεί εξυπηρετητή με μειωμένη ταχύτητα κατά ένα παράγοντα ίσο με την πιθανότητα σταθερής κατάστασης με την οποία το μειωμένο υποσύστημα SDPS με δείκτη \mathbf{n}_s είναι άδειο από πελάτες γρήγορων κλάσεων.

Η νοητική απλότητα που εισάγεται από την τροποποίηση των ρυθμών εξυπηρέτησης υπέρ των γρήγορων κλάσεων μπορεί να τονιστεί περισσότερο όσον αφορά συστήματα SDPS που χαρακτηρίζονται από ομοιογενείς αφίξεις Poisson και ξεχωριστούς θαλάμους εξυπηρέτησης για τις δύο ομάδες κλάσεων \mathcal{K}_s και \mathcal{K}_f , δηλαδή χώρο καταστάσεων της μορφής (3.5). (Ωστόσο, διαμοιρασμός κάθε θαλάμου εξυπηρέτησης μεταξύ πελατών κλάσεων της ίδιας ομάδας μπορεί να επιτραπεί.) Λόγω του διαχωρισμού των θαλάμων εξυπηρέτησης όλοι

οι χώροι καταστάσεων $\hat{\mathcal{S}}^{(n_s)} = \hat{\mathcal{S}}^{*(n_s)}$ είναι ίσοι με \mathcal{S}_f , και έτσι όλα τα μειωμένα υποσυστήματα με δείκτη οποιοδήποτε n_s συμπίπτουν με το ίδιο υποσύστημα που αντιστοιχεί στο αρχικό όταν το τελευταίο εξυπηρετεί αποκλειστικά κλάσεις της γρήγορης ομάδας.

Σε αντιστοιχία του γεγονότος αυτού, η (6.3) απλοποιείται σε $S_i(t) \approx \hat{S}_i^{*(0)}(t) \sum_{n_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i^*(n_s) = \hat{S}_i^{*(0)}(t)$ και οι κατανομές του χρόνου απόκρισης για τις κλάσεις της γρήγορης ομάδας είναι προσεγγιστικά αναλλοίωτες στην παρουσία πελατών αργών κλάσεων. Επιπλέον, από την ομοιογένεια των αφίξεων Poisson, τη σύμπτωση των χώρων καταστάσεων $\hat{\mathcal{S}}^{(n_s)}$ και την (5.4), προκύπτει ότι $\bar{\lambda}_i^*(n_s) = \lambda_i \sum_{n_f \in \hat{\mathcal{S}}^{*(n_s+e_i)}} \hat{\pi}^{*(n_s)}(n_f) = \lambda_i \hat{\pi}^{*(n_s)} \mathbf{1} = \lambda_i$, οπότε και το συναθροιστικό σύστημα χαρακτηρίζεται από ομοιογενείς αφίξεις Poisson. Τέλος, η μορφή των ρυθμών εξυπηρέτησης για το συναθροιστικό σύστημα και η συνταύτιση όλων των υποσυστημάτων με δείκτη n_s υποδηλώνουν ότι $\bar{\phi}_i^*(n_s) = \hat{\pi}^{*(0)}(\mathbf{0}) \phi_i^*(n_s)$. Με άλλα λόγια, η επίδοση των πελατών αργών κλάσεων αντιστοιχεί σε εκείνη του αρχικού συστήματος αλλά με μειωμένη ταχύτητα εξυπηρέτησης, με τον συντελεστή μείωσης να είναι ίσος με την πιθανότητα με την οποία είναι άδειο το τροποποιημένο σύστημα το οποίο εξυπηρετεί τις κλάσεις των γρήγορων πελατών.

Το αποτέλεσμα αυτό είναι σε συμφωνία με την κοινή λογική, με την έννοια ότι μεγαλύτερη πρόσβαση στον εξυπηρετητή των πελατών των γρήγορων κλάσεων ενισχύει την επίδοση τους χωρίς δυσανάλογο αντίκτυπο στην ούτως ή άλλως περιορισμένη επίδοση των πελατών των αργών κλάσεων. Πράγματι, για ένα τέτοιο σχήμα πρόσβασης στον εξυπηρετητή, εκπεφρασμένο μέσω του φορμαλισμού που αφορά την τροποποίηση υπέρ των κλάσεων της γρήγορης ομάδας, έγινε εφικτό να αποδείξουμε ότι οι πελάτες των γρήγορων κλάσεων όντως ευνοούνται, σε σημείο που η παρουσία αργών πελατών παραμένει σχεδόν ανεπιρρέαστη.

Φυσικά, η εν λόγω τροποποίηση έχει αντίκτυπο στην επίδοση των πελατών των αργών κλάσεων, επίδραση που εκφράζεται από τη μείωση στην ταχύτητα του εξυπηρετητή στο αντίστοιχο μοντέλο SDPS κατά τον παράγοντα $\hat{\pi}^{*(0)}(\mathbf{0})$. Ωστόσο, η επίδοση που σχετίζεται με το χρόνο απόκρισης για αυτό του τύπου τους πελάτες είναι από μόνη της χαμηλή, και συγκεκριμένα έχουμε δει ότι η εκθετική φθίση των πιθανοτήτων ουράς της κατανομής του χρόνου απόκρισης είναι αργή, με ρυθμό τάξης $O(\epsilon)$. Αυτό συμβαίνει εξαιτίας της αργής δυναμικής τους που αντικατοπτρίζεται στο γεγονός ότι οι μέσες απαιτήσεις εξυπηρέτησης $1/\mu_i$ των πελατών αυτών των κλάσεων είναι μεγάλες, τάξης $O(1/\epsilon)$. Με εφαρμογή της Πρότασης 2 στο συναθροιστικό μοντέλο βλέπουμε ότι για κλάσεις του \mathcal{K}_s , ο ασυμπτωτικός ρυθμός φθίσης των πιθανοτήτων ουράς του χρόνου απόκρισης είναι τάξης $O(\hat{\pi}^{*(0)}(\mathbf{0})\mu_i) = O(\hat{\pi}^{*(0)}(\mathbf{0})\epsilon)$ και έτσι,

η τροποποίηση υπέρ των γρήγορων κλάσεων παρόλο που μειώνει το ρυθμό φθίσης, δε μεταβάλλει την τάξη μεγέθους του. Επίσης, η διαστατοποίηση του συστήματος για καλύτερη επίδοση των γρήγορων κλάσεων οδηγεί συνήθως σε μεγαλύτερη πιθανότητα $\hat{\pi}^{*(0)}(\mathbf{0})$, υποβοηθώντας έτσι και την επίδοση των αργών κλάσεων.

Κλείνουμε την ενότητα αυτή με μία εκτίμηση του λάθους που εισάγεται κατά τη σύγκριση της προσέγγισης που αφορά το διαχωρισμό χρονικής κλίμακας με την προσέγγιση της τροποποίησης υπέρ των γρήγορων κλάσεων. Από τις (6.1) και (6.2), έχουμε ότι $\hat{M}^{(n_s)} = \hat{M}^{*(n_s)} + O(\delta)$. Εφόσον και οι δύο πίνακες είναι μη αναγώγιμοι απειροστικοί γεννήτορες, έχουν και οι δύο μία απλή μηδενική ιδιοτιμή, με τη μία ιδιοτιμή να είναι «διαταραχή» της άλλης με μηδενικό όμως, όρο διαταραχής. Τα αντίστοιχα ιδιοδιανύσματα είναι τα διανύσματα πιθανοτήτων σταθερής κατάστασης. Στην περίπτωση αυτή, από την καθιερωμένη θεωρία διαταραχών ιδιοτιμών έχουμε ότι $\hat{\pi}^{(n_s)} = \hat{\pi}^{*(n_s)} + O(\delta)$. Από την (4.5) (όπως αυτή εφαρμόζεται στα μειωμένα συστήματα με δείκτη n_s), την (5.2) και την (6.2), έπεται ότι $\hat{\nu}_i^{(n_s)} = \hat{\nu}_i^{*(n_s)} + O(\delta)$. Ομοίως, $\hat{B}_i^{(n_s)} = \hat{B}_i^{*(n_s)} + O(\delta)$ και μία εφαρμογή του Λήμματος 2 δείχνει ότι $\exp\{-\hat{B}_i^{(n_s)}t\} = \exp\{-\hat{B}_i^{*(n_s)}t\} + O(\delta)$, για κάθε $t \geq 0$, ενώ με χρήση του Θεωρήματος 2, $\hat{S}_i^{(n_s)}(t) = \hat{\nu}_i^{(n_s)} e^{-\hat{B}_i^{(n_s)}t} \mathbf{1} = (\hat{\nu}_i^{*(n_s)} + O(\delta))(e^{-\hat{B}_i^{*(n_s)}t} + O(\delta))\mathbf{1} = \hat{\nu}_i^{*(n_s)} e^{-\hat{B}_i^{*(n_s)}t} \mathbf{1} + O(\delta) = \hat{S}_i^{*(n_s)}(t) + O(\delta)$. Σύντομα σε αυτήν την ενότητα θα δείξουμε επίσης ότι $\bar{\pi} = \bar{\pi}^* + O(\delta)$, και έτσι από την (5.6) έχουμε ότι $\bar{\gamma}_i(\mathbf{n}_s) = \bar{\gamma}_i^*(\mathbf{n}_s) + O(\delta)$. Σε συνδυασμό με το Θεώρημα 3, εν τέλει παίρνουμε $S_i(t) = \sum_{\mathbf{n}_s \in \mathcal{S}^{(i)}} \bar{\gamma}_i^*(\mathbf{n}_s) \hat{S}_i^{*(n_s)}(t) + O(\delta) + O(\epsilon)$. Έτσι, η προσέγγιση της (6.3) εισάγει ένα σφάλμα τάξης $O(\delta + \epsilon) = O(\max\{\delta, \epsilon\})$.

Σχετικά με την ασυμπτωτική συμπεριφορά ουράς της προσέγγισης αυτής, σημειώνουμε ότι η σχέση $\hat{B}_i^{(n_s)} = \hat{B}_i^{*(n_s)} + O(\delta)$, σε συνδυασμό με το Λήμμα 1, υποδηλώνει ότι $\sigma_1(\hat{B}_i^{(n_s)}) = \sigma_1(\hat{B}_i^{*(n_s)}) + O(\delta)$. Μία διαταραχή ίδιας τάξης εφαρμόζεται στα αντίστοιχα αριστερά και δεξιά ιδιοδιανύσματα και με χρήση της (4.16), $a_1(\hat{B}_i^{(n_s)}) = a_1(\hat{B}_i^{*(n_s)}) + O(\delta)$. Λαμβάνοντας υπόψη τα παραπάνω καθώς και την ανάλυση στο τέλος της Ενότητας 5.2, συμπεραίνουμε ότι η προσέγγιση (6.3) διαταράσσει τον ασυμπτωτικό εκθετικό ρυθμό φθίσης των πιθανοτήτων ουράς του χρόνου απόκρισης και τον αντίστοιχο συντελεστή κατά ένα σφάλμα που είναι και πάλι τάξης $O(\max\{\delta, \epsilon\})$. Συγκεκριμένα, προκύπτει ότι $S_i(t) \sim (a_i^{*\infty} + O(\max\{\delta, \epsilon\}))e^{-(\sigma_i^{*\infty} + O(\max\{\delta, \epsilon\}))t}$, όπου τα $a_i^{*\infty}$ και $\sigma_i^{*\infty}$ ορίζονται σε αναλογία με την (5.8), για το σύστημα που προκύπτει μετά τις αντικαταστάσεις των ρυθμών που σχετίζονται με την τροποποίηση υπέρ των γρήγορων κλάσεων.

Σχετικά με το συναθροιστικό σύστημα, οι (5.4) και (6.1), σε συνδυασμό με το ότι τα $\lambda_i(\mathbf{n}_s, \mathbf{n}_f)$ και μ_i είναι τάξης $O(\epsilon)$ για $i \in \mathcal{K}_s$ και το γεγονός ότι $\hat{\pi}^{(n_s)} = \hat{\pi}^{*(n_s)} + O(\delta)$ (όπως είδαμε νωρίτερα), μας οδηγούν στο γεγονός πως

$\bar{M} = \bar{M}^* + O(\delta\epsilon)$, όπου και οι δύο πίνακες έχουν στοιχεία τάξης $O(\epsilon)$. Μεταβάλλοντας την κλίμακα της σχέσης αυτής κατά ϵ^{-1} (μία πράξη που μεταβάλλει την κλίμακα των ιδιοτιμών αλλά αφήνει ανεπηρέαστα τα ιδιοδιανύσματα) και με χρήση των απειροστικών γεννητόρων τάξης $O(1)$ που εμφανίζονται και τώρα διαφέρουν κατά ένα όρο $O(\delta)$, καταλήγουμε στο ότι $\bar{\pi} = \bar{\pi}^* + O(\delta)$, όπως ισχυριστήκαμε παραπάνω. Κάνοντας χρήση της (4.5), όπως αυτή εφαρμόζεται στο συναθροιστικό σύστημα και απλοποιώντας του όρους τάξης $O(\epsilon)$ στον αριθμητή και τον παρονομαστή (που εμφανίζονται λόγω της παρουσίας των μικρών ρυθμών $\bar{\lambda}_i(\mathbf{n}_s) = O(\epsilon) = \bar{\lambda}_i^*(\mathbf{n}_s) + O(\delta\epsilon)$), έχει κανείς ότι $\bar{\nu}_i = \bar{\nu}_i^* + O(\delta)$. Με επιχειρήματα παρόμοια με αυτά που χρησιμοποιήθηκαν για τον \bar{M} , μπορεί να δει κανείς ότι $\bar{B}_i = \bar{B}_i^* + O(\delta\epsilon)$ και με χρήση του Λήμματος 2 και του Θεωρήματος 2,

$$\bar{S}_i(t) = \bar{\nu}_i e^{-\bar{B}_i t} \mathbf{1} = (\bar{\nu}_i^* + O(\delta))(e^{-\bar{B}_i^* t} + O(\delta\epsilon)) \mathbf{1} = \bar{\nu}_i^* e^{-\bar{B}_i^* t} \mathbf{1} + O(\delta) = \bar{S}_i^*(t) + O(\delta).$$

Το αποτέλεσμα αυτό και το Θεώρημα 4 δείχνουν ότι $S_i(t) = \bar{S}_i^*(t) + O(\max\{\delta, \epsilon\})$, δηλαδή, η συνδυαστική προσέγγιση για τις κλάσεις με αργή δυναμική εισάγει επίσης ένα σφάλμα τάξης $O(\max\{\delta, \epsilon\})$. Για να εξετάσουμε την ασυμπτωτική συμπεριφορά της προσέγγισης, υπενθυμίζουμε ότι $\bar{B}_i = \bar{B}_i^* + O(\delta\epsilon)$ και ότι και οι δύο πίνακες είναι τάξης $O(\epsilon)$. Έτσι διαιρώντας με ϵ και εφαρμόζοντας ανάλυση διαταραχών των ιδιοτιμών, έχουμε $\sigma_1(\bar{B}_i) = \sigma_1(\bar{B}_i^*) + O(\delta\epsilon)$, ενώ τα αντίστοιχα ιδιοδιανύσματα διαφέρουν κατά $O(\delta)$ και έτσι από την (4.16), $a_1(\bar{B}_i) = a_1(\bar{B}_i^*) + O(\delta)$. Συνδυάζοντας τις εκφράσεις αυτές με τις αντίστοιχες διαταραχές που παρήχθησαν στο τέλος της Ενότητας 5.3, βλέπουμε ότι $a_1(B_i) = a_1(\bar{B}_i^*) + O(\max\{\delta, \epsilon\})$ και ότι $\sigma_1(B_i) = \sigma_1(\bar{B}_i^*) + O(\epsilon \max\{\delta, \epsilon\})$, με την τελευταία περίπτωση να αντιστοιχεί σε σφάλμα τάξης $O(\max\{\delta, \epsilon\})$, εφόσον και οι δύο ιδιοτιμές είναι τάξης $O(\epsilon)$. Τα αποτελέσματα αυτά διασφαλίζουν επίσης ότι η ουρά της κατανομής της συνδυαστικής προσέγγισης συμπεριφέρεται ορθώς και για της κλάσεις με αργή δυναμική.

6.2 Περαιτέρω χαρακτηρισμός της συμπεριφοράς ουράς για κλάσεις με γρήγορη δυναμική

Στη συνέχεια παραθέτουμε έναν πιο λεπτομερή χαρακτηρισμό της ουράς της κατανομής του χρόνου απόκρισης για πελάτες γρήγορων κλάσεων μέσω του καθορισμού του πληθυσμού n_s από πελάτες αργών κλάσεων για τον οποίο ισχύει η ελαχιστοποίηση στην (5.8). Το πρόβλημα εξαρτάται από διάφορες παραμέτρους του συστήματος, συμπεριλαμβανομένης της μορφής του θαλάμου εξυπηρέτησης και ιδιοτήτων των συναρτήσεων ρυθμών εξυπηρέτησης. Για συ-

στήματα που χαρακτηρίζονται από ομοιογενείς αφίξεις Poisson, μονότονες συναρτήσεις ρυθμών εξυπηρέτησης και ξεχωριστό θάλαμο εξυπηρέτησης για κάθε κλάση (δηλ., με χώρο καταστάσεων της μορφής (3.2)) είναι εφικτό να συγκριθούν οι CPDF που αντιστοιχούν σε διαφορετικά υποσυστήματα μέσω μίας ιδιότητας στοχαστικής διάταξης. Συγκεκριμένα, θα κάνουμε χρήση του παρακάτω αποτελέσματος το οποίο βασίζεται στο Λήμμα 1 του [13].

Λήμμα 4. Έστω ένα σύστημα SDPS πολλαπλών κλάσεων με ομοιογενείς αφίξεις Poisson, εκθετικά κατανομημένες απαιτήσεις εξυπηρέτησης, μονότονες συναρτήσεις ρυθμών εξυπηρέτησης, χώρο καταστάσεων \mathcal{S} της μορφής (3.2) και διάνυσμα καταστάσεων $\mathbf{N}(t)$ κατά τη χρονική στιγμή t . Έστω ότι $\tilde{\mathbf{N}}(t)$ είναι το διάνυσμα καταστάσεων ενός άλλου συστήματος SDPS το οποίο παράγεται από το πρώτο με μεταβολή των ρυθμών εξυπηρέτησης σε $\tilde{\phi}_i(\cdot)$, $i \in \mathcal{K}$, έτσι ώστε $\tilde{\phi}_i(\mathbf{n}) \leq \phi_i(\mathbf{n})$, για όλα τα $i \in \mathcal{K}$ και $\mathbf{n} \in \mathcal{S}$. Τότε, η συνθήκη $\tilde{\mathbf{N}}(0) \geq \mathbf{N}(0)$ σημαίνει ότι $\tilde{\mathbf{N}}(t) \geq_{st} \mathbf{N}(t)$ για όλα τα $t \geq 0$.

Απόδειξη. Εφόσον οι απαιτήσεις εξυπηρέτησης είναι εκθετικά κατανομημένες, το υπόλοιπο της απαίτησης εξυπηρέτησης κάποιου πελάτη έχει την ίδια κατανομή με την απαίτηση εξυπηρέτησης ενός νεοεισερχόμενου πελάτη της ίδιας κλάσης. Με εφαρμογή επιχειρημάτων δειγματικής διαδρομής (αγ.: sample path) παρατηρούμε ότι η σχέση των συναρτήσεων ρυθμών εξυπηρέτησης, η μονοτονία και οι αρχικές συνθήκες οδηγούν στο ότι $\tilde{\phi}_i(\tilde{\mathbf{N}}(0)) \leq \phi_i(\tilde{\mathbf{N}}(0)) \leq \phi_i(\mathbf{N}(0))$. Δηλαδή, το δεύτερο σύστημα είναι βραδύτερο και κρατάει τους πελάτες του τουλάχιστον όσο και το πρώτο σύστημα. Έτσι οι αναχωρήσεις λαμβάνουν χώρα κατά τρόπο που διατηρεί τη διάταξη των διανυσμάτων κατάστασης. Οι αφίξεις (οι οποίες εξαρτώνται από την κατάσταση του συστήματος) διατηρούν τη διάταξη των διανυσμάτων κατάστασης, δεδομένου ότι είτε γίνονται αποδεκτές είτε αποκλείονται και στα δυο συστήματα. Αν συμβεί αποκλεισμός στο ένα σύστημα (έστω τη χρονική στιγμή t_o για κάποιο πελάτη της κλάσης j), το σύστημα αυτό απαραίτητα θα είναι το βραδύτερο. Τότε είναι φανερό ότι $\tilde{N}_j(t_o^-) = L_j = \tilde{N}_j(t_o^+)$ ενώ $N_j(t_o^-) < L_j$ και $N_j(t_o^+) = N_j(t_o^-) + 1 \leq L_j = \tilde{N}_j(t_o^+)$. Έτσι η διάταξη αμέσως μετά τον αποκλεισμό διατηρείται και ο πελάτης που μόλις εισέρχεται στο σύστημα είναι στοχαστικά ισοδύναμος με ένα αντίστοιχο ήδη υπάρχων πελάτη ίδιας κλάσης στο άλλο σύστημα, ολοκληρώνοντας την απόδειξη. \square

Πρέπει να αναφερθεί ότι το Λήμμα 1 στο [13] σχετίζεται με συστήματα που διαθέτουν άπειρους χώρους καταστάσεων, στα οποία δεν υπάρχει αποκλεισμός πελατών και η ιδιότητα διάταξης ισχύει με πιθανότητα 1 και για τυχαία κατανομημένες απαιτήσεις εξυπηρέτησης. Η υπόθεση του Λήμματος 4 σχετικά με χώρο καταστάσεων της μορφής (3.2) είναι ουσιώδης. Για να γίνει αυτό

κατανοητό, ας θεωρήσουμε ένα πλήρως διαμοιραζόμενο θάλαμο εξυπηρέτησης, δηλαδή χώρο καταστάσεων της μορφής (3.3) με $L = 1$ με δύο κλάσεις πελατών και ας υποθέσουμε ότι τα δυο συστήματα ξεκινούν άδεια και στη συνέχεια εισέρχεται ένας πελάτης της κλάσης 1. Αν αυτός ο πελάτης έχει ολοκληρώσει την εξυπηρέτησή του μόνο στο ταχύτερο σύστημα πριν από μία νέα άφιξη κλάσης 2, το βραδύτερο σύστημα θα περιέχει ένα πελάτη κλάσης 1 και το ταχύτερο ένα πελάτη κλάσης 2, αναιρώντας τη διάταξη μεταξύ των διανυσμάτων κατάστασης.

Με χρήση της σχέσης της κατανομής του χρόνου απόκρισης με τη διαδικασία ανταμοιβής Markov του Κεφαλαίου 4 μαζί με την ιδιότητα στοχαστικής διάταξης που μόλις συζητήσαμε, είναι δυνατό να λάβουμε μία διάταξη των CPDF που αντιστοιχούν σε μειωμένα υποσυστήματα με δείκτη που αντιστοιχεί σε πληθυσμό πελατών αργών κλάσεων \mathbf{n}_s . Συγκεκριμένα, έχουμε την ακόλουθη:

Πρόταση 3. *Δεδομένων ομοιογενών αφίξεων Poisson, μονότονων συναρτήσεων ρυθμών εξυπηρέτησης και χώρο καταστάσεων της μορφής (3.2), για κάθε κλάση $i \in \mathcal{K}_f$ και κάθε $\mathbf{n}_{s,1}, \mathbf{n}_{s,2} \in \overline{\mathcal{S}}^{(i)}$, τέτοια ώστε $\mathbf{n}_{s,1} \leq \mathbf{n}_{s,2}$, ισχύει ότι $\hat{S}_i^{(\mathbf{n}_{s,1})}(t) \leq \hat{S}_i^{(\mathbf{n}_{s,2})}(t)$, για όλα τα $t \geq 0$.*

Απόδειξη. Για κάθε τυχαίο σύστημα SDPS που ικανοποιεί τις υποθέσεις της πρότασης (συμπεριλαμβανομένων των εκθετικών απαιτήσεων εξυπηρέτησης), θεωρούμε το αντίστοιχο τροποποιημένο σύστημα με ένα επιπρόσθετο μόνιμο πελάτη από την κλάση i καθώς και τη σχετική με αυτό διαδικασία ανταμοιβής Markov του Κεφαλαίου 4 και ορίζουμε

$$h_i(\mathbf{m}; t) \triangleq \mathbb{E} \left[e^{-\mu_i \int_0^t \phi_i^{(i)}(\mathbf{N}^{(i)}(u)) du} \mid \mathbf{N}^{(i)}(0) = \mathbf{m} \right].$$

Με χρήση της (4.10), $S_i(t) = \mathbb{E} \left[h_i(\mathbf{N}^{(i)}(0); t) \right]$. Επίσης η συνάρτηση $h_i(\cdot; t)$ είναι μη φθίνουσα. Αυτό μπορεί να φανεί θεωρώντας ένα άλλο σύστημα SDPS ταυτοτικά ίσο με εκείνο που περιέχει το μόνιμο πελάτη και συμβολίζοντας τις ποσότητες που αντιστοιχούν σε αυτό προσθέτοντας έναν τόνο. Τότε, για κάθε $\mathbf{N}'^{(i)}(0) = \mathbf{m}' \geq \mathbf{m} = \mathbf{N}^{(i)}(0)$, το Λήμμα 4 διασφαλίζει ότι $\mathbf{N}'^{(i)}(t) \geq_{\text{st}} \mathbf{N}^{(i)}(t)$ και η μονοτονία των ρυθμών εξυπηρέτησης μας δίνει ότι $-\int_0^t \phi_i^{(i)}(\mathbf{N}'^{(i)}(u)) du = -\int_0^t \phi_i^{(i)}(\mathbf{N}^{(i)}(u)) du \geq_{\text{st}} -\int_0^t \phi_i^{(i)}(\mathbf{N}^{(i)}(u)) du$. Εφόσον η $e^{\mu_i x}$ είναι αύξουσα συνάρτηση, ο ισχυρισμός ότι η $h_i(\cdot; t)$ είναι μη φθίνουσα αποδεικνύεται παίρνοντας μέσες τιμές σε κάθε μέρος.

Ας αντιστοιχίσουμε τώρα το αρχικό σύστημα SDPS με ένα σύστημα του οποίου οι ρυθμοί εξυπηρέτησης σχετίζονται με εκείνους του αρχικού όπως στο Λήμμα 4. Τα αντίστοιχα τροποποιημένα συστήματα που περιλαμβάνουν ένα

μόνιμο πελάτη κλάσης i χαρακτηρίζονται από ρυθμούς εξυπηρέτησης που φέρουν την ίδια συσχέτιση. Θεωρώντας ότι $\mathbf{N}^{(i)}(0) = \mathbf{m} = \tilde{\mathbf{N}}^{(i)}(0)$, το λήμμα διασφαλίζει και πάλι ότι $-\int_0^t \tilde{\phi}_i^{(i)}(\tilde{\mathbf{N}}^{(i)}(u)) du \geq_{st} -\int_0^t \phi_i^{(i)}(\mathbf{N}^{(i)}(u)) du$, έτσι ώστε $\tilde{h}_i(\mathbf{m}; t) \geq h_i(\mathbf{m}; t)$, για όλα τα \mathbf{m} και t . Μία άλλη συνέπεια του λήμματος (αυτή τη φορά με εφαρμογή σε συστήματα χωρίς μόνιμο πελάτη) είναι ότι στην σταθερή κατάσταση $\tilde{\mathbf{N}}^{(i)}(0) \geq_{st} \mathbf{N}^{(i)}(0)$ και έτσι

$$\tilde{S}_i(t) = \mathbb{E} \left[\tilde{h}_i(\tilde{\mathbf{N}}^{(i)}(0); t) \right] \geq \mathbb{E} \left[h_i(\tilde{\mathbf{N}}^{(i)}(0); t) \right] \geq \mathbb{E} \left[h_i(\mathbf{N}^{(i)}(0); t) \right] = S_i(t),$$

όπου η δεύτερη ανισότητα έπεται από τη στοχαστική διάταξη των αρχικών διανυσμάτων κατάστασης και τον μη φθίνων χαρακτήρα της $h_i(\cdot; t)$.

Η απόδειξη της πρότασης ολοκληρώνεται βάζοντας τα δυο υποσυστήματα που σχετίζονται με τα $\mathbf{n}_{s,1}$ και $\mathbf{n}_{s,2}$ στη θέση των δύο γενικών συστημάτων που συζητήθηκαν παραπάνω. Αυτό είναι δυνατό επειδή από τη μονοτονία των ρυθμών εξυπηρέτησης έπεται ότι $\phi_i^{(n_{s,2})}(\mathbf{n}_f) = \phi_i(\mathbf{n}_{s,2}, \mathbf{n}_f) \leq \phi_i(\mathbf{n}_{s,1}, \mathbf{n}_f) = \phi_i^{(n_{s,1})}(\mathbf{n}_f)$, για όλα τα $\mathbf{n}_f \geq \mathbf{0}$. \square

Δεδομένης της μορφής του χώρου καταστάσεων, προκύπτει άμεσα ότι για κάθε $\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}$, $\mathbf{n}_s \leq \mathbf{L}_s$, όπου τα στοιχεία του \mathbf{L}_s είναι ίσα με L_i , $i \in \mathcal{K}_s$. Έτσι η Πρόταση 3, το Θεώρημα 3 και το γεγονός ότι τα βάρη $\bar{\gamma}_i(\mathbf{n}_s)$ είναι θετικά και αθροίζουν στη μονάδα, οδηγούν στο ότι

$$\bar{\gamma}_i(\mathbf{L}_s) \hat{S}_i^{(\mathbf{L}_s)}(t) + O(\epsilon) \leq S_i(t) \leq \hat{S}_i^{(\mathbf{L}_s)}(t) + O(\epsilon).$$

Τα όρια αυτά ισχύουν για κάθε $t \geq 0$ και προφανώς ισχύουν και καθώς $t \rightarrow \infty$, δείχνοντας ότι η ουρά της κατανομής του χρόνου απόκρισης για πελάτες γρήγορων κλάσεων χαρακτηρίζεται από το υποσύστημα που αντιστοιχεί στο μέγιστο δυνατό πληθυσμό μόνιμα παρόντων πελατών των αργών κλάσεων.

Όταν ο θάλαμος εξυπηρέτησης είναι κοινός για όλους τους πελάτες ανεξαρτήτως κλάσης, δεν μπορούμε να εφαρμόσουμε πλέον το Λήμμα 4. Ωστόσο, για τη συνήθη περίπτωση ενός πλήρως διαμοιραζόμενου θαλάμου εξυπηρέτησης με μέγεθος L και αντίστοιχο χώρο καταστάσεων της μορφής (3.3), είναι δυνατό να εφαρμοστεί η Πρόταση 2, υπό την προϋπόθεση ότι η κλάση i με γρήγορη δυναμική, της οποίας μελετάται η επίδοση, χαρακτηρίζεται από ρυθμούς εξυπηρέτησης που εξαρτώνται αποκλειστικά από το συνολικό αριθμό πελατών που είναι παρόντες στο σύστημα, δηλαδή της μορφής $\phi_i(\mathbf{n}) = g_i(|\mathbf{n}|)$, όπου η συνάρτηση $g_i(\cdot)$ είναι μη αύξουσα. (Η συγκεκριμένη απαίτηση μονοτονίας ικανοποιείται πάντοτε στα συστήματα πρακτικού ενδιαφέροντος.) Οι άλλες κλάσεις (του \mathcal{K}_f ή/και του \mathcal{K}_s) δεν είναι απαραίτητο να έχουν συναρτήσεις ρυθμών εξυπηρέτησης αυτής της ειδικής μορφής.

Με εφαρμογή της Πρότασης 2 στο αρχικό, «πλήρες» σύστημα, παρατηρούμε ότι η ιδιοτιμή $\sigma_1(B_i)$, η οποία ελέγχει τη συμπεριφορά της ουράς της κατανομής του χρόνου απόκρισης για την κλάση i , ικανοποιεί ότι $\sigma_1(B_i) \geq \mu_i g_i(L)$. Επιπλέον, με εφαρμογή της πρότασης αυτής σε κάθε μειωμένο υποσύστημα έχουμε ότι $\sigma_1(\hat{B}_i^{(n_s)}) \geq \mu_i g_i(L)$ για όλα τα n_s . Τέλος, όταν $|n_s| = L - 1$, ο χώρος $\hat{S}^{(n_s)(i)}$ περιέχει το διάνυσμα $\mathbf{0}$ και τα άνω και κάτω φράγματα της Πρότασης 2 συμπίπτουν στο $\mu_i g_i(L)$. Έτσι, $\sigma_i^\infty = \mu_i g_i(L)$ και $a_i^\infty = \sum_{n_s: |n_s|=L-1} \bar{\gamma}_i(n_s) a_1(\hat{B}_i^{(n_s)})$, δείχνοντας ότι η συμπεριφορά ουράς της κατανομής του χρόνου απόκρισης χαρακτηρίζεται και πάλι από το υποσύστημα που αντιστοιχεί στο μεγαλύτερο δυνατό πληθυσμό μόνιμα παρόντων πελατών των αργών κλάσεων. Η ανάλυση επίσης υποδεικνύει ότι όσον αφορά χαμηλές επιδόσεις σε σχέση με τους χρόνους απόκρισης, δεν ενδεικνύεται να διατίθεται ένας κοινός θάλαμος εξυπηρέτησης για φορτία κίνησης που υπόκεινται σε διαχωρισμό χρονικής κλίμακας, επειδή οι πελάτες των κλάσεων με αργή δυναμική τείνουν να καταλαμβάνουν ολόκληρο το θάλαμο και να μειώσουν έτσι την επίδοση των κλάσεων με γρήγορη δυναμική στο χειρότερο δυνατό βαθμό.

Σαν τελευταία περίπτωση, διαπραγματευόμαστε συστήματα που χαρακτηρίζονται από τροποποίηση των ρυθμών εξυπηρέτησης υπέρ των γρήγορων κλάσεων, όπως εκείνα που μελετήθηκαν στην Ενότητα 6.1. Όπως έχουμε δει, όταν οι δύο ομάδες κλάσεων \mathcal{K}_s και \mathcal{K}_f έχουν ξεχωριστούς θαλάμους εξυπηρέτησης (που αντιστοιχούν σε χώρο καταστάσεων της μορφής (3.5)), τότε η κατανομές του χρόνου απόκρισης για τις κλάσεις με γρήγορη δυναμική παραμένουν προσεγγιστικά αναλλοίωτες στην παρουσία πελατών αργών κλάσεων. Εδώ θα μελετήσουμε ένα σύστημα που επιτρέπει το διαμοιρασμό του θαλάμου εξυπηρέτησης. Ο διαμοιρασμός επιτρέπει στους πελάτες με αργή δυναμική πρόσβαση σε ολόκληρο το θάλαμο εξυπηρέτησης: όποιο κομμάτι του θαλάμου παραμένει άδειο, διατίθεται στους πελάτες των γρήγορων κλάσεων σύμφωνα με κάποια προκαθορισμένη αναλογία. Το παραπάνω αποτελεί ένα λογικό σχήμα διαμοιρασμού δεδομένου ότι οι πελάτες των αργών κλάσεων δεν επηρεάζουν σχεδόν καθόλου τους ρυθμούς εξυπηρέτησης των γρήγορων κλάσεων.

Πιο λεπτομερώς, ο διαμοιρασμός του θαλάμου εξυπηρέτησης λαμβάνει χώρα ως ακολούθως: Το συνολικό μέγεθος L του θαλάμου εξυπηρέτησης διαχωρίζεται σε ξεχωριστά κομμάτια, με μεγέθη L_i , $i \in \mathcal{K}_f$, τέτοια ώστε να ισχύει ότι $\sum_{i \in \mathcal{K}_f} L_i = L$. Έστω ότι $0 < \xi_i \triangleq L_i/L < 1$, για $i \in \mathcal{K}_f$. Όταν υπάρχουν ήδη n_s πελάτες αργών κλάσεων στο σύστημα, νέοι πελάτες της κλάσης $i \in \mathcal{K}_f$ γίνονται αποδεκτοί έως ότου ο συνολικός τους πληθυσμός φτάσει το κατώφλι $\lfloor \xi_i(L - |n_s|) \rfloor$. Ένας νέος πελάτης από κλάση του \mathcal{K}_s γίνεται αποδεκτός επιπλέον του πληθυσμού (n_s, n_f) όταν $L - |n_f| - |n_s| > 0$ και $\lfloor \xi_j(L - 1 - |n_s|) \rfloor \geq n_j$ για

όλα τα $j \in \mathcal{K}_f$. Η δεύτερη συνθήκη είναι αναγκαία ώστε η πολιτική αποδοχής πελατών των γρήγορων κλάσεων να είναι συνεπής με την άφιξη πελάτη αργής κλάσης, χωρίς να χρειάζεται να «εκδιωχθούν» πελάτες από κλάσεις του \mathcal{K}_f . Μεταξύ των άλλων, το παράδειγμα αυτό χρησιμεύει για να δείξουμε πως το σχετικό θεωρητικό πλαίσιο που έχουμε ήδη υποθέσει μπορεί να συμπεριλάβει πιο περίπλοκα σχήματα θαλάμου εξυπηρέτησης.

Για το σχήμα που μόλις περιγράφηκε, $\bar{\mathcal{S}} = \{ \mathbf{n}_s \geq \mathbf{0} \mid |\mathbf{n}_s| \leq L \}$, ενώ $\bar{\mathcal{S}}^{(i)} = \{ \mathbf{n}_s \geq \mathbf{0} \mid |\mathbf{n}_s| \leq \lfloor L(1 - 1/L_i) \rfloor \}$, για $i \in \mathcal{K}_f$. Λόγω της τροποποίησης υπέρ των γρήγορων κλάσεων, για όλα τα $\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}$, τα αντίστοιχα μειωμένα υποσυστήματα έχουν τους ίδιους ρυθμούς αφίξεων και εξυπηρέτησης και χώρους καταστάσεων $\hat{\mathcal{S}}^{(\mathbf{n}_s)} = \times_{i \in \mathcal{K}_f} \{ 0 \leq n_i \leq \lfloor \xi_i(L - |\mathbf{n}_s|) \rfloor \}$, οι οποίοι είναι της μορφής (3.2). Επιπλέον, το μέγεθος κάθε θαλάμου ανά κλάση αυξάνεται καθώς το $|\mathbf{n}_s|$ μειώνεται. Υποθέτουμε ότι οι αφίξεις Poisson είναι ομοιογενείς και ότι οι προσεγγίσεις της τροποποίησης υπέρ των γρήγορων κλάσεων $\phi_i^*(\cdot)$ στους ρυθμούς εξυπηρέτησης (βλ. π.χ., (6.1)) είναι μονότονες. (Οι υποθέσεις αυτές είναι απαραίτητες μόνο για κλάσεις του \mathcal{K}_f .) Με αναφορά στην απόδειξη του Λήμματος 4, είναι εύκολο να παρατηρήσουμε ότι αν κάποιος αλλάξει το δεύτερο σύστημα SDPS εκεί, ώστε αυτό να χαρακτηρίζεται από τους ίδιους ρυθμούς εξυπηρέτησης όπως το πρώτο σύστημα, η στοχαστική διάταξη $\tilde{N}(t) \geq_{st} N(t)$ διατηρείται για όλα τα t . Από τα επιχειρήματα στην απόδειξη της Πρότασης 3, αυτό μας δίνει ότι $\tilde{S}_i(t) \geq S_i(t)$, για κάθε κλάση i . Με εφαρμογή του αποτελέσματος αυτού στα μειωμένα υποσυστήματα με δείκτη \mathbf{n}_s , έχει κανείς ότι αν $|\mathbf{n}_{s,1}| \leq |\mathbf{n}_{s,2}|$ τότε $\hat{S}^{(\mathbf{n}_{s,1})}(t) \geq \hat{S}^{(\mathbf{n}_{s,2})}(t)$. Οπότε, με χρήση της (6.3) και την ανάλυση σφάλματος της Ενότητας 6.2,

$$\bar{\gamma}_i^*(\mathbf{0})\hat{S}_i^{*(\mathbf{0})}(t) + O(\max\{\delta, \epsilon\}) \leq S_i(t) \leq \hat{S}_i^{*(\mathbf{0})}(t) + O(\max\{\delta, \epsilon\}), \quad t \geq 0, \quad i \in \mathcal{K}_f.$$

Όπως βλέπουμε, η ουρά των κατανομών του χρόνου απόκρισης των κλάσεων με γρήγορη δυναμική χαρακτηρίζεται από το υποσύστημα που αντιστοιχεί σε $\mathbf{n}_s = \mathbf{0}$, το οποίο είναι ισοδύναμο με το αρχικό σύστημα SDPS όταν αυτό εξυπηρετεί μόνο πελάτες από κλάσεις του \mathcal{K}_f . Το ίδιο φαινόμενο προέκυψε όταν η αργή και η γρήγορη ομάδα είχαν ξεχωριστούς θαλάμους εξυπηρέτησης, παρόλο που στην περίπτωση εκείνη η κατά προσέγγιση ισότητα του $S_i(t)$ με το $\hat{S}_i^{(\mathbf{0})}(t)$ επεκτεινόταν σε όλα τα $t \geq 0$ αντί να παραμένει ασυμπτωτική. Κατά συνέπεια, όταν το σύστημα χαρακτηρίζεται από τροποποίηση των ρυθμών εξυπηρέτησης υπέρ των γρήγορων κλάσεων, απεριόριστη πρόσβαση στο θάλαμο εξυπηρέτησης πελατών των αργών κλάσεων δε μειώνει την επίδοση πελατών από κλάσεις με γρήγορη δυναμική. Αντίθετα, όταν η συγκεκριμένη τροποποίηση απουσιάζει, όπως είδαμε προηγουμένως, αυτή η απεριόριστη πρόσβαση στον εξυπηρετητή των πελατών από τις αργές κλάσεις μπορεί πράγματι

να αποβεί επιζήμια για την επίδοση των γρήγορων κλάσεων.

Αξίζει να αναφερθεί ότι όταν το \mathcal{K}_f περιέχει μία μόνο κλάση, το σχήμα διαμοιρασμού του θαλάμου εξυπηρέτησης που μόλις εξετάσαμε μετατρέπεται σε πλήρη διαμοιρασμό του θαλάμου εξυπηρέτησης μεταξύ όλων των κλάσεων (δηλαδή και με αργή και με γρήγορη δυναμική). Συνεπώς, τα αποτελέσματα που συζητήθηκαν παραπάνω εφαρμόζονται και σε αυτή την περίπτωση.



Κεφάλαιο 7

Επέκταση των Αποτελεσμάτων για Γενικότερες Κατανομές Απαιτήσεων Εξυπηρέτησης

Η υπόθεση εκθετικά κατανεμημένων απαιτήσεων εξυπηρέτησης είναι ουσιώδης προκειμένου να εκφραστεί η κατανομή του χρόνου απόκρισης με τη μορφή του εκθετικού του πίνακα στο Θεώρημα 2. Πράγματι, το τελευταίο βήμα της (4.10) βασίζεται στην υπόθεση ότι $\Pr\{T_i > x\} = e^{-\mu_i x}$. Παρόλα αυτά, η ουσία του αποτελέσματος ισχύει επίσης, αν και σε μία κάπως πιο περίπλοκη μορφή, για την πιο γενική περίπτωση κατανομών απαιτήσεων εξυπηρέτησης οι οποίες έχουν τυχαία μαρκοβιανή μορφή τύπου-φάσης (αγ.: phase-type). (Οι κατανομές αυτές είναι πυκνό υποσύνολο όλων των συναρτησεων κατανομών.) Εν γένει, διαφορετικές κλάσεις πελατών μπορεί να αντιστοιχούν σε διαφορετικές κατανομές τύπου-φάσης.

Η πρώτη επιλογή σε αυτήν την πιο γενική περίπτωση είναι ότι ο χώρος καταστάσεων των απειροστικών γεννητόρων M και $M^{(i)}$, καθώς και όλων των σχετικών με αυτούς πινάκων και διανυσμάτων, πρέπει να επεκταθεί ούτως ώστε όχι μόνο να διατηρεί απλώς το συνολικό αριθμό πελατών από κάθε κλάση που είναι παρόντες στο σύστημα, αλλά επίσης και το πόσοι από αυτούς τους πελάτες καταλαμβάνουν καθεμία από τις φάσεις της αντιστοιχίας κατανομής απαιτήσεων εξυπηρέτησης. Σε αναλογία, η δομή των πινάκων M και $M^{(i)}$ γίνεται περισσότερο πολύπλοκη από εκείνη των (3.6) και (3.8), προκειμένου να εκφράσει την πιο περίπλοκη μορφή των γεγονότων ολοκλήρωσης εξυπηρέτησης, όπως αυτά ελέγχονται από τις κατανομές τύπου-φάσης. Ας σημειωθεί ότι παρά το ότι η κατάσταση του συστήματος είναι πιο περίπλοκη, οι συναρτήσεις των ρυθμών εξυπηρέτησης $\phi_i(\cdot)$ συνεχίζουν να εξαρτώνται μόνο από το

συνολικό πληθυσμό πελατών από κάθε κλάση. Συνεπώς, στην περίπτωση του εκτεταμένου χώρου καταστάσεων, ο διαγώνιος πίνακας $R^{(i)}$ στην (4.2) έχει ίσα στοιχεία για όλες τις καταστάσεις, οδηγώντας στο ίδιο διάνυσμα για το συνολικό πληθυσμό n . Επιπλέον, παρόλο που η επέκταση του χώρου καταστάσεων έχει καθαρό αντίκτυπο στην υπολογιστική πολυπλοκότητα όλων των σχετικών τεχνικών αριθμητικής επίλυσης, δεν έχει τόσο μεγάλο αντίκτυπο εννοιολογικά, καθώς το σύστημα παραμένει μαρκοβιανό και οι M και $M^{(i)}$ συνεχίζουν να είναι καλώς ορισμένοι απειροστικοί γεννήτορες.

Υποθέτουμε ένα συγκεκριμένο παράδειγμα, στο οποίο οι απαιτήσεις εξυπηρέτησης για κάποια κλάση i έχουν υπερεκθετική κατανομή με m_i φάσεις, έτσι ώστε $\Pr\{T_i > x\} = \sum_{j=1}^{m_i} \omega_{ij} e^{-\mu_{ij}x}$ και $\sum_{j=1}^{m_i} \omega_{ij} = 1$, αντιστοιχώντας σε μέση απαίτηση εξυπηρέτησης ίση με $\sum_{j=1}^{m_i} \omega_{ij} / \mu_{ij}$. Εξαιτίας του μαρκοβιανού χαρακτήρα του συστήματος, μπορεί κανείς να ακολουθήσει όλα τα βήματα της απόδειξης του Θεωρήματος 2 μέχρι την τελευταία ισότητα εντός της (4.10), η οποία λόγω της υπερεκθετικής κατανομής δίνει τώρα

$$S_i(t) = \sum_{j=1}^{m_i} \omega_{ij} \nu_i \exp\{-B_i(\mu_{ij})t\} \mathbf{1},$$

όπου $B_i(s) \triangleq sR^{(i)} - M^{(i)}$. Έπεται ότι η $S_i(t)$ εκφράζεται ακόμα ως ένας γραμμικός συνδυασμός εκφράσεων εκθετικών πινάκων και οι πίνακες σε αυτά τα εκθετικά έχουν την ίδια μορφή όπως αυτή προκύπτει από την (4.9) στο Θεώρημα 2. Ωστόσο, υπενθυμίζουμε ότι η δομή του $M^{(i)}$ τώρα είναι πιο περίπλοκη από την (3.8), λόγω της επέκτασης του χώρου καταστάσεων. Η δομή αυτή εξαρτάται και από την υπερεκθετική κατανομή που σχετίζεται με την κλάση i , η οποία εισήχθη εδώ ρητά, αλλά και από τις λεπτομέρειες των υπολοίπων κατανομών των απαιτήσεων εξυπηρέτησης οι οποίες είναι τύπου-φάσης και σχετίζονται με τις άλλες κλάσεις $l \neq i$.

Ας θεωρήσουμε τώρα ένα άλλο παράδειγμα, στο οποίο οι απαιτήσεις εξυπηρέτησης για την κλάση i είναι κατανομημένες κατά Erlang με m_i στάδια, καθένα με ρυθμό μ_i , για συνολική μέση απαίτηση εξυπηρέτησης ίση προς m_i/μ_i . Στην περίπτωση αυτή έχουμε ότι $\Pr\{T_i > t\} = \sum_{j=0}^{m_i-1} e^{-\mu_i t} (\mu_i t)^j / j!$ και το τελευταίο βήμα στην (4.10) μας δίνει

$$S_i(t) = \sum_{j=0}^{m_i-1} (-1)^j \frac{\mu_i^j}{j!} \nu_i \left(\left. \frac{d^j}{ds^j} \exp\{-B_i(s)t\} \right|_{s=\mu_i} \right) \mathbf{1},$$

με τον $B_i(s)$ όπως προηγουμένως (αλλά σημειώνοντας ότι ο $M^{(i)}$ είναι διαφορετικός από εκείνον του προηγούμενου παραδείγματος, αντικατοπτρίζοντας

τη διαφορετική μορφή της κατανομής των απαιτήσεων εξυπηρέτησης). Οι παράγωγοι των πινάκων $\frac{d^j}{ds^j} e^{-B_i(s)t} \Big|_{s=\mu_i}$ συνδέονται στενά με τον $e^{-B_i(\mu_i)t}$. Ακριβέστερα, $\frac{d^j}{ds^j} e^{-B_i(s)t} = O(t^j e^{-B_i(s)t})$, για όλα τα s και όλα τα $t \geq 0$. Προκειμένου να παρατηρηθεί αυτό, εφαρμόζουμε την ταυτότητα πινάκων (βλ. [57, εξ. 1.3])

$$e^{-Ht} - e^{-H't} = \int_0^t e^{-H'(t-u)} (H' - H) e^{-Hu} du \quad (7.1)$$

με $H = B_i(s + \Delta s)$ και $H' = B_i(s)$, διαιρούμε με Δs και αφήνουμε το $\Delta s \rightarrow 0$, ώστε να λάβουμε

$$-\frac{d}{ds} e^{-B_i(s)t} = \int_0^t e^{-B_i(s)(t-u)} R^{(i)} e^{-B_i(s)u} du. \quad (7.2)$$

Εφόσον ο πίνακας $-B_i(s)$ είναι ουσιαστικά μη αρνητικός για κάθε τιμή του s , το Θεώρημα 3.12 στο [7, σελ. 146] μας δίνει ότι $e^{-B_i(s)t} \geq 0$ για κάθε $t \geq 0$. Για το λόγο αυτό, το αριστερό μέρος της (7.2) είναι επίσης μη αρνητικό. Επιπλέον, $\phi_{i, \min} I \leq R^{(i)} \leq \phi_{i, \max} I$, $\phi_{i, \min} \triangleq \min_{n \in \mathcal{S}; n_i > 0} \phi_i(\mathbf{n})$ και το $\phi_{i, \max}$ ορίζεται παρόμοια. Συνδυάζοντας τα γεγονότα αυτά και την αναπαράσταση μέσω ολοκληρώματος της (7.2), έχουμε ότι $\phi_{i, \min} t e^{-B_i(s)t} \leq -\frac{d}{ds} e^{-B_i(s)t} \leq \phi_{i, \max} t e^{-B_i(s)t}$, αποδεικνύοντας τον ισχυρισμό σχετικά με την πρώτη παράγωγο. Περαιτέρω παραγωγή της (7.2) ως προς s μας δίνει

$$(-1)^{j+1} \frac{d^{j+1}}{ds^{j+1}} e^{-B_i(s)t} = \sum_{l=0}^j \binom{j}{l} \int_0^t (-1)^l \frac{d^l}{ds^l} e^{-B_i(s)(t-u)} R^{(i)} (-1)^{j-l} \frac{d^{j-l}}{ds^{j-l}} e^{-B_i(s)u} du$$

και επαγωγική εφαρμογή της προηγούμενης συλλογιστικής οδηγεί τελικά στο ότι

$$(\phi_{i, \min} t)^j \exp\{-B_i(s)t\} \leq (-1)^j \frac{d^j}{ds^j} \exp\{-B_i(s)t\} \leq (\phi_{i, \max} t)^j \exp\{-B_i(s)t\},$$

για όλα τα $j \geq 0$, όλα τα s και κάθε $t \geq 0$, ολοκληρώνοντας την απόδειξη του ισχυρισμού.

Συνδυασμός των φραγμάτων αυτών με τις εκφράσεις για την $S_i(t)$ μας οδηγεί στο

$$\left(\sum_{j=0}^{m_i-1} \frac{(\phi_{i, \min} \mu_i t)^j}{j!} \right) \nu_i \exp\{-B_i(\mu_i)t\} \mathbf{1} \leq S_i(t) \leq \left(\sum_{j=0}^{m_i-1} \frac{(\phi_{i, \max} \mu_i t)^j}{k!} \right) \nu_i \exp\{-B_i(\mu_i)t\} \mathbf{1},$$

έτσι ώστε η μορφή της $S_i(t)$ «παραλληλίζει» τη μορφή Erlang της κατανομής των απαιτήσεων εξυπηρέτησης στην ίδια κλάση. Η εκθετική φθίση της $S_i(t)$ κυβερνάται από μία έκφραση του εκθετικού ενός πίνακα, όπως εκείνο του Θεωρήματος 2. Ο αντίστοιχος πίνακας $B_i(\mu_i)$ είναι της μορφής (4.9), αλλά τώρα μ_i

είναι ο ρυθμός εξυπηρέτησης σε κάθε στάδιο της κατανομής Erlang και όχι το αντίστροφο του πλήρους μέσου της απαίτησης εξυπηρέτησης.

Στη γενική περίπτωση, η κατανομή των απαιτήσεων εξυπηρέτησης της κλάσης i θα έχει τυχαία μορφή τύπου-φάσης, έτσι $\Pr\{T_i > t\} = \mathbf{w}_i \exp\{-\Sigma_i t\} \mathbf{1}$, όπου Σ_i είναι ο πίνακας που ελέγχει τις μεταβάσεις μεταξύ φάσεων και \mathbf{w}_i είναι το αρχικό διάνυσμα πιθανοτήτων. Θεωρώντας την κανονική μορφή Jordan του Σ_i όταν η τελευταία ισότητα εφαρμόζεται στην (4.10), βλέπουμε ότι η $S_i(t)$ είναι ένας γραμμικός συνδυασμός από εκφράσεις, η καθεμία από τις οποίες, έχει τη μορφή που συναντήσαμε είτε στο παράδειγμα της υπερκθετικής κατανομής είτε σε εκείνο της κατανομής Erlang που συζητήθηκε νωρίτερα. Οι παράμετροι μ_{ij} στην έκφραση για το πρώτο παράδειγμα, όπως επίσης και οι μ_i στο δεύτερο, πρέπει τώρα να αντικατασταθούν από ιδιοτιμές του Σ_i (οι οποίες εγγυημένα έχουν θετικό πραγματικό μέρος και η ιδιοτιμή με το μικρότερο πραγματικό μέρος είναι πραγματική). Όλες οι ιδιοτιμές του Σ_i που αντιστοιχούν σε διαγώνιους στοιχειώδεις υποπίνακες Jordan (δηλαδή έχουν πλήρες σύνολο ιδιοδιανυσμάτων χωρίς κύρια διανύσματα) συνεισφέρουν στην κατανομή του χρόνου απόκρισης καθαρά εκθετικές υποεκφράσεις, όπως εκείνες που προέκυψαν στην ειδική περίπτωση της υπερκθετικής κατανομής. Οι ιδιοτιμές που αντιστοιχούν στους υποπίνακες Jordan και σχετίζονται με κύρια διανύσματα συνεισφέρουν υποεκφράσεις που συμπεριλαμβάνουν τις παραγώγους του εκθετικού του πίνακα, όπως στην ειδική περίπτωση της κατανομής Erlang. Όπως δείξαμε, εκφράσεις αυτής της μορφής είναι φραγμένες από αντίστοιχες απλούστερες εκφράσεις με μορφή δύναμης-εκθετικού.

Προκειμένου να συζητηθεί η ασυμπτωτική συμπεριφορά της $S_i(t)$, έστω ότι $\mu_i^* \triangleq \sigma_1(\Sigma_i)$ είναι η ιδιοτιμή του Σ_i με το ελάχιστο πραγματικό μέρος. Από την προηγούμενη ανάλυση, αν όλοι οι υποπίνακες Jordan που περιλαμβάνουν αυτήν την ιδιοτιμή είναι διαγώνιοι, τότε $\Pr\{T_i > t\} = O(e^{-\mu_i^* t})$ και $S_i(t) = \Pr\{V_i > t\} = O(e^{-\sigma_1(B_i(\mu_i^*))t})$, καθώς $t \rightarrow \infty$. Αν αυτή η ιδιοτιμή χαρακτηρίζεται από υποπίνακες Jordan που σχετίζονται με κύρια διανύσματα, έστω ότι το $m_i > 1$ συμβολίζει το μέγεθος του μεγαλύτερου τέτοιου υποπίνακα Jordan. Τότε, $\Pr\{T_i > t\} = O(t^{m_i-1} e^{-\mu_i^* t})$ και $S_i(t) = \Pr\{V_i > t\} = O(t^{m_i-1} e^{-\sigma_1(B_i(\mu_i^*))t})$, καθώς $t \rightarrow \infty$. Έτσι, για τυχαία τύπου-φάσης κατανομή απαιτήσεων εξυπηρέτησης, $S_i(t) = \Pr\{V_i > t\} = O(\Pr\{T_i > \zeta_i t\})$, καθώς το $t \rightarrow \infty$, όπου $\zeta_i \triangleq \sigma_1(B_i(\mu_i^*)) / \mu_i^*$ και $\mu_i^* = -\lim_{t \rightarrow \infty} t^{-1} \log \Pr\{T_i > t\} = \sigma_1(\Sigma_i)$. Με άλλα λόγια, τα συστήματα SDPS με μαρκοβιανές απαιτήσεις εξυπηρέτησης τύπου-φάσης υπακούουν σε μία μορφή της προσέγγισης μειωμένου ρυθμού εξυπηρέτησης, με συντελεστή κλίμακας ίσο με ζ_i . Όπως αναφέρθηκε στο Κεφάλαιο 3, η προσέγγιση μειωμένου ρυθμού εξυπηρέτησης ελέγχει την ουρά των κατανομών του χρόνου απόκρισης σε συστήματα SDPS που χαρακτηρίζονται από κα-

τανομές απαιτήσεων εξυπηρέτησης βαριάς ουράς [18]. Επίσης σημειώνουμε ότι στο περιβάλλον των κατανομών τύπου-φάσης, μία ακόμα πιο απλή σχέση συνδέει τις λογαριθμικές ασύμπτωτες των δυο κατανομών, συγκεκριμένα έχουμε $\lim_{t \rightarrow \infty} t^{-1} \log \Pr\{V_i > t\} = \sigma_1(B_i(\mu_i^*)) = \zeta_i \lim_{t \rightarrow \infty} t^{-1} \log \Pr\{T_i > t\}$.

Εφόσον οι συναρτήσεις των ρυθμών εξυπηρέτησης $\phi_i(\cdot)$ εξαρτώνται ακόμα μόνο από το διάνυσμα κατάστασης \mathbf{n} , η Πρόταση 2 συνεχίζει να εφαρμόζεται για το φράξιμο του $\sigma_1(B_i(\mu_i^*))$, δεδομένου ότι η παράμετρος μ_i εκεί αντικαθίσταται τώρα από τα μ_i^* . Οι πιθανότητες $\pi^{(i)}(\mathbf{n})$ στην πρόταση υποδηλώνουν ακόμα τις πιθανότητες σταθερής κατάστασης με τις οποίες υπάρχουν οι πληθυσμοί πελατών στο σύστημα. Ωστόσο, δεν αντιστοιχούν πλέον σε απλές καταστάσεις της εκτεταμένης αλυσίδας Markov $M^{(i)}$ αλλά τώρα είναι ίσες με αθροίσματα των πιθανοτήτων σταθερής κατάστασης κατάλληλων τέτοιων απλών καταστάσεων. Στην περίπτωση που το σύστημα διαθέτει την ιδιότητα ισορροπίας, οι πιθανότητες $\pi^{(i)}(\mathbf{n})$ είναι αναλλοίωτες στις λεπτομέρειες των κατανομών των απαιτήσεων εξυπηρέτησης πέρα από τη μέση τιμή τους και έτσι είναι ίσες με τις πιθανότητες σταθερής κατάστασης της απλούστερης αλυσίδας Markov στην (3.8). Οι παράμετροι μ_i εκεί εκφράζουν τώρα τις μέσες απαιτήσεις εξυπηρέτησης των αντίστοιχων κλάσεων. Επιπλέον, αυτές οι πιθανότητες μπορούν να εκφραστούν σε κλειστή μορφή, όπως συζητήθηκε στην Ενότητα 4.2.

Παρακάτω συζητούμε φαινόμενα διαχωρισμού χρονικής κλίμακας. Η συνθήκη για την εφαρμοσιμότητα του καθεστώτος αυτού στην πιο γενική περίπτωση κατανομών τύπου-φάσης παράγεται από την (5.1), με αντικατάσταση των ρυθμών ολοκλήρωσης εξυπηρέτησης μ_i από τις νόρμες των πινάκων τύπου-φάσης $\|\Sigma_i\|$, για οποιαδήποτε επιλογή νόρμας πίνακα $\|\cdot\|$, ομοιόμορφα για όλες τις κλάσεις. Εφόσον η έκφραση για την $S_i(t)$ είναι ένας γραμμικός συνδυασμός εκφράσεων εκθετικών πινάκων ή/και των παραγώγων τους, μπορεί να δει κανείς ότι οι προσεγγίσεις των κατανομών του χρόνου απόκρισης των Θεωρημάτων 3 και 4 για κλάσεις με γρήγορη και αργή δυναμική αντίστοιχα συνεχίζουν να ισχύουν. Αυτό μπορεί να αποδειχθεί με επιχειρήματα όμοια με εκείνα που χρησιμοποιήθηκαν για την εκθετική περίπτωση, με εφαρμογή των Λημμάτων 2 και 3 μαζί με τις παραλλαγές αυτών για τις παραγώγους των εκθετικών των πινάκων. Οι τελευταίες παράγονται εφαρμόζοντας την (7.2) και το αντίστοιχο κομμάτι της για παραγώγους ανώτερης τάξης. Τα αποτελέσματα της Ενότητας 6.1 που σχετίζονται με το συνδυαστική επίδραση του διαχωρισμού χρονικής κλίμακας και της τροποποίησης υπέρ των γρήγορων κλάσεων επίσης συνεχίζουν να βρίσκονται σε ισχύ.

Σημειώνουμε ότι παρά την επέκταση του χώρου καταστάσεων που απαιτείται για να διαπραγματευθούμε τις κατανομές τύπου-φάσης, το αναλλοίωτο των ρυθμών αφίξεων Poisson $\lambda_i(\cdot)$ και οι ρυθμοί εξυπηρέτησης $\phi_i(\cdot)$ από τις συ-

γκεκριμένες φάσεις που καταλαμβάνονται κάθε στιγμή, οδηγούν σε μία απλοποίηση των μοντέλων που προκύπτουν από την αποσύνθεση και τη συνάθροιση. Ειδικότερα, τα μειωμένα υποσυστήματα που ελέγχουν τις γρήγορες κλάσεις έχουν ακόμα δείκτη το υποδιάνυσμα n_s , προσμετρώντας τους μόνιμους πελάτες των αργών κλάσεων που είναι παρόντες. Ωστόσο, οι πίνακες και τα διανύσματα που σχετίζονται με κάθε τέτοιο μειωμένο υποσύστημα, διατρέχουν ένα εκτεταμένο χώρο καταστάσεων που συσχετίζεται με τις κατανομές τύπου-φάσης του χρόνου απόκρισης των γρήγορων κλάσεων.

Παρόμοια, το συναθροιστικό σύστημα χαρακτηρίζεται από έναν εκτεταμένο χώρο καταστάσεων, αυτή τη φορά αντίστοιχο με τις κατανομές τύπου-φάσης του χρόνου απόκρισης των αργών κλάσεων. Η εξίσωση (5.4) για τους ρυθμούς αφίξεων Poisson και τους ρυθμούς εξυπηρέτησης του συναθροιστικού συστήματος συνεχίζει να εφαρμόζεται, με τις ποσότητες $\hat{\pi}^{(n_s)}(n_f)$ εκεί και πάλι να εκφράζουν τις πιθανότητες σταθερής κατάστασης με τις οποίες πληθυσμοί των πελατών των γρήγορων κλάσεων υπάρχουν στο μειωμένο υποσύστημα με δείκτη n_s . Σχετικά με την Πρόταση 2, αυτή είναι πλέον ένα άθροισμα πιθανοτήτων σταθερής κατάστασης κατάλληλων στοιχειωδών καταστάσεων στον εκτεταμένο χώρο καταστάσεων για αυτό το υποσύστημα. Για συναρτήσεις ρυθμών εξυπηρέτησης με την ιδιότητα της ισορροπίας (του αρχικού, «πλήρους» συστήματος και έτσι επίσης για τις συναρτήσεις όλων των μειωμένων υποσυστημάτων), ο υπολογισμός των πιθανοτήτων αυτών απλοποιείται. Παρόμοια, η εξίσωση (5.6) και πάλι ελέγχει τους συντελεστές που πολλαπλασιάζουν τις CPDF του χρόνου απόκρισης στο Θεώρημα 3 ενώ το $\bar{\pi}(n_s)$ εκεί εκφράζει την στάσιμη πιθανότητα ύπαρξης πληθυσμού n_s στο συναθροιστικό σύστημα.

Κλείνουμε σημειώνοντας ότι δεν μεταφέρονται όλα τα αποτελέσματα του Κεφαλαίου 5 στη γενικότερη περίπτωση των κατανομών τύπου-φάσης. Συγκεκριμένα, το Λήμμα 4 (άρα και η Πρόταση 3) δεν μπορεί να εφαρμοστεί εγγυημένα επειδή όταν οι απαιτήσεις εξυπηρέτησης δεν είναι αυστηρά εκθετικές, το υπόλοιπο της απαίτησης εξυπηρέτησης δεν είναι στοχαστικά ισοδύναμο με μία πλήρη απαίτηση εξυπηρέτησης για την ίδια κλάση. Έτσι, για ξεχωριστούς θαλάμους εξυπηρέτησης της μορφής (3.2), όταν οι απαιτήσεις εξυπηρέτησης δεν είναι καθαρά εκθετικές, δεν μπορούμε να εγγυηθούμε πλέον ότι η κατανομή του χρόνου απόκρισης για μία γρήγορη κλάση θα κυριαρχείται από την κατανομή που αντιστοιχεί στο μειωμένο υποσύστημα με δείκτη το μέγιστο δυνατό πληθυσμό πελατών των αργών κλάσεων. Ωστόσο, στη βάση του γενικού χαρακτήρα αυτού του αποτελέσματος, εικάζεται ότι μάλλον είναι δυνατό να γενικευθεί για απαιτήσεις εξυπηρέτησης τύπου-φάσης, τουλάχιστον με μία ασυμπτωτική έννοια, ελέγχοντας την ουρά της κατανομής του χρόνου απόκρισης.



Κεφάλαιο 8

Μέτρα Επίδοσης και Διαχωρισμός Χρονικής Κλίμακας στον Ευκαιριακό Χρονοπρογραμματισμό

ΜΕΤΑ από την εκτεταμένη ανάλυση του γενικότερου προβλήματος των Κεφαλαίων 3-7 που αναφέρεται σε συστήματα SDPS, επιστρέφουμε στο πρόβλημα του ευκαιριακού χρονοπρογραμματισμού σε ασύρματα δίκτυα τηλεπικοινωνιών, όπως αυτό διατυπώθηκε στο Κεφάλαιο 2. Συνεχίζουμε με την εφαρμογή των αποτελεσμάτων που προέκυψαν ώστε να καταλήξουμε σε χρήσιμα μέτρα επίδοσης και να διερευνήσουμε θέματα δικαιοσύνης μέσω του διαχωρισμού χρονικής κλίμακας. Η αρχή γίνεται σε αυτό το κεφάλαιο του οποίου ο σκοπός είναι διττός. Αρχικά, για μία γενική μορφή της συνάρτησης που εισήχθη στο Κεφάλαιο 2 ως *κέρδος διαφορισμού πολλαπλών χρηστών* (δηλαδή χωρίς ακόμα αναφορά σε συγκεκριμένο αλγόριθμο χρονοπρογραμματισμού), θα δοθεί η συσχέτιση με το γενικό πρόβλημα συστημάτων SDPS και θα παρουσιαστούν τα βασικά μέτρα επίδοσης. Έπειτα, τα αποτελέσματα αυτά εξειδικεύονται για τη σημαντική περίπτωση κατά την οποία οι κλάσεις χαρακτηρίζονται από μεγάλες διαφορές στο μέσο ρυθμό μετάδοσης.

Στα πλαίσια της μελέτης που ακολουθεί κάνουμε τη συνήθη για τηλεπικοινωνιακά συστήματα υπόθεση ότι τα τερματικά της κλάσης i , $i \in \mathcal{K}$, ενεργοποιούνται (δηλαδή επιζητούν εξυπηρέτηση από το δίκτυο σύμφωνα με τον ορισμό της ενεργοποίησης που δόθηκε στην Ενότητα 2.2) ακολουθώντας μία μη εξαρτώμενη από την κατάσταση του συστήματος διαδικασία αφίξεων Poisson με μέσο ρυθμό λ_i . Παράλληλα, υποθέτουμε ότι τα ενεργά τερματικά διαθέτουν απαιτήσεις εξυπηρέτησης εκθετικά κατανομημένες (σε μονάδες δεδομένων) με μέση τιμή \bar{v} , ανεξαρτήτως της κλάσης στην οποία ανήκει το καθένα. Αυτού

του τύπου η μοντελοποίηση αναφέρεται σε τερματικά με παρόμοιες εφαρμογές, τα οποία διαφοροποιούνται μόνο εξαιτίας του μέσου ρυθμού μετάδοσης (π.χ., διαφορετική απόσταση από το σταθμό βάσης). Ωστόσο η ανάλυση μπορεί να γενικευθεί ώστε να συμπεριλάβει διαφορετικά μεγέθη ροών για κάθε κλάση. Ακολουθώντας τις εκφράσεις για τα κέρδη διαφορισμού πολλαπλών χρηστών $\phi_i(\mathbf{n})$ (βλ. π.χ., την (2.5) και δεδομένου του ορισμού του \tilde{r}_i ως μέσου ρυθμού μετάδοσης ενός και μόνο τερματικού κλάσης i , ο μέσος ρυθμός μετάδοσης κάποιου τερματικού κλάσης i όταν το σύστημα βρίσκεται στην κατάσταση \mathbf{n} , είναι $r_i(\mathbf{n}) = \tilde{r}_i \phi_i(\mathbf{n})$. Κατά συνέπεια, ο χρόνος ολοκλήρωσης της μετάδοσης για ένα τερματικό κλάσης i είναι επίσης εκθετικά κατανομημένος με ρυθμό $r_i(\mathbf{n})/\bar{v} = \tilde{r}_i \phi_i(\mathbf{n})/\bar{v}$. Επιπρόσθετα, σύμφωνα και με τον ορισμό του στην Ενότητα 3.3, ο κανονικοποιημένος μέσος ρυθμός μετάδοσης ενός τερματικού που εξυπηρετείται μόνο του στο σύστημα είναι $\mu_i = \tilde{r}_i/\bar{v}$. Λόγω του ορισμού του, το $1/\mu_i$ δηλώνει τη μέση απαίτηση εξυπηρέτησης για τερματικό κλάσης i εκπεφρασμένη σε μονάδες χρόνου.

Ας σημειωθεί ότι η χρήση ίδιου συμβολισμού με τις αντίστοιχες μεταβλητές του γενικού προβλήματος των Κεφαλαίων 3-7 έχει επιλεγεί σκόπιμα ώστε να καταστήσει ευκολότερο και πιο άμεσο τον παραλληλισμό των αντίστοιχων ποσοτήτων και εννοιών. Για παράδειγμα, ο μέσος ρυθμός μετάδοσης ενός τερματικού (πελάτη) στην κατάσταση \mathbf{n} , είναι σύμφωνα με τα παραπάνω $\mu_i \phi_i(\mathbf{n})$ και προέρχεται από δύο παράγοντες, τον κανονικοποιημένο μέσο ρυθμό μετάδοσης και το κέρδος διαφορισμού πολλαπλών χρηστών. Η σύνδεση με τις αντίστοιχες ποσότητες της Ενότητας 3.3 είναι πλέον προφανής.

Οι νέες ενεργοποιήσεις τερματικών υπόκεινται σε έλεγχο αποδοχής κλήσεων. Σχετικά με την αποδοχή νέων τερματικών, θα αποφύγουμε στο σημείο αυτό τον επαναπροσδιορισμό του χώρου καταστάσεων που προκύπτει από τη μορφή του εκάστοτε θαλάμου εξυπηρέτησης καθώς και των αντίστοιχων συνολικών συνόλων, με δεδομένο ότι οι χώροι καταστάσεων θεωρούνται ακριβώς όπως έχουν οριστεί στην Ενότητα 3.3 και συγκεκριμένα στις (3.1)-(3.5).

8.1 Συνολικά αποτελέσματα επίδοσης

Πλέον είμαστε σε θέση να δώσουμε αποτελέσματα για μέσες ποσότητες που αποτελούν σημαντικά μέτρα επίδοσης. Τα αποτελέσματα περιλαμβάνουν τη συνολική ρυθμαπόδοση, την πιθανότητα αποκλεισμού και το μέσο αριθμό ενεργών τερματικών κάθε κλάσης. Παράλληλα, παρέχονται αποτελέσματα για την κατανομή του χρόνου απόκρισης ενός ενεργού τερματικού.

Υπενθυμίζουμε ότι σύμφωνα με την ανάλυση που προηγήθηκε, η κατάσταση του συστήματος εξελίσσεται σύμφωνα με μία αλυσίδα Markov συνεχούς

χρόνου, που περιγράφεται από τον πίνακα ρυθμών μεταβάσεων M (βλ. π.χ., την (3.6)). Ο M χαρακτηρίζεται από το διάνυσμα πιθανοτήτων σταθερής κατάστασης π για το οποίο ισχύει $\pi M = \mathbf{0}$ και $\pi \mathbf{1} = 1$. Με χρήση των πιθανοτήτων σταθερής κατάστασης, η πιθανότητα αποκλεισμού ενός τερματικού κλάσης i προκύπτει ως

$$P_{B,i} = \sum_{\mathbf{n} \in \mathcal{C}_i} \pi(\mathbf{n}), \quad i \in \mathcal{K}. \quad (8.1)$$

Όταν ο θάλαμος εξυπηρέτησης είναι κοινός και άρα, ο δυνατός χώρος καταστάσεων χαρακτηρίζεται από ένα κοινό σύνορο (βλ. π.χ., την (3.4)), η πιθανότητα αποκλεισμού δεν εξαρτάται από την κλάση, με άλλα λόγια, $P_B = \sum_{\mathbf{n} \in \mathcal{C}} \pi(\mathbf{n})$. Ο μέσος αριθμός ενεργών τερματικών κλάσης i είναι

$$E[N_i] = \sum_{\mathbf{n} \in \mathcal{S}} n_i \pi(\mathbf{n}), \quad i \in \mathcal{K}. \quad (8.2)$$

Με εφαρμογή του νόμου του Little, δηλαδή διαιρώντας τον μέσο αριθμό ενεργών τερματικών κλάσης i με το μέσο ρυθμό αποδεκτών αφίξεων της ίδιας κλάσης, οδηγούμαστε στο μέσο χρόνο απόκρισης για τη συγκεκριμένη κλάση, δηλαδή

$$E[V_i] = \frac{E[N_i]}{\lambda_i(1 - P_{B,i})}, \quad i \in \mathcal{K}.$$

Ο μέσος χρόνος απόκρισης $E[V_i]$ συνδέεται άμεσα με τη ρυθμαπόδοση της κλάσης i , εφόσον η τελευταία ορίζεται ως ο λόγος του μέσου μεγέθους της ροής προς το μέσο χρόνο που απαιτείται για να εξυπηρετηθεί. Έτσι,

$$\tau_i = \frac{\bar{v}}{E[V_i]} = \frac{\lambda_i \bar{v} (1 - P_{B,i})}{E[N_i]} = \frac{\tilde{r}_i \rho_i (1 - P_{B,i})}{E[N_i]}, \quad i \in \mathcal{K}.$$

Εξάλλου, η συνολική ρυθμαπόδοση του συστήματος (όσον αφορά ένα σταθμό βάσης) είναι:

$$\tau = \sum_{\mathbf{n} \in \mathcal{S}} \pi(\mathbf{n}) \sum_{j \in \mathcal{K}} n_j \tilde{r}_j \phi_j(\mathbf{n}) = \bar{v} \sum_{j \in \mathcal{K}} \lambda_j (1 - P_{B,j}), \quad (8.3)$$

όπου η δεύτερη ισότητα οφείλεται στο γεγονός ότι σε ένα σταθερό σύστημα ο μέσος ρυθμός εισόδου πρέπει να ισούται με το μέσο ρυθμό εξόδου.

Τέλος, με άμεση εφαρμογή του Θεωρήματος 2 η κατανομή του χρόνου απόκρισης στο σύστημα είναι

$$S_i(t) = \nu_i \exp\{-B_i t\} \mathbf{1}, \quad i \in \mathcal{K}, \quad (8.4)$$

όπου $B_i \triangleq \mu_i R^{(i)} - M^{(i)}$ και $R^{(i)} \triangleq \text{diag}_{\mathbf{n} \in \mathcal{S} - \mathcal{C}_i} \{\phi_i(\mathbf{n} + \mathbf{e}_i)\}$. Ο $M^{(i)}$ είναι ο πίνακας ρυθμών μεταβάσεων με ένα επιπλέον μόνιμο τερματικό («μόνιμο πελάτη» κατά

τα λεγόμενα των προηγούμενων κεφαλαίων) στο σύστημα και έχει τη μορφή

$$M^{(i)}(\mathbf{n}, \mathbf{n}') = \begin{cases} \lambda_j \mathbf{I}_{\mathbf{n} + \mathbf{e}_i \notin \mathcal{C}_j}, & \mathbf{n}' = \mathbf{n} + \mathbf{e}_j, \quad j \in \mathcal{K}, \\ (n_i + 1) \mu_j \phi_j(\mathbf{n} + \mathbf{e}_i), & \mathbf{n}' = \mathbf{n} - \mathbf{e}_j, \quad j \in \mathcal{K}. \end{cases}$$

Για το ομοιογενές πρότυπο αφίξεων που θεωρούμε εδώ, τα στοιχεία του διανύσματος ν_i είναι ίσα με $\nu_i(\mathbf{n}) = \pi(\mathbf{n}) / (1 - P_{B,i})$, $\mathbf{n} \in \mathcal{S} - \mathcal{C}_i$. Επίσης, από τη συζήτηση της Ενότητας 4.1, η ασυμπτωτική συμπεριφορά της $S_i(t)$ καθώς $t \rightarrow \infty$ ελέγχεται από τη μικρότερη ιδιοτιμή του B_i , δηλαδή

$$\lim_{t \rightarrow \infty} \frac{\log S_i(t)}{t} = -\sigma_1(B_i). \quad (8.5)$$

8.2 Αποτελέσματα επίδοσης για κλάσεις με μεγάλες διαφορές ρυθμών εξυπηρέτησης

Τώρα επιστρέφουμε στα φαινόμενα διαχωρισμού χρονικής κλίμακας, τα οποία κάνουν την εμφάνισή τους όταν οι κλάσεις τερματικών παρουσιάζουν μεγάλες διαφορές στο σηματοθορυβικό λόγο και κατ' επέκταση στο μέσο ρυθμό μετάδοσης. Με αυτό τον τρόπο δίνεται ένα χειροπιαστό παράδειγμα για την αξία της ανάλυσης των Κεφαλαίων 5 και 6 καθώς και των αποτελεσμάτων της Ενότητας 2.4. Φαινόμενα διαχωρισμού των κλάσεων ανάλογα με την ισχύ του σήματος (σύμφωνα με την ανάλυση του Κεφαλαίου 2) ή αντίστοιχα, διαχωρισμού χρονικής κλίμακας (σύμφωνα με την πιο θεωρητική ματιά των Κεφαλαίων 5,6) εμφανίζονται όταν $\max_{i \in \mathcal{K}_s} \tilde{\gamma}_i / \min_{j \in \mathcal{K}_f} \tilde{\gamma}_j = \epsilon$, καθώς ο παράγων σύζευξης $\epsilon \rightarrow 0$, όπου υπενθυμίζουμε ότι γ_i είναι το μέσο SNR της κλάσης i . Λόγω της υπόθεσης γραμμικότητας μεταξύ σηματοθορυβικού λόγου και επιτεύξιμου ρυθμού μετάδοσης, η προηγούμενη ισότητα μεταφράζεται σε $\max_{i \in \mathcal{K}_s} \mu_i / \min_{j \in \mathcal{K}_f} \mu_j = O(\epsilon)$, δηλώνοντας ότι $\mu_i = O(1)$, $i \in \mathcal{K}_f$ και $\mu_i = O(\epsilon)$, $i \in \mathcal{K}_s$ καθώς το $\epsilon \rightarrow 0$. Με αυτό το δεδομένο, προκειμένου να διατηρηθεί η ομαλή λειτουργία του δικτύου υποθέτουμε ότι τα τερματικά των κλάσεων με ασθενές μέσο σήμα (αργές κλάσεις) εμφανίζονται επίσης σπάνια, δηλαδή $\lambda_i = \check{\lambda}_i \epsilon$, $i \in \mathcal{K}_s$ για κάποια τιμή $\check{\lambda}_i$ που είναι ανεξάρτητη του ϵ , ενώ $\lambda_i = O(1)$ όταν $i \in \mathcal{K}_f$.

Όπως έχουμε ήδη δείξει, όταν η δυναμική του συστήματος μπορεί να διαχωριστεί σε διαφορετικές χρονικές κλίμακες, ο συνολικός πίνακας ρυθμών μεταβάσεων M διαθέτει δομή NCD. Συνοψίζοντας τα αποτελέσματα της σχετικής ανάλυσης, οι υποπίνακες του M ελέγχουν τα υποσυστήματα όπου συμβαίνουν μόνο μεταβάσεις τερματικών με ισχυρό μέσο σήμα (γρήγορες κλάσεις) και το

κάθε υποσύστημα διαθέτει \mathbf{n}_s μόνιμα τερματικά από τις αργές κλάσεις. Οι υποπίνακες $\hat{M}^{(n_s)}$, $\mathbf{n}_s \in \bar{\mathcal{S}}$ διαθέτουν μη μηδενικά μη διαγώνια στοιχεία

$$\hat{M}^{(n_s)}(\mathbf{n}_f, \mathbf{n}'_f) = \begin{cases} \lambda_j \mathbf{I}_{(\mathbf{n}_s, \mathbf{n}_f) \notin \mathcal{C}_j}, & \mathbf{n}'_f = \mathbf{n}_f + \mathbf{e}_j, \quad j \in \mathcal{K}_f, \\ n_j \mu_j \phi_j(\mathbf{n}_s, \mathbf{n}_f), & \mathbf{n}'_f = \mathbf{n}_f - \mathbf{e}_j, \quad j \in \mathcal{K}_f, \end{cases} \quad (8.6)$$

όπου τα διαγώνια στοιχεία προκύπτουν από την απαίτηση ότι κάθε γραμμή των $\hat{M}^{(n_s)}$ πρέπει να έχει μηδενικό άθροισμα.

Αφότου η παραπάνω διαδικασία αποσύνθεσης ολοκληρωθεί, πρέπει να δημιουργηθεί ο συναθροιστικός πίνακας \bar{M} , ο οποίος είναι ο πίνακας ρυθμών μεταβάσεων που περιγράφει την εξέλιξη των τερματικών που ανήκουν στις αργές κλάσεις. Ο πίνακας \bar{M} δίνεται από τη σχέση

$$\bar{M}(\mathbf{n}_s, \mathbf{n}'_s) = \begin{cases} \bar{\lambda}_i(\mathbf{n}_s) \mathbf{I}_{\mathbf{n}_s \notin \bar{\mathcal{C}}_i}, & \mathbf{n}'_s = \mathbf{n}_s + \mathbf{e}_i, \quad i \in \mathcal{K}_s, \\ n_i \mu_i \bar{\phi}_i(\mathbf{n}_s), & \mathbf{n}'_s = \mathbf{n}_s - \mathbf{e}_i, \quad i \in \mathcal{K}_s, \end{cases} \quad (8.7)$$

όπου

$$\bar{\lambda}_i(\mathbf{n}_s) = \lambda_i \sum_{\mathbf{n}_f \in \mathcal{S}^{(\mathbf{n}_s + \mathbf{e}_i)}} \hat{\pi}^{(n_s)}(\mathbf{n}_f), \quad i \in \mathcal{K}_s \quad (8.8)$$

και

$$\bar{\phi}_i(\mathbf{n}_s) = \sum_{\mathbf{n}_f \in \mathcal{S}^{(n_s)}} \hat{\pi}^{(n_s)}(\mathbf{n}_f) \phi_i(\mathbf{n}_s, \mathbf{n}_f), \quad i \in \mathcal{K}_s. \quad (8.9)$$

Λόγω της υπόθεσης στην αρχή αυτού του κεφαλαίου ότι η διαδικασία αφίξεων Poisson δεν εξαρτάται από την κατάσταση του συστήματος, για την περίπτωση που ο χώρος καταστάσεων μπορεί να διαχωριστεί μεταξύ των ομάδων των κλάσεων (όπως π.χ., στην (3.5)) και άρα $\sum_{\mathbf{n}_f \in \mathcal{S}^{(\mathbf{n}_s + \mathbf{e}_i)}} \hat{\pi}^{(n_s)}(\mathbf{n}_f) = 1$, οι συναθροιστικοί ρυθμοί αφίξεων επίσης παραμένουν ανεξάρτητοι από την κατάσταση, δηλαδή

$$\bar{\lambda}_i(\mathbf{n}_s) = \lambda_i, \quad i \in \mathcal{K}_s, \quad \forall \mathbf{n}_s. \quad (8.10)$$

Ωστόσο, η περίπτωση είναι διαφορετική όταν ο θάλαμος εξυπηρέτησης είναι κοινός για την αργή και γρήγορη ομάδα κλάσεων, αφού $\bar{\lambda}_i(\mathbf{n}_s) = \lambda_i (1 - \hat{P}_{B,i}^{(n_s)})$, $i \in \mathcal{K}_s$, $\forall \mathbf{n}_s$, όπου $\hat{P}_{B,i}^{(n_s)} = \sum_{\mathbf{n}_f \in \bar{\mathcal{C}}_i^{(n_s)}} \hat{\pi}^{(n_s)}(\mathbf{n}_f)$, $\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}$, $i \in \mathcal{K}_f$ είναι η πιθανότητα αποκλεισμού σε ένα υποσύστημα με μεταβάσεις μόνο μεταξύ γρήγορων τερματικών και \mathbf{n}_s μόνιμο πληθυσμό από τις αργές κλάσεις.

Εκ κατασκευής, ο \bar{M} διαθέτει το δικό του διάνυσμα πιθανοτήτων σταθερής κατάστασης, με στοιχεία $\bar{\pi}(\mathbf{n}_s)$, $\mathbf{n}_s \in \bar{\mathcal{S}}$. Σύμφωνα με την ανάλυση της Ενότητας 5.1, το διάνυσμα πιθανοτήτων σταθερής κατάστασης του αρχικού, «πλήρους» συστήματος μπορεί να προσεγγιστεί με σφάλμα που παραμένει τάξης $O(\epsilon)$ καθώς $\epsilon \rightarrow 0$. Ειδικότερα,

$$\pi(\mathbf{n}_s, \mathbf{n}_f) = \bar{\pi}(\mathbf{n}_s) \hat{\pi}^{(n_s)}(\mathbf{n}_f) + O(\epsilon), \quad \mathbf{n}_s \in \bar{\mathcal{S}}, \quad \mathbf{n}_f \in \mathcal{S}^{(n_s)}, \quad \epsilon \rightarrow 0. \quad (8.11)$$

Με χρήση της (8.11), συνεχίζουμε με τη διατύπωση των μεγεθών επίδοσης της προηγούμενης ενότητας, όπως αυτά εφαρμόζονται στο περιβάλλον χρονικού διαχωρισμού. Η συνολική ρυθμαπόδοση προκύπτει άμεσα από την (8.3), με αντικατάσταση των $\pi(\mathbf{n}_s, \mathbf{n}_f)$ από την (8.11),

$$\tau = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}} \bar{\pi}(\mathbf{n}_s) \hat{\tau}^{(\mathbf{n}_s)} + O(\epsilon), \quad \epsilon \rightarrow 0. \quad (8.12)$$

Η $\hat{\tau}^{(\mathbf{n}_s)}$ είναι η συνολική ρυθμαπόδοση ενός υποσυστήματος με τη μόνιμη παρουσία \mathbf{n}_s τερματικών των αργών κλάσεων, δηλαδή

$$\hat{\tau}^{(\mathbf{n}_s)} = \sum_{j \in \mathcal{K}_f} \tilde{r}_j \sum_{\mathbf{n}_f \in \hat{\mathcal{C}}^{(\mathbf{n}_s)}} \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f) n_j \phi_j(\mathbf{n}_s, \mathbf{n}_f), \quad \mathbf{n}_s \in \bar{\mathcal{S}}. \quad (8.13)$$

Προφανώς, εφόσον το ότι $j \in \mathcal{K}_s$ σημαίνει ότι το \tilde{r}_j είναι τάξης $O(\epsilon)$ καθώς $\epsilon \rightarrow 0$, μόνο η ρυθμαπόδοση των γρήγορων κλάσεων συμμετέχει στην (8.13).

Με χρήση της (8.1), η πιθανότητα αποκλεισμού για τερματικό μίας γρήγορης κλάσης i , μπορεί να γραφεί ως $P_{B,i} = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}} \bar{\pi}(\mathbf{n}_s) \hat{P}_{B,i}^{(\mathbf{n}_s)} + O(\epsilon)$ καθώς $\epsilon \rightarrow 0$, όπου $\hat{P}_{B,i}^{(\mathbf{n}_s)} \triangleq \sum_{\mathbf{n}_f \in \hat{\mathcal{C}}_i^{(\mathbf{n}_s)}} \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f)$, $\mathbf{n}_s \in \bar{\mathcal{S}}$. Η πιθανότητα αποκλεισμού διατηρεί την εξάρτησή της από τη μορφή του θαλάμου εξυπηρέτησης και άρα του συνοριακού συνόλου καταστάσεων. Κατά συνέπεια, όταν όλες οι κλάσεις υπόκεινται σε ένα κοινό σόνορο L , (λόγω π.χ., χώρου καταστάσεων της μορφής (3.3)) $\hat{\mathcal{C}}_i^{(\mathbf{n}_s)} = \{\mathbf{n}_f \mid |\mathbf{n}_f| = L - |\mathbf{n}_s|\} = \hat{\mathcal{C}}^{(\mathbf{n}_s)}$, $\forall i \in \mathcal{K}$. Επίσης, όταν ο \mathcal{S} είναι μερικώς διαχωρισμένος (όπως π.χ., στην (3.5)), ισχύει πάντα ότι $\hat{\mathcal{C}}_i^{(\mathbf{n}_s)} = \{\mathbf{n}_f \mid |\mathbf{n}_f| = L_f\}$, $i \in \mathcal{K}_f$.

Ο μέσος αριθμός ενεργών τερματικών από την κλάση i , $i \in \mathcal{K}_f$, είναι

$$\mathbb{E}[N_i] = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}} \bar{\pi}(\mathbf{n}_s) \mathbb{E}[\hat{N}_i^{(\mathbf{n}_s)}] + O(\epsilon), \quad \epsilon \rightarrow 0. \quad (8.14)$$

όπου

$$\mathbb{E}[\hat{N}_i^{(\mathbf{n}_s)}] = \sum_{\mathbf{n}_f \in \hat{\mathcal{C}}^{(\mathbf{n}_s)}} n_i \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f), \quad \mathbf{n}_s \in \bar{\mathcal{S}}, \quad i \in \mathcal{K}_f.$$

Δεδομένου του ορισμού του $\hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f)$, το $\mathbb{E}[\hat{N}_i^{(\mathbf{n}_s)}]$ υποδηλώνει τον μέσο αριθμό τερματικών κλάσης i σε ένα υποσύστημα που περιέχει \mathbf{n}_s μόνιμα παρόντα τερματικά αργών κλάσεων.

Σε ό,τι αφορά την κατανομή του χρόνου απόκρισης ενός τερματικού κλάσης i , υπενθυμίζουμε ότι κατά την αποσύνθεση του συστήματος σε επιμέρους υποσυστήματα, εμφανίζονται οι ποσότητες $\bar{\gamma}_i(\mathbf{n}_s)$ της (5.6), οι οποίες για το ομοιογενές μοντέλο αφίξεων Poisson παίρνουν τη μορφή

$$\bar{\gamma}_i(\mathbf{n}_s) = \frac{\bar{\pi}(\mathbf{n}_s)(1 - \hat{P}_{B,i}^{(\mathbf{n}_s)})}{\sum_{\mathbf{n}'_s \in \bar{\mathcal{S}}^{(i)}} \bar{\pi}(\mathbf{n}'_s)(1 - \hat{P}_{B,i}^{(\mathbf{n}'_s)})}. \quad (8.15)$$

Με χρήση του Θεωρήματος 3, η κατανομή του χρόνου απόκρισης λαμβάνεται από ένα ζυγισμένο άθροισμα των κατανομών που αντιστοιχούν σε υποσυστήματα γρήγορων τερματικών, το καθένα με διαφορετικό συνδυασμό μόνιμα παρόντων αργών τερματικών. Οι συντελεστές του άθροίσματος είναι οι κανονικοποιημένες πιθανότητες της (8.15), δηλαδή

$$S_i(t) = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i(\mathbf{n}_s) \hat{S}_i^{(\mathbf{n}_s)}(t) + O(\epsilon), \quad \epsilon \rightarrow 0. \quad (8.16)$$

Η $\hat{S}_i^{(\mathbf{n}_s)}(t)$ είναι ακριβώς η κατανομή του χρόνου απόκρισης ενός συστήματος που χαρακτηρίζεται από \mathbf{n}_s τερματικά αργών κλάσεων και είναι περιορισμένο στο χώρο καταστάσεων $\hat{\mathcal{S}}^{(\mathbf{n}_s), (i)}$, $i \in \mathcal{K}_f$,

Με χρήση της (8.16), ο μέσος χρόνος απόκρισης της κλάσης i , $i \in \mathcal{K}_f$, προκύπτει άμεσα με ολοκλήρωση και των δύο μερών αυτής στο διάστημα $[0, \infty)$. Έπεται ότι,

$$\mathbb{E}[V_i] = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i(\mathbf{n}_s) \mathbb{E}[\hat{V}_i^{(\mathbf{n}_s)}] + O(\epsilon), \quad (8.17)$$

όπου

$$\mathbb{E}[\hat{V}_i^{(\mathbf{n}_s)}] \triangleq \hat{\nu}_i^{(\mathbf{n}_s)} (\hat{B}_i^{(\mathbf{n}_s)})^{-1} \mathbf{1}.$$

Με δεδομένα τα αποτελέσματα της Ενότητας 4.1, ο $\hat{B}_i^{(\mathbf{n}_s)}$, $\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}$, $i \in \mathcal{K}_f$ είναι ομαλός και μη αναγωγίμος (βλ. Λήμμα 1). Επιπλέον, για το ομοιογενές πρότυπο αφίξεων που θεωρούμε εδώ, $\hat{\nu}_i^{(\mathbf{n}_s)} = (1 - \hat{P}_{B,i}^{(\mathbf{n}_s)})^{-1} \hat{\pi}^{(\mathbf{n}_s)}$, $\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}$, $i \in \mathcal{K}_f$.

Όπως έχουμε δει, η ασυμπτωτική συμπεριφορά της $S_i(t)$ για τις γρήγορες κλάσεις σχετίζεται άμεσα με την ελάχιστη ιδιοτιμή του B_i καθώς και τον πληθυσμό πελατών των αργών κλάσεων \mathbf{n}_s που αντιστοιχεί στον υποπίνακα με την ελάχιστη ιδιοτιμή. Κατά συνέπεια, αν για κάποια διαμόρφωση του χώρου καταστάσεων ή/και των κερδών διαφορισμού πολλαπλών χρηστών είναι γνωστό το ποιος υποπίνακας $\hat{B}_i^{(\mathbf{n}'_s)}$ εμφανίζει την ελάχιστη ιδιοτιμή,

$$\mathbf{n}'_s = \arg \min_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{\sigma_1(\hat{B}_i^{(\mathbf{n}_s)})\}, \quad (8.18)$$

τότε προφανώς η διαμόρφωση με \mathbf{n}'_s τερματικά αποτελεί το χείριστο δυνατό σενάριο. Η συγκεκριμένη παρατήρηση θα χρησιμοποιηθεί στα επόμενα κεφάλαια για τους αλγορίθμους BR και NBR.

Συνεχίζουμε με την περίπτωση των αργών κλάσεων. Δεδομένης της (8.11), η πιθανότητα αποκλεισμού για τερματικό κλάσης i , $i \in \mathcal{K}_s$ είναι

$$\bar{P}_{B,i} = \sum_{\mathbf{n}_s \in \bar{\mathcal{C}}_i} \bar{\pi}(\mathbf{n}_s) \sum_{\mathbf{n}_f \in \mathcal{C}_i^{(\mathbf{n}_s)}} \hat{\pi}^{(\mathbf{n}_s)}(\mathbf{n}_f) + O(\epsilon), \quad \epsilon \rightarrow 0, \quad i \in \mathcal{K}_s. \quad (8.19)$$

όπου

$$\mathcal{C}_i^{(n_s)} = \{\mathbf{n}_f \geq \mathbf{0} \mid (\mathbf{n}_s, \mathbf{n}_f) \in \mathcal{S}, (\mathbf{n}_s, \mathbf{n}_f) + \mathbf{e}_i \notin \mathcal{S}\}, \quad i \in \mathcal{K}_s.$$

Όταν ο θάλαμος εξυπηρέτησης είναι κοινός για όλες τις κλάσεις, το συνολικό σύνολο δεν εξαρτάται από την κλάση i , δηλαδή $\bar{\mathcal{C}}_i = \bar{\mathcal{C}}, \forall i \in \mathcal{K}_s$. Κατά συνέπεια, η πιθανότητα αποκλεισμού για κάθε αργή κλάση δίνεται από τη σχέση $\bar{P}_B = \sum_{\mathbf{n}_s \in \bar{\mathcal{C}}} \bar{\pi}(\mathbf{n}_s)(1 - \hat{P}_B^{(n_s)}) + O(\epsilon)$, καθώς $\epsilon \rightarrow 0$. Επίσης, η εξίσωση (8.19) μπορεί να πάρει πιο απλή μορφή αν θεωρήσουμε θάλαμο εξυπηρέτησης με διαχωρισμό μεταξύ των ομάδων των κλάσεων, επειδή σε αυτή την περίπτωση $\sum_{\mathbf{n}_f \in \mathcal{C}_i^{(n_s)}} \hat{\pi}^{(n_s)}(\mathbf{n}_f) = 1$. Με εφαρμογή αυτής στην (8.19), έχουμε ότι $P_{B,i} = \sum_{\mathbf{n}_s \in \mathcal{C}_{s,i}} \bar{\pi}(\mathbf{n}_s) + O(\epsilon)$, καθώς $\epsilon \rightarrow 0$. Επιπρόσθετα, ο μέσος αριθμός τερματικών μίας αργής κλάσης i στο σύστημα είναι $E[N_i] = \sum_{\mathbf{n}_s \in \bar{\mathcal{S}}} n_i \bar{\pi}(\mathbf{n}_s) + O(\epsilon)$, $\epsilon \rightarrow 0$.

Τέλος, η κατανομή του χρόνου απόκρισης για μια αργή κλάση ($i \in \mathcal{K}_s$), προκύπτει από το Θεώρημα 4 και την αναφέρουμε επίσης εδώ για λόγους πληρότητας,

$$S_i(t) = \bar{S}_i(t) + O(\epsilon), \quad \epsilon \rightarrow 0, \quad (8.20)$$

όπου $\bar{B}_i \triangleq \mu_i \bar{R}^{(i)} - \bar{M}^{(i)}$, $\bar{R}^{(i)} \triangleq \text{diag}_{\mathbf{n}_s \in \bar{\mathcal{S}}^{(i)}} \{\bar{\phi}_i(\mathbf{n}_s)\}$ και $\bar{\nu}_i = (1 - P_{B,i})^{-1} \bar{\pi}$. Με εφαρμογή του νόμου του Little, ο μέσος χρόνος απόκρισης δίνεται από τη σχέση

$$\bar{V}_i = \frac{\bar{N}_i}{\lambda_i(1 - \bar{P}_{B,i})}, \quad i \in \mathcal{K}_s. \quad (8.21)$$



Κεφάλαιο 9

Σύγκριση των Αλγορίθμων BR και NBR

9.1 Αποτελέσματα επίδοσης του NBR

Σ ΥΝΕΧΙΖΟΥΜΕ στο κεφάλαιο αυτό με την παράθεση αποτελεσμάτων για τους αλγόριθμους BR και NBR στα πλαίσια της έως τώρα ανάλυσης. Παράλληλα, επιχειρούμε μία σύγκριση ως προς μέσα μέτρα επίδοσης αλλά και τη δικαιοσύνη που μπορούν να διατηρήσουν οι BR και NBR σε περιβάλλον διαχωρισμού χρονικής κλίμακας. Ξεκινούμε τη μελέτη με τον NBR για τον οποίο η ιδιότητα ισορροπίας που χαρακτηρίζει τα κέρδη διαφορισμού πολλαπλών χρηστών παρέχει απλούστερες εκφράσεις. Όπως έχει ήδη αναφερθεί (βλ. π.χ., (2.7)), για τον NBR ισχύει ότι $\phi_i(\mathbf{n}) = \phi(|\mathbf{n}|)$, $\forall i \in \mathcal{K}$, $\mathbf{n} \in \mathcal{S}$. Για τον ορισμό της ιδιότητας ισορροπίας και σχετικές ιδιότητες που απορρέουν από αυτή παραπέμπουμε στην Ενότητα 4.2 όπου έχει γίνει η σχετική ανάλυση.

Στην ειδική περίπτωση του NBR, συνδυασμός των (2.7) και (4.20) μας οδηγεί στο ότι

$$\frac{\phi_i(\mathbf{n} - \mathbf{e}_j)}{\phi_i(\mathbf{n})} = \frac{\phi_j(\mathbf{n} - \mathbf{e}_i)}{\phi_j(\mathbf{n})} = \frac{g(|\mathbf{n}| - 1)}{g(|\mathbf{n}|)}, \quad \forall \mathbf{n} : n_i, n_j > 0, i, j \in \mathcal{K}$$

και άρα, ένα σύστημα που χαρακτηρίζεται από κέρδη διαφορισμού πολλαπλών χρηστών της μορφής (2.7) πράγματι διαθέτει την ιδιότητα ισορροπίας. Η συνάρτηση ισορροπίας μπορεί να γραφεί ως $G(|\mathbf{n}|) \triangleq (\prod_{l=1}^{|\mathbf{n}|} g(l))^{-1}$ και σύμφωνα με την (4.21), το διάνυσμα πιθανοτήτων σταθερής κατάστασης είναι

$$\pi(\mathbf{n}) = \pi(\mathbf{0})G(|\mathbf{n}|) \prod_{l=1}^k \frac{\rho_l^{n_l}}{n_l!}, \quad \pi(\mathbf{0}) = \left(\sum_{\mathbf{n} \in \mathcal{S}} G(|\mathbf{n}|) \prod_{l=1}^k \frac{\rho_l^{n_l}}{n_l!} \right)^{-1}, \quad \mathbf{n} \in \mathcal{S}. \quad (9.1)$$

Με δεδομένο το αποτέλεσμα αυτό, είμαστε σε θέση να δώσουμε μία έκφραση για το διάνυσμα πιθανοτήτων σταθερής κατάστασης του συνολικού αριθμού ενεργών τερματικών στο σύστημα, ανεξαρτήτου κλάσης. Προς το σκοπό αυτό, ορίζουμε το σύνολο $\mathcal{S}_{[l]} \triangleq \{\mathbf{n} \mid |\mathbf{n}| = l\}$, το οποίο είναι το σύνολο διανυσμάτων κατάστασης με άθροισμα ενεργών τερματικών ίσο με l . Μετά από απλούς υπολογισμούς και με κατάλληλη εφαρμογή του αριστερού μέρους της (9.1), η πιθανότητα σταθερής κατάστασης να υπάρχουν στο σύστημα l τερματικά ενεργά είναι

$$\Pr\{|\mathbf{n}| = l\} = \sum_{\mathbf{n} \in \mathcal{S}_{[l]}} \pi(\mathbf{n}) = \pi(\mathbf{0})G(l)\frac{\rho^l}{l!}, \quad (9.2)$$

όπου $\rho = \sum_l \rho_l$. Επίσης, η συνθήκη κανονικοποίησης στο δεξί μέρος της (9.1) συνεπάγεται ότι $\pi(\mathbf{0})^{-1} = \sum_{m=0}^L \sum_{\mathbf{n}:|\mathbf{n}|=m} \pi(\mathbf{n}) = \sum_{m=0}^L \frac{G(m)\rho^m}{m!}$. Σημειώνουμε πως σε ό,τι ακολουθεί σχετικά με αποτελέσματα που αναφέρονται στον NBR, θεωρούμε θάλαμο εξυπηρέτησης της μορφής (3.3) για κάποιο $L \geq 1$. Η απλή μορφή του τελευταίου μέρους της (9.2) υποστηρίζει την επιλογή κοινού θαλάμου εξυπηρέτησης για όλες τις κλάσεις.

Για μεγαλύτερη απλοποίηση στο συμβολισμό, έστω ότι η πιθανότητα σταθερής κατάστασης να υπάρχουν l συνολικά ενεργά τερματικά, σε σύστημα με συνολική ένταση κίνησης ίση με ρ και μέγεθος θαλάμου εξυπηρέτησης L , συμβολίζεται με

$$\eta(l; \rho, L) \triangleq \sum_{\mathbf{n}:|\mathbf{n}|=l} \pi(\mathbf{n}) = \Pr\{|\mathbf{n}| = l\} = \pi(\mathbf{0})G(l)\frac{\rho^l}{l!}. \quad (9.3)$$

Όταν σε όλες τις κλάσεις διατίθεται κοινός θάλαμος εξυπηρέτησης, παρατηρούμε ότι $\mathcal{S} = \bigcup_{l=0}^L \mathcal{S}_{[l]}$ και $\mathcal{S}_{[l]} \cap \mathcal{S}_{[m]} = \emptyset, \forall l, m$. Κατά συνέπεια, η πιθανότητα αποκλεισμού, P_B είναι ανεξάρτητη της κλάσης και με χρήση της (9.3) έπεται ότι $P_B = \eta(L; \rho, L)$.

Το παραπάνω αποτελεί ένα αξιοσημείωτο αποτέλεσμα, το οποίο δηλώνει σαφώς ότι η πιθανότητα αποκλεισμού για τερματικό οποιασδήποτε κλάσης επηρεάζεται μόνο από το συνολικό φορτίο κίνησης και το μέγιστο αριθμό τερματικών που μπορεί να εξυπηρετήσει ταυτόχρονα το σύστημα. Επίσης, με χρήση της (9.2), οι πιθανότητες σταθερής κατάστασης του συστήματος είναι ακριβώς οι ίδιες όπως σε κάποιο αντίστοιχο σύστημα μίας κλάσης το οποίο χαρακτηρίζεται από το ίδιο μέγεθος θαλάμου εξυπηρέτησης, κέρδη διαφορισμού πολλαπλών χρηστών $\phi(|\mathbf{n}|)$ και ένταση κίνησης ίση με το άθροισμα των εντάσεων κίνησης που συνεισφέρουν όλες οι κλάσεις, $\rho = \sum_i \rho_i$.

Όσον αφορά τη συνολική ρυθμαπόδοση μίας κλάσης, η ανάλυση δε διαφέρει από εκείνη που έχουμε ακολουθήσει ως τώρα. Η ρυθμαπόδοση της κλάσης i , $i \in \mathcal{K}$, λαμβάνεται με εφαρμογή του αποτελέσματος της (9.2) και της αθροιστικής μορφής των κερδών διαφορισμού πολλαπλών χρηστών, δηλαδή

$$\tau_i = \sum_{\mathbf{n}:n_i>0} \tilde{r}_i g(|\mathbf{n}|) n_i \pi(\mathbf{n}) = (1 - P_B) \tilde{r}_i \rho_i, \quad (9.4)$$

ενώ η συνολική ρυθμαπόδοση δίνεται από τη σχέση

$$\tau = \sum_{i \in \mathcal{K}} \tau_i = (1 - P_B) \sum_{i \in \mathcal{K}} \tilde{r}_i \rho_i. \quad (9.5)$$

Επίσης, το δεξι σκέλος της (9.5) είναι πάντα ίσο με $\sum_{i \in \mathcal{K}} (1 - P_B) \lambda_i \bar{v}$, εφόσον το τελευταίο ισούται με τη συνολική κίνηση που γίνεται αποδεκτή από το σύστημα. Το αποτέλεσμα για τη συνολική ρυθμαπόδοση υποδηλώνει ότι ο NBR προσδίδει δικαιοσύνη με την έννοια ότι η μέση ρυθμαπόδοση κάθε κλάσης είναι ανάλογη με το μέσο επιτεύξιμο ρυθμό εξυπηρέτησης που την χαρακτηρίζει. Έτσι, ενώ οι πόροι του δικτύου κατανέμονται ανεξάρτητα του \tilde{r}_i (βλ. π.χ., (2.4) όπως εφαρμόζεται για $a_i = \tilde{r}_i^{-1}$), η ρυθμαπόδοση των γρήγορων κλάσεων δεν επηρεάζεται σε σημαντικό βαθμό.

Μία άλλη σημαντική απόρροια της ιδιότητας ισορροπίας είναι ότι ο μέσος αριθμός τερματικών κάθε κλάσης είναι ανάλογος με την ένταση κίνησης που χαρακτηρίζει τη συγκεκριμένη κλάση. Πράγματι, με εφαρμογή της (9.1) στην (8.2), έχουμε

$$\mathbb{E}[N_i] = \sum_{\mathbf{n}:0 < n_i \leq L} n_i \pi(\mathbf{n}) = \frac{\rho_i}{\rho} \sum_{m'=1}^L m' \eta(m'; \rho, L) = \frac{\rho_i}{\rho} \mathbb{E}[N].$$

Καταλήγουμε έτσι στο ότι το $\mathbb{E}[N_i]$ είναι ευθέως ανάλογο του $\mathbb{E}[N]$ με παράγοντα αναλογίας ίσο με το λόγο της έντασης κίνησης της κλάσης i προς το συνολικό άθροισμα εντάσεων κίνησης όλων των κλάσεων.

Με άμεση εφαρμογή του αποτελέσματος αυτού στο νόμο του Little, έχουμε ότι

$$\mathbb{E}[V_i] = \frac{\mathbb{E}[N_i]}{\lambda_i(1 - P_B)} = \frac{\rho_i}{\rho} \frac{\mathbb{E}[N]}{\lambda_i(1 - P_B)} = \frac{1}{\mu_i} \frac{\mathbb{E}[N]}{\rho(1 - P_B)} = \frac{\bar{v}}{\tilde{r}_i} \frac{\mathbb{E}[N]}{\rho(1 - P_B)}. \quad (9.6)$$

Το πρώτο κλάσμα στο τελευταίο μέρος της 9.6 ισούται με την κανονικοποιημένη μέση απαίτηση εξυπηρέτησης $1/\mu_i$, δηλαδή το μέσο χρόνο μετάδοσης όταν το τερματικό εξυπηρετείται μόνο του στο σύστημα. Κατά συνέπεια, για ένα σύστημα που χαρακτηρίζεται από κοινό θάλαμο εξυπηρέτησης για όλες τις

κλάσεις και λειτουργεί σύμφωνα με τον αλγόριθμο NBR, η μόνη επιπλοκή που συνεισφέρει η ταυτότητα της κλάσης στο μέσο χρόνο απόκρισης είναι εξαιτίας του μέσου επιτεύξιμου ρυθμού εξυπηρέτησης $\tilde{r}_i, i \in \mathcal{K}$. Όπως γίνεται αντιληπτό από το δεξί σκέλος της (9.6), ο μέσος χρόνος απόκρισης είναι μία συνάρτηση της συνολικής έντασης κίνησης και του συνολικού μεγέθους του θαλάμου εξυπηρέτησης, του τελευταίου μέσω του όρου της πιθανότητας αποκλεισμού. Η μέχρι τώρα ανάλυση δείχνει πως ο NBR διαχειρίζεται τις κλάσεις που έχουν διαφορετικό μέσο ρυθμό εξυπηρέτησης με δίκαιο τρόπο. Οι μέσες ποσότητες που ενδιαφέρουν παραμένουν αναλλοίωτες προς τα χαρακτηριστικά του ρυθμού εξυπηρέτησης πέρα από τη μέση του τιμή και το ίδιο ισχύει για το μέσο χρόνο απόκρισης. Για το λόγο αυτό, ο NBR πράγματι αποτελεί το πιο κατάλληλο σχήμα χρονοπρογραμματισμού, όταν η δικαιοσύνη μεταξύ των κλάσεων αποτελεί το κύριο ζητούμενο.

Συνεχίζουμε επιστρέφοντας στο περιβάλλον διαχωρισμού των κλάσεων σε γρήγορες και αργές κλάσεις. Ο βασικός σκοπός της ανάλυσης στο κομμάτι αυτό είναι να μελετηθεί η επίδραση του σχήματος διαχωρισμού του θαλάμου εξυπηρέτησης, τόσο για μέσα όσο και για ασυμπτωτικά αποτελέσματα. Στην περίπτωση αυτή, η δυναμική των τερματικών των γρήγορων κλάσεων περιγράφεται από μειωμένα υποσυστήματα, ένα για κάθε δυνατό συνδυασμό ενεργών τερματικών από τις αργές κλάσεις. Τα υποσυστήματα ελέγχονται από τους πίνακες ρυθμών μεταβάσεων $\hat{M}^{(n_s)}, n_s \in \bar{\mathcal{S}}$ (βλ. π.χ., (3.8) με την αντικατάσταση $l = n_s$).

Το διάνυσμα πιθανοτήτων σταθερής κατάστασης εντός του υποσυστήματος που χαρακτηρίζεται από n_s τερματικά αργών κλάσεων υπολογίζεται απευθείας με χρήση της ιδιότητας ισορροπίας για κάθε υποσύστημα. Με άμεση εφαρμογή της έκφρασης για τα $\pi(n_s, n_f)$, όπως αυτή δίνεται από τη δεξιά πλευρά της (9.1) για την ειδική περίπτωση του NBR, οδηγούμαστε στις ποσότητες που δίνονται από τις (8.13), (8.14) και (8.16). Αξίζει να σημειωθεί ότι σύμφωνα με το αποτέλεσμα στην (9.6), παρόλο που οι γρήγορες κλάσεις έχουν ρυθμαπόδοση ανάλογη του μέσου ρυθμού εξυπηρέτησης, ο μέσος χρόνος απόκρισης επηρεάζεται αισθητά, εφόσον ο υπολογισμός του ρ περιλαμβάνει την ένταση κίνησης όλων των κλάσεων.

Λαμβάνοντας υπόψη την ανάλυση της Ενότητας 5.3, η δυναμική των αργών κλάσεων ελέγχεται από το συναθροιστικό πίνακα ρυθμών μεταβάσεων, \bar{M} . Η μορφή του χώρου καταστάσεων σε συνδυασμό με την (8.8) υποδηλώνουν ότι οι ρυθμοί αφίξεων του \bar{M} έχουν τη μορφή

$$\bar{\lambda}_i(n_s) = \lambda_i \sum_{n_f \in \bar{\mathcal{S}}^{(n_s+e_i)}} \hat{\pi}^{(n_s)}(n_f) = \lambda_i(1 - \hat{P}_B^{(n_s)}), \quad n_s \in \bar{\mathcal{S}}^{(i)}.$$

Τα συναθροιστικά κέρδη διαφορισμού πολλαπλών χρηστών για τον \bar{M} προκύπτουν συνδυάζοντας την (2.7) με την (8.9), δηλαδή

$$\bar{g}(\mathbf{n}_s) = \left(\sum_{\mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s)} \pi(\mathbf{n}_s, \mathbf{n}_f) \right)^{-1} \sum_{\mathbf{n}_f \in \hat{\mathcal{S}}(\mathbf{n}_s)} \pi(\mathbf{n}_s, \mathbf{n}_f) g(|\mathbf{n}_s| + |\mathbf{n}_f|),$$

για $\mathbf{n}_s \in \bar{\mathcal{S}} : n_i > 0, i \in \mathcal{K}_s$. Μία εφαρμογή της παραπάνω στις (8.19), (8.20) και (8.21) αρκεί για να πάρουμε τις αντίστοιχες ποσότητες. Επίσης, από την (9.4) έπεται ότι η ρυθμαπόδοση των αργών κλάσεων παραμένει αμελητέα, όπως αναμενόταν.

Ωστόσο, τα αποτελέσματα σχετικά με την ασυμπτωτική συμπεριφορά του χρόνου απόκρισης των γρήγορων κλάσεων δεν είναι ενθαρρυντικά. Η περίπτωση που μελετούμε εδώ, δηλαδή σύστημα SDPS με κοινό θάλαμο εξυπηρέτησης μεγέθους L για όλες τις κλάσεις και κέρδη διαφορισμού πολλαπλών χρηστών που εξαρτώνται μόνο από το άθροισμα των ενεργών τερματικών, έχει αναλυθεί στο Κεφάλαιο 6. Το αποτέλεσμα μας δείχνει ότι $\sigma_1(B_i) = \mu_i g(L)$ και με χρήση της (8.5), $\lim_{t \rightarrow \infty} t^{-1} \log S_i(t) = -\mu_i g(L)$. Παράλληλα, η ανάλυση του Κεφαλαίου 6, δηλώνει ότι το παραπάνω σενάριο λαμβάνει χώρα όταν τα τερματικά των γρήγορων κλάσεων εξυπηρετούνται από κοινού με οποιοδήποτε συνδυασμό $L - 1$ τερματικών των αργών κλάσεων (ανεξαρτήτου της ταυτότητας της κλάσης). Για το λόγο αυτό τα αργά τερματικά τείνουν να καταλαμβάνουν όλο το διαθέσιμο χώρο υποβαθμίζοντας την απόδοση των γρήγορων στο χειρότερο δυνατό επίπεδο. Κατά συνέπεια, προκειμένου να αποφευχθεί η κατάσταση αυτή, ενδείκνυται ο περιορισμός του αριθμού των τερματικών από τις αργές κλάσεις που μπορεί να δεχτεί ο θάλαμος εξυπηρέτησης.

9.2 Αποτελέσματα επίδοσης του BR

Όταν το σύστημα λειτουργεί σύμφωνα με τον αλγόριθμο BR, τα κέρδη διαφορισμού πολλαπλών χρηστών δε διαθέτουν πλέον την αθροιστική μορφή της (2.7). Σαν συνέπεια, δεν ισχύει πλέον η ιδιότητα ισορροπίας και άρα, ούτε και οι συνδεδεμένες με αυτή ειδικές μορφές των εξισώσεων που εγγυώνται τη δικαιοσύνη μεταξύ των κλάσεων. Ωστόσο, σε περιβάλλον που δεν ισχύει ο διαχωρισμός χρονικής κλίμακας, τα γενικευμένα αποτελέσματα της Ενότητας 8 παραμένουν σε ισχύ για τον BR, δεδομένου ότι θεωρούμε κέρδη διαφορισμού πολλαπλών χρηστών της μορφής (2.6). Επιπλέον, σε περιβάλλον διαχωρισμού των κλάσεων όπως εκείνο που έχει ήδη αναλυθεί, το Θεώρημα 1 παρέχει πρόσθετες πληροφορίες για τα κέρδη του BR, οι οποίες σε συνδυασμό με την ανάλυση των Κεφαλαίων 5 και 6, οδηγούν σε χρήσιμα συμπεράσματα.

Το Θεώρημα 1 φανερώνει ότι σε περιβάλλον στο οποίο οι κλάσεις διαχωρίζονται σε μία ομάδα χαμηλού SNR (αργού ρυθμού μετάδοσης) και σε μια ομάδα υψηλού SNR (γρήγορου ρυθμού μετάδοσης), οι κλάσεις της γρήγορης ομάδας παραμένουν ανεπηρέαστες από την παρουσία τερματικών των αργών κλάσεων. Από την άλλη πλευρά, τα τερματικά των αργών κλάσεων λαμβάνουν πόρους εξυπηρέτησης πρακτικά μόνο όταν απουσιάζουν τα τερματικά των γρήγορων κλάσεων. Οι παραπάνω ιδιότητες μπορούν να αποδοθούν με την ακόλουθη μορφή της συνάρτησης κερδών διαφορισμού πολλαπλών χρηστών,

$$\phi_i(\mathbf{n}_s, \mathbf{n}_f) = \begin{cases} \phi_i(\mathbf{0}, \mathbf{n}_f), & i \in \mathcal{K}_f, \mathbf{n}_s = \mathbf{0}, \\ \phi_i(\mathbf{0}, \mathbf{n}_f) + o(\epsilon), & i \in \mathcal{K}_f, \mathbf{n}_s \neq \mathbf{0}, \\ \phi_i(\mathbf{n}_s, \mathbf{0}), & i \in \mathcal{K}_s, \mathbf{n}_f = \mathbf{0}, \\ o(\epsilon), & i \in \mathcal{K}_s, \mathbf{n}_f \neq \mathbf{0}, \end{cases} \quad (9.7)$$

καθώς $\epsilon \rightarrow 0$ (βλ. π.χ., το Θεώρημα 1 για λεπτομερή ορισμό των συναρτήσεων σφάλματος).

Η ειδική μορφή των $\phi_i(\mathbf{n}_s, \mathbf{n}_f)$, όπως αυτή εκφράζεται στην (9.7), έχει ήδη μελετηθεί στην Ενότητα 6.1 (με την ονομασία εκεί «ρυθμοί εξυπηρέτησης προνομιακοί για τις γρήγορες κλάσεις»). Σε αυτή την περίπτωση, η δυναμική των γρήγορων κλάσεων ελέγχεται από επί μέρους υποσυστήματα που αναφέρονται σε μειωμένο χώρο καταστάσεων, καθένα από τα οποία χαρακτηρίζεται από διαφορετικό συνδυασμό μόνιμων τερματικών («πελατών» στο πλαίσιο του Κεφαλαίου 5) από τις αργές κλάσεις. Κάθε μειωμένο υποσύστημα μπορεί να περιγραφεί αποτελεσματικά από μία αλυσίδα Markov συνεχούς χρόνου με απειροστικό γεννήτορα $\hat{M}^{(n_s)}$, $\mathbf{n}_s \in \bar{\mathcal{S}}$ (βλ. π.χ., (8.6)). Με αντικατάσταση των κερδών διαφορισμού πολλαπλών χρηστών στην (8.6) από τα αντίστοιχα κέρδη που προκύπτουν από το Θεώρημα 1, παρατηρούμε ότι οι ρυθμοί ολοκλήρωσης της εξυπηρέτησης που υπάρχουν σε κάθε $\hat{M}^{(n_s)}$, επηρεάζονται μόνο από ποσότητες που σχετίζονται με τα τερματικά των γρήγορων κλάσεων, δίνοντας $\hat{M}^{(n_s)} \approx \hat{M}^{(0)}$, για όλα τα $\mathbf{n}_s \in \bar{\mathcal{S}}$. Το αποτέλεσμα αυτό επιτρέπει μεγαλύτερη υπολογιστική και νοητική απλότητα καθώς η εξέλιξη των γρήγορων κλάσεων περιγράφεται από ένα και μόνο υποσύστημα, ανεξάρτητο του \mathbf{n}_s .

Όπως αναμένεται και διαισθητικά, ο αλγόριθμος BR ευνοεί τα τερματικά με καλές συνθήκες ασύρματου καναλιού (υψηλό μέσο SNR), ενώ επιβάλλει ένα είδος «ποινής» στα τερματικά με άσχημες συνθήκες μετάδοσης (χαμηλό μέσο SNR). Η διαφοροποίηση αυτή μπορεί να ποσοτικοποιηθεί με τη βοήθεια του ασυμπτωτικού ρυθμού φθίσης της κατανομής του χρόνου απόκρισης. Για διαχωρισμένους θαλάμους εξυπηρέτησης μεταξύ γρήγορων και αργών κλάσεων (όπως της μορφής (3.5)), η κατανομή του χρόνου απόκρισης των τερματικών

των γρήγορων κλάσεων κληρονομεί την ανεξαρτησία από την παρουσία των αργών τερματικών. Με εφαρμογή των αποτελεσμάτων της Ενότητας 6.1 παρατηρούμε ότι $S_i(t) \approx \hat{S}_i^{(0)}(t) \sum_{n_s \in \bar{\mathcal{S}}^{(i)}} \bar{\gamma}_i(\mathbf{n}_s)$, $i \in \mathcal{K}_f$ και με χρήση του ορισμού του $\bar{\gamma}_i(\mathbf{n}_s)$ στην (8.15) και της μορφής του θαλάμου εξυπηρέτησης, το παραπάνω παίρνει τη μορφή $S_i(t) \approx \hat{S}_i^{(0)}(t)$. Κατά συνέπεια έχουμε αποδείξει ότι τα τερματικά των γρήγορων κλάσεων εξελίσσονται στο σύστημα σαν αυτό να ήταν άδειο από τερματικά αργών κλάσεων και έτσι, η κατανομή του χρόνου απόκρισης των γρήγορων κλάσεων αντικατοπτρίζει το γεγονός της ενίσχυσης τους λόγω του αλγορίθμου χρονοπρογραμματισμού.

Τα προηγούμενα επιχειρήματα μπορούν να επεκταθούν ώστε να καλύψουν τα μέσα μεγέθη επίδοσης υπό την εφαρμογή του BR. Η μορφή της (8.14) για το μέσο αριθμό τερματικών κλάσης i , η οποία αναπτύχθηκε για τη γενική περίπτωση διαχωρισμού των κλάσεων, μπορεί τώρα να γραφεί ως

$$E[N_i] \approx E[\hat{N}_i^{(0)}], \quad i \in \mathcal{K}_f.$$

Ο μέσος χρόνος εξυπηρέτησης μιας γρήγορης κλάσης κληρονομεί επίσης την ανεξαρτησία από την παρουσία τερματικών των αργών κλάσεων. Εφαρμογή της (8.17) μας δείχνει ότι

$$E[V_i] \approx E[\hat{V}_i^{(0)}], \quad i \in \mathcal{K}_f.$$

Επίσης, η πιθανότητα αποκλεισμού για την ίδια κλάση είναι $P_{B,i} \approx \hat{P}_{B,i}^{(0)} = \sum_{n_f \in \hat{\mathcal{C}}_i^{(0)}} \hat{\pi}^{(0)}(\mathbf{n}_f)$. Με άλλα λόγια, όταν εφαρμόζεται ο BR και ο θάλαμος εξυπηρέτησης είναι διαχωρισμένος μεταξύ των γρήγορων και των αργών κλάσεων, η επίδοση των γρήγορων κλάσεων μπορεί να προσεγγιστεί από το υποσύστημα που εξυπηρετεί μόνο τερματικά από τις γρήγορες κλάσεις και είναι άδειο από τερματικά των αργών κλάσεων.

Παρόλα αυτά, η αναβάθμιση της επίδοσης των γρήγορων κλάσεων, μειονεκτεί λόγω της αδικίας που υφίστανται τα τερματικά των αργών κλάσεων. Πράγματι, σε αντίθεση με την περίπτωση του NBR, τα μέσα μεγέθη επίδοσης (π.χ., μέση ρυθμαπόδοση, μέσος χρόνος απόκρισης κ.λπ.) εξαρτώνται από το μέσο ρυθμό μετάδοσης της κλάσης, γεγονός που συνεισφέρει στην αδικία ως προς τις αργές κλάσεις. Ίσης σημασίας είναι η διαπίστωση ότι ο BR δε μεταχειρίζεται όλες τις γρήγορες κλάσεις με τον ίδιο τρόπο. Αντίθετα, τείνει να ευνοήσει μόνο τα τερματικά της γρηγορότερης κλάσης. Ως συνέπεια, ακόμα και αν ένα τερματικό ανήκει σε κλάση της γρήγορης ομάδας, μπορεί και πάλι να αντιμετωπίσει τα παραπάνω ανεπιθύμητα αποτελέσματα έλλειψης δικαιοσύνης.

Για την περίπτωση κοινού θαλάμου εξυπηρέτησης για όλες τις κλάσεις, η μόνη επίδραση που μπορεί να επιφέρει η παρουσία τερματικών των αργών

κλάσεων είναι η κατάληψη ελεύθερων θέσεων στο θάλαμο εξυπηρέτησης και άρα ο περιορισμός του διαθέσιμου χώρου για τερματικά των γρήγορων κλάσεων. Λαμβάνοντας υπόψη το σχόλιο αυτό, το διάνυσμα πιθανοτήτων σταθερής κατάστασης που αντιστοιχεί στον $\hat{M}^{(n_s)}, \hat{\pi}^{(n_s)}$, μπορεί να εξαρτάται από το n_s μόνο σε ένα σχήμα του θαλάμου εξυπηρέτησης που επιτρέπει το διαμοιρασμό του θαλάμου μεταξύ των κλάσεων.

Η εξέλιξη των τερματικών των αργών κλάσεων ελέγχεται από τον συναθροιστικό πίνακα \bar{M} (βλ. π.χ., (8.7)-(8.9)). Για την περίπτωση αυτή, το Θεώρημα 1 υποδηλώνει ότι για τις αργές κλάσεις τα $\phi_i(n_s, n_f)$ είναι τάξης $O(1)$ καθώς $\epsilon \rightarrow 0$, μόνο όταν $n_f = \mathbf{0}$, δηλαδή μόνο κατά την απουσία τερματικών από τις γρήγορες κλάσεις. Με αντικατάσταση των κερδών διαφορισμού πολλαπλών χρηστών της (8.9) από τα αντίστοιχα του Θεωρήματος 1, έχουμε ότι

$$\bar{\phi}_i(n_s) = \hat{\pi}^{(n_s)}(\mathbf{0})\phi_i(n_s, \mathbf{0}) + h^{(s)}(\epsilon; n_s, n_f), \quad \epsilon \rightarrow 0.$$

Ισοδύναμα, τα κέρδη διαφορισμού πολλαπλών χρηστών του συναθροιστικού πίνακα είναι ίσα με τα αντίστοιχα κέρδη ενός συστήματος που εξυπηρετεί μόνο τερματικά αργών κλάσεων, πολλαπλασιασμένα με την πιθανότητα ότι δεν υπάρχει κάποιο τερματικό γρήγορης κλάσης ενεργό. Αξίζει να σημειωθεί ότι η επίδοση των αργών κλάσεων σε κάθε περίπτωση είναι χαμηλή· η επιπλέον συνέπεια που υπεισέρχεται λόγω του BR είναι να μειωθεί η ήδη χαμηλή επίδοση κατά ένα παράγοντα $\hat{\pi}^{(n_s)}(\mathbf{0})$, ο οποίος ωστόσο, δεν εξαρτάται από το ϵ .

Από τη στιγμή που τα κέρδη διαφορισμού πολλαπλών χρηστών $\bar{\phi}_i(n_s)$ είναι γνωστά, οι εκφράσεις του Κεφαλαίου 8 μπορούν να χρησιμοποιηθούν ώστε να παρέχουν τις αντίστοιχες μετρικές επίδοσης. Είναι φανερό ότι κατά την εφαρμογή του BR, τα ενεργά τερματικά των αργών κλάσεων έχουν πρόσβαση στο δίκτυο μόνο απουσία γρήγορων τερματικών, ένα γεγονός που εκφράζεται μέσω της παρουσίας του $\hat{\pi}^{(n_s)}(\mathbf{0})$ σε όλες τις σχετικές ποσότητες. Όταν όμως επιπλέον, ο θάλαμος εξυπηρέτησης έχει μορφή ώστε να διαχωρίζει τις κλάσεις της γρήγορης ομάδας από εκείνες της αργής, οι ρυθμοί αφίξεων Poisson στο συναθροιστικό σύστημα συνεχίζουν να μην εξαρτώνται από την κατάσταση του συστήματος. Έτσι, η δυναμική των αργών κλάσεων εξελίσσεται ακριβώς όπως σε ένα σύστημα που εξυπηρετεί αποκλειστικά τερματικά αργών κλάσεων αλλά με μειωμένο ρυθμό κατά παράγοντα $\hat{\pi}^{(n_s)}(\mathbf{0})$. Αντίθετα για κοινό θάλαμο εξυπηρέτησης, κατά τη διαδικασία συνάθροισης θα πρέπει να ληφθεί υπόψη ότι οι συναθροιστικοί ρυθμοί αφίξεων εξαρτώνται από την κατάσταση του συστήματος (βλ. π.χ., την ανάλυση που ακολουθεί την (8.10)).

Τα αποτελέσματα της διερεύνησης της παρούσας ενότητας υποδεικνύουν ότι κατά την εφαρμογή του BR, η πρόσβαση στο σύστημα τερματικών των αργών κλάσεων δεν επηρεάζει σχεδόν καθόλου την επίδοση των γρήγορων

κλάσεων, για διαχωρισμένο θάλαμο εξυπηρέτησης. Πιο συγκεκριμένα, σε τέτοιου είδους περιβάλλον ο ασυμπτωτικός ρυθμός φθίσης της κατανομής του χρόνου απόκρισης παραμένει ανεξάρτητος από την παρουσία αργών τερματικών. Αυτό συμβαίνει εξαιτίας της ειδικής μορφής των κερδών διαφορισμού πολλαπλών χρηστών. Το κύριο μειονέκτημα του BR είναι ότι συμπεριφέρεται με άδικο τρόπο και μειώνει περαιτέρω την επίδοση τόσο των αργών κλάσεων αλλά και των υπολοίπων κλάσεων της γρήγορης ομάδας πέραν της γρηγορότερης όλων. Υπό αυτή την έννοια, εμφανίζεται η ανάγκη πρότασης ενός νέου σχήματος που συνδυάζει τα επιθυμητά χαρακτηριστικά του BR και του NBR, διατηρώντας δηλαδή τις καλές ασυμπτωτικές ιδιότητες του BR για τις γρήγορες κλάσεις, αλλά προσδίδοντας και ένα βαθμό δικαιοσύνης μεταξύ όλων των κλάσεων. Αυτό αποτελεί το θέμα του επόμενου κεφαλαίου.



Κεφάλαιο 10

Εισαγωγή του Αλγορίθμου SG-PF

ΑΚΟΛΟΥΘΕΙ η παρουσίαση ενός νέου αλγορίθμου που αποτελείται από συνδυασμό των BR και NBR και για το λόγο αυτό αποκαλείται αλγόριθμος Αναλογικά Δίκαιος - Κατηγορίας Ισχύος (αγ.: Strength Group - Proportional Fair (SG-PF)). Όπως είδαμε στο προηγούμενο κεφάλαιο, δύο είναι οι κύριες συνιστώσες σχετικά με τα θέματα επίδοσης και δικαιοσύνης που προκύπτουν από την ανάλυση των BR και NBR. Η πρώτη σημαντική παρατήρηση έγκειται στο ότι ο BR ενισχύει τα τερματικά των γρήγορων κλάσεων (και μάλιστα περισσότερο εκείνα της γρηγορότερης κλάσης), αγνοώντας πλήρως τα τερματικά των αργών κλάσεων κατά την εξυπηρέτηση των γρήγορων κλάσεων. Κατά συνέπεια, η γρηγορότερη κυρίως κλάση εμφανίζει ικανοποιητικό ασυμπτωτικό ρυθμό φθίσης της κατανομής του χρόνου απόκρισης με αντιστάθμισμα την αδικία έναντι των αργών κλάσεων. Δεύτερον, ο NBR αντιμετωπίζει όλες τις κλάσεις με δίκαιο τρόπο αλλά αποτυγχάνει να επωφεληθεί από τον υψηλό ρυθμό εξυπηρέτησης που εμφανίζουν οι γρήγορες κλάσεις. Ο αλγόριθμος SG-PF που προτείνεται και αναλύεται στη συνέχεια, συνδυάζει τις ευνοϊκές ασυμπτωτικές ιδιότητες της κατανομής του χρόνου απόκρισης που προκύπτουν από τη χρήση του BR μαζί με την αίσθηση δικαιοσύνης που προσδίδει η χρήση του NBR.

Ο SG-PF βρίσκει εφαρμογή σε περιβάλλον διαχωρισμού των κλάσεων καθώς διατηρεί τη διαφοροποίηση των κλάσεων σε αργή και γρήγορη ομάδα σύμφωνα με τον μέσο ρυθμό μετάδοσης της καθεμίας και λαμβάνει την απόφαση χρονοπρογραμματισμού σύμφωνα με την ακόλουθη λογική: Προκειμένου να διατηρηθούν τα θετικά χαρακτηριστικά του BR (δηλαδή ο ασυμπτωτικός ρυθμός φθίσης της κατανομής του χρόνου απόκρισης), τα τερματικά των αργών κλάσεων ανταγωνίζονται για να λάβουν πόρους του συστήματος μόνο όταν απουσιάζουν τα τερματικά των γρήγορων κλάσεων. Με άλλα λόγια, όταν στο

σύστημα υπάρχουν τερματικά των γρήγορων κλάσεων ο SG-PF αγνοεί πλήρως εκείνα των αργών κλάσεων. Η κίνηση αυτή αποσκοπεί στο να διατηρήσει τον κύριο διαχωρισμό που επιβάλλει ο BR. Ωστόσο, προκειμένου να αυξηθεί κάπως η δικαιοσύνη μεταξύ των γρήγορων κλάσεων, τα τερματικά που ανήκουν σε αυτές (όταν είναι ενεργά) εξυπηρετούνται με εφαρμογή του NBR. Κατά τον ίδιο τρόπο, όταν δεν υπάρχει κανένα ενεργό τερματικό γρήγορης κλάσης τα τερματικά των αργών ομάδων εξυπηρετούνται και πάλι με χρήση του NBR. Η μορφή του θαλάμου εξυπηρέτησης όπως έχουμε δει παίζει σημαντικό ρόλο σε τέτοιου είδους συστήματα. Στο προηγούμενο κεφάλαιο έγινε φανερό ότι η χειρότερη δυνατή περίπτωση για τον NBR προκύπτει όταν ο θάλαμος εξυπηρέτησης είναι γεμάτος από τερματικά των αργών κλάσεων. Για το λόγο αυτό, κατά τη λειτουργία του SG-PF προτείνεται ο διαχωρισμός του θαλάμου εξυπηρέτησης μεταξύ των κλάσεων της αργής και της γρήγορης ομάδας, με ξεχωριστά όρια για κάθε ομάδα, L_s και L_f , αντίστοιχα.

Το συγκεκριμένο σχήμα συνδυάζει τις επιθυμητές ιδιότητες των δύο βασικών αλγορίθμων. Εφόσον η δικαιοσύνη αποτελεί μία σημαντική παράμετρο κατά τη λήψη της απόφασης του τερματικού που θα εξυπηρετηθεί, η εφαρμογή του NBR εντός κάθε ομάδας κλάσεων εγγυάται ότι τα ενεργά τερματικά επιτυγχάνουν ρυθμαπόδοση ανάλογη του μέσου ρυθμού εξυπηρέτησής τους. Επιπρόσθετα, η βασική λογική του BR διατηρείται, καθώς τα τερματικά των αργών κλάσεων μπορούν να εξυπηρετηθούν μόνο κατά την απουσία γρήγορων τερματικών.

Στην παραπάνω περιγραφή έχουμε κάνει τη θεώρηση ότι η κατηγοριοποίηση των τερματικών σε ομάδες λαμβάνει χώρα στο σταθμό βάσης απευθείας κατά την ενεργοποίηση του τερματικού. Δηλαδή κατά την ενεργοποίηση, θεωρούμε ότι ο σταθμός βάσης διαθέτει «άμεση» γνώση για τον επιτεύξιμο ρυθμό μετάδοσης του τερματικού. Σε ένα πραγματικό σύστημα η διαδικασία αυτή απαιτεί μία εκτίμηση της απόστασης του τερματικού από το σταθμό βάσης. Η χρήση της συγκεκριμένης πληροφορίας επιτρέπει την κατηγοριοποίηση ενός τερματικού που μόλις έχει γίνει ενεργό, εφόσον υπάρχει αντιστοιχία μεταξύ της απόστασης από το σταθμό βάσης και τον μέσο ρυθμό μετάδοσης.

Η ιδέα της εκτίμησης της απόστασης ενός κινητού τερματικού έχει ξαναχρησιμοποιηθεί στο παρελθόν. Για μία αναφορά που εστιάζει στην εκτίμηση της θέσης ενός τερματικού σε ασύρματα δίκτυα επικοινωνιών τρίτης γενιάς αναφέρουμε εδώ το [64]. Εξάλλου, μελέτες που χρησιμοποιούν τον υπολογισμό της θέσης ενός τερματικού περιλαμβάνονται στα [30], [43] και [63]. Παρατηρούμε λοιπόν ότι η εκτίμηση της απόστασης από το σταθμό βάσης είναι τεχνολογικά εφικτή και υλοποιήσιμη.

Κατά την εφαρμογή του SG-PF, τα κέρδη διαφορισμού πολλαπλών χρη-

στών λαμβάνονται από την (2.8), με κατάλληλο περιορισμό του δείκτη i στα \mathcal{K}_s ή \mathcal{K}_f (αντιστοιχώντας στο σύνολο όπου ανήκει το i) δηλαδή,

$$\phi_i^{SG-PF}(\mathbf{n}_s, \mathbf{n}_f) = \begin{cases} \int_0^\infty x \hat{F}(x)^{|\mathbf{n}_f|-1} d\hat{F}(x) = g(|\mathbf{n}_f|), & i \in \mathcal{K}_f, \mathbf{n}_f \neq \mathbf{0} \\ \int_0^\infty x \hat{F}(x)^{|\mathbf{n}_s|-1} d\hat{F}(x) = g(|\mathbf{n}_s|), & i \in \mathcal{K}_s, \mathbf{n}_f = \mathbf{0}. \end{cases}$$

Έτσι για τις γρήγορες κλάσεις τα κέρδη διαφορισμού πολλαπλών χρηστών διαθέτουν την αθροιστική μορφή της (2.7), όπου τώρα το \mathbf{n} μειώνεται σε \mathbf{n}_f . Η παραπάνω μορφή των κερδών διαφορισμού και ο διαχωρισμένος θάλαμος εξυπηρέτησης που έχει υποθεθεί, έχουν ως αποτέλεσμα ότι τα τερματικά των γρήγορων κλάσεων δεν έχουν πλέον καμία αλληλεπίδραση με εκείνα των αργών. Κατά συνέπεια, το υποσύστημα που ελέγχει την εξέλιξη των γρήγορων κλάσεων δίνεται από την εξίσωση

$$\hat{M}^{(0)}(\mathbf{n}_f, \mathbf{n}'_f) = \begin{cases} \lambda_j \mathbf{I}_{(\mathbf{0}, \mathbf{n}_f) \notin \mathcal{C}_j}, & \mathbf{n}'_f = \mathbf{n}_f + \mathbf{e}_j, \quad j \in \mathcal{K}_f, \\ n_j \mu_j g(|\mathbf{n}_f|), & \mathbf{n}'_f = \mathbf{n}_f - \mathbf{e}_j, \quad j \in \mathcal{K}_f, \end{cases}$$

η οποία δεν είναι τίποτα περισσότερο από την (8.6), όπως αυτή εφαρμόζεται για την περίπτωση του SG-PF. Το γεγονός αυτό επιτρέπει την άμεση εξαγωγή εκφράσεων για τις μέσες ποσότητες επίδοσης που έχουμε δει ως τώρα, (π.χ., μέση ρυθμαπόδοση, μέσος χρόνος απόκρισης κ.λπ.) ακολουθώντας ακριβώς την ίδια μεθοδολογία με εκείνη του Κεφαλαίου 9 και με τις κατάλληλες αντικαταστάσεις $|\mathbf{n}|$ σε $|\mathbf{n}_f|$ και \mathcal{K} σε \mathcal{K}_f όπου είναι απαραίτητο.

Η πιθανότητα αποκλεισμού για τις ισχυρές/γρήγορες κλάσεις μπορεί να βρεθεί μέσω της (9.3) ως $P_B = \eta(L_f; \rho_f, L_f)$, όπου $\rho_f \triangleq \sum_{i \in \mathcal{K}_f} \rho_i$. Όπως αναμένουμε, η πιθανότητα αποκλεισμού αντιστοιχεί στο υποσύστημα που εξυπηρετεί αποκλειστικά τερματικά των γρήγορων κλάσεων. Δεδομένης της (9.5), η συνολική ρυθμαπόδοση των γρήγορων κλάσεων δίνεται από την

$$\tau_f = \sum_{i \in \mathcal{K}_f} \tau_i = (1 - P_B) \sum_{i \in \mathcal{K}_f} \tilde{r}_i \rho_i.$$

Επίσης, σύμφωνα με την ανάλυση της Ενότητας 9.1, ο μέσος αριθμός ενεργών τερματικών από τις ισχυρές κλάσεις είναι $E[N_i] = \frac{\rho_i}{\rho_f} E[N_f]$, όπου

$$E[N_f] = \sum_{m'=1}^{L_f} m' \eta(m'; \rho_f, L_f).$$

Ο μέσος χρόνος απόκρισης για τερματικό που ανήκει σε κάποια γρήγορη κλάση λαμβάνεται με εφαρμογή της (9.6) στο μειωμένο υποσύστημα το οποίο εξυπηρετεί αποκλειστικά τερματικά γρήγορων κλάσεων ενώ είναι άδειο από τερματικά αργών κλάσεων. Έτσι, όταν το μέγεθος του θαλάμου εξυπηρέτησης μειωθεί σε L_f και η συνολική ένταση κίνησης των γρήγορων κλάσεων είναι ρ_f ,

από την (9.6) έπεται ότι

$$E[V_i] = \frac{\bar{v}}{\tilde{r}_i \rho_f (1 - P_B)}, \quad i \in \mathcal{K}_f.$$

Σε αυτή την περίπτωση, μόνο οι γρήγορες κλάσεις συμμετέχουν στο μέσο χρόνο απόκρισης, σε αντίθεση με την (9.6), όπου οι κλάσεις και της γρήγορης και της αργής ομάδας λαμβάνονται υπόψη. Το αποτέλεσμα αυτό φανερώνει την έννοια δικαιοσύνης που εισάγει ο SG-PF εφόσον οι γρήγορες κλάσεις δεν επηρεάζονται από το φορτίο κίνησης των αργών κλάσεων.

Η κατανομή του χρόνου απόκρισης για κάποια γρήγορη κλάση μπορεί να δοθεί με χρήση των αποτελεσμάτων της Ενότητας 8.2. Για το σκοπό αυτό πρέπει να κατασκευαστεί ο πίνακας B_i με χρήση των κερδών διαφορισμού πολλαπλών χρηστών $\phi_i^{SG-PF}(\mathbf{0}, \mathbf{n}_f)$, $i \in \mathcal{K}_f$, $\mathbf{n}_f \neq \mathbf{0}$ και θεωρώντας μέγιστο μέγεθος θαλάμου εξυπηρέτησης L_f . Με εφαρμογή της (8.4) οδηγούμαστε στο

$$S_i(t) = \hat{\nu}_i^{(0)} \exp\{-\hat{B}_i^{(0)} t\} \mathbf{1}, \quad i \in \mathcal{K}_f,$$

ενώ η (8.5) υποδεικνύει ότι η ασυμπτωτική συμπεριφορά του χρόνου απόκρισης ελέγχεται από την ελάχιστη ιδιοτιμή του $\hat{B}_i^{(0)}$, δηλαδή

$$\lim_{t \rightarrow \infty} t^{-1} \log S_i(t) = -\sigma_1(\hat{B}_i^{(0)}), \quad i \in \mathcal{K}_f.$$

Τώρα στρέφουμε την προσοχή μας στις κλάσεις της αργής ομάδας και την επίδοσή τους κατά την εφαρμογή του SG-PF. Από την ανάλυση του Κεφαλαίου 8, η δυναμική και άρα και η κατανομή του χρόνου απόκρισης ελέγχονται από τους συναθροιστικούς πίνακες $\bar{M}^{(i)}$ και \bar{B}_i , $i \in \mathcal{K}_s$ αντιστοίχα. Βέβαια θα πρέπει να σημειωθεί ότι τώρα οι κανονικοποιημένοι ρυθμοί εξυπηρέτησης $\bar{\mu}_i$ πρέπει να μειωθούν κατά έναν παράγοντα ίσο με την πιθανότητα σταθερής κατάστασης ενός συστήματος άδειου από τερματικά που ανήκουν στις γρήγορες κλάσεις, δηλαδή $\bar{\mu}_i = \hat{\pi}^{(0)}(\mathbf{0}) \mu_i = \eta(0; \rho_f, L_f) \mu_i$. Τα κέρδη διαφορισμού $\bar{g}(|\mathbf{n}_s|)$ παραμένουν ως έχουν και αντιστοιχούν ακριβώς σε εκείνα ενός συστήματος που εξυπηρετεί αποκλειστικά $|\mathbf{n}_s|$ τερματικά. Επίσης, λόγω της υπόθεσης διαχωρισμένου θαλάμου εξυπηρέτησης μεταξύ γρήγορων και αργών κλάσεων και της (8.8), ισχύει πάντα ότι οι συναθροιστικοί ρυθμοί αφίξεων είναι ίσοι με τους μη εξαρτώμενους από την κατάσταση ρυθμούς, λ_i . Τα παραπάνω είναι αρκετά ώστε να μετατρέψουν τη γενική μορφή του συναθροιστικού πίνακα της (8.7) σε

$$\bar{M}(\mathbf{n}_s, \mathbf{n}'_s) = \begin{cases} \lambda_i \mathbf{I}_{\mathbf{n}_s \notin \bar{\mathcal{C}}_j}, & \mathbf{n}'_s = \mathbf{n}_s + \mathbf{e}_i, \quad i \in \mathcal{K}_s, \\ n_i \mu_i \eta(0; \rho_f, L_f) g(|\mathbf{n}_s|), & \mathbf{n}'_s = \mathbf{n}_s - \mathbf{e}_i, \quad i \in \mathcal{K}_s. \end{cases}$$

Όταν εφαρμόζεται ο SG-PF, οι κανονικοποιημένοι ρυθμοί εξυπηρέτησης του \bar{M} ελαττώνονται κατά ένα παράγοντα ίσο με την πιθανότητα να είναι το σύστημα

άδειο από τερματικά γρήγορων κλάσεων. Σε αυτή την περίπτωση οι αργές κλάσεις συνεχίζουν να λαμβάνουν μικρό ποσοστό εξυπηρέτησης, μικρότερο από εκείνο που θα είχαν αν εφαρμοζόταν ο BR κατά $\eta(0; \rho_f, L_f)$. Τα μέσα μεγέθη επίδοσης των ασθενών κλάσεων μπορούν να δοθούν με μορφή που περιέχει την ένταση κίνησης $\bar{\rho}_i$, όπου $\bar{\rho}_i \triangleq \lambda_i / \bar{\mu}_i = \rho_i / \eta(0; \rho_f, L_f)$. Έτσι, ορίζοντας τα $\rho_s, \bar{\rho}_s$ ως το άθροισμα των φορτίων των ασθενών κλάσεων και το «αυξημένο» φορτίο αντίστοιχα, καταλήγουμε στο ότι $\bar{\rho}_s = \rho_s / \eta(0; \rho_f, L_f)$, το οποίο σημαίνει ότι το φορτίο κίνησης των αργών κλάσεων φαίνεται να αυξάνεται κατά ένα παράγοντα ίσο με $\eta(0; \rho_f, L_f)$.

Ο αλγόριθμος NBR στον οποίο καταλήγει ο SG-PF όταν στο σύστημα υπάρχουν μόνο τερματικά των ασθενών κλάσεων, εγγυάται τις ίδιες ιδιότητες ως προς τη δικαιοσύνη όπως αυτές που περιγράφηκαν προηγουμένως για τις ισχυρές κλάσεις. Στην περίπτωση αυτή η πιθανότητα αποκλεισμού είναι $P_B = \eta(L_s; \bar{\rho}_s, L_s)$, αντικατοπτρίζοντας το γεγονός ότι το σύστημα συμπεριφέρεται ως να ήταν αποκλεισμένο από τις γρήγορες κλάσεις αλλά με μειωμένο ρυθμό εξυπηρέτησης. Κατ' αντιστοιχία, η μέση ρυθμαπόδοση για τις αργές κλάσεις είναι

$$\tau_s = \sum_{i \in \mathcal{K}_s} \tau_i = (1 - P_B) \sum_{i \in \mathcal{K}_s} \tilde{r}_i \bar{\rho}_i,$$

δηλώνοντας ότι η ρυθμαπόδοση κάθε κλάσης παραμένει ανάλογη προς το μέσο ρυθμό μετάδοσης. Ο μέσος αριθμός ενεργών τερματικών από όλες τις αργές κλάσεις είναι $E[N_s] = \sum_{m'=1}^{L_s} m' \eta(m'; \bar{\rho}_s, L_s)$. Ωστόσο ο λόγος του μέσου αριθμού ενεργών τερματικών μιας δεδομένης κλάσης, $E[N_i]$ προς το συνολικό μέσο αριθμό των αργών κλάσεων $E[N_s]$ δε μεταβάλλεται, αφού $E[N_i] = \frac{\bar{\rho}_i}{\bar{\rho}_s} E[N_s] = \frac{\rho_i}{\rho_s} E[N_s]$. Και σε αυτή την περίπτωση η (9.6) δηλώνει ότι ο μέσος χρόνος εξυπηρέτησης ενός τερματικού αργής κλάσης είναι

$$E[V_i] = \frac{\bar{v}}{\tilde{r}_i \bar{\rho}_s (1 - P_B)}, \quad i \in \mathcal{K}_s.$$

Τέλος, ο ασυμπτωτικός ρυθμός φθίσης της κατανομής του χρόνου απόκρισης για μια αργή κλάση ελέγχεται από τη μικρότερη ιδιοτιμή του συναθροιστικού πίνακα $\bar{M}^{(i)}$, δηλαδή

$$\lim_{t \rightarrow \infty} t^{-1} \log S_i(t) = -\sigma_1(\bar{M}^{(i)}), \quad i \in \mathcal{K}_s.$$



Κεφάλαιο 11

Αριθμητικά Αποτελέσματα

11.1 Αριθμητικά αποτελέσματα για την επαλήθευση του μοντέλου SDPS

ΠΑΡΑΚΑΤΩ παραθέτουμε μία συλλογή αριθμητικών αποτελεσμάτων που διευκρινίζει τις έννοιες των Κεφαλαίων 4-6 σχετικά με το γενικευμένο πρόβλημα που αντιστοιχεί σε ένα σύστημα SDPS και φανερώνει την αξία της προσέγγισης μέσω της διαδικασίας αποσύνθεσης και συνάθροισης που ακολουθήθηκε εκεί για την περίπτωση του διαχωρισμού χρονικής κλίμακας.

Εκινάμε με ένα απλό σύστημα που εξυπηρετεί δυο κλάσεις πελατών, αριθμημένες με 1 και 2 αντίστοιχα. Οι αφίξεις στο σύστημα συμβαίνουν σύμφωνα με μία χρονικά ομοιογενή διαδικασία Poisson, με ρυθμούς $\lambda_1 = 1$ και $\lambda_2 = 15$ πελάτες/sec και διαθέτουν εκθετικά κατανομημένες απαιτήσεις εξυπηρέτησης με παραμέτρους $\mu_1 = 1$ και $\mu_2 = 15 \text{ sec}^{-1}$. Το σύστημα αυτό χαρακτηρίζεται από διαχωρισμό χρονικής κλίμακας με τις κλάσεις 1 και 2 να αντιστοιχούν σε αργή και γρήγορη δυναμική. Ο παράγοντας σύζευξης είναι $\epsilon = \lambda_1/\lambda_2 = \mu_1/\mu_2 \approx 0.067$. Επιπλέον, θεωρούμε διαχωρισμένους θαλάμους εξυπηρέτησης (δηλ. χώρο καταστάσεων της μορφής (3.2)) με μεγέθη $L_1 = L_2 = 3$. Επιλέγουμε την πολιτική DPS για τους ρυθμούς εξυπηρέτησης, με χρήση βαρών $w_1 = 1$ και $w_2 = 2$, ώστε $\phi_i(\mathbf{n}) = w_i/(w_1 n_1 + w_2 n_2)$, $i = 1, 2$.

Όταν υπολογίζουμε την κατανομή του χρόνου απόκρισης για τη γρήγορη κλάση 2, χωρίς να λάβουμε υπόψη το διαχωρισμό χρονικής κλίμακας, ο αντίστοιχος πίνακας B_2 του Θεωρήματος 2 έχει διάσταση ίση με $(L_1 + 1)L_2 = 12$ (μία από τις διαθέσιμες θέσεις καταλαμβάνεται από τον μόνιμο πελάτη), ενώ η προσέγγιση του Θεωρήματος 3 θεωρεί καθένα από τα $L_1 + 1 = 4$ μειωμένα υποσυστήματα με δείκτη $n_s = n_1$. Κάθε τέτοιο υποσύστημα αντιστοιχεί σε χώρο κα-

ταστάσεων μεγέθους $\text{card}(\hat{\mathcal{S}}^{(n_1)}) = L_2 + 1 = 4$ και οι αντίστοιχοι πίνακες $\hat{B}_2^{(n_1)}$ έχουν διάσταση ίση με $L_2 = 3$. Προκειμένου να συγκριθεί ο ρυθμός φθίσης της ακριβούς κατανομής του χρόνου απόκρισης, $S_2(t)$, με εκείνη της προσέγγισης στο Θεώρημα 3, το πάνω αριστερά κομμάτι του Πίνακα 11.1 παρουσιάζει τις ιδιοτιμές με το μικρότερο πραγματικό μέρος $\sigma_1(\hat{B}_2^{(n_1)})$ των μειωμένων υποσυστημάτων, τις αντίστοιχες ιδιοτιμές $\sigma_j(B_2)$ του «πλήρους» πίνακα B_2 και τα σχετικά σφάλματα $(\sigma_j(B_2) - \sigma_1(\hat{B}_2^{(n_1)}))/\sigma_j(B_2)$. (Για όλα τα αποτελέσματα αυτής της ενότητας, οι υπόλοιπες ιδιοτιμές με μεγαλύτερο πραγματικό μέρος προσεγγίζονται με σχετικό σφάλμα μικρότερο από εκείνο των ιδιοτιμών που παρουσιάζονται.)

Γρήγορη δυναμική (κλάση 2)							
n_1	$\sigma_1(\hat{B}_2^{(n_1)})$	$\sigma_j(B_2)$	σχετ. σφάλμα	t (sec)	$S_2^{\text{appr}}(t)$	$S_2(t)$	σχετ. σφάλμα
3	3.7750	3.9240	0.0380	0.20	$4.192 \cdot 10^{-1}$	$4.192 \cdot 10^{-1}$	$< 10^{-4}$
2	4.4091	5.6047	0.2133	0.40	$1.833 \cdot 10^{-1}$	$1.836 \cdot 10^{-1}$	$1.6 \cdot 10^{-3}$
1	5.3338	6.6924	0.2030	0.60	$8.124 \cdot 10^{-2}$	$8.154 \cdot 10^{-2}$	$3.7 \cdot 10^{-3}$
0	6.8169	7.9462	0.1421	0.80	$3.637 \cdot 10^{-2}$	$3.651 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$

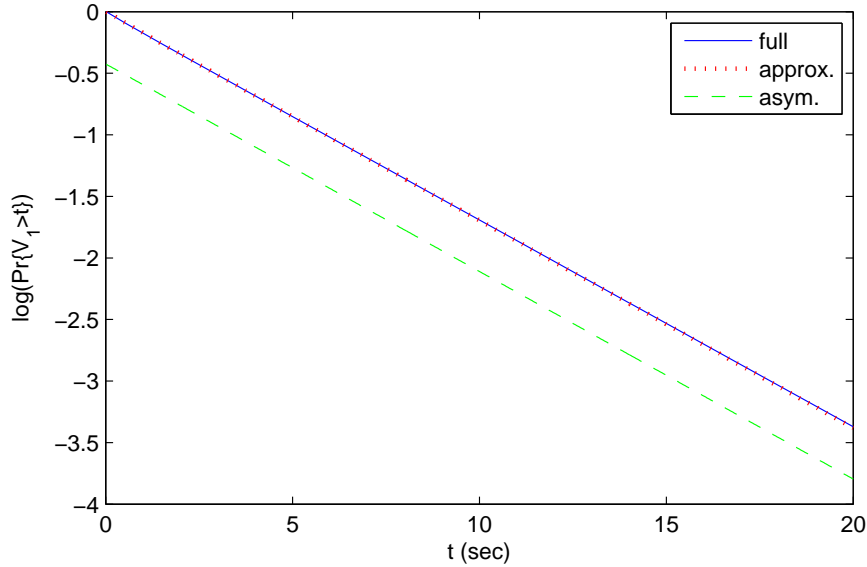
Αργή δυναμική (κλάση 1)							
	$\sigma_j(\bar{B}_1)$	$\sigma_j(B_1)$	σχετ. σφάλμα	t (sec)	$\bar{S}_1(t)$	$S_1(t)$	σχετ. σφάλμα
	0.1686	0.1680	-0.0035	6.0	$3.592 \cdot 10^{-1}$	$3.595 \cdot 10^{-1}$	$8.3 \cdot 10^{-4}$
	1.0831	1.0778	-0.0048	12.0	$1.306 \cdot 10^{-1}$	$1.312 \cdot 10^{-1}$	$4.6 \cdot 10^{-3}$
	1.9362	1.9209	-0.0079	18.0	$4.750 \cdot 10^{-2}$	$4.789 \cdot 10^{-2}$	$8.1 \cdot 10^{-3}$

Πίνακας 11.1: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, ρυθμοί εξυπηρέτησης DPS

Το σύστημα που μελετάμε εδώ χαρακτηρίζεται από ομοιογενείς αφίξεις Poisson, μονότονους ρυθμούς εξυπηρέτησης και ξεχωριστό θάλαμο εξυπηρέτησης για κάθε κλάση. Για το λόγο αυτό, η ανάλυση του Κεφαλαίου 5 καταλήγει στο ότι ο ρυθμός φθίσης της $S_2(t)$ προσεγγίζεται από την $\sigma_1(\hat{B}_2^{(3)})$, αντιστοιχώντας στο μειωμένο υποσύστημα με το μέγιστο δυνατό πληθυσμό πελατών με αργή δυναμική, $n_1 = 3$. Τα δεδομένα στον Πίνακα 11.1 επαληθεύουν το γεγονός αυτό.

Το πάνω δεξιά μέρος του Πίνακα 11.1 περιέχει τις τιμές της ακριβούς κατανομής $S_2(t)$, της προσέγγισής της σύμφωνα με το Θεώρημα 3 και του σχετικού σφάλματος μεταξύ των δύο, για κάποιες μικρές τιμές του χρόνου t . Επίσης, το Σχήμα 11.2 παρουσιάζει την τάση αυτών των συναρτήσεων για πολύ μεγαλύτερο εύρος χρονικών τιμών, σε ημιλογαριθμική κλίμακα. Τέσσερις καμπύλες περιέχονται στο συγκεκριμένο σχήμα: Η ακριβής CPDF $S_2(t)$, η προσέγγισή της σύμφωνα με το Θεώρημα 3 (η οποία δύσκολα διακρίνεται από την ακριβή καμπύλη, καθώς οι δύο αλληλοεπικαλύπτονται), η πιο χονδροειδής προσέγγιση που προκύπτει με τη χρήση μόνο της $\hat{S}_2^{(3)}(t)$, που αντιστοιχεί στο κυρίαρχο υποσύστημα με δείκτη $n_1 = 3$ και η ασυμπτωτική μορφή της πλήρους

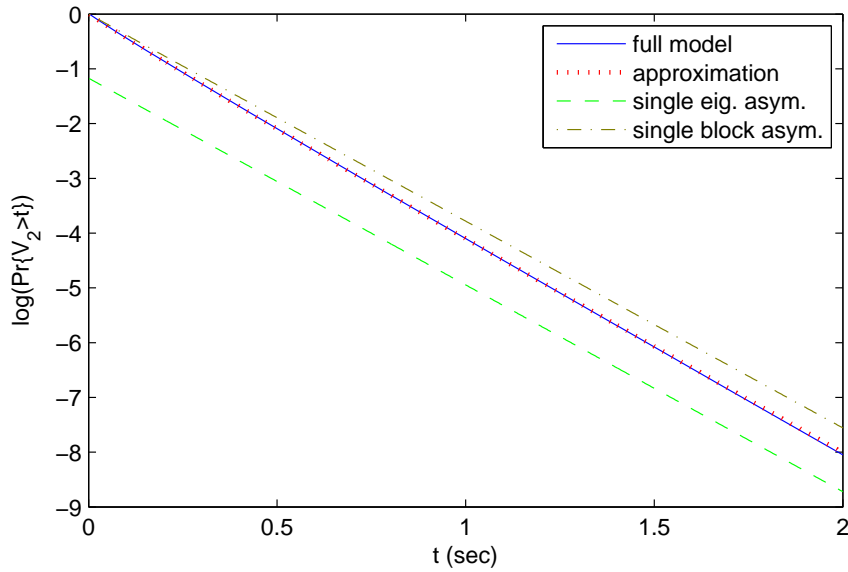
προσέγγισης του Θεωρήματος 3, $\bar{\gamma}_2(3)a_1(\hat{B}_2^{(3)})e^{-\sigma_1(\hat{B}_2^{(3)})t}$. Όπως παρατηρούμε, οι προσεγγίσεις ισχύουν για όλες τις χρονικές στιγμές και η ασυμπτωτική μορφή παρακολουθεί το ρυθμό φθίσης της ακριβούς καμπύλης.



Σχήμα 11.1: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, ρυθμοί εξυπηρέτησης DPS - Κλάση 1

Η αργή δυναμική της κλάσης 1 ελέγχεται από το συναθροιστικό σύστημα, το οποίο διαθέτει χώρο καταστάσεων μεγέθους $\text{card}(\bar{\mathcal{S}}) = L_1 + 1 = 4$. Ο αντίστοιχος πίνακας \bar{B}_1 έχει διάσταση $L_1 = 3$ και οι ιδιοτιμές του εμφανίζονται στο κάτω αριστερά μέρος του Πίνακα 11.1, σε συνδυασμό με τις μικρότερες ιδιοτιμές του B_1 (με μέγεθος 12). Τα σχετικά σφάλματα δίνονται επίσης. Είναι εύκολο να παρατηρήσει κανείς ότι η $\sigma_1(B_1)$, η οποία καθορίζει τον ασυμπτωτικό εκθετικό ρυθμό φθίσης της $S_1(t)$, προσεγγίζεται με τάξη $O(\epsilon^2)$, ακριβώς όπως προβλέπει η θεωρητική ανάλυση. Το κάτω δεξιό μέρος του πίνακα περιέχει τις τιμές της ακριβούς CPDF $S_1(t)$, της προσέγγισής της $\bar{S}(t)$ σύμφωνα με το Θεώρημα 4, και το σχετικό σφάλμα μεταξύ των δύο για κάποιες τιμές του χρόνου t . Οι ακριβείς μορφές παρουσιάζονται στο Σχήμα 11.1, που περιέχει καμπύλες (σε ημιλογαριθμική κλίμακα) για την $S_1(t)$, την προσέγγισή της $\bar{S}(t)$ και την ασυμπτωτική μορφή της τελευταίας $a_1(\bar{B})e^{-\sigma_1(\bar{B})t}$. Και πάλι, η προσέγγιση ταιριάζει σχεδόν πλήρως με την ακριβή καμπύλη και η ασυμπτωτική μορφή ακολουθεί ικανοποιητικά τον ασυμπτωτικό ρυθμό φθίσης.

Για το δεύτερο παράδειγμα, διατηρούμε τις δύο κλάσεις πελατών με τα ίδια χαρακτηριστικά αφίξεων και απαιτήσεων εξυπηρέτησης, αλλά αυτή τη φορά υποθέτουμε κοινό θάλαμο εξυπηρέτησης με μέγεθος $L = 3$ και ρυθμούς εξυπηρέτησης $\phi_i(\mathbf{n}) = 1/|\mathbf{n}|$, για $i = 1, 2$. Τα αποτελέσματα που σχετίζονται με



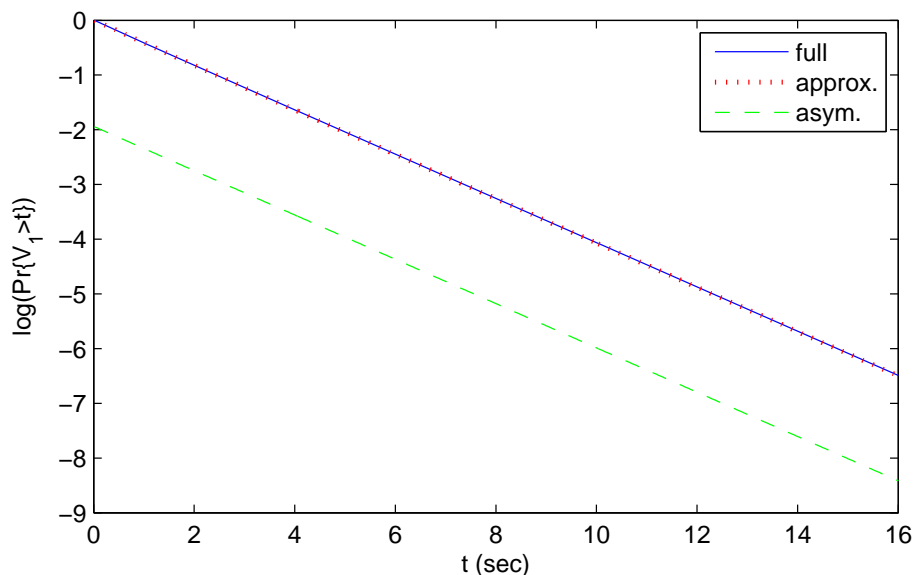
Σχήμα 11.2: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, ρυθμοί εξυπηρέτησης DPS - Κλάση 2

το σύστημα αυτό εμφανίζονται στον Πίνακα 11.2 και τα Σχήματα 11.3 και 11.4, με τον ίδιο τρόπο παρουσίασης όπως παραπάνω. Παρατηρούμε ότι και πάλι οι προσεγγίσεις είναι της ίδιας καλής ποιότητας. Αξίζει να σημειωθεί ότι τα μειωμένα υποσυστήματα έχουν άνισους χώρους καταστάσεων και η διάσταση των πινάκων $\hat{B}_2^{(n_1)}$ επίσης διαφέρει ανάλογα και είναι ίση με $L - n_1$, για πληθυσμό $n_1 = 0, \dots, L - 1$. Εφόσον ο χώρος εξυπηρέτησης είναι πλήρως διαμοιραζόμενος και η κλάση με γρήγορη δυναμική έχει ρυθμούς εξυπηρέτησης που εξαρτώνται από το διάνυσμα κατάστασης μόνο μέσω του αθροίσματος των στοιχείων του $|\mathbf{n}|$, τα αποτελέσματα της Ενότητας 5 προβλέπουν ότι το μειωμένο υποσύστημα που διαθέτει την ελάχιστη ιδιοτιμή για το $\sigma_1(\hat{B}_2^{(n_1)})$ είναι εκείνο που αντιστοιχεί στην τιμή $n_1 = 2$ και ότι $\sigma_1(\hat{B}_2^{(n_1)}) = \mu_2/3 = 5 \text{ sec}^{-1}$. Τα δεδομένα του πίνακα επιβεβαιώνουν ακριβώς τη συγκεκριμένη πρόβλεψη. Ας σημειωθεί επίσης ότι, ενώ το συναθροιστικό σύστημα που αντιστοιχεί στο προηγούμενο παράδειγμα χαρακτηριζόταν από ομοιογενείς αφίξεις Poisson (λόγω του γεγονότος ότι ο χώρος καταστάσεων στο παράδειγμα εκείνο ήταν της μορφής (3.2), η οποία αποτελεί ειδική περίπτωση της (3.5)), τώρα το συναθροιστικό σύστημα χαρακτηρίζεται από αφίξεις Poisson με ρυθμούς που εξαρτώνται από την κατάσταση, η οποία με χρήση της (5.4) είναι ίση με $\bar{\lambda}_1(n_1) = \lambda_1(1 - \hat{\pi}^{(n_1)}(L - n_1))$.

Στη συνέχεια παρουσιάζουμε ένα παράδειγμα που χαρακτηρίζεται από την επιπλέον τροποποίηση των ρυθμών εξυπηρέτησης υπέρ των γρήγορων κλάσεων. Το σύστημα εξυπηρετεί τρεις κλάσεις με ομοιογενείς αφίξεις Poisson $\lambda_1 = 0.5$, $\lambda_2 = 1$ και $\lambda_3 = 15$ πελάτες/sec και παραμέτρους των απαιτήσεων εξυ-

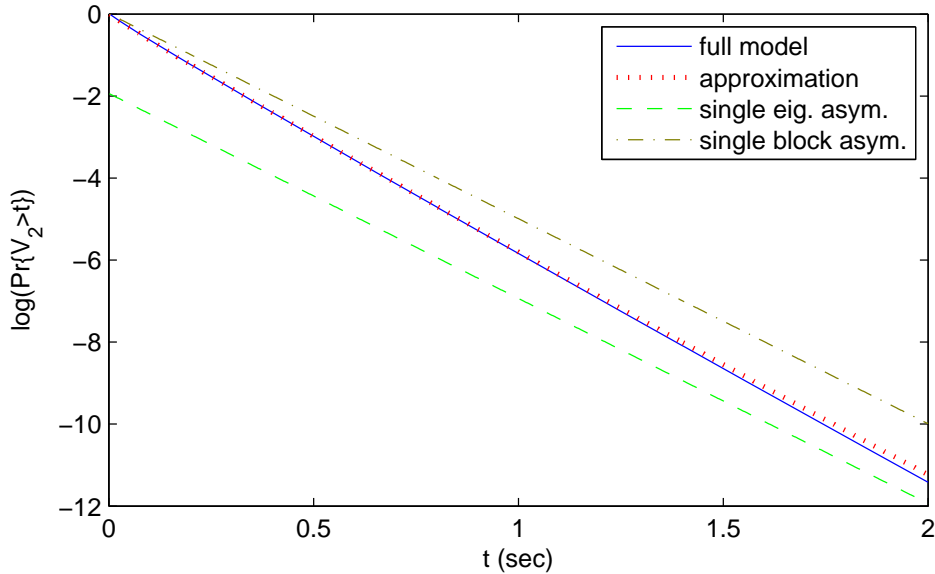
Γρήγορη δυναμική (κλάση 2)							
n_1	$\sigma_1(\hat{B}_2^{(n_1)})$	$\sigma_j(B_2)$	σχετ. σφάλμα	t (sec)	$S_2^{\text{appr}}(t)$	$S_2(t)$	σχετ. σφάλμα
2	5.0000	5.4372	0.080	0.15	0.3873	0.3880	$1.80 \cdot 10^{-3}$
1	5.5700	6.2179	0.104	0.30	0.1600	0.1602	$1.25 \cdot 10^{-3}$
0	6.8169	7.2679	0.062	0.45	0.0670	0.0669	$-1.49 \cdot 10^{-3}$
Αργή δυναμική (κλάση 1)							
	$\sigma_j(\bar{B}_1)$	$\sigma_j(B_1)$	σχετ. σφάλμα	t (sec)	$\bar{S}_1(t)$	$S_1(t)$	σχετ. σφάλμα
	0.4043	0.4049	$1.48 \cdot 10^{-3}$	2.50	0.3555	0.3558	$8.43 \cdot 10^{-4}$
	0.9740	1.0000	$2.60 \cdot 10^{-2}$	5.00	0.1286	0.1289	$2.33 \cdot 10^{-3}$
	1.5359	1.5951	$3.71 \cdot 10^{-2}$	7.50	0.0467	0.0469	$4.26 \cdot 10^{-3}$

Πίνακας 11.2: Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS



Σχήμα 11.3: Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 1

πηρέτησης $\mu_1 = 0.5$, $\mu_2 = 1$ και $\mu_3 = 15 \text{ sec}^{-1}$. Οι κλάσεις 1 και 2 έχουν αργή δυναμική, ενώ η κλάση 3 έχει γρήγορη. Ο παράγοντας σύζευξης ισούται με $\lambda_2/\lambda_3 = \mu_2/\mu_3 \approx 0.067$, δηλαδή έχει την ίδια τιμή με προηγουμένως. Θεωρούμε τρόπο εξυπηρέτησης που περιγράφεται μέσω της πολιτικής DPS, με βάρη ίσα με $w_1 = 0.5$, $w_2 = 1$ και $w_3 = 15$. Όταν τα βάρη αυτά κανονικοποιηθούν κατάλληλα, παίρνουν τη μορφή $w_i = \bar{w}_i \delta$, $i = 1, 2$ και $w_3 = 1$, με $\bar{w}_1 = 0.5$, $\bar{w}_2 = 1$ και $\delta = 1/15 = \epsilon$. Έτσι, οι ρυθμοί εξυπηρέτησης διαθέτουν την τροποποίηση υπέρ των γρήγορων κλάσεων, όπως ορίζεται από την (6.1), με $\phi_i^*(n_1, n_2) = \bar{w}_i / (\bar{w}_1 n_1 + \bar{w}_2 n_2)$, για $i = 1, 2$ και $\phi_3^*(n_3) = 1/n_3$. Ο χώρος καταστάσεων της γρήγορης κλάσης 3 έχει μέγεθος $L_f = 3$ και είναι διαχωρισμένος από το θάλαμο εξυπηρέτησης για τις κλάσεις με αργή δυναμική. Αυτός ο δευ-



Σχήμα 11.4: Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 2

τερος θάλαμος με μέγεθος $L_s = 3$, διαμοιράζεται πλήρως μεταξύ κλάσεων 1 και 2.

Εφόσον ο χώρος καταστάσεων του συστήματος είναι της μορφής (3.5), σύμφωνα με την ανάλυση της Ενότητας 6.1 όλα τα μειωμένα υποσυστήματα συμπίπτουν με ένα κοινό σύστημα, το οποίο είναι ισοδύναμο με το αρχικό όταν αυτό δέχεται και εξυπηρετεί αποκλειστικά πελάτες της κλάσης 3. Στην ειδική περίπτωση που εξετάζουμε εδώ, πρόκειται για ένα EPS σύστημα μίας κλάσης με πεπερασμένο θάλαμο εξυπηρέτησης μεγέθους $L_f = 3$. Οι συγκεκριμένες προβλέψεις επιβεβαιώνονται από τα δεδομένα του Πίνακα 11.3, τα οποία δείχνουν ότι όλα τα μειωμένα υποσυστήματα έχουν ίσες ιδιοτιμές $\sigma_1(\hat{B}_3^{(n_s)})$. Από τη στιγμή που τα μειωμένα συστήματα συμπίπτουν, η πλήρης προσέγγιση της $S_3(t)$ σύμφωνα με το Θεώρημα 3 και η πιο χονδροειδής προσέγγιση η οποία χρησιμοποιεί απλώς ένα μειωμένο υποσύστημα, επίσης συμπίπτουν στην περίπτωση αυτή, όπως φαίνεται στο Σχήμα 11.6.

Το συναθροιστικό σύστημα που αντιστοιχεί σε αυτό το παράδειγμα αποτελείται από ένα σύστημα που εξυπηρετεί δυο κλάσεις με αργή δυναμική, με πολιτική DPS χρησιμοποιώντας βάρη $\bar{w}_{1,2}$ και διαθέτει εξυπηρετητή μειωμένου ρυθμού κατά παράγοντα ίσο με την πιθανότητα με την οποία το EPS σύστημα για την κλάση 3 είναι κενό. Ο Πίνακας 11.3 και το Σχήμα 11.5 παρουσιάζουν τα δεδομένα που σχετίζονται με την κλάση 1. Τα αντίστοιχα δεδομένα της κλάσης 2 παραλείπονται εδώ καθώς εμφανίζουν ακριβώς τα ίδια χαρακτηριστικά.

Ας σημειωθεί ότι οι προσεγγίσεις που σχετίζονται με αυτό το σύστημα

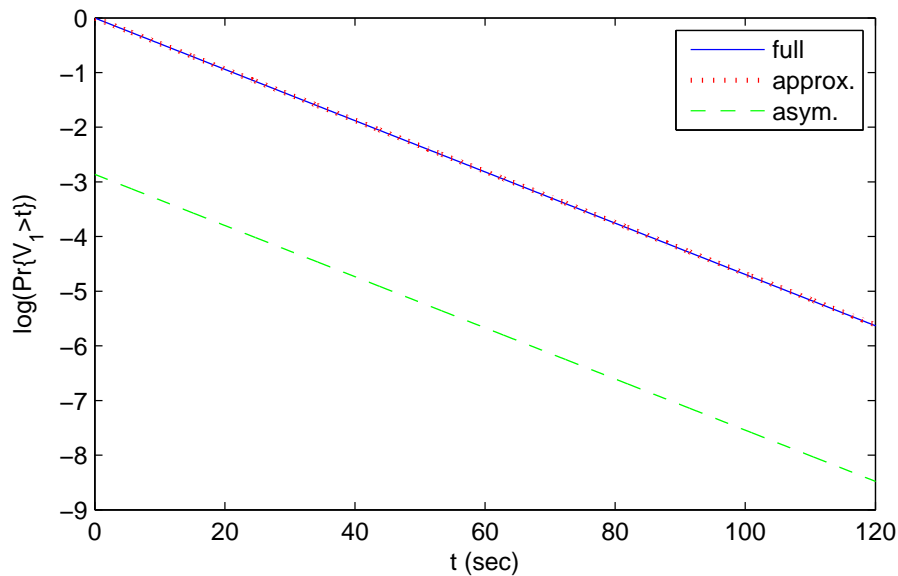
Γρήγορη δυναμική (κλάση 3)							
(n_1, n_2)	$\sigma_1(\hat{B}_3^{(n_1, n_2)})$	$\sigma_j(B_3)$	σχετ. σφάλμα	t (sec)	$S_3^{\text{approx.}}(t)$	$S_3(t)$	σχετ. σφάλμα
(3,0)	6.8169	6.1592	-0.107	0.05	0.6446	0.6651	0.031
(2,1)	6.8169	6.2572	-0.089	0.10	0.4404	0.4652	0.053
(1,2)	6.8169	6.3712	-0.007	0.15	0.3080	0.3325	0.074
(0,3)	6.8169	6.6205	-0.030	0.20	0.2176	0.2398	0.093
(2,0)	6.8169	7.7557	0.121	0.25	0.1543	0.1736	0.111
(1,1)	6.8169	7.9207	0.139	0.30	0.1096	0.1260	0.130
(0,2)	6.8169	8.1451	0.163	0.35	0.0779	0.0915	0.149
(1,0)	6.8169	8.1685	0.165	0.40	0.0554	0.0664	0.166
(0,1)	6.8169	8.3123	0.180	0.45	0.0394	0.0483	0.184
(0,0)	6.8169	8.4570	0.194	0.50	0.0280	0.0351	0.202
Αργή δυναμική (κλάση 1)							
	$\sigma_j(\bar{B}_1)$	$\sigma_j(B_1)$	σχετ. σφάλμα	t (sec)	$\bar{S}_1(t)$	$S_1(t)$	σχετ. σφάλμα
	0.0391	0.0394	0.008	15	0.5506	0.5479	-0.005
	0.1056	0.1042	-0.013	30	0.3052	0.3024	-0.009
	0.1697	0.1649	-0.029	45	0.1695	0.1672	-0.014
	1.3046	1.3125	0.006	60	0.0942	0.0925	-0.018
	1.7455	1.7353	-0.006	75	0.0524	0.0512	-0.023
	2.2101	2.1800	-0.014	90	0.0291	0.0283	-0.028

Πίνακας 11.3: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης

συνδυάζουν το διαχωρισμό της χρονικής κλίμακας και την τροποποίηση των ρυθμών εξυπηρέτησης υπέρ των γρήγορων κλάσεων με αποτέλεσμα να περιλαμβάνουν ένα σφάλμα τάξης $O(\max\{\delta, \epsilon\})$, εκτός από τις μικρές ιδιοτιμές που προσεγγίζονται από το συναθροιστικό σύστημα, για τις οποίες το σφάλμα είναι τάξης $O(\epsilon \max\{\delta, \epsilon\})$. Στο παράδειγμα αυτό ισχύει ότι $\delta = O(\epsilon)$ (καθώς οι δύο παράμετροι είναι ίσες), και έτσι οι παραπάνω τάξεις μειώνονται στα «συνηθισμένα» $O(\epsilon)$ και $O(\epsilon^2)$.

Στο επόμενο παράδειγμα θεωρούμε το σύστημα που συζητήθηκε παραπάνω, αλλά τώρα με κοινό θάλαμο εξυπηρέτησης μεγέθους $L = 3$ που επιτρέπει πλήρη διαμοίραση του χώρου μεταξύ όλων των κλάσεων. Αυτή τη φορά τα μειωμένα υποσυστήματα που αντιστοιχούν στα υποδιανύσματα των αργών κλάσεων $\mathbf{n}_s = (n_1, n_2)$ δε συμπίπτουν, επειδή παρόλο που χαρακτηρίζονται από κοινούς ρυθμούς αφίξεων και εξυπηρέτησης, έχουν διαφορετικούς χώρους καταστάσεων. Το γεγονός αυτό αντικατοπτρίζεται στις τιμές των $\sigma_1(\hat{B}_3^{(n_s)})$ στον Πίνακα 11.4, οι οποίες τώρα εξαρτώνται από το \mathbf{n}_s . Μία άλλη επιπλοκή του κοινού θαλάμου εξυπηρέτησης είναι ότι το συναθροιστικό σύστημα που αντιστοιχεί στην περίπτωση αυτή, χαρακτηρίζεται από ρυθμούς αφίξεων Poisson που εξαρτώνται από την κατάσταση, όπως καθορίζονται από την (5.4).

Από τη στιγμή που υπάρχει μόνο μία κλάση με γρήγορη δυναμική, ο

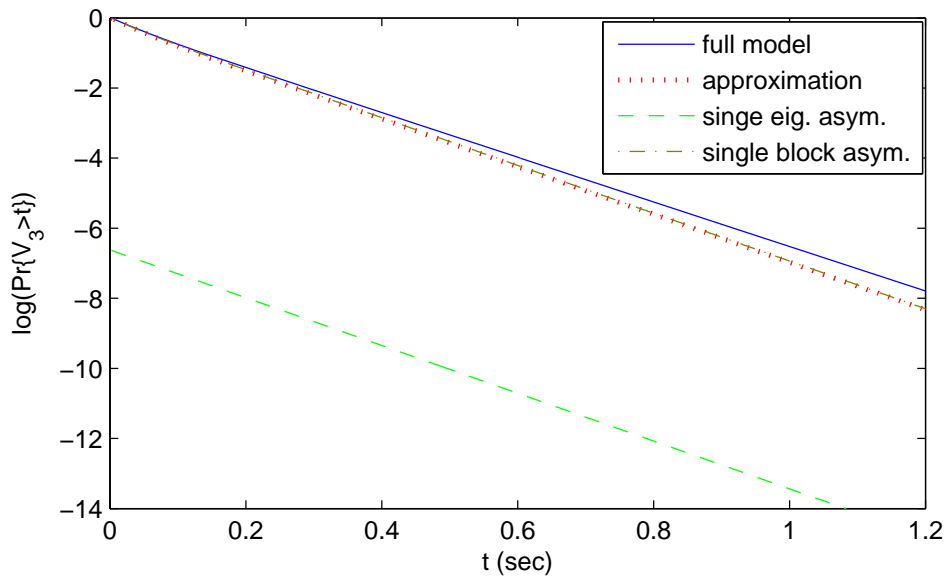


Σχήμα 11.5: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης - Κλάση 1

πλήρης διαμοιρασμός του θαλάμου εξυπηρέτησης εμπίπτει στην ειδική περίπτωση διαμοιρασμού που επιτρέπει πλήρη πρόσβαση στο θάλαμο για πελάτες αργής δυναμικής αλλά διατηρεί ξεχωριστά κομμάτια του θαλάμου για πελάτες γρήγορης δυναμικής, όπως προτείνεται στην Ενότητα 6.2. Σύμφωνα με την εκεί ανάλυση, αναμένουμε ότι το μειωμένο υποσύστημα με την ελάχιστη ιδιοτιμή $\sigma_1(\hat{B}_3^{(n_s)})$, το οποίο ελέγχει τη συμπεριφορά ουράς της $S_3(t)$, είναι εκείνο που αντιστοιχεί στο $n_s = (0, 0)$, δηλαδή όταν η κλάση 3 εξυπηρετείται σε αποκλειστικότητα. Τα δεδομένα στον Πίνακα 11.4 συμφωνούν με αυτή την πρόβλεψη. Μάλιστα, τα δεδομένα του πίνακα σε συνδυασμό με τα Σχήματα 11.7 και 11.8, παρουσιάζουν την καλή ποιότητα των προσεγγίσεων, όπως στις προηγούμενες περιπτώσεις.

Τα μέχρι τώρα αριθμητικά παραδείγματα έχουν παραμείνει περιορισμένα σε μικρό μέγεθος του θαλάμου εξυπηρέτησης προκειμένου να παραμείνει ο αριθμός των δυνατών καταστάσεων του συστήματος επίσης μικρός. Ο λόγος που επιδιώχθηκε αυτό ήταν το ότι σε διαφορετική περίπτωση το μέγεθος των αντίστοιχων πινάκων θα ήταν υπέρμετρα μεγάλο δυσκολεύοντας την παρουσίαση αλλά και τη σύγκριση των αποτελεσμάτων. Σε καμία περίπτωση ωστόσο, αυτό δεν πρέπει να εκληφθεί ως αδυναμία να παραχθούν αποτελέσματα για μεγαλύτερους χώρους καταστάσεων και να τα προσεγγιστούν σύμφωνα με τη μεθοδολογία που έχει αναπτυχθεί.

Προκειμένου να γίνει αυτό πιο φανερό, παρουσιάζουμε εδώ ένα ακόμη



Σχήμα 11.6: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης - Κλάση 3

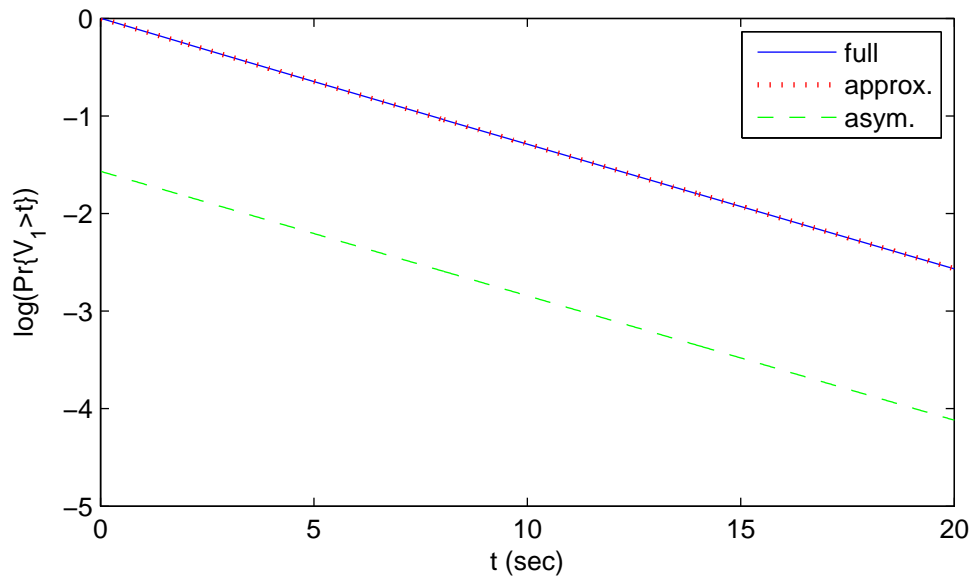
αριθμητικό παράδειγμα, το οποίο είναι αντίστοιχο με τη δεύτερη περίπτωση της τρέχουσας ενότητας. Το παράδειγμα που ακολουθεί, αντιστοιχεί σε ένα σύστημα δύο κλάσεων (δηλ., κλάση 1 (αργή) και κλάση 2 (γρήγορη)) και κοινό θάλαμο εξυπηρέτησης για όλες τις κλάσεις με μέγεθος $L = 10$. Οι δύο κλάσεις χαρακτηρίζονται από ρυθμούς αφίξεων και εξυπηρέτησης $\lambda_1 = 1$, $\lambda_2 = 30$ πελάτες/sec και $\mu_1 = 1.2$, $\mu_2 = 36 \text{ sec}^{-1}$ αντίστοιχα για κάθε κλάση, οδηγώντας σε παράγοντα σύζευξης $\epsilon = 1/30 \approx 0.033$. Οι ρυθμοί εξυπηρέτησης είναι και πάλι εκείνοι ενός συστήματος EPS, δηλαδή $\phi_i(\mathbf{n}) = 1/|\mathbf{n}|$, για $i = 1, 2$. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 11.5 με ακριβώς αντίστοιχη μορφή όπως στον Πίνακα 11.2. Είναι εύκολο να παρατηρήσει κανείς ότι τα αποτελέσματα βρίσκονται σε συμφωνία με τις θεωρητικές προβλέψεις. Ο παραπάνω ισχυρισμός ενισχύεται από τα Σχήματα 11.9 και 11.10 που ακολουθούν και μπορούν να παραλληλιστούν με τις καμπύλες των αντίστοιχων Σχημάτων 11.3 και 11.4.

Γρήγορη δυναμική (κλάση 3)							
n_1	$\sigma_1(\hat{B}_2^{(n_1)})$	$\sigma_j(B_2)$	σχετ. σφάλμα	t (sec)	$S_2^{\text{approx.}}(t)$	$S_2(t)$	σχετ. σφάλμα
(0,0)	6.8169	7.3126	0.068	0.05	0.5248	0.5462	0.039
(0,1)	9.5096	9.4133	-0.010	0.10	0.2856	0.3068	0.069
(1,0)	9.5096	9.5963	0.009	0.15	0.1600	0.1759	0.090
(0,2)	15.000	13.3545	-0.123	0.20	0.0920	0.1026	0.103
(1,1)	15.000	13.7129	-0.094	0.25	0.0543	0.0608	0.107
(2,0)	15.000	14.0939	-0.064	0.30	0.0327	0.0366	0.107
Αργή δυναμική (κλάση 1)							
	$\sigma_j(\bar{B}_1)$	$\sigma_j(B_1)$	σχετ. σφάλμα	t (sec)	$\bar{S}_1(t)$	$S_1(t)$	σχετ. σφάλμα
	0.1275	0.1276	$7.84 \cdot 10^{-4}$	4.5	0.5572	0.5568	$-7.18 \cdot 10^{-4}$
	0.3383	0.3370	$-3.86 \cdot 10^{-3}$	9.0	0.3127	0.3124	$-9.60 \cdot 10^{-4}$
	0.5955	0.5946	$-1.51 \cdot 10^{-3}$	13.5	0.1759	0.1756	$-1.71 \cdot 10^{-4}$
	1.0482	1.0204	$-2.72 \cdot 10^{-2}$	18.0	0.0990	0.0989	$-1.01 \cdot 10^{-4}$
	1.4147	1.3709	$-3.19 \cdot 10^{-2}$	22.5	0.0558	0.0557	$-1.80 \cdot 10^{-4}$
	1.9198	1.8277	$-5.04 \cdot 10^{-2}$	27.0	0.0314	0.0313	$-3.19 \cdot 10^{-4}$

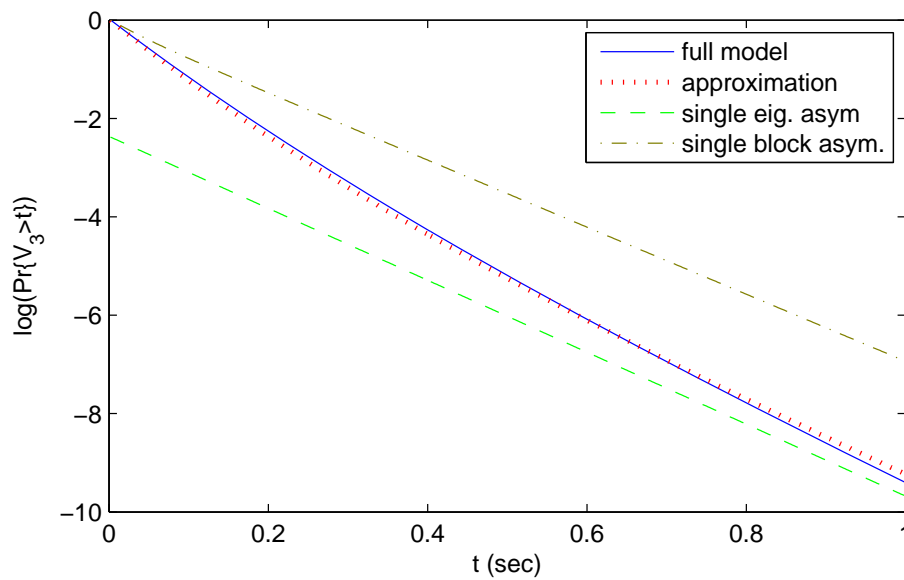
Πίνακας 11.4: Κοινός θάλαμος εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης

Γρήγορη δυναμική (κλάση 2)							
n_1	$\sigma_1(\hat{B}_2^{(n_1)})$	$\sigma_j(B_2)$	σχετ. σφάλμα	t (sec)	$S_2^{\text{appr}}(t)$	$S_2(t)$	σχετ. σφάλμα
9	3.6000	3.8910	0.075	0.15	0.5235	0.5236	$2 \cdot 10^{-4}$
8	3.6424	4.1665	0.126	0.30	0.2819	0.2820	$4 \cdot 10^{-4}$
7	3.6966	4.4435	0.168	0.45	0.1529	0.1529	$< 10^{-4}$
6	3.7676	4.7212	0.202	0.60	0.0832	0.0832	$< 10^{-4}$
5	3.8638	4.9986	0.227	0.75	0.0454	0.0453	$-2.2 \cdot 10^{-3}$
4	3.9988	5.2750	0.242	0.90	0.0248	0.0247	$-4.9 \cdot 10^{-3}$
3	4.1961	5.5506	0.244	1.05	0.0136	0.0135	$-7.4 \cdot 10^{-3}$
2	4.4959	5.8287	0.229	1.20	0.0074	0.0074	$< 10^{-4}$
1	4.9641	6.1355	0.191	1.35	0.0041	0.0041	$< 10^{-4}$
0	5.6851	6.6089	0.134	1.50	0.0022	0.0022	$< 10^{-4}$
Αργή δυναμική (κλάση 1)							
	$\sigma_j(\bar{B}_1)$	$\sigma_j(B_1)$	σχετ. σφάλμα	t (sec)	$\bar{S}_1(t)$	$S_1(t)$	σχετ. σφάλμα
	0.1387	0.1388	0.0007	2.50	0.7016	0.7010	$-9 \cdot 10^{-4}$
	0.3592	0.3635	0.0118	5.00	0.4942	0.4941	$-2 \cdot 10^{-4}$
	0.5859	0.5952	0.0156	7.50	0.3487	0.3487	$< 10^{-4}$
	0.8183	0.8339	0.0187	10.0	0.2462	0.2463	$4 \cdot 10^{-4}$
	1.0543	1.0774	0.0214	12.5	0.1739	0.1741	$1.1 \cdot 10^{-3}$
	1.2911	1.3226	0.0238	15.0	0.1229	0.1230	$8 \cdot 10^{-4}$
	1.5257	1.5661	0.0258	17.5	0.0868	0.0870	$2.3 \cdot 10^{-4}$
	1.7555	1.8048	0.0273	20.0	0.0614	0.0615	$1.6 \cdot 10^{-3}$
	1.9792	2.0365	0.0281	22.5	0.0434	0.0434	$< 10^{-4}$
	2.1970	2.2612	0.0284	25.0	0.0306	0.0307	$3.2 \cdot 10^{-3}$

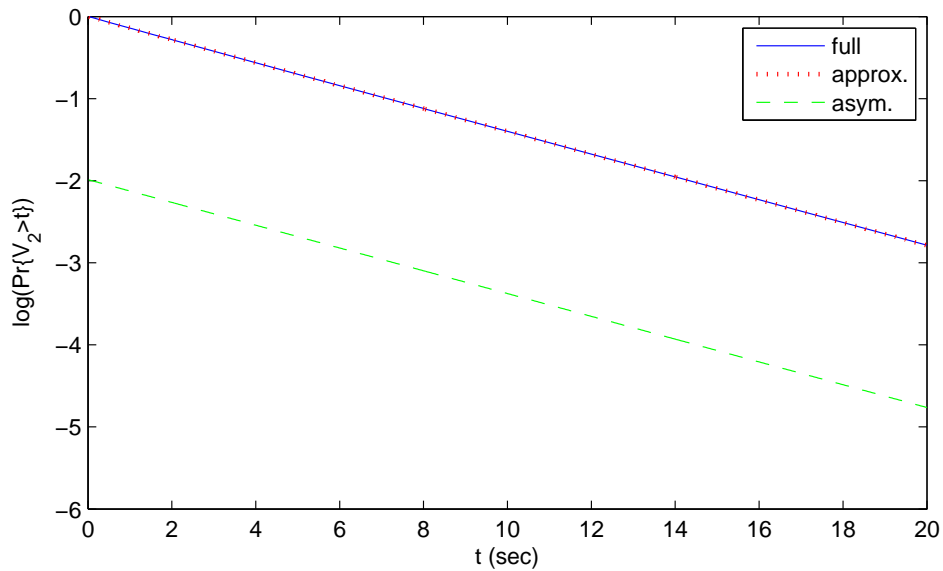
Πίνακας 11.5: Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS, $L = 10$



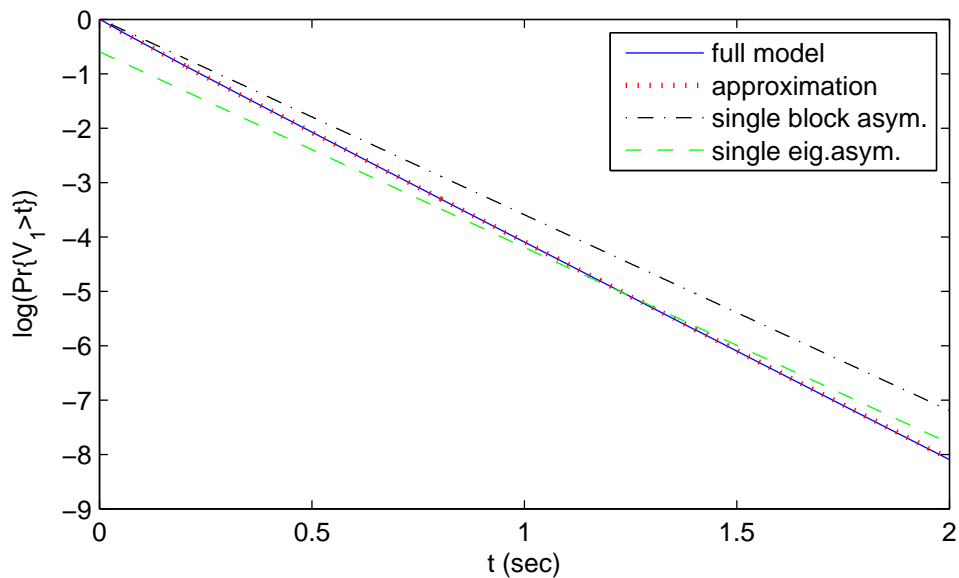
Σχήμα 11.7: Κοινός θάλαμος εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης - Κλάση 1



Σχήμα 11.8: Κοινός θάλαμος εξυπηρέτησης, τροποποιημένοι ρυθμοί εξυπηρέτησης υπέρ της γρήγορης κλάσης - Κλάση 3



Σχήμα 11.9: Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 1 - $L = 10$



Σχήμα 11.10: Κοινός θάλαμος εξυπηρέτησης, ρυθμοί εξυπηρέτησης EPS - Κλάση 2 - $L = 10$

11.2 Αριθμητικά αποτελέσματα για τη σύγκριση του SG-PF με τους BR και NBR

Συνεχίζουμε εγκαταλείποντας το γενικευμένο πρόβλημα των Κεφαλαίων 4-6 και επιστρέφουμε στο δίκτυο του Κεφαλαίου 2 για την παρουσίαση και ανάλυση αριθμητικών αποτελεσμάτων που έχουν στόχο τη σύγκριση του SG-PF με τους BR και NBR σε περιβάλλον διαχωρισμού χρονικής κλίμακας. Σε όλες τις ακόλουθες περιπτώσεις θεωρούμε ένα σύστημα που εξυπηρετεί τέσσερις κλάσεις τερματικών, με δείκτη $i = 1, \dots, 4$. Εκτός από μία περίπτωση στην οποία θα αναφερθεί η αλλαγή, θεωρούμε ότι τα τερματικά υπόκεινται σε σκέδαση Rayleigh. Οι κλάσεις 1, 2 διαθέτουν μέσους ρυθμούς μετάδοσης $\tilde{r}_1 = 0.08 \text{ Mb/s}$ και $\tilde{r}_2 = 0.12 \text{ Mb/s}$ (συνθέτοντας την αργή ομάδα κλάσεων), ενώ οι άλλες δύο κλάσεις (3, 4) χαρακτηρίζονται από μέσους ρυθμούς μετάδοσης $\tilde{r}_3 = 1.2 \text{ Mb/s}$ και $\tilde{r}_4 = 1.8 \text{ Mb/s}$, (γρήγορη ομάδα αντίστοιχα). Οι απαιτήσεις εξυπηρέτησης θεωρούνται εκθετικά κατανομημένες με μέση τιμή $\bar{v} = 0.4 \text{ Mb}$. Η επιλογή των ρυθμών μετάδοσης αντιστοιχεί σε ένα τυπικό μοντέρνο δίκτυο 3G και παρόλο που ενδιάμεσες τιμές είναι επιτεύξιμες, οι παραπάνω επιλέχθηκαν ώστε να τονίσουν τις συνέπειες ενός περιβάλλοντος που χαρακτηρίζεται από έντονες διαφορές στο μέσο SNR κάθε κλάσης. Επίσης, έστω ότι οι αφίξεις κάθε κλάσης λαμβάνουν χώρα σύμφωνα με μία ομοιογενή διαδικασία Poisson με ρυθμούς $\lambda_1 = 0.02\lambda$, $\lambda_2 = 0.08\lambda$, $\lambda_3 = 0.2\lambda$ και $\lambda_4 = 0.7\lambda$, με την επιλογή $\lambda = 2$ τερματικά/sec. Έτσι, η «απαισιόδοξη» (δηλ., η μέγιστη δυνατή) τιμή του παράγοντα σύζευξης που χρησιμοποιείται για την ποσοτικοποίηση του διαχωρισμού κλίμακας είναι $\epsilon = 0.12/1.2 = 0.1$.

Το πρώτο παράδειγμα αναφέρεται σε διαχωρισμένους θαλάμους εξυπηρέτησης μεταξύ γρήγορων και αργών κλάσεων, όπως υποδηλώνεται από την εξίσωση (3.5) με μεγέθη $L_s = 7$ και $L_f = 4$. Ο διαμοιρασμός του κάθε κομματιού του θαλάμου εξυπηρέτησης για τερματικά διαφορετικής κλάσης αλλά της ίδιας ομάδας επιτρέπεται, δηλαδή οι χώροι \mathcal{S}_s και \mathcal{S}_f έχουν την αθροιστική μορφή που δίνεται στην (3.3). Ο Πίνακας 11.6 παρουσιάζει τα αποτελέσματα για την περίπτωση που συζητούμε. Για καθένα από τους τρεις αλγορίθμους BR, NBR και SG-PF παρατίθεται η μέση ρυθμαπόδοση, ο μέσος χρόνος απόκρισης και η ελάχιστη ιδιοτιμή του B_i , $i = 1, \dots, 4$, η οποία ελέγχει την ασυμπτωτική συμπεριφορά της κατανομής του χρόνου απόκρισης. Τα μεγέθη των θαλάμων εξυπηρέτησης έχουν επιλεγεί ώστε να αποφεύγεται η περίπτωση κατά την οποία η πιθανότητα αποκλεισμού ξεπερνά το 1%. Για ευκολότερη σύγκριση, ο μέσος χρόνος απόκρισης και η μικρότερη ιδιοτιμή του B_i έχουν κανονικοποιηθεί ως προς τη μέση απαίτηση εξυπηρέτησης της κλάσης, (δηλ., $1/\mu_i$ και το αντίστροφο αυτού, αντίστοιχα).

BR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0158	0.0635	0.1593	0.5577
$E[V_i] \mu_i$	2.8957	2.3852	1.2173	1.1428
$\sigma_1(B_i)/\mu_i$	0.3167	0.3784	0.7865	0.8308
$\Pr_{B,i}$	0.0099		0.0042	
NBR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0160	0.0639	0.1585	0.5548
$E[V_i] \mu_i$	1.5570	1.5570	1.5509	1.5509
$\sigma_1(B_i)/\mu_i$	0.5993	0.5773	0.3665	0.3655
$\Pr_{B,i}$	0.0011		0.0092	
SG-PF	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0158	0.0635	0.1593	0.5576
$E[V_i] \mu_i$	2.4941	2.4941	1.1625	1.1625
$\sigma_1(B_i)/\mu_i$	0.3698	0.3569	0.8278	0.8093
$\Pr_{B,i}$	0.0095		0.0042	

Πίνακας 11.6: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7$, $L_f = 4$

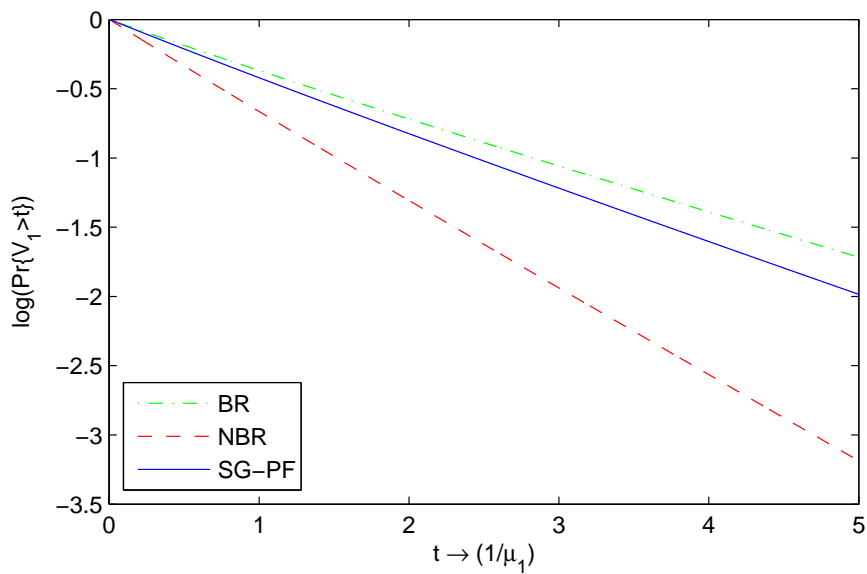
Τα αποτελέσματα του Πίνακα 11.6 δείχνουν ότι η καλύτερη επίδοση όταν χρησιμοποιείται ο BR, όσον αφορά τον κανονικοποιημένο μέσο χρόνο απόκρισης, εμφανίζεται στη γρηγορότερη από όλες τις κλάσεις (κλάση 4), όπως άλλωστε αναμένουμε. Επίσης βλέπουμε ότι η κλάση 3 εμφανίζει μεγαλύτερη μέση καθυστέρηση (κατά 6.5% περίπου). Σε αντίθεση, όπως έχουμε δει και αναλυτικά, ο BR παραμελεί τις αργές κλάσεις και εξυπηρετεί τα τερματικά που ανήκουν σε αυτές μόνο απουσία τερματικών των γρήγορων κλάσεων. Ως αποτέλεσμα, οι αργές κλάσεις εμφανίζουν πολύ μεγαλύτερο μέσο χρόνο απόκρισης, όπως βλέπουμε και στα αριθμητικά αποτελέσματα. Παρατηρούμε ακόμη, ότι ο NBR πράγματι φέρεται με δικαιοσύνη μεταξύ των κλάσεων που ανήκουν σε διαφορετικές ομάδες. Ο μέσος χρόνος απόκρισης για την περίπτωση εδώ, είναι ακριβώς ίσος για κλάσεις της ίδιας ομάδας και πολύ κοντά μεταξύ της αργής και της γρήγορης ομάδας. Αυτό αποτελεί ένα αναμενόμενο αποτέλεσμα, εφόσον ο NBR εκ κατασκευής διανέμει τους πόρους του συστήματος με δίκαιο τρόπο. Ωστόσο, τα κύρια μειονεκτήματα του NBR είναι η αύξηση του μέσου χρόνου απόκρισης των γρήγορων κλάσεων και η επιδείνωση της ασυμπτωτικής μορφής της κατανομής του χρόνου απόκρισης για τις κλάσεις αυτές, δηλαδή μείωση των $\sigma_1(B_i)$, $i = 3, 4$. Αναφέρουμε επίσης ότι για το ίδιο μέγεθος θαλάμου εξυπηρέτησης L_f , η πιθανότητα αποκλεισμού των γρήγορων κλάσεων είναι περισσότερο από το διπλάσιο εκείνης που αντιστοιχεί στον BR, αναδεικνύοντας έτσι ένα ακόμη ανεπιθύμητο φαινόμενο.

Το κάτω μέρος του Πίνακα 11.6 αναφέρεται στον SG-PF. Ο συγκεκριμένος αλγόριθμος αντισταθμίζει τις διαφορές μεταξύ κλάσεων της ίδιας ομάδας, χωρίς να επιφέρει σημαντική μείωση της επίδοσης σε σύγκριση με τον BR. Για παράδειγμα, καταλήγει σε κανονικοποιημένο μέσο χρόνο απόκρισης για τη γρηγορότερη κλάση που είναι μόνο 1.72% μεγαλύτερος από τον αντίστοιχο του BR, αλλά το αποτέλεσμα αυτό ισχύει και για τις δύο γρήγορες κλάσεις. Παράλληλα, η επίδοση της πιο αργής κλάσης (κλάση 1), η οποία είναι η πιο αδικημένη από τον BR, τώρα έχει βελτιωθεί και η μέση καθυστέρησή της είναι ίση με εκείνη της κλάσης 2. Έτσι, παρά το ότι και ο SG-PF παραμελεί τα τερματικά των αργών κλάσεων, δε μειώνει περισσότερο από τον BR την επίδοσή τους. Σημειώνεται επίσης και μία μικρή ελάττωση της πιθανότητας αποκλεισμού των αργών κλάσεων.

Το σημαντικό θέμα της ασυμπτωτικής συμπεριφοράς της κατανομής του χρόνου απόκρισης αναλύεται ξεχωριστά για κάθε κλάση, μέσω σύγκρισης μεταξύ των τριών αλγορίθμων χρονοπρογραμματισμού. Ας θεωρήσουμε αρχικά την κλάση 4, η οποία είναι η ταχύτερη όλων. Όπως αναμένουμε από τη θεωρητική ανάλυση που έχει προηγηθεί, κατά την εφαρμογή του BR αυτή η κλάση πρέπει να εμφανίζει την καλύτερη συμπεριφορά σε σχέση με τους SG-PF και NBR όσον αφορά την ουρά της κατανομής του χρόνου απόκρισης. Το Σχήμα 11.14 παρουσιάζει την κατανομή του χρόνου απόκρισης για την κλάση 4, με τον οριζόντιο άξονα σε κανονικοποιημένες μονάδες χρόνου ($1/\mu_4$) και επαληθεύει αυτή ακριβώς την πρόβλεψη. Επίσης παρατηρούμε ότι ο SG-PF εμφανίζει σχεδόν το ίδιο αποτέλεσμα όπως ο BR, μία παρατήρηση που έρχεται να επιβεβαιώσει το γεγονός πως ο SG-PF δεν έχει σημαντικό αντίκτυπο στην κατανομή του χρόνου απόκρισης των γρήγορων κλάσεων, ενώ η ουρά της κατανομής που προκύπτει για τον NBR είναι εμφανώς χειρότερη. Το Σχήμα 11.13 δίνει την ανάλογη σύγκριση που αφορά την κλάση 3, η οποία επίσης ανήκει στη γρήγορη ομάδα κλάσεων αλλά είναι λίγο πιο αργή από την κλάση 4. Αξίζει να σημειωθεί ότι στην περίπτωση αυτή, ο SG-PF εμφανίζει βελτίωση της ουράς σε σχέση με τον BR, εφόσον εκ κατασκευής τείνει να εξομαλύνει τις διαφορές μεταξύ κλάσεων της ίδιας ομάδας. Κατά συνέπεια, με χαμηλό κόστος σε όρους επιδείνωσης της ουράς, η εφαρμογή του SG-PF μπορεί να προσδώσει την έννοια της δικαιοσύνης μεταξύ των γρήγορων κλάσεων.

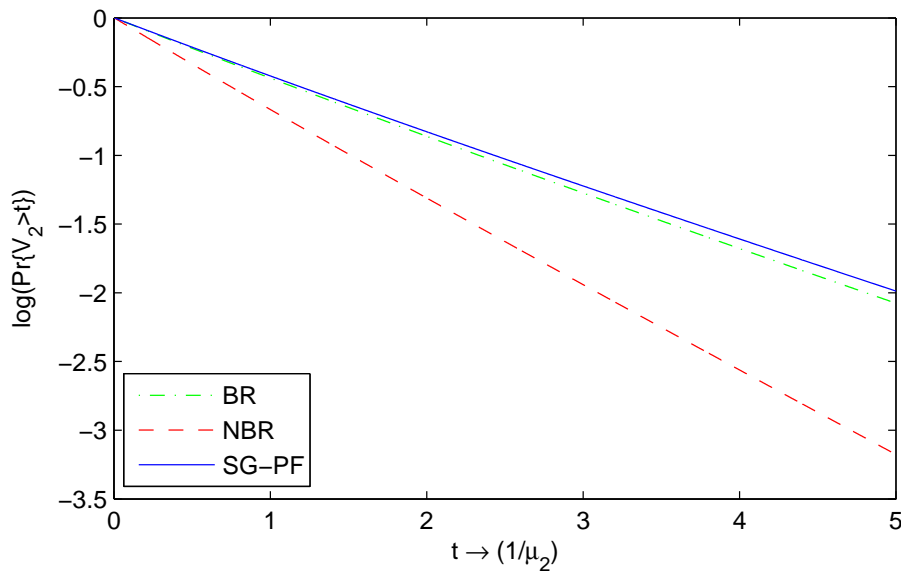
Τα Σχήματα 11.11-11.12 απεικονίζουν τις αντίστοιχες κατανομές που αφορούν τις αργές κλάσεις για την περίπτωση που συζητούμε εδώ. Προφανώς ο πιο αποτελεσματικός αλγόριθμος τώρα είναι ο NBR, κάτι που άλλωστε αναμένουμε και θεωρητικά. Ο SG-PF είναι και πάλι πολύ κοντά στα αποτελέσματα του BR, ένα ενδιαφέρον αποτέλεσμα το οποίο δηλώνει ότι η εφαρμογή του SG-PF αντί του BR δεν ελαττώνει περαιτέρω την επίδοση των τερματικών των αργών κλά-

σεων. Επιπλέον, με αναφορά στον SG-PF, τα Σχήματα 11.11-11.12 αποκαλύπτουν ότι η διαφορά της ουράς της κατανομής του χρόνου απόκρισης μεταξύ των αργών κλάσεων εξομαλύνεται σε σχέση με εκείνη που προκύπτει από τον BR, με βελτίωση της κλάσης 1 (η οποία είναι η βραδύτερη όλων) και με μία πολύ μικρή επιδείνωση της κλάσης 2.



Σχήμα 11.11: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7$, $L_f = 4$ - Κλάση 1

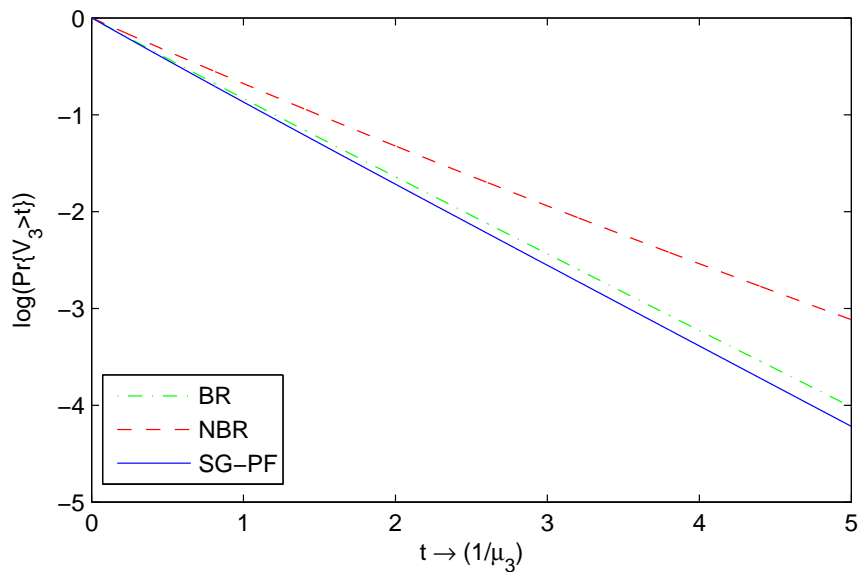
Το δεύτερο αριθμητικό αποτέλεσμα εστιάζει κυρίως στην επίδραση της αλλαγής του μεγέθους του θαλάμου εξυπηρέτησης των αργών κλάσεων. Ο Πίνακας 11.7 παρουσιάζει την περίπτωση ενός συστήματος που χαρακτηρίζεται από διαδικασία αφίξεων και μέγεθος θαλάμου εξυπηρέτησης των γρήγορων κλάσεων ακριβώς ίδια όπως στο προηγούμενο παράδειγμα, αλλά με μειωμένο μέγεθος θαλάμου εξυπηρέτησης των αργών κλάσεων, $L_s = 3$. Όπως είναι αναμενόμενο η πιθανότητα αποκλεισμού των αργών κλάσεων αυξάνεται σημαντικά. Η μείωση του χώρου καταστάσεων των αργών κλάσεων δεν έχει επίδραση στην επίδοση των γρήγορων κλάσεων για τους BR και SG-PF, αλλά βελτιώνει λίγο την επίδοση κατά την εφαρμογή του NBR. Ωστόσο τα συμπεράσματα του προηγούμενου παραδείγματος που αφορούν τις γρήγορες κλάσεις συνεχίζουν να βρίσκονται σε ισχύ. Σχετικά με τις αργές κλάσεις, παρατηρούμε ότι η συνολική τάση είναι παρόμοια με το προηγούμενο παράδειγμα αλλά ο μέσος χρόνος απόκρισης μειώνεται για όλους τους αλγορίθμους. Αυτό είναι κάτι το αναμενόμενο εφόσον λιγότερα τερματικά από τις αργές κλάσεις παράλληλα στο σύστημα έχουν ως αποτέλεσμα μικρότερο διαμοιρασμό των πόρων του δικτύου. Αντίστοιχα σχήματα με εκείνα του προηγούμενου παραδείγματος, που



Σχήμα 11.12: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7$, $L_f = 4$ - Κλάση 2

παρουσιάζουν την κατανομή του χρόνου απόκρισης για κάθε κλάση και κάθε αλγόριθμο έχουν επίσης παραχθεί. Δεν παρατίθενται εδώ για την οικονομία του χώρου και επειδή ταυτόχρονα η μορφή τους είναι ακριβώς ίδια όπως στα Σχήματα 11.11-11.14, καταλήγοντας σε ακριβώς αντίστοιχα συμπεράσματα.

Στη συνέχεια αφήνουμε την περίπτωση διαχωρισμένων θαλάμων εξυπηρέτησης και θεωρούμε ένα σύστημα για το οποίο ο θάλαμος εξυπηρέτησης είναι κοινός για όλες τις κλάσεις, με μέγεθος, $L = 7$. Η πρόθεσή μας είναι να συγκριθεί η επίδοση των τριών αλγορίθμων με την προηγούμενη περίπτωση όπου υπήρχαν ξεχωριστοί θάλαμοι εξυπηρέτησης για κάθε ομάδα κλάσεων. Τα αριθμητικά αποτελέσματα παρουσιάζονται στον Πίνακα 11.8. Σε ό,τι αφορά τις μέσες τιμές των χρόνων απόκρισης των γρήγορων κλάσεων, μία σύγκριση με τα αποτελέσματα για τον διαχωρισμένο θάλαμο που δίνεται στον Πίνακα 11.7 που χαρακτηρίζεται από το ίδιο μέγεθος (δηλ., $L_f + L_s = 7$), αποκαλύπτει ότι ο διαχωρισμός δεν έχει επίπτωση όταν εφαρμόζονται οι BR ή SG-PF, ενώ υπάρχει μία μικρή επιδείνωση για τον NBR. Επίσης, ο κοινός θάλαμος εξυπηρέτησης έχει ως αποτέλεσμα την εξίσωση των κανονικοποιημένων μέσων χρόνων απόκρισης για όλες τις κλάσεις, όπως άλλωστε αναμένουμε και θεωρητικά. Παρόλα αυτά, η μικρότερη ιδιοτιμή του B_i για όλες τις κλάσεις, $\sigma_1(B_i)$, (επίσης κανονικοποιημένη ως προς μ_i , $i = 1, \dots, 4$), είναι σε κάθε περίπτωση χειρότερη, ένα γεγονός που αποτρέπει τη χρήση αυτού του τύπου θαλάμου εξυπηρέτησης. Επίσης, κάτι που συνεισφέρει στον ισχυρισμό αυτό, είναι η αύξηση της πιθανότητας αποκλεισμού για τις γρήγορες κλάσεις. Ακολουθώντας την τακτική του

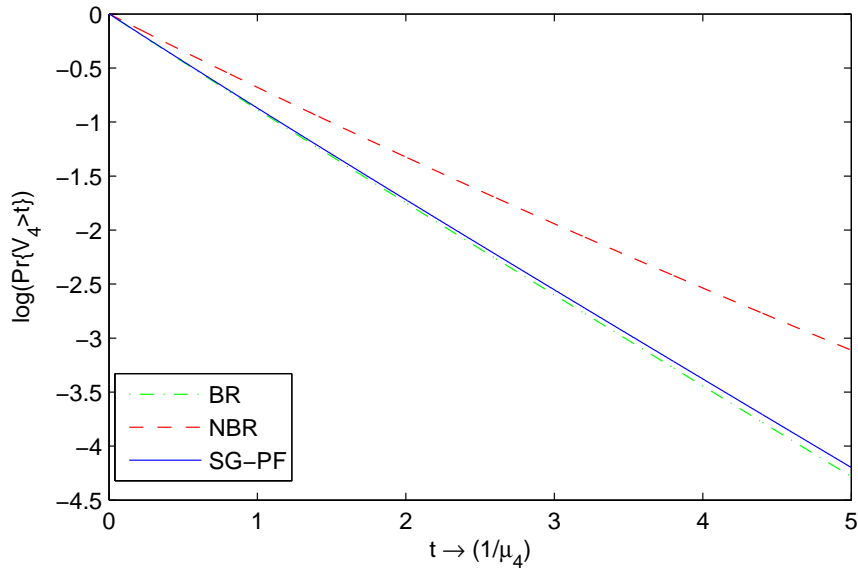


Σχήμα 11.13: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7$, $L_f = 4$ - Κλάση 3

προηγούμενου παραδείγματος, παραλείπουμε τα αντίστοιχα σχήματα για την κατανομή του χρόνου απόκρισης και αναφερόμαστε στα Σχήματα 11.11-11.14, που παρουσιάζουν ακριβώς την ίδια μορφή.

Το επόμενο αριθμητικό αποτέλεσμα αποτελείται από την προηγούμενη περίπτωση για κοινό θάλαμο εξυπηρέτησης και ακριβώς ίδιες συνθήκες κίνησης, αλλά με την αλλαγή ότι για τη συγκεκριμένη περίπτωση και μόνο, θεωρούμε σκέδαση με κατανομή Nakagami με παράμετρο $m = 2$ αντί για σκέδαση Rayleigh (η οποία είναι ειδική περίπτωση της σκέδασης Nakagami για $m = 1$). Μία μελέτη για το συμβιβασμό μεταξύ δικαιοσύνης και ρυθμιστικής για σκέδαση Nakagami στο [5], έχει δείξει ότι τα παραπάνω παραμένουν σχεδόν ανεπηρέαστα από την παράμετρο Nakagami και τις παραμέτρους συσχέτισης πολλαπλής διαδρομής (αγ.: multipath correlation). Το γεγονός αυτό αναδεικνύεται και από τα εδώ αποτελέσματα στον Πίνακα 11.9. Παρόλο που η συνολική επίδοση επιδεινώνεται λόγω της ελάττωσης των κερδών διαχωρισμού πολλαπλών χρηστών, (βλ., π.χ., τον Πίνακα 11.10 για τις κλάσεις 3 και 4 κατά την εφαρμογή του BR), τα κύρια χαρακτηριστικά και των τριών αλγορίθμων όπως έχουν ήδη αναλυθεί, εξακολουθούν να υφίστανται.

Τα παραδείγματα που ακολουθούν, διευκρινίζουν περισσότερο την επίδραση του μεγέθους του θαλάμου εξυπηρέτησης στα μέτρα επίδοσης που μας ενδιαφέρουν. Σε ό,τι ακολουθεί, συνεχίζουμε για την περίπτωση των τεσσάρων κλάσεων με τις ίδιες συνθήκες κίνησης, ακριβώς όπως εισήχθησαν στην αρχή της συγκεκριμένης ενότητας.



Σχήμα 11.14: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 7$, $L_f = 4$ - Κλάση 4

Ακολουθεί η περίπτωση για διαχωρισμένους θαλάμους εξυπηρέτησης μεταξύ γρήγορων και αργών κλάσεων. Παρατηρώντας τον Πίνακα 11.11 βλέπουμε ότι κατά την εφαρμογή του BR, ο μέσος χρόνος απόκρισης παραμένει σχεδόν αναλλοίωτος ως προς το μέγεθος του θαλάμου εξυπηρέτησης, L_f . Ταυτόχρονα, η μικρότερη ιδιοτιμή του B_i , άμεσα συνδεδεμένη με την ασυμπτωτική συμπεριφορά της κατανομής του χρόνου απόκρισης της κλάσης i , διακρίνεται από παρόμοια τάση για τις γρήγορες κλάσεις, $i = 3, 4$. Στα πλαίσια της μελέτης μας εδώ, μπορούμε να θεωρήσουμε ότι η επιλογή $L_f \geq 9$ δεν επηρεάζει περαιτέρω το χρόνο απόκρισης των γρήγορων κλάσεων. Ωστόσο, ο χρόνος απόκρισης των αργών κλάσεων εξαρτάται από το μέγεθος του διατιθέμενου θαλάμου εξυπηρέτησης, L_s . Προκειμένου να εμβαθύνουμε στην επίδραση της μεταβολής του L_s , επιλέγουμε την τιμή του που δεν επιτρέπει το μέσο χρόνο απόκρισης της κλάσης 2 να υπερβεί 2.4 φορές τη μέση απαίτηση εξυπηρέτησης (δηλ. $2.4 \cdot (1/\mu_2)$). Όπως βλέπουμε από το πάνω μέρος του Πίνακα 11.11 αυτό συμβαίνει για $L_s = 7$.

Έχοντας τα παραπάνω υπόψη επιχειρούμε να συγκρίνουμε την επίδοση του NBR όσον αφορά παρόμοια μεγέθη του θαλάμου εξυπηρέτησης. Τα αντίστοιχα αποτελέσματα για παρόμοια μεγέθη του θαλάμου εξυπηρέτησης και εφαρμογή του NBR, παρουσιάζονται στον Πίνακα 11.12. Εφόσον κατά την περίπτωση αυτή οι δύο ομάδες κλάσεων δεν συμπεριφέρονται ανεξάρτητα αλλά αντίθετα, η επιλογή των L_f και L_s επηρεάζει το χρόνο απόκρισης και της άλλης ομάδας, έχουμε συμπεριλάβει παραδείγματα με αρκετούς διαφορετικούς

BR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0134	0.0536	0.1593	0.5577
$E[V_i] \mu_i$	2.3240	2.0600	1.2173	1.1428
$\sigma_1(B_i)/\mu_i$	0.4225	0.4745	0.7865	0.8308
$\Pr_{B,i}$	0.1622		0.0042	
NBR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0146	0.0585	0.1587	0.5554
$E[V_i] \mu_i$	1.4438	1.4438	1.4985	1.4985
$\sigma_1(B_i)/\mu_i$	0.6788	0.6723	0.4599	0.4578
$\Pr_{B,i}$	0.0856		0.0083	
SG-PF	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0126	0.0535	0.1593	0.5576
$E[V_i] \mu_i$	2.1264	2.1264	1.1625	1.1625
$\sigma_1(B_i)/\mu_i$	0.4598	0.4582	0.8278	0.8093
$\Pr_{B,i}$	0.1469		0.0042	

Πίνακας 11.7: Διαχωρισμένοι θάλαμοι εξυπηρέτησης, $L_s = 3$, $L_f = 4$

συνδυασμούς για τα μεγέθη των χώρων, L_f και L_s . Η περίπτωση που αντιστοιχεί σε $L_s = 7$ and $L_f = 9$, η οποία επισημάνθηκε προηγουμένως για τον BR, καταλήγει σε μέσο χρόνο απόκρισης για όλες τις κλάσεις που υπερβαίνει 1.56 φορές την αντίστοιχη μέση κανονικοποιημένη απαίτηση εξυπηρέτησης.

Η περίπτωση που αναφέρεται στον SG-PF για διαχωρισμένους θαλάμους εξυπηρέτησης παρουσιάζεται στον Πίνακα 11.13. Μπορούμε να δούμε ότι η επίδοση των γρήγορων κλάσεων παραμένει χωρίς μεταβολή, όπως κατά την εφαρμογή BR. Το μέγεθος του θαλάμου εξυπηρέτησης των γρήγορων κλάσεων $L_f = 9$ έχει ως αποτέλεσμα ένα λίγο μεγαλύτερο μέσο χρόνο απόκρισης αλλά με όφελος τη δικαιοσύνη μεταξύ των γρήγορων κλάσεων. Οι δύο γρήγορες κλάσεις έχουν ίσο μέσο κανονικοποιημένο χρόνο απόκρισης με ασυμπτωτική συμπεριφορά πολύ κοντά σε εκείνη του BR. Ωστόσο αυτή τη φορά προκύπτει ένα πλεονέκτημα για τις αργές κλάσεις. Προκειμένου για παράδειγμα να επιτευχθεί μέσος χρόνος απόκρισης 2.4 φορές τη μέση απαίτηση εξυπηρέτησης της κλάσης 2, απαιτείται θάλαμος εξυπηρέτησης με μέγεθος $L_s = 5$, το οποίο είναι μικρότερο από το αντίστοιχο του BR.

Ακολουθώντας την έως τώρα τακτική, παρουσιάζουμε και την αντίστοιχη περίπτωση που αναφέρεται σε κοινό θάλαμο εξυπηρέτησης. Έτσι παραθέτουμε τους Πίνακες 11.14-11.16, στους οποίους παρουσιάζονται και οι τρεις αλγόριθμοι για μεταβλητά μεγέθη του κοινού θαλάμου εξυπηρέτησης. Τα μεγέθη για τους BR και SG-PF έχουν επιλεγεί στο εύρος που αφήνουν το μέσο χρόνο από-

BR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0158	0.0632	0.1578	0.5522
$E[V_i] \mu_i$	2.8040	2.3240	1.2153	1.1414
$\sigma_1(B_i)/\mu_i$	0.3358	0.4032	0.7702	0.8102
$\Pr_{B,i}$	0.0139		0.0139	
NBR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0159	0.0634	0.1586	0.5551
$E[V_i] \mu_i$	1.5409	1.5409	1.5409	1.5409
$\sigma_1(B_i)/\mu_i$	0.5983	0.5983	0.3704	0.3704
$\Pr_{B,i}$	0.0088		0.0088	
SG-PF	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0158	0.0632	0.1578	0.5524
$E[V_i] \mu_i$	2.4252	2.4253	1.1611	1.1611
$\sigma_1(B_i)/\mu_i$	0.3903	0.3903	0.8139	0.7853
$\Pr_{B,i}$	0.0136		0.0136	

Πίνακας 11.8: Κοινός θάλαμος εξυπηρέτησης, $L = 7$

κρισης των γρήγορων κλάσεων (και την αντίστοιχη ελάχιστη ιδιοτιμή) σχεδόν ανεπηρέαστο. Επίσης, τα αποτελέσματα που αντιστοιχούν σε μέγεθος $L = 16$ μπορούν άμεσα να συγκριθούν με εκείνα της προηγούμενης περίπτωσης για διαχωρισμένο θάλαμο όπου $L_s = 7$ και $L_f = 9$.

Σε σχέση με τον NBR, παρουσιάζονται αποτελέσματα για μεγέθη θαλάμου εξυπηρέτησης $L = 8$ ως $L = 11$. Οι συγκεκριμένες τιμές επιλέχθηκαν ώστε να είναι κατά το δυνατό πιο κοντά σε εκείνες για την περίπτωση διαχωρισμένων θαλάμων. Ο NBR πράγματι προσδίδει δικαιοσύνη μεταξύ όλων των κλάσεων με κόστος τον αυξημένο μέσο χρόνο απόκρισης για τις γρήγορες κλάσεις. Παρόλα αυτά, αποτελεί την ενδεδειγμένη λύση αν το ζητούμενο είναι η βελτίωση της επίδοσης των αργών κλάσεων.

BR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0156	0.0624	0.1557	0.5448
$E[V_i] \mu_i$	3.7088	2.6833	1.3175	1.1909
$\sigma_1(B_i)/\mu_i$	0.2496	0.3438	0.6610	0.7158
$\Pr_{B,i}$	0.0272		0.0272	
NBR	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0157	0.0627	0.1568	0.5488
$E[V_i] \mu_i$	1.8480	1.8480	1.8480	1.8480
$\sigma_1(B_i)/\mu_i$	0.4988	0.4812	0.3003	0.3003
$\Pr_{B,i}$	0.0200		0.0200	
SG-PF	κλάση 1	κλάση 2	κλάση 3	κλάση 4
τ_i (Mb/s)	0.0156	0.0624	0.1558	0.5454
$E[V_i] \mu_i$	2.8975	2.8975	1.2253	1.2253
$\sigma_1(B_i)/\mu_i$	0.3200	0.3113	0.7246	0.6743
$\Pr_{B,i}$	0.0260		0.0260	

Πίνακας 11.9: Κοινός θάλαμος εξυπηρέτησης, $L = 7$ - Σκέδαση Nakagami με $m = 2$

	$\phi_3(0, 0, n_3, n_4)$			$\phi_4(0, 0, n_3, n_4)$	
(n_3, n_4)	$m = 1$	$m = 2$	(n_3, n_4)	$m = 1$	$m = 2$
(1,0)	1.0000	1.0000	(0,1)	1.0000	1.0000
(1,1)	0.6400	0.5248	(1,1)	0.8400	0.8208
(1,2)	0.4637	0.3404	(2,1)	0.7425	0.7187
(1,3)	0.3599	0.2454	(3,1)	0.6744	0.6497
(1,4)	0.2920	0.1885	(4,1)	0.6232	0.5985
(1,5)	0.2443	0.1512	(5,1)	0.5826	0.5586
(1,6)	0.2091	0.1250	(6,1)	0.5494	0.5261
(2,0)	0.7500	0.6875	(0,2)	0.7500	0.6875

Πίνακας 11.10: Κέρδη διαφορισμού πολλαπλών χρηστών για σκέδαση Rayleigh ($m = 1$) και σκέδαση Nakagami με $m = 2$

L_s	$E[V_1] \mu_1$	$E[V_2] \mu_2$	$\sigma_1(B_1)/\mu_1$	$\sigma_1(B_2)/\mu_2$	$\Pr_{B,1,2}$
4	2.5751	2.2091	0.3757	0.4348	0.0856
5	2.7372	2.2999	0.3473	0.4093	0.0427
6	2.8400	2.3561	0.3286	0.3910	0.0203
7	2.9002	2.3884	0.3162	0.3779	0.0092
8	2.9329	2.4057	0.3080	0.3683	0.0040
9	2.9496	2.4144	0.3027	0.3613	0.0017
10	2.9575	2.4185	0.2993	0.3563	$6.73 \cdot 10^{-4}$
L_f	$E[V_3] \mu_3$	$E[V_4] \mu_4$	$\sigma_1(B_3)/\mu_3$	$\sigma_1(B_4)/\mu_4$	$\Pr_{B,3,4}$
6	1.2246	1.1474	0.7716	0.8128	$1.44 \cdot 10^{-4}$
7	1.2249	1.1476	0.7702	0.8102	$2.44 \cdot 10^{-5}$
8	1.2250	1.1476	0.7698	0.8092	$3.94 \cdot 10^{-6}$
9	1.2250	1.1476	0.7696	0.8088	$6.09 \cdot 10^{-7}$
10	1.2250	1.1476	0.7696	0.8087	$9.09 \cdot 10^{-8}$
11	1.2250	1.1476	0.7696	0.8086	$1.31 \cdot 10^{-8}$
15	1.2250	1.1476	0.7696	0.8086	$4.43 \cdot 10^{-12}$

Πίνακας 11.11: Διαχωρισμένοι θάλαμοι εξυπηρέτησης - BR

(L_s, L_f)	$E[V_1] \mu_1$	$E[V_2] \mu_2$	$\sigma_1(B_1)/\mu_1$	$\sigma_1(B_2)/\mu_2$	$\Pr_{B,1,2}$
(5,6)	1.5408	1.5408	0.6173	0.6023	0.0119
(5,8)	1.5410	1.5410	0.6172	0.6023	0.0119
(5,9)	1.5410	1.5410	0.6172	0.6023	0.0119
(5,10)	1.5410	1.5410	0.6172	0.6023	0.0119
(6,9)	1.5554	1.5554	0.6044	0.5855	0.0040
(7,9)	1.5612	1.5612	0.5972	0.5751	0.0013
(10,9)	1.5644	1.5644	0.5906	0.5628	$3.04 \cdot 10^{-5}$
(15,9)	1.5645	1.5645	0.5899	0.5605	$3.31 \cdot 10^{-8}$
(L_s, L_f)	$E[V_3] \mu_3$	$E[V_4] \mu_4$	$\sigma_1(B_3)/\mu_3$	$\sigma_1(B_4)/\mu_4$	$\Pr_{B,3,4}$
(5,6)	1.5545	1.5545	0.3610	0.3593	$4.65 \cdot 10^{-4}$
(5,8)	1.5553	1.5553	0.3601	0.3582	$1.84 \cdot 10^{-5}$
(5,9)	1.5554	1.5554	0.3601	0.3581	$3.39 \cdot 10^{-6}$
(5,10)	1.5554	1.5554	0.3600	0.3580	$6.02 \cdot 10^{-7}$
(6,9)	1.5612	1.5612	0.3282	0.3267	$3.58 \cdot 10^{-6}$
(7,9)	1.5634	1.5634	0.3022	0.3010	$3.68 \cdot 10^{-6}$
(10,9)	1.5644	1.5644	0.2463	0.2456	$3.75 \cdot 10^{-6}$
(15,9)	1.5645	1.5645	0.1913	0.1910	$3.75 \cdot 10^{-6}$

Πίνακας 11.12: Διαχωρισμένοι θάλαμοι εξυπηρέτησης - NBR

L_s	$E[V_1] \mu_1$	$E[V_2] \mu_2$	$\sigma_1(B_1)/\mu_1$	$\sigma_1(B_2)/\mu_2$	$\Pr_{B,1,2}$
3	2.1298	2.1298	0.4613	0.4576	0.1625
4	2.2881	2.2881	0.4233	0.4174	0.0846
5	2.3940	2.3940	0.3980	0.3898	0.0421
6	2.4599	2.4599	0.3809	0.3703	0.0200
10	2.5336	2.5336	0.3529	0.3341	$6.65 \cdot 10^{-4}$
L_f	$E[V_3] \mu_3$	$E[V_4] \mu_4$	$\sigma_1(B_3)/\mu_3$	$\sigma_1(B_4)/\mu_4$	$\Pr_{B,3,4}$
7	1.1681	1.1681	0.8139	0.7853	$2.55 \cdot 10^{-5}$
8	1.1682	1.1682	0.8135	0.7840	$4.17 \cdot 10^{-6}$
9	1.1682	1.1682	0.8134	0.7835	$6.55 \cdot 10^{-7}$
10	1.1682	1.1682	0.8134	0.7833	$9.94 \cdot 10^{-8}$
15	1.1682	1.1682	0.8133	0.7832	$5.36 \cdot 10^{-12}$

Πίνακας 11.13: Διαχωρισμένοι θάλαμοι εξυπηρέτησης - SG-PF

L	$E[V_1] \mu_1$	$E[V_2] \mu_2$	$\sigma_1(B_1)/\mu_1$	$\sigma_1(B_2)/\mu_2$	\Pr_B
7	2.8036	2.3210	0.3358	0.4032	0.0137
9	2.9214	2.3955	0.3112	0.3743	0.0029
12	2.9598	2.4193	0.2980	0.3548	$2.10 \cdot 10^{-4}$
15	2.9636	2.4216	0.2949	0.3477	$1.17 \cdot 10^{-5}$
16	2.9637	2.4217	0.2946	0.3467	$4.25 \cdot 10^{-6}$
L	$E[V_3] \mu_3$	$E[V_4] \mu_4$	$\sigma_1(B_3)/\mu_3$	$\sigma_1(B_4)/\mu_4$	\Pr_B
7	1.2153	1.1414	0.7702	0.8102	0.0139
9	1.2227	1.1461	0.7696	0.8088	0.0029
12	1.2248	1.1475	0.7696	0.8086	$2.10 \cdot 10^{-4}$
15	1.2250	1.1476	0.7696	0.8086	$1.17 \cdot 10^{-5}$
16	1.2250	1.1476	0.7696	0.8086	$4.25 \cdot 10^{-6}$

Πίνακας 11.14: Κοινός θάλαμος εξυπηρέτησης - BR

L	$E[V_1] \mu_1$	$E[V_2] \mu_2$	$\sigma_1(B_1)/\mu_1$	$\sigma_1(B_2)/\mu_2$	\Pr_B
8	1.5531	1.5531	0.6048	0.5852	0.0038
9	1.5593	1.5593	0.5985	0.5764	0.0016
10	1.5622	1.5622	0.5947	0.5704	$6.31 \cdot 10^{-4}$
11	1.5635	1.5635	0.5925	0.5666	$2.46 \cdot 10^{-4}$
12	1.5641	1.5641	0.5913	0.5641	$9.33 \cdot 10^{-5}$
15	1.5644	1.5644	0.5900	0.5611	$4.44 \cdot 10^{-6}$
16	1.5645	1.5645	0.5900	0.5608	$1.55 \cdot 10^{-6}$
L	$E[V_3] \mu_3$	$E[V_4] \mu_4$	$\sigma_1(B_3)/\mu_3$	$\sigma_1(B_4)/\mu_4$	\Pr_B
8	1.5531	1.5531	0.3397	0.3397	0.0038
9	1.5593	1.5593	0.3143	0.3143	0.0016
10	1.5622	1.5622	0.2929	0.2929	$6.31 \cdot 10^{-4}$
11	1.5635	1.5635	0.2745	0.2745	$2.46 \cdot 10^{-4}$
12	1.5641	1.5641	0.2586	0.2586	$9.33 \cdot 10^{-5}$
15	1.5644	1.5644	0.2212	0.2212	$4.44 \cdot 10^{-6}$
16	1.5645	1.5645	0.2113	0.2113	$1.55 \cdot 10^{-6}$

Πίνακας 11.15: Κοινός θάλαμος εξυπηρέτησης - NBR

L	$E[V_1] \mu_1$	$E[V_2] \mu_2$	$\sigma_1(B_1)/\mu_1$	$\sigma_1(B_2)/\mu_2$	\Pr_B
7	2.4253	2.4253	0.3903	0.3814	0.0135
9	2.5080	2.5080	0.3656	0.3519	0.0028
12	2.5345	2.5345	0.3519	0.3324	$2.08 \cdot 10^{-4}$
15	2.5371	2.5371	0.3484	0.3257	$1.18 \cdot 10^{-5}$
16	2.5372	2.5372	0.3481	0.3248	$4.31 \cdot 10^{-6}$
L	$E[V_3] \mu_3$	$E[V_4] \mu_4$	$\sigma_1(B_3)/\mu_3$	$\sigma_1(B_4)/\mu_4$	\Pr_B
7	1.1611	1.1611	0.8139	0.7853	0.0135
9	1.1665	1.1665	0.8134	0.7835	0.0028
12	1.1680	1.1680	0.8133	0.7832	$2.08 \cdot 10^{-4}$
15	1.1682	1.1682	0.8133	0.7832	$1.18 \cdot 10^{-5}$
16	1.1682	1.1682	0.8133	0.7832	$4.31 \cdot 10^{-6}$

Πίνακας 11.16: Κοινός θάλαμος εξυπηρέτησης - SG-PF



Κεφάλαιο 12

Συμπεράσματα και Κατευθύνσεις Μελλοντικής Έρευνας

ΚΑΝΟΝΤΑΣ έναν επίλογο, στο κεφάλαιο αυτό παραθέτουμε τα συμπεράσματα που προέκυψαν κατά την εκπόνηση της διατριβής και παράλληλα αναφέρουμε πιθανά θέματα μελλοντικής έρευνας η οποία μπορεί να βασιστεί στην παρούσα ερευνητική προσπάθεια.

12.1 Συμπεράσματα

Στα πλαίσια της διατριβής τέθηκε το θέμα του ευκαιριακού χρονοπρογραμματισμού σε ασύρματα δίκτυα επικοινωνιών που υποστηρίζουν τη συγκεκριμένη δυνατότητα και αναλύθηκαν οι βασικοί αλγόριθμοι BR και NBR. Ακολούθησε η μοντελοποίηση των παραπάνω αλγορίθμων μέσω της έννοιας του *κέρδους διαφορισμού πολλαπλών χρηστών*, ως ένα γενικό σύστημα SDPS με ταχύτητα του εξυπηρετητή που εξαρτάται από το πλήθος των ενεργών τερματικών. Για το σκοπό αυτό κρίθηκε σκόπιμο να μελετηθεί ξεχωριστά το γενικό μαθηματικό πρόβλημα της κατανομής του χρόνου απόκρισης σε συστήματα SDPS και στη συνέχεια να εφαρμοστούν τα αποτελέσματα στο πρόβλημα των αλγορίθμων χρονοπρογραμματισμού.

Κάνοντας χρήση της θεωρίας αλυσίδων Markov συνεχούς χρόνου, έγινε εφικτός ο υπολογισμός της κατανομής του χρόνου απόκρισης ενός «πελάτη» σε σύστημα αναμονής SDPS πεπερασμένου μεγέθους, με πολλαπλές κλάσεις πελατών και απαιτήσεις εξυπηρέτησης εκθετικά κατανεμημένες, μέσω του εκθετικού ενός κατάλληλου πίνακα τύπου M. Η μέθοδος που ακολουθήθηκε περιλαμβάνει τη διασύνδεση του ποσού εξυπηρέτησης κάποιου πελάτη με το ποσό εξυπηρέτησης που λαμβάνει ένας «μόνιμος» πελάτης της ίδιας κλάσης και την

αντιστοίχιση του ποσού αυτού με την εξέλιξη μίας διαδικασίας ανταμοιβής Markov. Ο πίνακας τύπου M διαθέτει φασματικές ιδιότητες που βοηθούν στην περαιτέρω ανάλυση των αποτελεσμάτων, ενώ για συστήματα που διαθέτουν την ιδιότητα ισορροπίας, (όπως π.χ., κατά την εφαρμογή του NBR σε ασύρματα ευκαιριακά δίκτυα) αποδείχτηκε ότι είναι διαγωνιοποιήσιμος.

Με αναφορά πάντα στο γενικό μαθηματικό πρόβλημα που έχει τεθεί, η έκφραση του εκθετικού του πίνακα αποδίδει πλήρως τη μαρκοβιανή δομή του συστήματος, γεγονός που διευκολύνει την εξειδίκευση του προβλήματος σε συστήματα που χαρακτηρίζονται από διαχωρισμό χρονικής κλίμακας. Ο διαχωρισμός χρονικής κλίμακας αποτελεί μία συχνή και ενδιαφέρουσα κατάσταση σε ρεαλιστικά συστήματα, όπως για παράδειγμα στην περίπτωση κλάσεων τερματικών με μεγάλες διαφορές του μέσου ρυθμού εξυπηρέτησης σε ασύρματα δίκτυα ευκαιριακού χρονοπρογραμματισμού που μελετώνται στην παρούσα διατριβή. Προς το σκοπό αυτό, έγινε χρήση της δομής NCD της υποκείμενης αλυσίδας Markov, η οποία αναδύεται λόγω του διαχωρισμού χρονικής κλίμακας και αναπτύχθηκαν προσεγγιστικά αποτελέσματα για την κατανομή του χρόνου απόκρισης για τις κλάσεις με γρήγορη και με αργή δυναμική. Οι προσεγγίσεις αυτές προκύπτουν από μειωμένα υποσυστήματα τα οποία εξυπηρετούν μόνο πελάτες κλάσεων της ίδιας χρονικής κλίμακας και κατά συνέπεια, εκτός από το υπολογιστικό όφελος που προσφέρουν, παρέχουν και επιπλέον ευκολία για την κατανόηση των αποτελεσμάτων, χρήσιμη για την επιπρόσθετη διαίσθηση της λειτουργίας του συστήματος.

Με κίνητρο τη μορφή των κερδών διαφορισμού πολλαπλών χρηστών που ισχύει για τον BR σε περιβάλλον διαχωρισμού χρονικής κλίμακας, μελετήθηκε στο γενικότερο και πάλι πλαίσιο η περίπτωση κατά την οποία η παρουσία αργών πελατών δεν επηρεάζει το ρυθμό εξυπηρέτησης των γρήγορων, ενώ αντίθετα, οι πελάτες με αργή δυναμική λαμβάνουν μη αμελητέο ποσο εξυπηρέτησης όταν απουσιάζουν πελάτες γρήγορων κλάσεων. Ο συνδυασμός του διαχωρισμού χρονικής κλίμακας και επιπλέον τροποποίησης των ρυθμών εξυπηρέτησης οδήγησε σε απλούστερες εκφράσεις των τελικών αποτελεσμάτων. Παράλληλα, λόγω της συσχέτισης της κατανομής του χρόνου απόκρισης με την αντίστοιχη διαδικασία ανταμοιβής Markov, η ανάλυση κατευθύνθηκε προς τον προσδιορισμό του συνδυασμού πελατών των αργών κλάσεων, που όταν εμφανίζεται, αποτελεί το χειρότερο δυνατό σενάριο για την ασυμπτωτική συμπεριφορά ουράς της κατανομής του χρόνου απόκρισης των πελατών με γρήγορη δυναμική. Η σχετική ανάλυση συνέβαλε στο προσδιορισμό της επίδρασης της μορφής του θαλάμου εξυπηρέτησης (δηλαδή του σχήματος ελέγχου αποδοχής) στην ασυμπτωτική συμπεριφορά της κατανομής του χρόνου απόκρισης.

Κλείνοντας τη μελέτη του ευρύτερου θεωρητικού προβλήματος, παρου-

σιάστηκε ένα γενικότερο πλαίσιο που αφορά όχι μόνο εκθετικά κατανεμημένες απαιτήσεις εξυπηρέτησης αλλά και απαιτήσεις που προέρχονται από κατανομή με τυχαία μαρκοβιανή μορφή τύπου-φάσης. Όπως αποδείχτηκε, υπό το πρίσμα της γενίκευσης αυτής η βασική ουσία των αποτελεσμάτων παραμένει, με αυξημένη ωστόσο πολυπλοκότητα.

Με την ολοκλήρωση του θεωρητικού κομματιού, ακολούθησε η εφαρμογή των θεωρητικών αποτελεσμάτων που παρήχθησαν στο ρεαλιστικό σύστημα ενός ασύρματου δικτύου που υποστηρίζει ευκαιριακό χρονοπρογραμματισμό. Οι εκφράσεις για τα μέσα μεγέθη επίδοσης σε απουσία διαχωρισμού χρονικής κλίμακας (ο οποίος στη συγκεκριμένη περίπτωση εκφράζεται ως διαχωρισμός των ρυθμών εξυπηρέτησης) αποδόθηκαν σε συμπαγή μορφή, ενώ σε περιβάλλον διαχωρισμού χρονικής κλίμακας αποδείχτηκε ότι ο BR ευνοεί τα τερματικά με καλές συνθήκες μετάδοσης σε βάρος των υπολοίπων. Σε αντίθεση, ο NBR αυξάνει την έννοια της δικαιοσύνης μεταξύ των τερματικών, καθώς οι μέσες μετρικές επίδοσης κάθε κλάσης παραμένουν αναλλοίωτες πέραν της μέσης τιμής τους, με κόστος τη μειωμένη συνολική επίδοση.

Κάνοντας επίσης χρήση των αποτελεσμάτων που αφορούν την επίδραση των τερματικών των αργών κλάσεων στην κατανομή του χρόνου απόκρισης των τερματικών με γρήγορη δυναμική, έγινε φανερό ότι κατά την εφαρμογή του BR, η απεριόριστη πρόσβαση στο σύστημα σε τερματικά αργών κλάσεων έχει αμελητέα επίδραση στην ασυμπτωτική συμπεριφορά ουράς της κατανομής του χρόνου απόκρισης των γρήγορων τερματικών. Αντίθετα, κατά την εφαρμογή του NBR, οι συνέπειες είναι μη αμελητέες και ως εκ τούτου προτείνεται ο διαχωρισμός του θαλάμου εξυπηρέτησης.

Η ανάλυση και η σύγκριση των BR και NBR οδήγησε στην πρόταση ενός νέου αλγορίθμου ο οποίος συνδυάζει τα θετικά χαρακτηριστικά των δύο παραπάνω. Ο SG-PF όπως ονομάστηκε, διατηρεί τα θετικά αποτελέσματα του BR όσον αφορά την ασυμπτωτική συμπεριφορά ουράς της κατανομής του χρόνου απόκρισης των τερματικών γρήγορων κλάσεων, ενώ παράλληλα, αυξάνει την έννοια της δικαιοσύνης μεταξύ τερματικών κλάσεων της ίδιας ομάδας. Προς το σκοπό αυτό δείξαμε ότι τα τερματικά των αργών κλάσεων μπορούν να παραμεληθούν πλήρως όταν υπάρχει στο σύστημα έστω και ένα ενεργό τερματικό από τις γρήγορες κλάσεις. Η συγκεκριμένη απόφαση δεν ελαττώνει αξιοσημείωτα την ήδη μικρή επίδοση των τερματικών των αργών κλάσεων.

12.2 Κατευθύνσεις Μελλοντικής Έρευνας

Στα πλαίσια της διατριβής επιλέχθηκαν οι αλγόριθμοι BR και NBR οι οποίοι αποτελούν χαρακτηριστικά παραδείγματα ευκαιριακού χρονοπρογραμ-

ματισμού και διαθέτουν συγκριτικά πλεονεκτήματα και μειονεκτήματα που αναλύθηκαν εκτενώς. Ωστόσο οι αλγόριθμοι ευκαιριακού χρονοπρογραμματισμού δεν περιορίζονται στα δύο βασικά αυτά σχήματα. Στη βιβλιογραφία έχουν προταθεί διάφορες παραλλαγές προκειμένου να υπάρξουν εγγυήσεις στην ποιότητα της προσφερόμενης υπηρεσίας για εφαρμογές πραγματικού χρόνου ή μη και η επίδοση αυτών των σχημάτων θα ήταν ιδιαίτερα ενδιαφέρον να μοντελοποιηθεί και να συγκριθεί κατά τον τρόπο που αναπτύχθηκε εδώ.

Μία άλλη οικογένεια αλγορίθμων συμπεριλαμβάνει στη λήψη της απόφασης για το ποιο τερματικό θα εξυπηρετηθεί και τον χρόνο αναμονής κάθε ενεργού τερματικού. Για παράδειγμα, αυτό μπορεί να πραγματοποιηθεί μέσω μιας γραμμικής σχέσης δίνοντας τον αλγόριθμο τροποποιημένης μέγιστης καθυστέρησης (αγ.: Modified-Largest Weighted Delay First (M-LWDF)) ή μέσω μιας εκθετικής συνάρτησης καταλήγοντας στον Εκθετικό Αλγόριθμο (αγ.: Exponential Rule). Η παρουσία του όρου της καθυστέρησης στον τρόπο λήψης της απόφασης, περιπλέκει αρκετά το πρόβλημα της εύρεσης της κατανομής του χρόνου απόκρισης και κατά συνέπεια αποτελεί ένα ανοιχτό πεδίο έρευνας.

Το ζήτημα του ευκαιριακού χρονοπρογραμματισμού έχει αναδυθεί ως πρόβλημα και σε νεότερα συστήματα ασύρματης δικτύωσης. Εξαιτίας της αυξανόμενης ανάγκης για παροχή καλύτερων υπηρεσιών σε χαμηλότερο κόστος, οι πάροχοι υιοθετούν όλο και συχνότερα τη χρήση ασύρματων σταθμών βάσης χαμηλής ισχύος γνωστών ως φεμτοκυψέλες (αγ.: femtocell). Οι φεμτοκυψέλες λειτουργούν στο ίδιο φάσμα με τη μακροκυψέλη (αγ.: macrocell) και κατά κανόνα συνδέονται με το υπόλοιπο δίκτυο μέσω της σταθερής ευρυζωνικής σύνδεσης, δημιουργώντας έτσι μία συνολική αρχιτεκτονική δύο επιπέδων. Παρά τα προφανή πλεονεκτήματα της συγκεκριμένης μεθόδου, προκύπτει ένα σοβαρό πρόβλημα λόγω αυξημένων παρεμβολών μεταξύ των φεμτοκυψελών αλλά και μεταξύ των φεμτοκυψελών και της μακροκυψέλης.

Με χρήση αποτελεσμάτων που αφορούν συστήματα αναμονής SDPS, μπορεί να επεκταθεί το παρόν μοντέλο σε μοντέρνα συστήματα τέταρτης γενιάς (4G) ώστε να περιγραφεί ένα δίκτυο δύο επιπέδων φεμτοκυψελών - μακροκυψέλης. Με τον τρόπο αυτό και λαμβάνοντας υπόψη την τοπολογία του δικτύου, μπορεί κανείς να προχωρήσει στη διαστασιολόγηση του δικτύου με βάση τον αριθμό των φεμτοκυψελών που μπορούν να συνυπάρχουν ώστε η παρεμβολή να παραμείνει εντός κάποιου ορίου αλλά και στον αριθμό των χρηστών που μπορούν να εξυπηρετούνται παράλληλα.

Θα πρέπει τέλος να αναφερθεί ότι κάποιες από τις παραδοχές που έγιναν κατά την ανάπτυξη της μοντελοποίησης, παρόλο που αντικατοπτρίζουν την πραγματικότητα σε βαθμό που να μην αλλοιώνουν την ουσία των αποτελεσμάτων, θα μπορούσαν ενδεχομένως να γίνουν πιο ελαστικές. Ως παρά-

δειγμα αναφέρουμε την υπόθεση γραμμικής σχέσης επιτεύξιμου ρυθμού μετάδοσης και SNR, η οποία κατά περίπτωση μπορεί να εκφραστεί λεπτομερέστερα μέσω μίας λογαριθμικής συνάρτησης ή ακόμα και μέσω μίας προκαθορισμένης αντιστοίχισης σε ένα προκαθορισμένο εύρος τιμών. Επιπλέον, θα μπορούσε να υποθέσει κανείς διαφορετικό μοντέλο σκέδασης για διαφορετικές κλάσεις τερματικών, δίνοντας κατά τον τρόπο αυτό μεγαλύτερη έμφαση στο περιβάλλον της ασύρματης ζεύξης. Οι παραπάνω περιορισμοί δημιουργούν ένα ιδιαίτερα σύνθετο πρόβλημα, το οποίο θα μπορούσε να αποτελέσει ένα επιπλέον αντικείμενο μελέτης.



Παράρτημα

I Απόδειξη ότι το γινόμενο $n_i \phi_i(\mathbf{n})$ είναι γνησίως αύξουσα συνάρτηση ως προς $n_i, i \in \mathcal{K}$

Έστω ότι X_{n_i} είναι ο μέγιστος ρυθμός μετάδοσης εντός της κλάσης i , δηλαδή $X_{n_i} \triangleq \max_{m=1, \dots, n_i} \tilde{R}_m$. Τότε έχουμε

$$\begin{aligned} n_i \phi_i(\mathbf{n}) &= \frac{1}{\tilde{r}_i} \int_0^\infty y \prod_{\substack{m \in \mathcal{K} \\ m \neq i}} (F_m(\frac{a_i y}{a_m}))^{n_m} n_i (F_i(y))^{n_i-1} dF_i(y) = \\ &= \frac{1}{\tilde{r}_i} \int_0^\infty y \prod_{\substack{m \in \mathcal{K} \\ m \neq i}} (F_m(\frac{a_i y}{a_m}))^{n_m} dF_i^{n_i}(y) = \frac{1}{\tilde{r}_i} \mathbb{E} \left[X_{n_i} \prod_{\substack{m \in \mathcal{K} \\ m \neq i}} (F_m(\frac{a_i X_{n_i}}{a_m}))^{n_m} \right]. \end{aligned}$$

Η X_{n_i} είναι στοχαστικά αύξουσα συνάρτηση και λόγω του γεγονότος αυτού, για κάθε αύξουσα συνάρτηση $h(x)$ ισχύει ότι

$$X_{n_1} \geq_{st} X_{n_2} \Rightarrow \mathbb{E} [h(X_{n_1})] \geq \mathbb{E} [h(X_{n_2})].$$

(Για αυστηρή απόδειξη του τελευταίου ισχυρισμού βλ. π.χ., [49] σελ.252.) Με την αντικατάσταση $h(x) = x \prod_{\substack{m \in \mathcal{K} \\ m \neq i}} (F_m(\frac{a_i x}{a_m}))^{n_m}$ η απόδειξη ολοκληρώνεται.

II Απόδειξη του Θεωρήματος 1

Μελετάμε αρχικά μία γρήγορη κλάση, $i \in \mathcal{K}_f$. Με εφαρμογή της (2.6),

$$|\phi_i(\mathbf{n}_s, \mathbf{n}_f) - \phi_i(\mathbf{0}, \mathbf{n}_f)| = \int_0^\infty J(x) d\hat{F}(x),$$

όπου $J(x) \triangleq x(\hat{F}(x))^{n_i-1} \prod_{\substack{m \in \mathcal{K}_f \\ m \neq i}} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m}))^{n_m} (1 - \prod_{m \in \mathcal{K}_s} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m}))^{n_m})$. Το παραπάνω ολοκλήρωμα μπορεί να χωριστεί στο ενδιαμέσο σημείο $x = \sqrt{\epsilon}$, ώστε να λάβουμε

$$|\phi_i(\mathbf{n}_s, \mathbf{n}_f) - \phi_i(\mathbf{0}, \mathbf{n}_f)| = \int_0^{\sqrt{\epsilon}} J(x) d\hat{F}(x) + \int_{\sqrt{\epsilon}}^{\infty} J(x) d\hat{F}(x). \quad (\text{II.1})$$

Από τον ορισμό του $\delta_1(\epsilon)$,

$$\begin{aligned} 1 - \prod_{m \in \mathcal{K}_s} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m}))^{n_m} &= 1 - \prod_{m \in \mathcal{K}_s} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m \epsilon}))^{n_m} \leq 1 - \prod_{m \in \mathcal{K}_s} (\hat{F}(\frac{\tilde{r}_i}{\tilde{r}_m \sqrt{\epsilon}}))^{n_m} \\ &\leq 1 - (\hat{F}(\frac{r}{\sqrt{\epsilon}}))^{|n_s|} = (1 - (1 - \delta_1(\epsilon))^{|n_s|}) = \delta_1(\epsilon) |n_s| + O(\delta_1(\epsilon)^2), \end{aligned}$$

για όλα τα $x \geq \sqrt{\epsilon}$. Έτσι, διαχωρίζοντας τους όρους εντός του γινομένου σε αργή και γρήγορη ομάδα κλάσεων, το δεύτερο ολοκλήρωμα στο δεξί μέρος της (II.1) μπορεί να φραχτεί άνω ως ακολούθως:

$$\begin{aligned} \int_{\sqrt{\epsilon}}^{\infty} J(x) d\hat{F}(x) &\leq \int_{\sqrt{\epsilon}}^{\infty} x(\hat{F}(x))^{n_i-1} \prod_{\substack{m \in \mathcal{K}_f \\ m \neq i}} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m}))^{n_m} (\delta_1(\epsilon) |n_s| + O(\delta_1(\epsilon)^2)) d\hat{F}(x) \\ &= \delta_1(\epsilon) |n_s| \int_{\sqrt{\epsilon}}^{\infty} x(\hat{F}(x))^{n_i-1} \prod_{\substack{m \in \mathcal{K}_f \\ m \neq i}} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m}))^{n_m} d\hat{F}(x) + O(\delta_1(\epsilon)^2) \\ &= \delta_1(\epsilon) |n_s| \phi_i(\mathbf{0}, \mathbf{n}_f) + O(\delta_1(\epsilon)^2). \end{aligned}$$

Για το πρώτο ολοκλήρωμα στο δεξί μέρος της (II.1), έχουμε ότι

$$\begin{aligned} \int_0^{\sqrt{\epsilon}} J(x) d\hat{F}(x) &\leq \int_0^{\sqrt{\epsilon}} x(\hat{F}(x))^{n_i-1} \prod_{\substack{m \in \mathcal{K}_f \\ m \neq i}} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m}))^{n_m} d\hat{F}(x) \\ &\leq \sqrt{\epsilon} \prod_{\substack{m \in \mathcal{K}_f \\ m \neq i}} (\hat{F}(\frac{\tilde{r}_i \sqrt{\epsilon}}{\tilde{r}_m}))^{n_m} (\hat{F}(\sqrt{\epsilon}))^{n_i} \leq \sqrt{\epsilon} \delta_2(\epsilon)^{|n_f|}, \end{aligned}$$

αφού η υπό ολοκλήρωση ποσότητα είναι μη αρνητική και το πρώτο κομμάτι της εκφώνησης έχει αποδειχθεί.

Ας θεωρήσουμε τώρα την περίπτωση για μία αργή κλάση, $i \in \mathcal{K}_s$. Όταν $\mathbf{n}_f = \mathbf{0}$, με εφαρμογή της (2.6), παίρνουμε

$$\phi_i(\mathbf{n}_s, \mathbf{n}_f) = \int_0^{\infty} x \prod_{\substack{m \in \mathcal{K}_s \\ m \neq i}} (\hat{F}(\frac{\tilde{r}_i x}{\tilde{r}_m}))^{n_m} (\hat{F}(x))^{n_i-1} d\hat{F}(x) = \phi_i(\mathbf{n}_s, \mathbf{0}), \quad \forall \epsilon > 0.$$

Όταν $\mathbf{n}_f > \mathbf{0}$, έχουμε ότι

$$\phi_i(\mathbf{n}_s, \mathbf{n}_f) = \int_0^\infty J(x) d\hat{F}(x). \quad (\text{II.2})$$

Και πάλι, θεωρούμε το ενδιαμέσο σημείο $x_0 = 1/\sqrt{\epsilon}$ και χωρίζουμε το ολοκλήρωμα της δεξιάς πλευράς της (II.2) σε δύο μέρη. Το πρώτο κομμάτι, έστω J_1 , είναι

$$J_1 = \int_0^{\frac{1}{\sqrt{\epsilon}}} x \prod_{\substack{m \in \mathcal{K}_s \\ m \neq i}} (\hat{F}(\frac{\check{r}_i x}{\check{r}_m}))^{n_m} (\hat{F}(x))^{n_i-1} \prod_{m \in \mathcal{K}_f} (\hat{F}(\frac{\check{r}_i \epsilon x}{\check{r}_m}))^{n_m} d\hat{F}(x).$$

Για $x \leq 1/\sqrt{\epsilon}$, ισχύει πάντα ότι $\hat{F}(\frac{\check{r}_i \epsilon x}{\check{r}_m}) \leq \delta_2(\epsilon)$ και έτσι

$$J_1 \leq \int_0^\infty x \prod_{\substack{m \in \mathcal{K}_s \\ m \neq i}} (\hat{F}(\frac{\check{r}_i x}{\check{r}_m}))^{n_m} (\hat{F}(x))^{n_i-1} (\delta_2(\epsilon))^{|n_f|} d\hat{F}(x) = (\delta_2(\epsilon))^{|n_f|} \phi_i(\mathbf{n}_s, \mathbf{0}).$$

Για το δεύτερο ολοκλήρωμα, έστω J_2 , έχουμε από την (II.2) ότι

$$J_2 = \int_{\frac{1}{\sqrt{\epsilon}}}^\infty x \prod_{\substack{m \in \mathcal{K}_s \\ m \neq i}} (\hat{F}(\frac{\check{r}_i x}{\check{r}_m}))^{n_m} (\hat{F}(x))^{n_i-1} \prod_{m \in \mathcal{K}_f} (\hat{F}(\frac{\check{r}_i \epsilon x}{\check{r}_m}))^{n_m} d\hat{F}(x),$$

που είναι ποσότητα άνω φραγμένη από το

$$\int_{\frac{1}{\sqrt{\epsilon}}}^\infty x d\hat{F}(x) = \frac{1}{\sqrt{\epsilon}} (1 - \hat{F}(\frac{1}{\sqrt{\epsilon}})) + \int_{\frac{1}{\sqrt{\epsilon}}}^\infty (1 - \hat{F}(u)) du = \delta_3(\epsilon),$$

ολοκληρώνοντας την απόδειξη του θεωρήματος.

III Απόδειξη της Πρότασης 1

Από την ανισότητα Chebyshev έχουμε ότι $1 - \hat{F}(x) \leq e^{-\theta_0 x} s(\theta_0)$, $\forall x$. Επίσης, λόγω του ορισμού του $\delta_1(\epsilon)$ (βλ. π.χ., την εκφώνηση του Θεωρήματος 1)

$$\delta_1(\epsilon) \leq e^{-\theta_0 \frac{x}{\sqrt{\epsilon}}} s(\theta_0).$$

Ομοίως, έχουμε ότι

$$\delta_3(\epsilon) \leq s(\theta_0) (\frac{1}{\sqrt{\epsilon}} + \frac{1}{\theta_0}) e^{-\frac{\theta_0}{\sqrt{\epsilon}}}.$$

Έτσι, τα $\delta_1(\epsilon)$, $\delta_3(\epsilon)$ είναι και τα δύο το πολύ τάξης $o(\epsilon^\kappa)$, $\forall \kappa > 0$, καθώς $\epsilon \rightarrow 0$, δηλαδή τα $\delta_1(\epsilon)$, $\delta_3(\epsilon)$ φθίνουν γρηγορότερα από οποιαδήποτε πολυωνυμική συνάρτηση του ϵ , καθώς $\epsilon \rightarrow 0$.

Επειδή από υπόθεση $\hat{F}(x) = O(x^a)$, για $a > 0$, έχουμε ότι $\delta_2(\epsilon) = O(\epsilon^{a/2})$, καθώς το $\epsilon \rightarrow 0$. Μία εφαρμογή της τελευταίας εξίσωσης στις εκφράσεις της εκφώνησης του Θεωρήματος 1 που περιλαμβάνουν το $\delta_2(\epsilon)$, ολοκληρώνει την απόδειξη.

IV Απόδειξη του Λήμματος 2

Εφόσον για οποιοδήποτε δυο νόρμες $\|\cdot\|_a$ και $\|\cdot\|_b$, ισχύει ότι $\|\cdot\|_a = O(\|\cdot\|_b)$ (βλ. π.χ., [20, Θεώρημα 4.4.2]), δηλαδή όλες οι νόρμες είναι ισοδύναμες με αυτήν την έννοια, μπορούμε χωρίς βλάβη της γενικότητας να κάνουμε χρήση της νόρμας $\|\cdot\|_\infty$.

Θεωρούμε την κανονική μορφή Jordan J του $-A$ (ο οποίος έχει ιδιοτιμές $-\sigma_j(A)$), με αντίστοιχα δεξιά ιδιοδιανύσματα και κύρια διανύσματα τις στήλες του πίνακα X . Τότε, $-A = XJX^{-1}$ και $e^{-At} = Xe^{Jt}X^{-1}$, οδηγώντας στο ότι $\|e^{-At}\|_\infty \leq c_1\|e^{Jt}\|_\infty$, για όλα τα $t \geq 0$, με $c_1 = \|X\|_\infty\|X^{-1}\|_\infty$. Έστω ότι p είναι ο αριθμός των στοιχειωδών υποπινάκων Jordan J_i του J , με μεγέθη m_i και αντίστοιχες ιδιοτιμές $-\sigma_i(A)$, $i = 1, \dots, p$. Έτσι, $\exp\{Jt\} = \text{diag}_i\{\exp\{J_it\}\}$ και $\|\exp\{Jt\}\|_\infty = \max_i\|\exp\{J_it\}\|_\infty$, ώστε το προηγούμενο φράγμα για όλα τα $t \geq 0$ μετατρέπεται σε

$$\|e^{-At}\|_\infty \leq c_1 \max_i \|e^{J_it}\|_\infty = c_1 \max_i |e^{-\sigma_i(A)t}| \sum_{j=0}^{m_i-1} t^j/j! = c_1 \max_i e^{-\text{Re}(\sigma_i(A))t} \sum_{j=0}^{m_i-1} t^j/j!,$$

όπου η πρώτη ισότητα ισχύει εξαιτίας της ειδικής μορφής των στοιχειωδών υποπινάκων Jordan και της επιλογής $\|\cdot\|_\infty$. Από τις υποθέσεις του λήμματος, κάθε ιδιοτιμή με πραγματικό μέρος ίσο με $\sigma_{\min}(A)$ έχει το δικό της ιδιοδιάνυσμα, και έτσι ο στοιχειώδης υποπίνακας Jordan που της αντιστοιχεί έχει μέγεθος ίσο με 1. Αν υπάρχουν υποπίνακες Jordan με $m_i > 1$, τότε απαραίτητα $\text{Re}\{\sigma_i(A)\} > \sigma_{\min}(A)$. Για κάθε $\alpha > \beta$ και κάθε $r > 0$, μπορεί κανείς να βρει μία σταθερά c_2 τέτοια ώστε $t^r e^{-\alpha t} \leq c_2 e^{-\beta t}$ για όλα τα $t \geq 0$, κατά συνέπεια, το εκθετικό φράγμα παίρνει τη μορφή

$$\|e^{-At}\|_\infty \leq c_3 e^{-\sigma_{\min}(A)t}, \quad \forall t \geq 0,$$

για κάποια κατάλληλα προσαρμοσμένη σταθερά c_3 .

Προκειμένου να εφαρμοστεί το αποτέλεσμα αυτό στον $-(A + E)$, υπενθυμίζεται ότι $\|E\|_\infty = O(\epsilon)$ και ότι όλες οι ιδιοτιμές του A που βρίσκονται στο $\sigma_{\min}(A)$, ακόμα και αν είναι πολλαπλές, έχουν το δικό τους ιδιοδιάνυσμα. Έτσι, για ϵ αρκετά μικρό, από την καθιερωμένη ανάλυση διαταραχών (βλ. π.χ., [58, Κεφ. 2]) προκύπτει ότι ο $A + E$ έχει ίσο αριθμό ιδιοτιμών που διαφέρουν κατά $O(\epsilon)$ από εκείνες του A που είναι οι μικρότερες και έχουν διακεκριμένα

ιδιοδιανύσματα. Ως αποτέλεσμα, ο $A + E$ ικανοποιεί τη συνθήκη ιδιοδιανυσμάτων του λήμματος και $\sigma_{\min}(A + E) = \sigma_{\min}(A) + \delta(E)$, με $\delta(E) = O(\epsilon)$. Έτσι, το αποτέλεσμα πράγματι εφαρμόζεται επίσης για τον $A + E$, δίνοντας

$$\|e^{-(A+E)t}\|_{\infty} \leq c_4 e^{-(\sigma_{\min}(A) + \delta(E))t}, \quad t \geq 0$$

(με διαφορετική σταθερά, αφού αυτή εξαρτάται από τον πίνακα της αριστερής πλευράς).

Εφαρμόζοντας την ταυτότητα της (7.1) με $H = A + E$ και $H' = A$, παίρνοντας νόρμες και αντικαθιστώντας τα φράγματα για τα δύο εκθετικά πινάκων, έχουμε

$$\begin{aligned} \|e^{-(A+E)t} - e^{-At}\|_{\infty} &= \left\| - \int_0^t e^{-A(t-s)} E e^{-(A+E)s} ds \right\|_{\infty} \leq \int_0^t \|e^{-A(t-s)} E e^{-(A+E)s}\|_{\infty} ds \\ &\leq \int_0^t \|e^{-A(t-s)}\|_{\infty} \|E\|_{\infty} \|e^{-(A+E)s}\|_{\infty} ds \leq c_5 \|E\|_{\infty} \int_0^t e^{-\sigma_{\min}(A)(t-s)} e^{-(\sigma_{\min}(A) + \delta(E))s} ds \\ &= c_5 \|E\|_{\infty} e^{-\sigma_{\min}(A)t} \frac{e^{-\delta(E)t} - 1}{-\delta(E)} = O(e^{-\sigma_{\min}(A)t} (e^{O(\epsilon)t} - 1)), \end{aligned}$$

όπου η τελευταία ισότητα ακολουθεί από το γεγονός ότι, εφόσον και η $\|E\|_{\infty}$ και το $-\delta(E)$ είναι τάξης $O(\epsilon)$, η $\|E\|_{\infty}/(-\delta(E))$ είναι φραγμένη από μία σταθερά καθώς $\epsilon \rightarrow 0$. Η δεύτερη ανισότητα στην εκφώνηση του λήμματος ακολουθεί από το γεγονός ότι $e^x - 1 \leq xe^x$.

V Απόδειξη του Λήμματος 3

Εφόσον κάθε \hat{A}_I , $I = 1, \dots, K$, είναι ένας μη αναγώγιμος απειροστικός γεννήτορας, ο $-\hat{A}_I$ διαθέτει ακριβώς μία απλή μηδενική ιδιοτιμή, με αριστερό και δεξί ιδιοδιάνυσμα $\hat{\pi}_I$ και $\mathbf{1}_I$, αντιστοίχα. (Για μεγαλύτερη σαφήνεια, σε αυτή την περίπτωση θα χρησιμοποιούμε για το διάνυσμα στήλη $\mathbf{1}$ δείκτη που δηλώνει το μέγεθός του έτσι ώστε να συνδέεται ευκολότερα με τον αντίστοιχο πίνακα.) Για το λόγο αυτό, ο $-\hat{A}_I$ μπορεί να γραφεί σε όρους της κανονικής του μορφής Jordan ως

$$-\hat{A}_I = \begin{bmatrix} \mathbf{1}_I & X_I \end{bmatrix} \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & J_I \end{bmatrix} \begin{bmatrix} \hat{\pi}_I \\ Y_I \end{bmatrix},$$

όπου ο X_I (αντίστ. Y_I) είναι ο (μη τετραγωνικός) πίνακας με στήλες (γραμμές) τα δεξιά (αριστερά) ιδιοδιανύσματα και κύρια διανύσματα που αντιστοιχούν στις μη μηδενικές ιδιοτιμές του $-\hat{A}_I$, κατάλληλα κανονικοποιημένα ώστε

$\begin{bmatrix} \hat{\pi}_I \\ Y_I \end{bmatrix} = [\mathbf{1}_I X_I]^{-1}$. Σημειώνουμε ότι ο $-\hat{A}_I$ είναι ένας μη ομαλός, μη αναγώγιμος πίνακας τύπου M, οπότε όλες οι ιδιοτιμές στον J_I έχουν αυστηρά θετικό πραγματικό μέρος.

Ορίζουμε τους (μη τετραγωνικούς εν γένει) μπλοκ διαγώνιους πίνακες $\Pi \triangleq \text{diag}_I\{\hat{\pi}_I\}$, $G \triangleq \text{diag}_I\{\mathbf{1}_I\}$, $X \triangleq \text{diag}_I\{X_I\}$, $Y \triangleq \text{diag}_I\{Y_I\}$ και τον τετραγωνικό μπλοκ διαγώνιο πίνακα $J \triangleq \text{diag}_I\{J_I\}$. Αναδιατάσσοντας τους δείκτες έτσι ώστε οι K μηδενικές ιδιοτιμές να εμφανίζονται πρώτες, ο πίνακας $\hat{A}_- \triangleq \text{diag}_I\{-\hat{A}_I\}$ μπορεί να γραφεί σε όρους της κανονικής του μορφής Jordan ως

$$\hat{A}_- = \text{diag}_I\{-\hat{A}_I\} = \begin{bmatrix} G & X \end{bmatrix} \begin{bmatrix} 0_{KK} & 0 \\ 0 & J \end{bmatrix} \begin{bmatrix} \Pi \\ Y \end{bmatrix},$$

όπου σε αναλογία με τους απλούς υποπίνακες, ο $Q \triangleq \begin{bmatrix} \Pi \\ Y \end{bmatrix}$ έχει για γραμμές τα αριστερά ιδιοδιανύσματα και κύρια διανύσματα του \hat{A}_- και ο $Q^{-1} = [G X]$, έχει στήλες ίσες με τα αντίστοιχα δεξιά ιδιοδιανύσματα και κύρια διανύσματα. Με αντιστροφή του μετασχηματισμού ομοιότητας,

$$\begin{bmatrix} 0_{KK} & 0 \\ 0 & J \end{bmatrix} = Q\hat{A}_-Q^{-1} = \begin{bmatrix} \Pi \\ Y \end{bmatrix} \hat{A}_- \begin{bmatrix} G & X \end{bmatrix} = \begin{bmatrix} \Pi\hat{A}_-G & \Pi\hat{A}_-X \\ Y\hat{A}_-G & Y\hat{A}_-X \end{bmatrix}. \quad (\text{V.1})$$

Διαιρούμε τις πρώτες K γραμμές του Q με ϵ και παίρνουμε τον πίνακα T ,

$$T \triangleq \begin{bmatrix} \frac{1}{\epsilon}\Pi \\ Y \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} \epsilon G & X \end{bmatrix}. \quad (\text{V.2})$$

Εφόσον $A = \hat{A}_- + O(\epsilon)$, μπορούμε να γράψουμε $A = \hat{A}_- + \epsilon C$, με $C = O(1)$ καθώς $\epsilon \rightarrow 0$ και

$$\begin{aligned} TAT^{-1} &= T\hat{A}_-T^{-1} + T(\epsilon C)T^{-1} = \begin{bmatrix} \Pi\hat{A}_-G & \frac{1}{\epsilon}\Pi\hat{A}_-X \\ \epsilon Y\hat{A}_-G & Y\hat{A}_-X \end{bmatrix} + \begin{bmatrix} \epsilon\Pi CG & \Pi CX \\ \epsilon^2 YCG & \epsilon YCX \end{bmatrix} \\ &= \begin{bmatrix} \bar{A} & \Pi CX \\ \epsilon^2 YCG & J + \epsilon YCX \end{bmatrix}, \end{aligned} \quad (\text{V.3})$$

όπου η δεύτερη ισότητα κάνει χρήση της (V.2), της τρίτης σχέσης της (V.1) και του γεγονότος ότι $\Pi AG = \epsilon\Pi CG = [\hat{\pi}_I A_{IJ} \mathbf{1}_J] = \bar{A}$. Το αποτέλεσμα μπορεί επίσης να δοθεί ως

$$TAT^{-1} = W + O(\epsilon^2), \quad W \triangleq \begin{bmatrix} \bar{A} & \Pi CX \\ 0 & J + \epsilon YCX \end{bmatrix}. \quad (\text{V.4})$$

Ας θυμηθούμε τώρα ότι $A = \hat{A}_- + O(\epsilon)$ και έτσι, ο A έχει K ιδιοτιμές τάξης $O(\epsilon)$, ενώ οι υπόλοιπες είναι τάξης $O(1)$. Εφόσον ο A είναι ένας ομαλός μη αναγώγιμος πίνακας τύπου M, όλες οι ιδιοτιμές του έχουν θετικό πραγματικό μέρος και εκείνη με το ελάχιστο πραγματικό μέρος, η οποία βρίσκεται μεταξύ αυτών που είναι τάξης $O(\epsilon)$, είναι απλή. Κατά συνέπεια, καθώς το $\epsilon \rightarrow 0$, ο πίνακας $\epsilon^{-1}A$ έχει K ιδιοτιμές τάξης $O(1)$, ενώ οι υπόλοιπες ιδιοτιμές μεγαλώνουν απεριόριστα. Η ιδιοτιμή με το ελάχιστο πραγματικό μέρος (συμβολίζεται ως $\sigma_1(\epsilon^{-1}A)$), είναι μεταξύ εκείνων που είναι τάξης $O(1)$ και είναι απλή, οπότε και καλώς διαχωρισμένη από τις υπόλοιπες. Εξαιτίας της μπλοκ τριγωνικής άνω δομής στην (V.4), οι ιδιοτιμές του $\epsilon^{-1}W$ είναι ακριβώς εκείνες του $\epsilon^{-1}\bar{A}$ και εκείνες του $\epsilon^{-1}(J + \epsilon YCX)$. Επίσης, επειδή $\bar{A} = O(\epsilon)$ και όλες οι ιδιοτιμές του J έχουν θετικό πραγματικό μέρος τάξης $O(1)$, προκύπτει ότι ο $\epsilon^{-1}W$ έχει K ιδιοτιμές τάξης $O(1)$, ακριβώς εκείνες του $\epsilon^{-1}\bar{A}$, ενώ οι υπόλοιπες μεγαλώνουν απεριόριστα. Από την (V.4), το ότι $\epsilon^{-1}TAT^{-1} = \epsilon^{-1}W + O(\epsilon)$ και τη συνέχεια των ιδιοτιμών, το αναλλοίωτό τους ως προς μετασχηματισμό ομοιότητας και την καθιερωμένη θεωρία διαταραχών, έχουμε ότι οι K μικρότερες ιδιοτιμές τάξης $O(1)$ του $\epsilon^{-1}A$ τείνουν στις ιδιοτιμές του $\epsilon^{-1}\bar{A}$, διαφέροντας από εκείνες κατά $O(\epsilon)$. Έτσι, οι ιδιοτιμές του $\epsilon^{-1}\bar{A}$ έχουν θετικό πραγματικό μέρος και μία από αυτές, που συμβολίζεται με $\sigma_1(\epsilon^{-1}\bar{A})$, τείνει στο $\sigma_1(\epsilon^{-1}A)$, οπότε είναι απλή και έχει το ελάχιστο πραγματικό μέρος μεταξύ των ιδιοτιμών του $\epsilon^{-1}\bar{A}$. Ένας τελευταίος πολλαπλασιασμός με ϵ δείχνει ότι όλες οι ιδιοτιμές του \bar{A} έχουν θετικό πραγματικό μέρος και ότι η $\sigma_1(\bar{A})$ είναι απλή και $\sigma_1(\bar{A}) = \sigma_1(A) + O(\epsilon^2)$.

Επιπρόσθετα, μπορούμε να χρησιμοποιήσουμε την προσεγγιστικά μπλοκ τριγωνική άνω μορφή του TAT^{-1} στο δεξί μέρος της (V.3) ώστε να πάρουμε τις ακόλουθες εκφράσεις για το αριστερό και δεξί ιδιοδιάνυσμα που αντιστοιχεί στη $\sigma_1(A)$:

$$\mathbf{x}_1(TAT^{-1}) = \begin{bmatrix} \mathbf{x}_1(\bar{A}) + O(\epsilon) \\ O(\epsilon^2) \end{bmatrix}, \quad \mathbf{y}_1(TAT^{-1}) = [\mathbf{y}_1(\bar{A}) + O(\epsilon) \quad \mathbf{u}], \quad \mathbf{u} = O(1). \quad (\text{V.5})$$

Θα εξηγήσουμε την παραγωγή της πρώτης από αυτές τις εκφράσεις, με τα επιχειρήματα για τις υπόλοιπες να είναι παρόμοια. Διαχωρίζοντας το ιδιοδιάνυσμα σε υποδιανύσματα σύμμορφα με τη μπλοκ δομή, όπως $\mathbf{x}_1(TAT^{-1}) \doteq \mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}$, και υπενθυμίζοντας ότι κάθε ιδιοδιάνυσμα μπορεί να κανονικοποιηθεί ώστε τα στοιχεία του να είναι (το πολύ) τάξης $O(1)$, η σχέση ιδιοτιμής-ιδιοδιανύσματος $TAT^{-1}\mathbf{z} = (\sigma_1(\bar{A}) + O(\epsilon^2))\mathbf{z}$ δίνει

$$(J + \epsilon YCX - \sigma_1(\bar{A})I)\mathbf{z}_2 = O(\epsilon^2) \quad \text{και} \quad (\bar{A} - \sigma_1(\bar{A})I)\mathbf{z}_1 = -\Pi CX\mathbf{z}_2 + O(\epsilon^2).$$

Επειδή η $\sigma_1(\bar{A})$ είναι καλώς διαχωρισμένη από τις ιδιοτιμές του J , του οποίου

όλες είναι τάξης $O(1)$, η πρώτη έκφραση είναι αντιστρεπτή και οδηγεί στο ότι $z_2 = O(\epsilon^2)$. Με αντικατάσταση αυτού του αποτελέσματος στη δεύτερη έκφραση και μετά διαιρώντας με ϵ , παίρνουμε $(\epsilon^{-1}\bar{A} - \epsilon^{-1}\sigma_1(\bar{A})I)z_1 = O(\epsilon)$. Ο μη ομαλός πίνακας στα αριστερά είναι τάξης $O(1)$ με μηδενικό χώρο διατρεχόμενο από το $\mathbf{x}_1(\bar{A})$. Έτσι, απαραίτητα η λύση σε αυτό το μη ομαλό σύστημα εξισώσεων είναι $z_1 = \mathbf{x}_1(\bar{A}) + O(\epsilon)$.

Με εφαρμογή της (V.5), προσαρμόζοντας τα ιδιοδιανύσματα κατάλληλα ώστε να ληφθεί υπόψη ο μετασχηματισμός ομοιότητας, με χρήση της (V.2) και με υπενθύμιση του ότι $\mathbf{b}G = \mathbf{b} \operatorname{diag}_I \{ \mathbf{1}_I \} = \beta_b$ και του γεγονότος ότι $Y\mathbf{1} = \mathbf{0}$ (από την ορθογωνιότητα), έχουμε

$$\begin{aligned} \mathbf{b}\mathbf{x}_1(A) &= \mathbf{b}T^{-1}\mathbf{x}_1(TAT^{-1}) = [\epsilon\beta_b \quad \mathbf{b}X] \begin{bmatrix} \mathbf{x}_1(\bar{A}) + O(\epsilon) \\ O(\epsilon^2) \end{bmatrix} = \epsilon(\beta_b\mathbf{x}_1(\bar{A}) + O(\epsilon)), \\ \mathbf{y}_1(A)\mathbf{1} &= \mathbf{y}_1(TAT^{-1})T\mathbf{1} = [\mathbf{y}_1(\bar{A}) + O(\epsilon), \quad \mathbf{u}] \begin{bmatrix} \frac{1}{\epsilon}\mathbf{1} \\ \mathbf{0} \end{bmatrix} = \frac{1}{\epsilon}(\mathbf{y}_1(\bar{A})\mathbf{1} + O(\epsilon)). \end{aligned}$$

Πολλαπλασιασμός των δύο αυτών εκφράσεων ολοκληρώνει την απόδειξη του πρώτου μέρους του λήμματος.

Για το δεύτερο μέρος του λήμματος, εφαρμόζουμε την ανάπτυξη σε σειρά του εκθετικού του πίνακα και παρατηρούμε ότι $e^{-At} = T^{-1}e^{-TAT^{-1}t}T = T^{-1}(e^{-Wt} + O(\epsilon^2))T$, όπου η δεύτερη ισότητα ακολουθεί από την (V.4) και την εφαρμογή του Λήμματος 2 στον $T^{-1}AT$, που είναι δυνατό επειδή το ότι η ελάχιστη ιδιοτιμή είναι απλή διασφαλίζει ότι η συνθήκη ιδιοδιανυσμάτων του λήμματος ικανοποιείται. Χρήση της (V.2) οδηγεί στο

$$\begin{aligned} \mathbf{b}e^{-At}\mathbf{1} &= \mathbf{b}T^{-1}e^{-TAT^{-1}t}T\mathbf{1} = [\epsilon\beta_b \quad \mathbf{b}X] (e^{-Wt} + O(\epsilon^2)) \begin{bmatrix} \frac{1}{\epsilon}\mathbf{1} \\ \mathbf{0} \end{bmatrix} \\ &= [\epsilon\beta_b \quad \mathbf{b}X] e^{-Wt} \begin{bmatrix} \frac{1}{\epsilon}\mathbf{1} \\ \mathbf{0} \end{bmatrix} + O(\epsilon). \end{aligned}$$

Επειδή ο W είναι μπλοκ άνω τριγωνικός, όλες οι δυνάμεις του είναι επίσης μπλοκ άνω τριγωνικές και θεωρώντας την ανάπτυξη σε σειρά για το εκθετικό του πίνακα, οδηγούμαστε στη μορφή

$$e^{-Wt} = \begin{bmatrix} e^{-\bar{A}t} & Z(t) \\ 0 & e^{-(J+\epsilon YCX)t} \end{bmatrix},$$

για κάποιο κατάλληλο πίνακα $Z(t)$. Αντικατάσταση αυτής της έκφρασης στην προηγούμενη εξίσωση οδηγεί στο επιθυμητό αποτέλεσμα.

Βιβλιογραφία

- [1] S. Aalto, U. Ayesta, S. Borst, V. Misra, and R. Núñez-Queija. Beyond processor sharing. *ACM SIGMETRICS Performance Evaluation Review*, 34(4):36-43, 2007.
- [2] M. Alouini and A. Goldsmith. Area spectral efficiency of cellular mobile radio systems. *IEEE Transactions on Vehicular Technology*, 48(4):1047--1066, 1999.
- [3] A. Alsawah and I. Fijalkow. Resource allocation in OFDMA downlink with reduced feedback overhead. In *Proc. of IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, pages 1--5, Cannes, France, 2008.
- [4] G. Aniba and S. Aissa. Fast packet scheduling assuring fairness and quality of service in HSDPA. In *Proc. of the Canadian Conference on Electrical and Computer Engineering*, volume 4, pages 2243--2246, Niagara Falls Ontario, Canada, May 2004.
- [5] M. Assaad and D. Zeghlache. Analytical model of HSDPA throughput under Nakagami fading channel. *IEEE Transactions on Vehicular Technology*, 58(2):610--624, 2009.
- [6] F. Berggren and R. Jantti. Multiuser scheduling over Rayleigh fading channels. In *Proc. of Global Telecommunications Conference (IEEE GLOBECOM '03)*, volume 1, pages 158--162, San Francisco, U.S.A., 2003.
- [7] A. Berman and R. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, 1979.
- [8] T. Bonald. A score-based opportunistic scheduler for fading radio channels. In *Proc. of European Wireless Conference (EWC '04)*, pages 283--292, Barcelona, Spain, February 2004.

-
- [9] T. Bonald. Flow-level performance analysis of some opportunistic scheduling algorithms. *European Transactions on Telecommunications*, 16:65-75, 2005.
- [10] T. Bonald, S. Borst, N. Hegde, and A. Proutière. Wireless data performance in multi-cell scenarios. In *Proc. of ACM SIGMETRICS Performance Evaluation Review*, volume 32, pages 378--388, New York, U.S.A., 2004.
- [11] T. Bonald and A. Proutière. Insensitivity in processor-sharing networks. *Performance Evaluation*, 49:193--209, 2002.
- [12] T. Bonald and A. Proutière. Wireless downlink data channels: User performance and cell dimensioning. In *Proc. of the 9th Annual International Conference on Mobile Computing and Networking (ACM MOBICOM '03)*, pages 339--352, San Diego, U.S.A., 2003.
- [13] T. Bonald and A. Proutière. On stochastic bounds for monotonic processor sharing networks. *Queueing Systems*, 47:81--106, 2004.
- [14] S. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. In *Proc. of IEEE INFOCOM '03*, volume 1, pages 321--331, San Francisco, U.S.A., 2003.
- [15] S. Borst, O. Boxma, and N. Hegde. Sojourn times in finite-capacity processor sharing queues. In *Proc. of Next Generation Internet Networks (EURO NGI '05)*, pages 53--60, Rome, Italy, 2005.
- [16] S. Borst, O. Boxma, J. Morrison, and R. Núñez-Queija. The equivalence between processor sharing and service in random order. *Operations Research Letters*, 31(4):254--262, 2003.
- [17] S. Borst, K. Clarkson, J. Graybeal, H. Viswanathan, and P. Whiting. User-level QoS and traffic engineering for 3G wireless 1xEV-DO systems. *Bell Labs Technical Journal*, 8(2):33--47, 2003.
- [18] S. Borst, R. Núñez-Queija, and B. Zwart. Sojourn time asymptotics in processor sharing queues. *Queueing Systems*, 53(1):31--51, 2006.
- [19] O. Boxma, N. Hegde, and R. Núñez-Queija. Exact and approximate analysis of sojourn times in finite discriminatory processor sharing queues. *International Journal of Electronics and Communications*, 60:109--115, 2006.

-
- [20] R. Bulirsch and J. Stoer. *Introduction to Numerical Analysis*. Springer-Verlag, 1980.
- [21] C.-J. Chen and L.-C. Wang. A unified capacity analysis for wireless systems with joint antenna and multiuser diversity in Nakagami fading channels. *IEEE Transactions on Communications*, 54(3):469--478, 2006.
- [22] S. Cheung, Van den Berg, and R. Boucherie. Decomposing the queue length distribution of processor-sharing models into queue lengths of permanent customer queues. *Performance Evaluation*, 62(1-4):100--116, 2005.
- [23] S. Cheung, Van den Berg, and R. Boucherie. Insensitive bounds for the moments of the sojourn time distribution in the M/G/1 processor-sharing queue. *Queueing Systems*, 53:7--18, 2006.
- [24] E. Coffman, Jr., R. Muntz, and H. Trotter. Waiting time distributions for processor sharing systems. *Journal of the ACM (JACM)*, 17(1):123--130, 1970.
- [25] J. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica*, 12:245--284, 1979.
- [26] P. Courtois. *Decomposability - Queueing and Computer System Applications*. Academic Press, 1977.
- [27] Van den Berg and O. Boxma. The M/G/1 queue with processor sharing and its relation to a feedback queue. *Queueing Systems*, 9(4):365--401, 1991.
- [28] R. Egorova, M. Mandjes, and B. Zwart. Sojourn time asymptotics in processor sharing queues with varying service rate. *Queueing Systems*, 56(3):169--181, 2007.
- [29] R. Egorova, B. Zwart, and O. Boxma. Sojourn time tails in the M/D/1 processor sharing queue. *Probability in the Engineering and Informational Sciences*, 20(3):429--446, 2006.
- [30] H. Fu and D. Kim. Analysis of throughput and fairness with downlink scheduling in WCDMA networks. *IEEE Transactions on Wireless Communications*, 5(8):2164--2174, 2006.
- [31] A. Gyasi-Agyei and S. Kim. Comparison of opportunistic scheduling policies in time-slotted AMC wireless networks. In *Proc. of 1st International Symposium on Wireless Pervasive Computing (IEEE ISWPC '06)*, pages 1--6, Phuket, Thailand, January 2006.
-

-
- [32] J. Holtzman. Asymptotic analysis of proportional fair algorithm. In *Proc. of 12th International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE PIMRC '01)*, pages F--33 -- F--37, San Diego, U.S.A., 2001.
- [33] M. Hu, J. Zhang, and J. Sadowsky. Traffic aided opportunistic scheduling for wireless networks: algorithms and performance bounds. *Computer Networks*, 46(4):505--518, 2004.
- [34] B. Iyer, L. Donatiello, and P. Heidelberger. Analysis of performability for stochastic models of fault-tolerant systems. *IEEE Transactions on Computers*, 100(35):902--907, 1986.
- [35] K. Khawam, D. Kofman, and E. Altman. The weighted proportional fair scheduler. In *Proc. of the 3rd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (ACM QShine '06)*, page 43, Waterloo, Canada, August 2006.
- [36] L. Kleinrock. Time-shared systems: A theoretical treatment. *Journal of the ACM (JACM)*, 14(2):242--261, 1967.
- [37] R. Knopp and P. Humblet. Information capacity and power control in single-cell multiuser communications. In *Proc. of International Conference on Communications (IEEE ICC '95)*, volume 1, pages 331--335, Seattle, U.S.A., June 1995.
- [38] S. Kulkarni and C. Rosenberg. Opportunistic scheduling policies for wireless systems with short term fairness constraints. In *Proc. of Global Telecommunications Conference (IEEE GLOBECOM '03)*, volume 1, pages 533--537, San Francisco, U.S.A., 2003.
- [39] V. Kulkarni, V. Nicola, and K. Trivedi. On modelling the performance and reliability of multimode computer systems. *The Journal of Systems and Software*, 6(1-2):175--182, 1986.
- [40] X. Liu, E.K.P. Chong, and N.B. Shroff. A framework for opportunistic scheduling in wireless networks. *Computer Networks*, 41(4):451--474, 2003.
- [41] Y. Liu and E. Knightly. Opportunistic fair scheduling over multiple wireless channels. In *Proc. of IEEE INFOCOM' 03*, volume 2, pages 1106--1115, San Francisco, U.S.A., 2003.
- [42] M. Mandjes and B. Zwart. Large deviations of sojourn times in processor sharing queues. *Queueing Systems*, 52(4):237--250, 2006.

-
- [43] M. McGuire, K. Plataniotis, and A. Venetsanopoulos. Location of mobile terminals using time measurements and survey points. *IEEE Transactions on Vehicular Technology*, 52(4):999--1011, 2003.
- [44] R. Núñez-Queija. Sojourn times in non-homogeneous QBD processes with processor sharing. *Stochastic Models*, 17(1):61--92, 2001.
- [45] T. Ott. The sojourn-time distribution in the M/G/1 queue with processor sharing. *Journal of Applied Probability*, 21(2):360--378, 1984.
- [46] S. Patil and G. de Veciana. Managing resources and quality of service in heterogeneous wireless systems exploiting opportunism. *IEEE/ACM Transactions on Networking*, 15(5):1046--1058, 2007.
- [47] P. Puri. A method for studying the integral functionals of stochastic processes with applications: I. Markov chain case. *Journal of Applied Probability*, 8(2):331--343, 1971.
- [48] J. Roberts. A survey on statistical bandwidth sharing. *Computer Networks*, 45(3):319--332, 2004.
- [49] S. Ross. *Stochastic Processes*. Wiley New York, 1983.
- [50] R. Schassberger. A new approach to the M/G/1 processor-sharing queue. *Advances in Applied Probability*, 16(1):202--213, 1984.
- [51] R. Serfozo. *Introduction to Stochastic Networks*. Springer, 1999.
- [52] S. Shakkottai, T. Rappaport, and P. Karlsson. Cross-layer design for wireless networks. *IEEE Communications Magazine*, 41(10):74--80, 2003.
- [53] S. Shakkottai and A. Stolyar. Scheduling algorithms for a mixture of real time and non-real-time data in HDR. In *Proc. of 17th International Teletraffic Congress (ITC-17)*, pages 793--804, Salvador de Bahia, Brazil, September 2001.
- [54] M. Sharif and B. Hassibi. Delay considerations for opportunistic scheduling in broadcast fading channels. *IEEE Transactions on Wireless Communications*, 6(9):3353, 2007.
- [55] G. Song and Y. Li. Asymptotic throughput analysis for channel-aware scheduling. *IEEE Transactions on Communications*, 54(10):1827--1834, 2006.

-
- [56] G. Van Kessel, R. Núñez-Queija, and S. Borst. Differentiated bandwidth sharing with disparate flow sizes. In *Proc. of IEEE INFOCOM '05*, volume 4, pages 2425 -- 2435, Miami, U.S.A., 2005.
- [57] C. Van Loan. The sensitivity of the matrix exponential. *SIAM Journal on Numerical Analysis*, 14(6):971--981, 1977.
- [58] J. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, 1965.
- [59] L. Yang and M. Alouini. Performance analysis of multiuser selection diversity. In *Proc. of International Conference on Communications (IEEE ICC '04)*, volume 5, pages 3066--3070, Paris, France, 2004.
- [60] L. Yang, M. Kang, and M. Alouini. On the capacity-fairness tradeoff in multiuser diversity systems. *IEEE Transactions on Vehicular Technology*, 56(4):1901, 2007.
- [61] S. Yashkov. A derivation of response time distribution for a M/G/1 processor sharing queue. *Problems of Control and Information Theory*, 12:133--148, 1983.
- [62] S. Yashkov. Processor-sharing queues: Some progress in analysis. *Queueing Systems*, 2(1):1--17, 1987.
- [63] Y. Zhao. Mobile phone location determination and its impact on intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 1(1):55 -- 64, 2000.
- [64] Y. Zhao. Standardization of mobile phone positioning for 3G systems. *IEEE Communications Magazine*, 40(7):108 -- 116, 2002.

Γλωσσάρι

admission control έλεγχος αποδοχής. 32

aggregation συναθροιστικό. 5

balanced networks δίκτυα με ιδιότητα ισορροπίας. 43

Best Rate (BR) αλγόριθμος μέγιστου ρυθμού. 10

Complementary Probability Distribution Function (CPDF) συμπληρωματική συνάρτηση κατανομής πιθανότητας. 35

decomposition αποσύνθεση. 4

Egalitarian Processor Sharing (EPS) ισόνομος μερισμός εξυπηρητητή. 24

embedding ενσωμάτωση. 36

Exponential Rule εκθετικός αλγόριθμος. 130

fast dynamics bias τροποποίηση υπέρ των γρήγορων. 30

femtocell φεμτοκυψέλη. 130

forward equations εξισώσεις προς τα εμπρός. 36

Generalized Processor Sharing (GPS) γενικευμένος μερισμός εξυπηρητητή. 25

heavy tailed με βαριά ουρά. 26

indicator function δείκτρια συνάρτηση. 16

irreducible μη αναγώγιμο. 33

Laplace-Stieltjes Transform (LST) μετασχηματισμός Laplace-Stieltjes. 25

light-tailed με ελαφριά ουρά. 26

Line Of Sight (LOS) γραμμή οπτικής επαφής. 13

macrocell μακροκυψέλη. 130

Markov Reward process διαδικασία Ανταμοιβής Markov. 27

maximal μεγιστοτικό. 41

Modified-Largest Weighted Delay First (M-LWDF) τροποποιημένη μέγιστη καθυστέρηση. 130

multipath correlation συσχέτιση πολλαπλής διαδρομής. 118

multipath fading πολλαπλή σκέδαση. 13

multiuser diversity gain κέρδος διαφορισμού πολλαπλών χρηστών. 15

Nearly Completely Decomposable (NCD) σχεδόν πλήρως αποσυζεύξιμο. 4

Normalized Best Rate (NBR) αλγόριθμος κανονικοποιημένου μέγιστου ρυθμού. 10

opportunistic scheduling ευκαιριακός χρονοπρογραμματισμός. 1

phase-type τύπου-φάσης. 71

Processor Sharing (PS) μερισμός εξυπηρητητή. 3

Proportional Fair (PF) αναλογικά δίκαιος. 3

reduced service rate approximation προσέγγιση μειωμένου ρυθμού εξυπηρέτησης. 26

regularly varying ομαλά μεταβαλλόμενο. 26

round robin κυκλική εναλλαγή. 11

sample path δειγματική διαδρομή. 65

service pool θάλαμος εξυπηρέτησης. 5

short term fairness δικαιοσύνη μικρού χρόνου. 24

Signal to Noise Ratio (SNR) σηματοθορυβικός λόγος. 2

sojourn time χρόνος απόκρισης. 3

square-root-invariant subexponential-concave τετραγωνικής-ρίζας-αναλλοίωτο υποεκθετικό-κοίλο. 26

starvation period περίοδος αποκλεισμού. 11

State Dependent Processor Sharing (SDPS) μερισμός εξυπηρετητή εξαρτημένος από την κατάσταση. 4

stationary increments στάσιμες ελάχιστες μεταβολές. 42

stationary stochastic process στάσιμη στοχαστική διαδικασία. 13

steady state probability vector διάνυσμα πιθανοτήτων σταθερής κατάστασης. 33

Strength Group - Proportional Fair (SG-PF) Αναλογικά Δίκαιος - Κατηγορίας Ισχύος. 5, 95

supported bound φραγμένη υποστήριξη. 26

tail ουρά της κατανομής. 26

throughput ρυθμαπόδοση. 3

time-scale separation διαχωρισμός χρονικής κλίμακας. 4

timeslot χρονοσχισμή. 1

Συμβολισμοί Μεγεθών

\tilde{r}	μέσος επιτεύξιμος ρυθμός μετάδοσης
\tilde{R}	στιγμιαίος ρυθμός μετάδοσης
k	συνολικός αριθμός κλάσεων
\mathcal{K}	σύνολο των κλάσεων
\mathcal{K}_s	υποσύνολο των αργών κλάσεων
\mathcal{K}_f	υποσύνολο των γρήγορων κλάσεων
$\tilde{\gamma}$	μέσο SNR
n_i	αριθμός πελατών (ενεργών τερματικών) της κλάσης i
\mathbf{n}	διάνυσμα κατάστασης, $\mathbf{n} = (n_1, \dots, n_k)$
\mathbf{n}_s	διάνυσμα κατάστασης σύμμορφο με το \mathcal{K}_s
\mathbf{n}_f	διάνυσμα κατάστασης σύμμορφο με το \mathcal{K}_f
$f(\cdot)$	συνάρτηση πυκνότητας πιθανότητας
K	παράγοντας Rice
m	παράμετρος Nakagami
$\phi_i(\mathbf{n})$	ρυθμός εξυπηρέτησης πελάτη της κλάσης i στην κατάσταση \mathbf{n}
\mathbf{I}_A	δείκτρια συνάρτηση του γεγονότος \mathcal{A}
$F(\cdot)$	συνάρτηση κατανομής πιθανότητας

$ \mathbf{n} $	άθροισμα στοιχείων του διανύσματος \mathbf{n}
ϵ	παράγοντας σύζευξης διαχωρισμού χρονικής κλίμακας
\tilde{r}	μέσος επιτεύξιμος ρυθμός μετάδοσης ανεξάρτητος του ϵ
$\lambda_i(\mathbf{n})$	μέσος ρυθμός αφίξεων της κλάσης i στην κατάσταση \mathbf{n}
$1/\mu_i$	μέση απαίτηση εξυπηρέτησης της κλάσης i εκπεφρασμένη σε μονάδες χρόνου
\mathcal{S}	χώρος καταστάσεων
\mathcal{C}_i	σύνολο των συνοριακών καταστάσεων της κλάσης i
L	μέγεθος θαλάμου εξυπηρέτησης
M	πίνακας ρυθμών μεταβάσεων καταστάσεων, $M = [M(\mathbf{n}, \mathbf{n}')]]$
π	διάνυσμα πιθανοτήτων σταθερής κατάστασης
$z^{(l)}$	το μέγεθος z που αντιστοιχεί στο τροποποιημένο σύστημα που περιέχει l μόνιμους πελάτες
$\mathbf{0}$	διάνυσμα κατάλληλου μεγέθους με όλα τα στοιχεία ίσα με μηδέν
$\mathbf{1}$	διάνυσμα κατάλληλου μεγέθους με όλα τα στοιχεία ίσα με μονάδα
\mathbf{e}_i	διάνυσμα κατάλληλου μεγέθους με όλα τα στοιχεία ίσα με μηδέν και μονάδα στη θέση i
$z^{(i)}$	το μέγεθος $z^{(e_i)}$
$Q_i(t)$	ποσό εξυπηρέτησης σε πελάτη κλάσης i στο χρονικό διάστημα $(0, t]$
ν_i	διάνυσμα αρχικών πιθανοτήτων της κλάσης i
$\lambda_{av,i}$	μέσος ρυθμός αφίξεων πελατών της κλάσης i που εισέρχονται στο σύστημα

$P_{B,i}$	πιθανότητα αποκλεισμού για πελάτη της κλάσης i
V_i	χρόνος απόκρισης πελάτη κλάσης i
$S_i(\cdot)$	CPDF του χρόνου απόκρισης ενός πελάτη κλάσης i
T_i	αρχική απαίτηση εξυπηρέτησης πελάτη κλάσης i
$\sigma_1(A)$	ιδιοτιμή του A με το μικρότερο πραγματικό μέρος
$\text{Re}\{\cdot\}$	πραγματικό μέρος
$\Phi(\cdot)$	συνάρτηση ισορροπίας
ρ_i	ένταση κίνησης της κλάσης i
\bar{z}	το μέγεθος z που αντιστοιχεί στο συναθροιστικό σύστημα
$\hat{z}^{(l)}$	το μέγεθος z που αντιστοιχεί στο τροποποιημένο σύστημα που περιέχει l μόνιμους αργούς πελάτες και μειωμένο χώρο καταστάσεων που περιλαμβάνει μόνο γρήγορους πελάτες
z^*	το μέγεθος z με επιπλέον τροποποίηση υπέρ των γρήγορων κλάσεων
δ	παράμετρος σύζευξης ρυθμών εξυπηρέτησης με ρυθμούς εξυπηρέτησης προνομιακούς για τις γρήγορες κλάσεις
$\ A\ $	νόρμα του πίνακα A
$\lfloor z \rfloor$	ο μεγαλύτερος φυσικός αριθμός μικρότερος από το z
Σ_i	πίνακας μεταβάσεων μεταξύ φάσεων της κλάσης i
\tilde{v}	μέση τιμή απαιτήσεων εξυπηρέτησης σε μονάδες δεδομένων
τ	ρυθμαπόδοση
$\mathcal{S}_{[z]}$	σύνολο διανυσμάτων κατάστασης με άθροισμα ενεργών τερματικών ίσο με z

$\eta(l; \rho, L)$

πιθανότητα σταθερής κατάστασης να υπάρχουν l συνολικά ενεργά τερματικά, σε σύστημα με συνολική ένταση κίνησης ίση με ρ και κοινό θάλαμο εξυπηρέτησης μεγέθους L

$\text{card}(\hat{\mathcal{S}})$

πληθικός αριθμός του συνόλου \mathcal{S}

Ευρετήριο

- BR, 10, 95, 113, 127
 διαχωρισμός χρονικής κλίμακας, 19, 89
 κέρδη διαφορισμού πολλαπλών χρηστών, 89
- Chebyshev
 ανισότητα, 135
- Exponential Rule, 130
- femtocell, 130
- HDR, 2, 23
- HSDPA, 2
- Jordan
 μορφή, 41, 74, 136, 137
 οποπίνακας, 74, 136
- Laplace-Stieltjes
 μετασχηματισμός, 25, 36
- Little
 νόμος του, 25, 39, 79, 87
- M-LWDF, 130
- macrocell, 130
- Markov
 αλυσίδα διακριτού χρόνου, 36
 αλυσίδα με μόνιμους πελάτες, 34
 αλυσίδα συνεχούς χρόνου, 33, 35, 79, 90
- διαδικασία ανταμοιβής, 27, 28, 35, 38, 66, 128
- maxSNR, 3
- NBR, 10, 85, 95, 113, 127
 διαχωρισμός χρονικής κλίμακας, 88
 κέρδη διαφορισμού πολλαπλών χρηστών, 17
- Nearly Completely Decomposable, 4, 29, 50, 54, 56, 128
- PASTA, 37
- Processor Sharing, 3, 24
 Discriminatory, 25, 101
 Egalitarian, 24, 106
 Generalized, 25
 State Dependent, 4, 25, 28, 51, 101, 127
- Proportional Fair, 3, 9
- SG-PF, 5, 95, 113
 κέρδη διαφορισμού πολλαπλών χρηστών, 97
- Signal to Noise Ratio, 2, 9, 12, 13, 19
- απαιτήσεις εξυπηρέτησης
 βαριάς ουράς, 26
 ελαφριάς ουράς, 26
 κατανομή, 31, 77
 τύπου-φάσης, 71
 μέση τιμή, 31, 77
 αποσύνθεση, 4, 50, 81

αφίξεις Poisson, 25, 31, 37, 52, 60, 101
 διάνυσμα
 αρχικών πιθανοτήτων, 36, 38, 53
 κατάστασης, 12, 31, 65, 86, 104
 πιθανοτήτων σταθερής κατάστασης,
 33, 34, 51, 63, 79, 81, 85, 88
 διαχωρισμός χρονικής κλίμακας, 4, 27,
 47, 80
 ένταση κίνησης, 43, 48, 86, 87, 97
 θάλαμος εξυπηρέτησης, 5, 32, 86, 91, 96
 ιδιότητα
 ισορροπίας, 4, 25, 28, 43, 85, 128
 μονοτονίας, 30, 31
 κέρδος διαφορισμού πολλαπλών
 χρηστών, 15, 31, 78
 μειωμένο υποσύστημα, 49, 52, 55, 61,
 76, 88, 97, 101, 102
 μόνιμος πελάτης, 4, 27, 33, 35, 127
 παράγοντας σύζευξης, 48
 πιθανότητα
 αποκλεισμού, 32, 79, 84, 86
 σταθερής κατάστασης, 27, 42, 43, 52,
 75
 πίνακας
 εκθετικό, 28
 μη αναγώγιμος, 33, 34, 39, 40, 51,
 63
 μικρότερη ιδιοτιμή, 39, 40
 φράγματα, 41
 νόρμα, 75, 136
 ομαλός, 39, 40
 ρυθμών μεταβάσεων, 33, 50, 51, 63,
 71, 79
 συναθροιστικός, 51, 81, 88, 92
 τύπου M, 4, 39, 128
 ρυθμαπόδοση, 3, 78, 82, 87, 113
 ρυθμός εξυπηρέτησης
 συνάρτηση, 31
 ρυθμός μετάδοσης
 μέσος, 10, 78
 στιγμιαίος, 10
 ρυθμός φθίσης
 κατανομής του χρόνου απόκρισης,
 26, 41, 80
 συναρτήσεων σφάλματος, 20
 σκέδαση
 Nakagami, 13
 Rayleigh, 13
 Rice, 13
 πολλαπλή, 13
 στάσιμη στοχαστική διαδικασία, 13
 συνάθροιση, 4, 50, 76
 συναθροιστικό σύστημα SDPS, 47, 57
 συνάρτηση
 ισορροπίας, 43, 85
 στοχαστικά αύξουσα, 133
 σύνολο κλάσεων, 12, 31
 τροποποίηση υπέρ των γρήγορων
 κλάσεων, 30, 59
 χρόνος απόκρισης, 3
 κατανομή, 3, 37, 79, 101
 αργών κλάσεων, 57, 84
 γρήγορων κλάσεων, 54, 82, 91
 τροποποιημένου συστήματος, 61
 μέσος, 83, 88, 97, 113

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

ΠΡΟΣΩΠΙΚΑ ΣΤΟΙΧΕΙΑ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ: Γιαννουλάκης Ιωάννης

ΗΜΕΡΟΜΗΝΙΑ ΓΕΝΝΗΣΕΩΣ: 12 Δεκεμβρίου 1979

ΔΙΕΥΘΥΝΣΗ ΚΑΤΟΙΚΙΑΣ: Μέτωνος 64, 15561 Χολαργός, Αθήνα

ΤΗΛΕΦΩΝΟ ΚΑΤΟΙΚΙΑΣ: 210 6544841

ΗΛ. ΤΑΧΥΔΡΟΜΕΙΟ: giannoul@iit.demokritos.gr

ΥΠΗΚΟΟΤΗΤΑ: Ελληνική

ΟΙΚΟΓΕΝΕΙΑΚΗ ΚΑΤΑΣΤΑΣΗ: Έγγαμος

ΑΚΑΔΗΜΑΪΚΕΣ ΣΠΟΥΔΕΣ

- Πτυχιούχος του τμήματος Φυσικής Πανεπιστημίου Ιωαννίνων. (1997 - 2001)
Βαθμός πτυχίου: «Λίαν Καλώς» 7,12
- Υποτροφία επιστημονικής συνεργασίας τετραετούς διάρκειας παράλληλα με την εκπόνηση της διδακτορικής διατριβής από το Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών του Ε.Κ.Ε.Φ.Ε «ΔΗΜΟΚΡΙΤΟΣ». (2002)
- Αποδοχή σε θέση υποψήφιου διδάκτορα από το τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. (2003)

ΣΥΓΓΡΑΦΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ

- Εκπόνηση πτυχιακής εργασίας με θέμα: «Τηλεχειρισμός μέσω τηλεφώνου». Κατά την πτυχιακή εργασία σχεδιάστηκε το ηλεκτρονικό κύκλωμα και κατασκευάστηκε εξ' ολοκλήρου στο εργαστήριο του πανεπιστημίου. (2001)

ΔΕΥΤΕΡΟΒΑΘΜΙΑ ΕΚΠΑΙΔΕΥΣΗ

- 1ο Λύκειο Χολαργού. Βαθμός απολυτηρίου: 17,8/20 (1994 - 1997)

ΕΠΑΓΓΕΛΜΑΤΙΚΗ ΕΜΠΕΙΡΙΑ

- Business College Administration (B.C.A.), Αθήνα. (2009)
Μερική απασχόληση ως καθηγητής πληροφορικής.
- Φροντιστήριο «Ορίζοντες», Χολαργός. (2003-2005)
Μερική απασχόληση ως καθηγητής μαθηματικών (Προετοιμασία των υποψηφίων για διαγωνισμό του Α.Σ.Ε.Π.)
- Συμμετοχή σε συγγραφή προτάσεων και παραδοτέων ερευνητικών έργων με συμμετοχή του Ε.Κ.Ε.Φ.Ε «ΔΗΜΟΚΡΙΤΟΣ»

ΞΕΝΕΣ ΓΛΩΣΣΕΣ - ΑΛΛΕΣ ΕΞΕΤΑΣΕΙΣ

- Αγγλικά: Certificate of Proficiency in English (2001)
- Συμμετοχή στις εξετάσεις του γενικού G.R.E. (2003)
Μαθηματικά: 780
Λογική: 790

ΔΙΑΚΡΙΣΕΙΣ / ΒΡΑΒΕΙΑ

- Υποτροφία από το Ίδρυμα Κρατικών Υποτροφιών (Ι.Κ.Υ.) για την κατάληψη της τρίτης κατά σειρά θέσης στη γενική βαθμολογία μεταξύ των φοιτητών του πρώτου έτους σπουδών. (1998)
- Υποτροφία επιστημονικής συνεργασίας τετραετούς διάρκειας παράλληλα με την εκπόνηση της διδακτορικής διατριβής, από το Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών του Ε.Κ.Ε.Φ.Ε «ΔΗΜΟΚΡΙΤΟΣ». (2002)

ΕΙΔΙΚΕΣ ΓΝΩΣΕΙΣ

- Χειρισμός Η/Υ (Win, Office, Latex), Προγραμματισμός Η/Υ (C, FORTRAN, Matlab, Mathematica)

ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ

- Αξιολόγηση άρθρων ως κριτής (peer reviewer) σε διεθνή επιστημονικά περιοδικά και συνέδρια.
- Μέλος της IEEE (Institute of Electrical and Electronics Engineers) (2005)

ΔΗΜΟΣΙΕΥΣΕΙΣ

- I. Giannoulakis, K. Kontovasilis, N. Mitrou, **Sojourn time distributions and time scale separation results for multiclass state-dependent processor sharing systems with a finite service pool and exponential service requirements**, *Performance Evaluation*, 68 (7) (2011): 586-613, doi:10.1016/j.peva.2011.03.008
- I. Giannoulakis, K. Kontovasilis, N. Mitrou, **Balancing the performance fairness tradeoff for wireless network access through channel-aware opportunistic schedulers under heterogeneous radio conditions**, *Tech. Report, 2012, submitted for publication.*
- I. Giannoulakis, K. Kontovasilis, N. Mitrou, **Performance, Dimensioning and Interference Tradeoffs for Two-tier Wireless Networks**, *accepted for publication in IEEE PIMRC 2013, London, UK.*