Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Ισοδυναμία και Αντικατάσταση Κυκλοφορίας σε Πολυπλέκτη: Θεωρία και Εφαρμογές

Αντώνης Δημάκης

Μεταπτυχιακή Εργασία

Ηράκλειο, Φεβρουάριος 1999

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Ισοδυναμία και Αντικατάσταση Κυκλοφορίας σε Πολυπλέκτη: Θεωρία και Εφαρμογές

Εργασία που υποβλήθηκε από τον
ΔΗΜΑΚΗ ΑΝΤΩΝΙΟ
ως μερική εκπλήρωση των απαιτήσεων
για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Αντώνης Δημάκης
Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Εισηγητική Επιτροπή:

Κώστας Κουρκουμπέτης
Καθηγητής

Απόστολος Τραγανίτης
Αναπληρωτής Καθηγητής

Γεώργιος Δ. Σταμούλης
Επίκουρος Καθηγητής

Δεκτή:

Πάνος Κωνσταντόπουλος
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Φεβρουάριος 1999

. . . στον Παντελή Μπάμπα που έχει πάει βόλτα. . . .

.

Ισοδυναμία και Αντικατάσταση Κυκλοφορίας σε Πολυπλέκτη: Θεωρία και Εφαρμογές

Αντώνης Δημάκης

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Περίληψη

Για έναν πολυπλέκτη ο οποίος τροφοδοτείται από μεγάλο αριθμό πηγών, εξάγουμε συνθήκες σύμφωνα με τις οποίες μία πηγή μπορεί να αντικαταστήσει ένα υποσύνολο των πηγών διατηρώντας ταυτόχρονα την τιμή της πιθανότητας υπερχείλισης του ενταμιευτή και τις πιο πιθανές χρονικές κλίμακες στις οποίες υπερχειλίζει ο ενταμιευτής. Η ισοδυναμία αυτή είναι ισχυρότερη από απλή ισότητα των ισοδυνάμων ευρών ζώνης, γιατί λαμβάνει υπόψιν και ιδιότητες του περιβάλλοντος στο οποίο γίνεται η πολυπλεξία, βάσει του σημείου λειτουργίας. Αυτό επιτρέπει την αντικατάσταση αυθαίρετα μεγάλου αριθμού πηγών κυκλοφορίας, χωρίς να αλλάζει το σημείο λειτουργίας του πολυπλέκτη όπως αυτό γίνεται αντιληπτό από τις υπόλοιπες πηγές. Η προτεινόμενη μεθοδολογία δεν βασίζεται σε μοντέλα κυκλοφορίας ενώ η απαραίτητη πληροφορία λαμβάνεται από αρχεία-ίχνη πραγματικής κυκλοφορίας. Μελετάμε την χρήση της κλασματικής κίνησης Brown για υποκατάσταση πραγματικών πηγών και παραθέτουμε θεωρητικά και πειραματικά αποτελέσματα τα οποία επιβεβαιώνουν την χρησιμότητα της μεθοδολογίας. Επιπλέον προτείνουμε μια διαφορετική εφαρμογή της μεθόδου αντικατάστασης, η οποία μπορεί να χρησιμοποιηθεί για τον ταχύ υπολογισμό του σημείου λειτουργίας του πολυπλέκτη και κατά συνέπεια για την εκτίμηση της πιθανότητας υπερχείλισης του ενταμιευτή και του ισοδυνάμου εύρους ζώνης, σε πραγματικό χρόνο. Πειραματικά δείχνουμε ότι η χρήση τέτοιων εκτιμήσεων είναι εφικτή στην πράξη σε μηχανισμούς ελέγχου κυκλοφορίας ενώ εξετάζεται η επίδοση ενός απλού αλγόριθμου ελέγχου αποδοχής κλήσεων ο οποίος βασίζεται στους εκτιμητές αυτούς.

Επόπτης Καθηγητής: Κώστας Κουρκουμπέτης
Καθηγητής
Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Traffic Equivalence and Substitution in a Multiplexer: Theory and Applications

Antonis Dimakis

Master of Science Thesis

Computer Science Department
University of Crete

Abstract

For a multiplexer fed with a large number of sources, we derive conditions under which a single source can be substituted for a given subset of the sources while preserving the buffer overflow probability and the dominant time scales of buffer overflows. This equivalence is stronger than simple effective bandwidth equality, because it takes into account the context in which multiplexing takes place, on the basis of the notion of operating point. This allows a substitution to be made for arbitrarily large proportions of the traffic without changing the operating point of the multiplexer as experienced by the rest of the traffic. It corresponds to defining a single source which is equivalent in a sense ‘‘local’’ to a given context, rather than equivalent in a sense which is ‘‘universal’’ to all contexts. The proposed methodology does not rely on traffic models and obtains the necessary information from the actual traffic traces. We study the case of fractional Brownian motion as a single source substitute and provide theoretical and experimental results that validate the approach. Furthermore we propose a different application of the substitution approach, which can be used towards the fast calculation of the operating point parameters of the multiplexer, and hence for the on-line estimation of the buffer overflow probability and the effective bandwidth of the traffic. Experimental results indicate that these estimators can be used for real-time traffic control while we evaluate the performance of a simple call admission control algorithm based on such estimators.

Supervisor: Costas Courcoubetis
Professor
Computer Science Department
University of Crete

Ευχαριστίες

Καταρχάς θα ήθελα να ευχαριστήσω τον επόπτη μου Κώστα Κουρκουμπέτη για την καθοδήγηση που μου παρείχε, για τις ιδέες του που αποτελούσαν μια διαρκή πηγή έμπνευσης για μένα καθώς και για την αγάπη που μου ενέπνευσε για το αντικείμενο. Επίσης ευχαριστώ τον Γιώργο Σταμούλη για την συνεχή καθοδήγησή του, για τις πολύτιμες γνώσεις που μου έδωσε σε αντικείμενα στα οποία δεν είχα υπόβαθρο, όπως επίσης και για τα σχόλια και συμβουλές του όπου μέσα από αυτά έμαθα πώς να εργάζομαι.

Ευχαριστώ επίσης τα μέλη του τμήματος Δικτύων και Τηλεπικοινωνιών του Ινστιτούτου Πληροφορικής στο Ίδρυμα Τεχνολογίας και Έρευνας, για την αμέριστη συμπαράσταση -υλική και ψυχική- που μου παρείχαν. Ειδικότερα ευχαριστώ τον Βασίλη Σύρρη για τον κώδικα καθώς και για τις τεχνικές συμβουλές που με υπομονή διαρκώς παρείχε.

Ευχαριστώ τους φίλους μου Θανάση Χρυσό, Χρήστο Μανωλάκη, Κώστα Αναγνωστάκη, και Φώτη Χαρμαντζή για την παρέα τους και τη συμβολή τους στην εργασία αυτή με πολλούς και διάφορους τρόπους.

Η εργασία αυτή δε θα μπορούσε να υπάρξει χωρίς τη συμπαράσταση της Νίκης η οποία τράβηξε αρκετά · υπόσχομαι να μην το ξανακάνω. Επίσης ευχαριστώ τους Μάνο, Μαρία και Παναγιώτη για όλα.

Τέλος, θα ήθελα να ευχαριστήσω την πλειοψηφία της Επιτροπής Ερευνών που χωρίς τη βοήθειά της η εργασία αυτή θα είχε εκπληρωθεί κάτω από ανθρώπινες συνθήκες.

Περιεχόμενα

Περίληψη **iii**

Abstract **v**

Ευχαριστίες **vii**

Περιεχόμενα **ix**

Κατάλογος Σχημάτων **xi**

1 Εισαγωγή **1**

1.1 Ευρυζώνια Τηλεπικοινωνιακά Δίκτυα - Υπηρεσίες ATM 1

1.2 Μοντέλα για την Κυκλοφορία - Επισκόπηση Βιβλιογραφίας 3

1.3 Προσφορά και Οργάνωση της Εργασίας 6

2 Υπόβαθρο: *Ασυμπτωτικό Πολλών Πηγών* **9**

2.1 Το Βασικό Μοντέλο 9

2.2 Ισοδύναμο Εύρος Ζώνης και το Ασυμπτωτικό Πολλών Πηγών 10

2.3 Ισοδύναμο Εύρος Ζώνης Πραγματικής Κυκλοφορίας 13

3 Ισοδυναμία και Αντικατάσταση Κυκλοφορίας **17**

3.1 Η Συνθήκη Ισοδυναμίας 17

3.2 Αντικατάσταση με πηγές Gaussian 21

3.3 Υψηλό Κέρδος Στατιστικής Πολυπλεξίας 21

3.4 Η Περίπτωση $B = 0$ 22

3.5 Κλασματική Κίνηση Brown 24

4 Πειραματική Αξιολόγηση της Μεθόδου Αντικατάστασης **29**

4.1 Ακρίβεια Αντικατάστασης 30

4.2 Χρονική Κλίμακα Υπερχείλισης 31

4.3 Αντικατάσταση βάσει μόνο του Ισοδυνάμου Εύρους Ζώνης 32

5 Υπολογισμός της Πιθανότητας Υπερχείλισης σε Πραγματικό Χρόνο 35

5.1 Ταχύς Υπολογισμός του Σημείου Λειτουργίας 35

5.2 Υπολογισμός της Πιθανότητας Υπερχείλισης σε ένα Σύνδεσμο ATM 38

5.3 Πειραματικά Αποτελέσματα 40

5.3.1 Πηγές ON/OFF Markov Modulated Fluids 40

5.3.2 Πραγματικές Πηγές Κυκλοφορίας 42

6 Ένας Απλός Αλγόριθμος CAC 47

6.1 Το Μοντέλο Κλήσεων 48

6.2 Αλγόριθμος CAC 48

6.3 Βέλτιστος αλγόριθμος CAC 49

6.4 Απλοί αλγόριθμοι CAC βασιζόμενοι σε πηγές ON/OFF 50

6.5 Πειραματικά αποτελέσματα 51

7 Επίλογος 55

7.1 Συμπεράσματα 55

7.2 Μελλοντικές Κατευθύνσεις 57

Βιβλιογραφία 59

Κατάλογος Σχημάτων

1.1 Ο ρυθμός μετάδοσης σε συνάρτηση του χρόνου, της ταινίας video Star Wars κωδικοποιημένης σύμφωνα με το πρότυπο MPEG-1. 3

2.1 Το βασικό μοντέλο ενός πολυπλέκτη ATM. 10

2.2 Ο ρυθμός μετάδοσης πληροφορίας ως συνάρτηση του χρόνου. Το μοντέλο αυτό έχει ανακατασκευαστεί από ένα αρχείο-ίχνος το οποίο προέρχεται από ένα μικρό τμήμα της ταινίας video Star Wars η οποία έχει κωδικοποιηθεί σύμφωνα με το πρότυπο MPEG-1. 14

2.3 Η συνάρτηση ισοδυνάμου εύρους ζώνης $\alpha(s,t)$ της ταινίας video Star Wars η οποία έχει κωδικοποιηθεί σύμφωνα με το πρότυπο MPEG-1. 15

3.1 Η πιθανότητα υπερχείλισης συνδέσμου χωρητικότητας 155 Mbps χωρίς ενταμιευτή, ως συνάρτηση του αριθμού των πηγών Star Wars που τροφοδοτούν τον σύνδεσμο. . . . 24

3.2 Η Φ_R ως συνάρτηση του μεγέθους B του ενταμιευτή για διάφορες πηγές MPEG-1. Σε κάθε περίπτωση η κυκλοφορία που τροφοδοτεί τον σύνδεσμο (με χωρητικότητα 155 Mbps), προέρχεται από υπέρθεση πολλών όμοιων αλλά ανεξάρτητων ταινιών video. Σε όλα τα πειράματα, ο βαθμός χρησιμοποίησης του συνδέσμου ήταν 88%. 28

4.1 Η BOP για διάφορα μοντέλα που χρησιμοποιήθηκαν για αντικατάσταση σε σύγκριση με την BOP των πραγματικών πηγών για σύνδεσμο 34 Mbps (α) και 155 Mbps (β), ως συνάρτηση του αριθμού των πηγών Star Wars των οποίων η υπέρθεση συνιστά την πραγματική κυκλοφορία εισόδου. 31

4.2 Η BOP που επιτυγχάνεται όταν αντικαθίστανται μέρος των πηγών εισόδου από πηγές βασισμένες σε μοντέλα fBm. Η BOP της αρχικής κυκλοφορίας αντιστοιχεί σε ‘‘αντικατάσταση’’ του 0% της κυκλοφορίας. 31

4.3 Απεικόνιση του πιθανότερου χρόνου στον οποίο γεμίζει ο ενταμιευτής, για την πραγματική κυκλοφορία και πηγές MMF και fBm οι οποίες αντικαθιστούν τις πραγματικές σε σύνδεσμο 34 Mbps με ενταμιευτή μεγέθους 500 κελιά. 32

4.4 Σύγκριση της BOP όταν δύο μοντέλα ON/OFF MMF αντικαθιστούν την πραγματική κυκλοφορία. Το ένα από αυτά βασίζεται στην έννοια της ισοδυναμίας κυκλοφορίας ενώ το άλλο βασίζεται σε ισότητα μόνο των ισοδυνάμων ευρών ζώνης και δε λαμβάνει υπόψιν τις συνθήκες πρώτης τάξεως. 33

5.1 Τα βήματα τα οποία ακολούθησε η μέθοδος γρήγορου υπολογισμού του σημείου λειτουργίας. 510 πηγές Star Wars πολυπλέκονται σε σύνδεσμο (C =) 155 Mbps με ενταμιευτή μεγέθους (B =) 2000 κελιών. Απεικονίζονται οι ισοσταθμικές καμπύλες της συνάρτησης $G_R(s,t)$ για τη συνολική πραγματική κυκλοφορία. Ενδεικτικά ελήφθησαν δύο σημεία εκκίνησης, το $s = 0.01$, $t = 300$ και το $s = 0.0001$, $t = 110$. Είναι φανερό ότι η σύγκλιση είναι γρήγορη και για τα δύο αρχικά σημεία. Το σημείο λειτουργίας του συστήματος είναι το $s = 0.006, t = 168$ ('1' στο σχήμα), ενώ υπάρχουν άλλα δύο σταθερά (σαγματικά) σημεία, το $s = 0.0042, t = 248$ και το $s = 0.0097, t = 82$ ('2' και '3' στο σχήμα). Η μέθοδος συνέκλινε στο σωστό σημείο λειτουργίας και για τα δύο σημεία εκκίνησης του αλγορίθμου. 37

5.2 Επίδοση της μεθόδου εκτίμησης σε πραγματικό χρόνο της BOP (a)-(b), του ισοδυνάμου εύρους ζώνης (c)-(d), της παραμέτρου χρόνου του συστήματος (e)-(f) και της παραμέτρου χώρου του συστήματος (g)-(h) για πηγές ON/OFF MMF. 41

5.3 Επίδοση της μεθόδου εκτίμησης σε πραγματικό χρόνο της BOP (a)-(b), του ισοδύναμου εύρους ζώνης (c)-(d), της παραμέτρου χρόνου του συστήματος (e)-(f) και της παραμέτρου χώρου του συστήματος (g)-(h) για πραγματική κυκλοφορία (MPEG-1). 43

5.4 Επίδοση της μεθόδου εκτίμησης σε πραγματικό χρόνο της BOP. (a) Οι δειγματικοί μέσοι των εκτιμήσεων ως συνάρτηση του χρόνου και (b) η εμπειρική κατανομή των εκτιμήσεων τη χρονική στιγμή t=100 sec. 45

5.5 (a) Οι δειγματικές τιμές του εκτιμητή της BOP $\hat{\Psi}_n$ ως συνάρτηση του χρόνου, για 2 ανεξάρτητα πειράματα. (b) Οι δειγματικές τιμές των εκτιμητών $\hat{\Psi}_n$ και $\hat{\Phi}_n$ ως συνάρτηση του χρόνου. Με 'actual BOP' απεικονίζεται η πραγματική BOP. 46

6.1 Η κατανομή κλήσεων στον πολυπλέκτη: (i) για τον βέλτιστο αλγόριθμο που βασίζεται σε πηγές ON/OFF και σε σύστημα χωρίς ενταμιευτή ('optimal ON/OFF-based CAC'), (ii) για τον αλγόριθμο που προτείνουμε ('CAC algorithm') και (iii) για τον βέλτιστο αλγόριθμο CAC ('optimal CAC'). 52

6.2 Η πιθανότητα υπερχείλισης του ενταμιευτή (BOP), για κάθε αριθμό κλήσεων. Απεικονίζεται η στάσιμη BOP μέσω της προσέγγισης Bahadur-Rao καθώς και η BOP που παρατηρήθηκε στην προσομοίωση. 53

Κεφάλαιο 1

Εισαγωγή

1.1 Ευρυζώνια Τηλεπικοινωνιακά Δίκτυα - Υπηρεσίες ATM

Ο Ασύγχρονος Τρόπος Μετάδοσης (ATM) και η εξέλιξη της τεχνολογίας των οπτικών μέσων μετάδοσης έχουν κάνει δυνατή την προσφορά πολλών διαφορετικών τηλεπικοινωνιακών υπηρεσιών μέσω ενός φυσικού δικτύου. Οι υπηρεσίες αυτές μπορεί να είναι πολύ απλές όπως ηλεκτρονικό ταχυδρομείο μέχρι και πολύπλοκες όπως διανομή τηλεόρασης υψηλής ευκρίνειας (HDTV). Για τη μετάδοση όλων αυτών των διαφορετικών υπηρεσιών έχουν προταθεί πέντε κλάσεις υπηρεσίας, βλ. [5], στις οποίες αντιστοιχίζονται οι υπηρεσίες αυτές οι οποίες θα μπορούσαν να χωριστούν σε δύο γενικές κατηγορίες: τις υπηρεσίες *εγγυημένης ποιότητας*, στην οποία ανήκουν οι κλάσεις CBR, rt-VBR και nrt-VBR και τις *ελαστικές* που είναι οι κλάσεις ABR και UBR.

Οι κλάσεις rt-VBR και nrt-VBR, με τις οποίες θα ασχοληθούμε στην παρούσα εργασία, δίνουν τη δυνατότητα στις εφαρμογές να μεταδοθούν διασφαλίζοντας ότι θα ικανοποιούνται οι απαιτήσεις τους σε ποιότητα υπηρεσίας (Quality of Service - QoS). Η ποιότητα υπηρεσίας εκφράζεται από μέγιστα φράγματα στην καθυστέρηση μεταφοράς (Cell Transfer Delay - CTD), στη μεταβλητότητα της καθυστέρησης μεταφοράς (Cell Delay Variation - CDV) και στην πιθανότητα απώλειας της πληροφορίας (Cell Loss Probability - CLP). Για την εξασφάλιση τέτοιων απαιτήσεων σε QoS, το δίκτυο πρέπει να αποδώσει τους πόρους του (χώρο ενταμίευσης και ταχύτητα μετάδοσης) στις εφαρμογές που το χρησιμοποιούν με έναν ελεγχόμενο τρόπο. Για το σκοπό αυτό η χρήση των πόρων του δικτύου από τις εφαρμογές ελέγχεται από τους μηχανισμούς *Ελέγχου Κυκλοφορίας και Συμφόρησης* [19] του δικτύου. Οι μηχανισμοί αυτοί ενεργούν ως εξής:

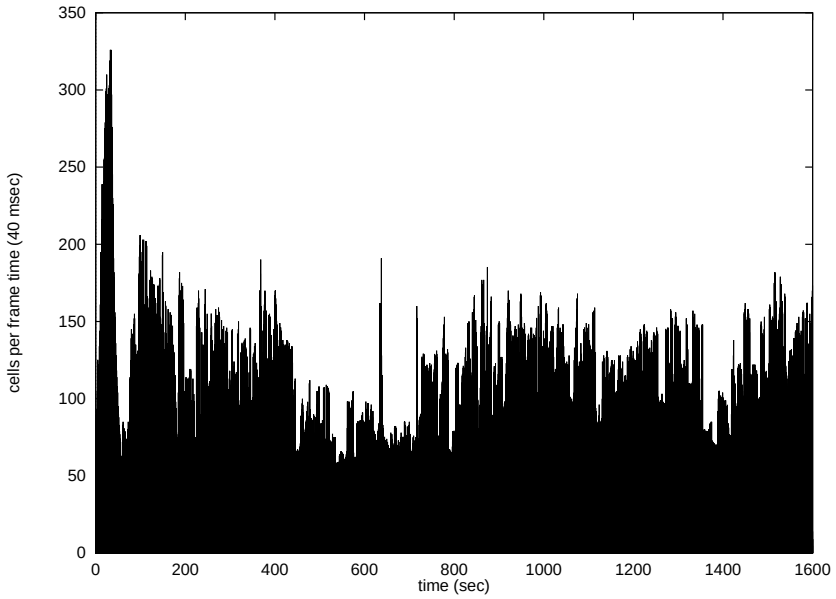
- Κατά την έναρξη της σύνδεσης, οι μηχανισμοί ενημερώνονται για τη συμπεριφορά της κυκλοφορίας που μεταδίδει η εφαρμογή και τις απαιτήσεις σε QoS μέσω του *Συμβολαίου*

Κυκλοφορίας.

- Κατά τη διάρκεια της σύνδεσης, οι μηχανισμοί ελέγχουν την τήρηση της συμπεριφοράς της εφαρμογής που συμφωνήθηκε στο Συμβόλαιο Κυκλοφορίας. Εάν υπάρχει απόκλιση από τη συμπεριφορά αυτή, μπορεί το δίκτυο να προκαλέσει απώλεια του μη συμμορφούμενου μέρους της πληροφορίας που μεταδίδει η εφαρμογή έτσι ώστε να ικανοποιηθούν τα Συμβόλαια Κυκλοφορίας και των υπολοίπων εφαρμογών που χρησιμοποιούν το δίκτυο.

Η απαίτηση σε πόρους από μια συγκεκριμένη εφαρμογή, μπορεί να μην είναι σταθερή με το χρόνο αλλά μεταβαλλόμενη. Ακόμα η εφαρμογή αυτή μπορεί να θέτει απαιτήσεις σε QoS τις οποίες το δίκτυο πρέπει να ικανοποιήσει. Αυτό ισχύει κυρίως για τις εφαρμογές που μεταδίδονται μέσω VBR στις οποίες εστιάζουμε στην παρούσα εργασία. Ο βέλτιστος σχεδιασμός και λειτουργία κατάλληλων μηχανισμών Ελέγχου Κυκλοφορίας και Συμφόρησης του δικτύου έτσι ώστε να μπορούν να ανταπεξέλθουν σε απαιτήσεις εφαρμογών αυτού του τύπου είναι ένα ανοιχτό ερευνητικό πεδίο. Το πρόβλημα που αντιμετωπίζουν οι μηχανισμοί αυτοί είναι ότι ο τρόπος με τον οποίο μεταβάλλονται οι απαιτήσεις σε πόρους μιας εφαρμογής, είναι άγνωστος στο δίκτυο. Έτσι, μια βέλτιστη απόδοση πόρων πρέπει ταυτόχρονα να ικανοποιεί τις απαιτήσεις σε QoS των εφαρμογών, χωρίς να γίνεται με συντηρητικό τρόπο. Η ανάγκη για αποδοτικούς μηχανισμούς Ελέγχου Κυκλοφορίας και Συμφόρησης ενδυναμώνεται λόγω του κέρδους πολυπλεξίας που παρουσιάζεται στα δίκτυα ATM. Τα κέρδη αυτά οφείλονται κυρίως στο ότι τον περισσότερο χρόνο οι εφαρμογές μεταδίδουν πληροφορία με πολύ μικρότερο ρυθμό από τον μέγιστο που πρόκειται να χρησιμοποιήσουν κάποια χρονική στιγμή. Για παράδειγμα, στο Σχήμα 1.1 απεικονίζεται η εξέλιξη του ρυθμού μετάδοσης μιας εφαρμογής κωδικοποιημένης εικόνας video. Συγκεκριμένα πρόκειται για κωδικοποίηση σύμφωνα με το πρότυπο MPEG, το οποίο αναμένεται να κυριαρχήσει στη χρήση των υπηρεσιών rt-VBR και nrt-VBR. Έστω ότι υπάρχει απαίτηση CLP$<10^{-6}$. Μια απλή στρατηγική χορήγησης πόρων θα ήταν η δέσμευση εύρους ζώνης ίσο με το μέγιστο ρυθμό μετάδοσης του Σχήματος 1.1 με αποτέλεσμα να μπορούν να εξυπηρετηθούν λίγες εφαρμογές. Για παράδειγμα εάν η χωρητικότητα ενός πολυπλέκτη είναι $C = 34$ Mbps, βάσει της στρατηγικής αυτής θα εξυπηρετούνταν μόνο 9 πηγές. Αντίθετα, το εύρος ζώνης που πραγματικά είναι αρκετό για τη μετάδοση των εφαρμογών και την ταυτόχρονη ικανοποίηση της απαίτησης CLP$<10^{-6}$, είναι πολύ μικρότερο από αυτό που δεσμεύεται με αποτέλεσμα, να μπορούσε να εξυπηρετηθεί μεγαλύτερο πλήθος εφαρμογών. Συγκεκριμένα, για $C = 34$ Mbps θα μπορούσαν να εξυπηρετηθούν 93 πηγές.

Από τα παραπάνω είναι φανερό ότι οι σχεδιαστές και οι μηχανικοί δικτύων καλούνται να απαντήσουν και να δώσουν λύσεις σε δύσκολα ερωτήματα τα οποία αφορούν την



Σχήμα 1.1: Ο ρυθμός μετάδοσης σε συνάρτηση του χρόνου, της ταινίας video Star Wars κωδικοποιημένης σύμφωνα με το πρότυπο MPEG-1.

πρόβλεψη της επίδοσης των δικτυακών συστημάτων έτσι ώστε να αποφεύγονται τόσο τα φαινόμενα συμφόρησης τα οποία οδηγούν σε υπερχείλιση των ενταμιευτών και άρα σε απώλεια κυκλοφορίας, όσο και η υποεκμετάλλευση των πόρων λόγω συντηρητικής πολιτικής απόδοσής τους. Για την αντιμετώπιση τέτοιων φαινομένων, χρησιμοποιούνται μαθηματικά μοντέλα για την πραγματική κυκλοφορία μέσω των οποίων είναι δυνατή η ανάλυση επίδοσης των τηλεπικοινωνιακών συστημάτων.

1.2 Μοντέλα για την Κυκλοφορία - Επισκόπηση Βιβλιογραφίας

Για την ανάλυση της επίδοσης ενός δικτύου χρησιμοποιούνται στοχαστικά μοντέλα τα οποία επιτρέπουν είτε την χρήση γνωστών μαθηματικών τεχνικών της θεωρίας ουρών, είτε την προσομοίωση και τον πειραματισμό.

Η κυκλοφορία μιας εφαρμογής, μοντελοποιείται σαν μια στοχαστική διαδικασία. Για παράδειγμα η $\{X_t, t \in T\}$ μπορεί να αντιπροσωπεύει τον φόρτο που εισάγει μια πηγή σε έναν πολυπλέκτη σε συνάρτηση με τον χρόνο. Εάν $T = Z$ τότε έχουμε μοντέλο διακριτού χρόνου, όπου η διαδικασία μπορεί να εκφράζει τον φόρτο που εισάγει μια πηγή στο υπ'αριθμών t χρονικό διάστημα. Εάν $T = R^+$ έχουμε μοντέλο συνεχούς χρόνου και η διαδικασία μπορεί να εκφράζει τον φόρτο στο χρονικό διάστημα $[0, t]$. Ο φόρτος μπορεί να εκφράζεται σε bytes, κελιά ATM κτλ. Είναι σύνηθες η διαδικασία $\{X_t\}$ να θεωρείται ότι έχει στάσιμες αυξήσεις,δηλαδή ότι ο τυχαίος φόρτος ίσης διάρκειας χρονικών διαστημάτων ακολουθεί την ίδια κατανομή.

Ο πιο συνηθισμένος τρόπος περιγραφής μιας στοχαστικής διαδικασίας είναι μέσω των ροπών της, και ιδιαίτερα των ροπών πρώτης και δεύτερης τάξης. Οι ροπές δεύτερης τάξης είναι η διασπορά και συνδιασπορά των τ.μ. $\{X_t, t \in T\}$. Οι συσχετίσεις δεύτερης τάξης (μεταξύ δύο χρονικών στιγμών) αυτών των τ.μ., αποτυπώνονται στη συνδιασπορά τους, όπου μοντελοποιείται και η εκρηκτικότητα των πραγματικών πηγών.

Έχει διαπιστωθεί, μέσω πειραματισμού και ανάλυσης, ότι σε πολλά απλά μοντέλα πηγών, τα χαρακτηριστικά εκείνα των πηγών που καθορίζουν την απόδοση της πολυπλεξίας τους σε έναν ενταμιευτή είναι οι δύο πρώτες ροπές ([16, 24, 17, 21]). Ή με άλλα λόγια, πηγές κυκλοφορίας με ίδιες δύο πρώτες ροπές, επιφέρουν σχεδόν την ίδια επίδοση εάν τροφοδοτήσουν τον ίδιο πολυπλέκτη. Στο γεγονός αυτό βασίζονται στη συντριπτική τους πλειοψηφία, οι προσεγγίσεις στη μοντελοποίηση κυκλοφορίας. Έτσι, αντί να μελετηθεί η επίδοση ενός πολυπλέκτη κάτω από πραγματική κυκλοφορία, μελετάται η επίδοσή του κάτω από απλά μοντέλα κυκλοφορίας, τα οποία έχουν (κατά το δυνατόν) ίδιες ροπές πρώτης και δεύτερης τάξης με την πραγματική κυκλοφορία. Η επίδοση ενός πολυπλέκτη μελετάται με τη χρήση αναλυτικών μεθόδων, των οποίων η μορφή και πολυπλοκότητα εξαρτάται από τα συγκεκριμένα μοντέλα που χρησιμοποιήθηκαν. Χρησιμοποιούνται απλά μοντέλα, έτσι ώστε τόσο να υπάρχουν αναλυτικές μέθοδοι επίλυσης τους, όσο και οι μέθοδοι αυτές να είναι υπολογιστικά δυνατές.

Πολλές φορές, δεν χρησιμοποιούνται οι ροπές αλλά κάποια άλλα στατιστικά μεγέθη. Τέτοια μπορεί να είναι ο λόγος μέγιστου ρυθμού κυκλοφορίας προς τον μέσο, ο μέσος χρόνος στον οποίο η κυκλοφορία έχει ρυθμό μεταξύ κάποιων τιμών, ο *δείκτης διασποράς των σημείων άφιξης* (index of dispersion of counts - IDC) κτλ. Ανεξάρτητα από ποιά συγκεκριμένα στατιστικά μεγέθη χρησιμοποιούνται κάθε φορά, οι προσεγγίσεις αυτές αποσκοπούν στην κατασκευή ενός μοντέλου κυκλοφορίας ‘‘καθολικής ισχύος’’. Δηλαδή, κάποιου μοντέλου το οποίο μπορεί να αντικαταστήσει την πραγματική κυκλοφορία (όσον αφορά την επίδοση) σε ένα οποιονδήποτε πολυπλέκτη, και με την παρουσία οποιασδήποτε άλλης πηγής κυκλοφορίας η οποία πολυπλέκεται μαζί.

Με βάση τα παραπάνω, το έργο της μοντελοποίησης κυκλοφορίας μπορεί να διαιρεθεί σε δύο μέρη: (α) την εκτίμηση διαφόρων στατιστικών μεγεθών της πραγματική κυκλοφορίας και (β) την επιλογή παραμέτρων των μοντέλων κυκλοφορίας έτσι ώστε να υπάρχει συμφωνία στα αντίστοιχα στατιστικά μεγέθη. Βασική προϋπόθεση για το (β) είναι να έχει επιλεγεί κατάλληλα το μοντέλο το οποίο πρόκειται να αντικαταστήσει την πραγματική κυκλοφορία. Δηλαδή, πρέπει οι παράμετροι που το χαρακτηρίζουν να παρέχουν αρκετή ευελιξία στη μοντελοποίηση της πραγματικής κυκλοφορίας. Πολλά μοντέλα έχουν προταθεί για διαφορετικό τύπο πραγματικής κυκλοφορίας, π.χ. συχνά χρησιμοποιούμενα μοντέλα

είναι [20, 12, 30, 27]:

- Αριθμός κατειλλημένων κυκλωμάτων (μοντέλο κλήσεων): μοντέλο Poisson
- Πηγές φωνής PCM: πηγές ON/OFF, Interrupted Poisson Process (IPP)
- Υπέρθεση πηγών φωνής: υπέρθεση ON/OFF Markov-Modulated Fluids (MMF), Markov-Modulated Poisson Process (MMPP)
- Πηγές video μεταβλητού ρυθμού: MMPP, Autoregressive Moving Average (ARMA),
- Πηγές δεδομένων: Αυτοπαρεμφερείς διαδικασίες π.χ. κλασματική κίνηση Brown (fractional Brownian motion - fBm), διαδικασίες Pareto-Modulated.

Οι προσεγγίσεις αυτές, είναι προσανατολισμένες στα χαρακτηριστικά των πηγών κυκλοφορίας και στην ουσία αγνοούν το ρόλο που μπορεί να παίξει στην επίδοση η συνολική σύνθεση της κυκλοφορίας, η χωρητικότητα και το μέγεθος ενταμιευτή ενός πολυπλέκτη. Στο Κεφάλαιο 2 θα δούμε ότι τα χαρακτηριστικά των πηγών που διαμορφώνουν την επίδοση εξαρτώνται από τα χαρακτηριστικά του πολυπλέκτη. Επιπλέον, μέσω του *ασυμπτωτικού πολλών πηγών* θα δούμε μια σχέση η οποία συνδέει τα χαρακτηριστικά των πηγών και του πολυπλέκτη με την επίδοση του συστήματος.

Δυστυχώς, τα απλά μοντέλα κυκλοφορίας που χρησιμοποιούνται στις παραπάνω προσεγγίσεις συχνά αποτυγχάνουν να αποδώσουν όλα τα σημαντικά χαρακτηριστικά της πραγματικής κυκλοφορίας. Για παράδειγμα, οι μηχανισμοί κωδικοποίησης που χρησιμοποιούνται για κυκλοφορία η οποία προέρχεται από εφαρμογές πολυμέσων (π.χ. MPEG), προκαλούν ιδιοτυπίες στη δομή της κυκλοφορίας και εκρηκτικότητα σε πολλές διακριτές χρονικές κλίμακες η οποία εξαρτάται από τη συγκεκριμένη δομή της ακολουθίας πλαισίων. Η δομή αυτή πιστεύεται ότι επηρεάζει σημαντικά την απόδοση των τηλεπικοινωνιακών συστημάτων, όπως την πιθανότητα υπερχείλισης των ενταμιευτών, ιδιαίτερα σε μικρού μεγέθους ενταμιευτές οι οποίοι χρησιμοποιούνται σε συστήματα τα οποία μεταφέρουν πληροφορία πραγματικού χρόνου.

Η εναλλακτική προσέγγιση στην ανάλυση επίδοσης μέσω μοντέλων πραγματικής κυκλοφορίας είναι η προσομοίωση. Σύμφωνα με την προσέγγιση αυτή, δημιουργείται τεχνητή κυκλοφορία η οποία μιμείται όσο το δυνατόν περισσότερο την πραγματική. Στη συνέχεια, η τεχνητή αυτή κυκλοφορία εισάγεται στον πολυπλέκτη τον οποίο παρακολουθούμε και εξετάζουμε την επίδοσή του. Το κύριο πρόβλημα αυτής της προσέγγισης είναι η δημιουργία τεχνητής κυκλοφορίας η οποία πρέπει να μιμείται πραγματική κυκλοφορία προερχόμενη από ένα μεγάλο αριθμό πηγών (εκατοντάδες ή χιλιάδες). Μια λύση η οποία ακολουθείται στην πράξη είναι η χρήση εξειδικευμένου υλικού το οποίο είναι προγραμματισμένο αποκλειστικά

να μιμείται τις πηγές σε πραγματικό χρόνο. Η δημοφιλής τακτική είναι οι πηγές να περιγράφονται με μηχανές καταστάσεων οι οποίες βασίζονται σε μοντέλα διαδικασιών Poisson οι οποίες είναι διαμορφωμένες σύμφωνα με διαδικασίες Markov (Markov Modulated Poisson Processes - MMPP). Αν και το κόστος αυτού του εξειδικευμένου υλικού είναι συχνά υψηλό, υπάρχουν ερωτηματικά ως προς την καταλληλότητα των μοντέλων MMPP για κυκλοφορία προερχόμενη από εφαρμογές πολυμέσων[1].

1.3 Προσφορά και Οργάνωση της Εργασίας

Στην παρούσα εργασία αναλύουμε μια νέα προσέγγιση στη μοντελοποίηση κυκλοφορίας χρησιμοποιώντας μια μόνο πηγή ισοδύναμη με πολλές πραγματικές πηγές. Η πηγή αυτή μπορεί να αντικαταστήσει ένα σύνολο πραγματικών πηγών χωρίς να διαταράσσει τα κυρίαρχα φαινόμενα που προκαλούν την υπερχείλιση των ενταμιευτών. Η προτεινόμενη μεθοδολογία μπορεί να χρησιμοποιεί μετρήσεις προερχόμενες από πραγματική κυκλοφορία χωρίς να προϋποθέτει κάποιο συγκεκριμένο μοντέλο.

Η έννοια της ισοδυναμίας κυκλοφορίας όπως τη χρησιμοποιούμε εδώ, ορίσθηκε στο [1], όπου περιλαμβάνονται και λίγα πρώτα αποτελέσματα από όπου διαφαίνεται η χρησιμότητα της έννοιας αυτής στη μοντελοποίηση κυκλοφορίας. Αν και ο συγγραφέας στο [1] δεν επεκτείνεται πολύ στην έννοια αυτή, εντούτοις έχουν σχηματιστεί οι ιδέες οι οποίες επιτρέπουν τον ορισμό ενός πλαισίου αντικατάστασης το οποίο περιγράφουμε στην παρούσα εργασία.

Η σημασία του σημείου λειτουργίας στην ανάλυση ενός πολυπλέκτη αποτυπώνεται και στο [9]. Συγκεκριμένα, μεταξύ διαφόρων αποτελεσμάτων οι συγγραφείς εξάγουν ότι εάν οι ροές κυκλοφορίας, οι ενταμιευτές και οι χωρητικότητες των συνδέσμων πολλών ετερογενών πολυπλεκτών αθροιστούν, τότε το επίπεδο ποιότητας υπηρεσίας που απαιτεί κάθε ροή εξακολουθεί να προσφέρεται με την προϋπόθεση ότι οι ροές (πριν αθροιστούν) ‘‘ταιριάζουν’’, δηλ. έχουν το ίδιο σημείο λειτουργίας. (Επίσης δίδεται ένα αντιπαράδειγμα για την αντίθετη περίπτωση.) Συνεπώς, το ασυμπτωτικό πολλών πηγών εφαρμόζεται και στο [9] και στην παρούσα εργασία, και επομένως η συμπεριφορά της κυκλοφορίας αναλύεται μόνο στις χρονικές κλίμακες εκείνες που δίδονται από το σημείο λειτουργίας. Όμως, η παρούσα εργασία εστιάζει σε άλλη κατεύθυνση, αφού ‘‘κατασκευάζουμε’’ μια νέα ροή κυκλοφορίας η οποία μπορεί να αντικαταστήσει την αθροιστική κυκλοφορία ανεξάρτητα του εάν αυτή προέρχεται από ροές που ‘‘ταιριάζουν’’ ή όχι. Στο Κεφάλαιο 3 θα αποδείξουμε ότι η αντικατάσταση την οποία χρησιμοποιούμε διατηρεί την ποιότητα υπηρεσίας εάν οι δύο ροές

[1] Λόγω της μη πεπερασμένης διασποράς των χρόνων παραμονής στις καταστάσεις [21], η οποία δεν επιτυγχάνεται με το υλικό αυτό.

''ταιριάζουν'' και έχουν ισότητα των ισοδυνάμων ευρών ζώνης τους.

Η καινοτομία της προσέγγισης έγκειται στο γεγονός ότι κατασκευάζει ένα ''τοπικό'' μοντέλο για ένα σύνολο πηγών, αντί ενός ''καθολικού''. Μέχρι τώρα έχει γίνει κατανοητό ότι δε συμβάλλουν όλα τα χαρακτηριστικά της κυκλοφορίας των πηγών εξίσου στα φαινόμενα τα οποία προκαλούν την υπερχείλιση στους ενταμιευτές [7, 33]. Ανάλογα με το μέγεθος του ενταμιευτή και τη σύνθεση των διαφορετικών τύπων κυκλοφορίας που χρησιμοποιούν ένα σύνδεσμο, διαφέρει η χρονική κλίμακα εκρηκτικότητας που κυρίως ευθύνεται για την υπερχείλιση. Το κίνητρο για την προσέγγισή μας είναι το εξής: προσεγγίζουμε την εκρηκτική συμπεριφορά των πηγών, μόνο στη σχετική με την υπερχείλιση χρονική κλίμακα. Για να το καταφέρουμε, χρησιμοποιούμε το *ασυμπτωτικό πολλών πηγών* (many sources asymptotic) το οποίο μας δίδει τον κυρίαρχο όρο της πιθανότητας απώλειας κελιού, όπως και τη χρονική κλίμακα στην οποία η συνεισφορά των πηγών στην υπερχείλιση είναι μέγιστη. Το αποτέλεσμα αυτό βασίζεται σε αποτελέσματα της θεωρίας *μεγάλων αποκλίσεων* (large deviations). Στο Κεφάλαιο 2, παρέχουμε μια εισαγωγή στο ασυμπτωτικό πολλών πηγών και στο ισοδύναμο εύρος ζώνης το οποίο εμπλέκεται στο ασυμπτωτικό, καθώς και ένα τρόπο υπολογισμού αυτών των μεγεθών βάσει αρχείων-ιχνών πραγματικής κυκλοφορίας.

Τα ασυμπτωτικά αυτά αποτελέσματα μας επιτρέπουν να ορίσουμε τυπικά ένα πλαίσιο, μέσα στο οποίο μπορούμε να ορίσουμε την ισοδυναμία κυκλοφορίας: μία πηγή μπορεί να αντικαταστήσει ένα σύνολο πηγών, εάν κατά την αντικατάσταση διατηρούνται ορισμένες σημαντικές ιδιότητες του όρου μεγάλων αποκλίσεων που δίδεται από το ασυμπτωτικό πολλών πηγών. Στο Κεφάλαιο 3 ορίζονται αυστηρά οι συνθήκες της ισοδυναμίας κυκλοφορίας, καθώς και μελετώνται διάφορα ζητήματα αντικατάστασης και κοινώς χρησιμοποιούμενα μοντέλα. Ειδικότερα, εξετάζουμε τις συνθήκες της αντικατάστασης σε συστήματα με μεγάλο βαθμό πολυπλεξίας και σε συστήματα χωρίς ενταμιευτή. Επιπλέον εξετάζουμε κατά πόσο μπορεί να χρησιμοποιηθεί η κλασματική κίνηση Brown (fractional Brownian motion - fBm) στην αντικατάσταση πολλών πηγών video. Θα δούμε ότι, κάτω από μη περιοριστικές συνθήκες για την πραγματική κυκλοφορία, μπορούμε πάντα να εξάγουμε ένα ισοδύναμο μοντέλο fBm.

Στο Κεφάλαιο 4, μελετάμε την απόδοση της μεθόδου αντικατάστασης όπου αντικαθιστούμε μεγάλα σύνολα πηγών video, κωδικοποιημένων σύμφωνα με το πρότυπο MPEG-1. Θα δούμε ότι τα πειράματα αυτά επαληθεύουν τη μέθοδό μας, παρέχοντας καλύτερες προσεγγίσεις της πραγματικής επίδοσης από αυτές που προσφέρουν άλλες κλασικές μέθοδοι. Επίσης υπογραμμίζεται η σημασία του ρόλου του σημείου λειτουργίας στην πιθανότητα υπερχείλισης ενός πολυπλέκτη, γεγονός το οποίο απλοποιεί εξαιρετικά την ανάλυση της απόδοσης και επιβεβαιώνει ξανά την ιδέα στην οποία βασίζεται η μέθοδος αντικατάστασης

που προτείνουμε.

Στο Κεφάλαιο 5 χρησιμοποιούμε τη μεθοδολογία που προτείνουμε, για ταχύ υπολογισμό της πιθανότητας υπερχείλισης του ενταμιευτή, καθώς και άλλων μεγεθών, όπως το σημείο λειτουργίας ενός πολυπλέκτη και το ισοδύναμο εύρος ζώνης των πηγών. Στη συνέχεια χρησιμοποιούμε τα αποτελέσματα αυτά, για την εκτίμηση των ποσοτήτων αυτών σε πραγματικό χρόνο. Επίσης αξιολογούμε πειραματικά τη μέθοδο χρησιμοποιώντας κυκλοφορία μοντέλων Markov, καθώς και πραγματική από πηγές video, όπου εξάγουμε ταχείς και συνεπής εκτιμητές γεγονός το οποίο επιτρέπει την εφαρμογή της μεθόδου αυτής στην πράξη.

Στο Κεφάλαιο 6 χρησιμοποιούμε τη μέθοδο εκτίμησης της πιθανότητας υπερχείλισης και του σημείου λειτουργίας ενός πολυπλέκτη σε πραγματικό χρόνο του Κεφαλαίου 5, ως τη βάση ενός απλού αλγόριθμου για *Έλεγχο Αποδοχής Κλήσεων* (Call Admission Control - CAC). Συγκρίνουμε την επίδοσή του με αυτήν του βέλτιστου αλγορίθμου CAC όπως και με την επίδοση του βέλτιστου αλγορίθμου που χρησιμοποιεί μοντέλα κλήσεων ON/OFF. Αλγόριθμοι CAC βασισμένοι σε μοντέλα ON/OFF, έχουν εκτενώς μελετηθεί στη βιβλιογραφία και έχουν προσφέρει στην κατανόηση των διαφόρων φαινομένων που καθορίζουν την επίδοση ενός αλγόριθμου CAC. Η επίδοση αυτών των αλγορίθμων για την περίπτωση πραγματικής (MPEG) κυκλοφορίας είναι εξαιρετικά χαμηλή ενώ απλοί αλγόριθμοι όπως αυτός που προτείνουμε μπορούν να επιτύχουν υψηλό βαθμό χρησιμοποίησης σε έναν πολυπλέκτη διατηρώντας ταυτόχρονα την πιθανότητα υπερχείλισης κάτω από μια μέγιστη αποδεκτή τιμή.

Τέλος στο Κεφάλαιο 6, κλείνουμε την εργασία με ένα απολογισμό τον αποτελεσμάτων και της πιθανής χρήσης τους. Επίσης δίδουμε κατευθύνσεις για μελλοντική έρευνα.

Κεφάλαιο 2

Υπόβαθρο: *Ασυμπτωτικό Πολλών Πηγών*

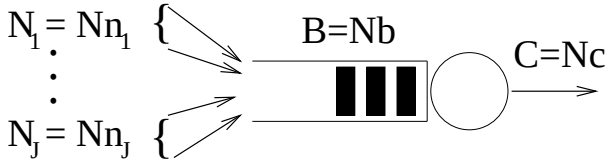
Στο κεφάλαιο αυτό παρουσιάζουμε ορισμένα βασικά αποτελέσματα της θεωρίας πολυπλεξίας. Στην Παράγραφο 2.1 περιγράφουμε το βασικό μοντέλο ενός πολυπλέκτη. Το μοντέλο αυτό είναι απλό αφού σκοπός μας είναι να επιδείξουμε τις βασικές έννοιες, όμως εύκολα μπορεί να επεκταθεί ώστε να περιλαμβάνει προτεραιότητες, βλ. [22]. Στην Παράγραφο 2.2 δίδουμε τον ορισμό του *ισοδυνάμου εύρους ζώνης* μιας πηγής και παρουσιάζουμε το *ασυμπτωτικό πολλών πηγών*. Στην τελευταία παράγραφο, περιγράφουμε τον τρόπο με τον οποίο μπορούμε να υπολογίσουμε το ισοδύναμο εύρος ζώνης μιας πραγματικής πηγής.

2.1 Το Βασικό Μοντέλο

Για να μπορέσουμε να αναλύσουμε μαθηματικά τα φαινόμενα σε ένα τηλεπικοινωνιακό δίκτυο, χρησιμοποιούμε τη θεωρία των δικτύων ουρών. Ένα δίκτυο ουρών αποτελείται από ένα δίκτυο το οποίο διατρέχουν πελάτες οι οποίοι απαιτούν κάποιο είδος επεξεργασίας από τους εξυπηρετητές στους κόμβους. Τυπικά, οι εξυπηρετητές έχουν πεπερασμένη χωρητικότητα και οι πελάτες μπορεί να χρειαστεί να αναμένουν σε μια ουρά περιμένοντας να εξυπηρετηθούν. Το κλασικό πρόβλημα στη θεωρία ουρών είναι η ανάλυση της συμπεριφοράς ενός μόνο συστατικού στοιχείου του δικτύου: ουρά με ένα μοναδικό εξυπηρετητή.

Η διαδικασία αφίξεων σε έναν ευρυζώνιο σύνδεσμο ATM θεωρούμε ότι προέρχεται από την υπέρθεση πηγών J διαφορετικών τύπων. Έστω ότι έχουμε $N_j = N n_j$ πηγές του τύπου j και $n = (n_1, \ldots, n_J)$ (οι συνιστώσες n_j δεν είναι απαραίτητα ακέραιοι αριθμοί). Μπορούμε να θεωρήσουμε ότι το σύστημα αυτό έχει N πηγές του ίδιου τύπου, όπου κάθε τέτοια πηγή προέρχεται από ανάμειξη μιας συγκεκριμένης αναλογίας των J τύπων πηγών και μπορεί

να χαρακτηρισθεί από το διάνυσμα n. Ο ευρυζώνιος σύνδεσμος έχει ένα διαμοιραζόμενο ενταμιευτή μεγέθους $B = Nb$ και χωρητικότητα $C = Nc$. Η παράμετρος N είναι η παράμετρος κλίμακας (μέγεθος του συστήματος) και οι παράμετροι b, c είναι ο ενταμιευτής και η χωρητικότητα ανά πηγή αντίστοιχα. Το μοντέλο αυτό απεικονίζεται στο Σχήμα 2.1.



Σχήμα 2.1: Το βασικό μοντέλο ενός πολυπλέκτη ATM.

2.2 Ισοδύναμο Εύρος Ζώνης και το Ασυμπτωτικό Πολλών Πηγών

$X_j[0,t]$ είναι ο συνολικός φόρτος ο οποίος παράγει μια πηγή του τύπου j στο χρονικό διάστημα $(0,t]$. Θεωρούμε ότι ο φόρτος $X_j[0,t]$ έχει στάσιμες αυξήσεις. Το *ισοδύναμο εύρος ζώνης* μιας πηγής τύπου j ορίζεται ως [22]

$$\alpha_j(s,t) = \frac{1}{st}\log \mathrm{E}\left[e^{sX_j[0,t]}\right] \ , \tag{2.1}$$

όπου s, t είναι παράμετροι του συστήματος οι οποίες ορίζονται βάσει του περιβάλλοντος των πηγών, δηλαδή τα χαρακτηριστικά της κυκλοφορίας η οποία πολυπλέκεται, η απαίτηση σε QoS και οι πόροι του συνδέσμου (χωρητικότητα και χώρος ενταμίευσης). Το ζεύγος (s,t) καλείται σημείο λειτουργίας του συστήματος. Ειδικότερα η *χρονική* παράμετρος t (η οποία μετράται σε μονάδες χρόνου π.χ. msec) αντιστοιχεί στην πιο πιθανή διάρκεια ενεργής περιόδου του ενταμιευτή πριν συμβεί υπερχείλιση (δηλ. ο χρόνος στον οποίο γεμίζει ο ενταμιευτής). Η *χωρική* παράμετρος s (η οποία μετράται π.χ. σε μονάδες kb^{-1}) αντιστοιχεί στο βαθμό πολυπλεξίας του συνδέσμου και εξαρτάται μεταξύ άλλων, από το μέγεθος του μέγιστου ρυθμού μετάδοσης των πηγών σχετικά με τη χωρητικότητα του συνδέσμου. Το ισοδύναμο εύρος ζώνης είναι αύξουσα συνάρτηση του s. Ειδικότερα για χωρητικότητες συνδέσμου πολύ μεγαλύτερες από τον μέγιστο ρυθμό μετάδοσης των πηγών, το s τείνει στο μηδέν και το $\alpha_j(s,t)$ συγκλίνει στον μέσο ρυθμό μετάδοσης της πηγής, ενώ για χωρητικότητες συνδέσμου όχι πολύ μεγαλύτερες από τον μέγιστο ρυθμό μετάδοσης των πηγών, το s μεγαλώνει και το $\alpha_j(s,t)$ συγκλίνει στη μέγιστη τιμή την οποία μπορεί να λάβει η τυχαία μεταβλητή $\frac{X_j[0,t]}{t}$.

Έστω $Q(Nc, Nb, Nn) = \mathrm{P(overflow)}$ η πιθανότητα ότι σε έναν ενταμιευτή άπειρου μεγέθους στον οποίο πολυπλέκονται $Nn = (Nn_1, \ldots, Nn_j)$ πηγές και εξυπηρετούνται με ρυθμό $C = Nc$ το μέγεθος της ουράς είναι πάνω από $B = Nb$. Για το $Q(Nc, Nb, Nn)$ ισχύει

το εξής:

$$\lim_{N\to\infty}\frac{1}{N}\log Q(Nc,Nb,Nn) = -\inf_t\sup_s\left[s(ct+b)-st\sum_{j=1}^{J}n_j\alpha_j(s,t)\right] = -I\,, \qquad (2.2)$$

όπου το I ονομάζεται η ασυμπτωτική *συνάρτηση ρυθμού*. Η τελευταία ισότητα αναφέρεται ως το *ασυμπτωτικό πολλών πηγών* (ή η παράσταση `infsup`) και έχει αποδειχθεί ότι ισχύει για διακριτό χρόνο στο [8], για συνεχή χρόνο στο [4] και για μια ειδική περίπτωση στο [34]. Λόγω της εξίσωσης (2.2), η πιθανότητα υπερχείλισης μπορεί να γραφεί $\mathrm{P(overflow)} = e^{-NI+o(N)}$, οπότε για μεγάλο N οδηγούμαστε στην εξής προσέγγιση:

$$\log \mathrm{P(overflow)} \approx -\inf_t\sup_s\left[s(B+Ct)-st\alpha_R(s,t)\right] = -\Phi_R\,, \qquad (2.3)$$

όπου $\alpha_R(s,t)$ είναι η συνάρτηση του ισοδυνάμου εύρους ζώνης της συνολικής εισερχόμενης κυκλοφορίας (λόγω της ανεξαρτησίας των εισερχόμενων πηγών, ισούται με το άθροισμα των ισοδυνάμων ευρών ζώνης όλων των πηγών). Το Φ_R μπορούμε να πούμε ότι είναι η *σταθμισμένη συνάρτηση ασυμπτωτικού ρυθμού*.

Εάν υπάρχει μια απαίτηση σε QoS του τύπου $\mathrm{P(overflow)} \le e^{-\gamma}$ τότε η περιοχή αποδοχής (τα σημεία n στα οποία η παραπάνω απαίτηση ικανοποιείται) για μεγάλο N προσεγγίζεται από ένα γραμμικό ώς προς N σύνορο [7]

$$\sum_{j=1}^{J}N_j\alpha_j(s^*,t^*) \le C+\frac{1}{t^*}\left(B-\frac{\gamma}{s^*}\right) = C^*\,, \qquad (2.4)$$

όπου (s^*,t^*) είναι ένα σημείο το οποίο βελτιστοποιεί την εξίσωση (2.2) και C^* είναι η ''ισοδύναμη χωρητικότητα'' του συστήματος στο παραπάνω σημείο λειτουργίας (s^*,t^*). Το σημείο αυτό είναι *σημείο σάγματος* της απεικόνισης $(s,t)\mapsto[s(B+Ct)-st\alpha_R(s,t)]$. Δηλαδή, οι παράγωγοι της συνάρτησης αυτής ως προς s και t μηδενίζονται στο (s^*,t^*), ενώ η Εσσιανή μήτρα δεν είναι ούτε θετικά, ούτε αρνητικά ημιορισμένη. Λόγω του ότι η Εσσιανή μήτρα είναι πίνακας 2×2, η δευτέρας τάξεως συνθήκη είναι ισοδύναμη με την ορίζουσα της Εσσιανής μήτρας να είναι αρνητική. Στο εξής, όταν αναφερόμαστε σε σαγματικά σημεία, θα θεωρούμε ότι αυτά αντιστοιχούν σε $\inf_t\sup_s$ και όχι σε $\sup_t\inf_s$, εκτός αν ρητά αναφερθούμε στο αντίθετο.

Η προσέγγιση (2.3) μπορεί να βελτιωθεί με τη χρήση του θεωρήματος Bahadur-Rao. Οι συγγραφείς του [26], επεκτείνουν την απόδειξη του ασυμπτωτικού πολλών πηγών στο [8], όπου αποδεικνύουν ότι καθώς $N\to\infty$ ισχύει το εξής:

$$\mathrm{P(overflow)} = \frac{1}{\sqrt{2\pi N\sigma^2 s^2}}e^{-NI}\left(1+O\left(\frac{1}{N}\right)\right)\,, \qquad (2.5)$$

όπου το (s,t) είναι σημείο που βελτιστοποιεί τη (2.2), ενώ το σ^2 δίδεται από την

$$\sigma^2 = \frac{\partial^2}{\partial s^2}\log \mathrm{E}\left[e^{sX_j[0,t]}\right] = \frac{M''(s)}{M(s)}-(ct+b)^2\,,$$

όπου $M(s) = \mathrm{E}\left[e^{sX_j[0,t]}\right]$ είναι η ροπογεννήτρια της τυχαίας μεταβλητής $X_j[0,t]$. Βασιζόμενοι στην (2.5) οδηγούμαστε στην εξής προσέγγιση:

$$\mathrm{P(overflow)} \approx \frac{1}{\sqrt{2\pi N\sigma^2 s^2}} e^{-NI} = e^{-NI-\frac{1}{2}\log\left(2\pi N\sigma^2 s^2\right)} . \tag{2.6}$$

Θα αναφερόμαστε στην (2.6) ως τη βελτίωση Bahadur-Rao στο ασυμπτωτικό πολλών πηγών. Ο όρος $\frac{1}{2}\log\left(2\pi N\sigma^2 s^2\right)$ μπορεί να προσεγγισθεί από τον $\frac{1}{2}\log\left(4\pi NI\right)$ [29]. Κατά συνέπεια η εξίσωση (2.6) δεν απαιτεί παραπάνω υπολογισμούς από ό,τι η (2.3).

Είναι εύκολο να δούμε, μετά από λίγη άλγεβρα, ότι η βελτίωση Bahadur-Rao στο ασυμπτωτικό πολλών πηγών (εξίσωση (2.6)) δίδει το ακόλουθο σύνορο (τοπικά σε μια γειτονία του σημείου (s,t) που βελτιστοποιεί τη 2.2) για την περιοχή αποδοχής:

$$\sum_{j=1}^{J} N_j\alpha_j(s,t) \le C + \frac{1}{t}\left(B - \frac{\gamma'}{s}\right) = C^*_{B-R} , \tag{2.7}$$

όπου $\gamma' = \gamma - \frac{1}{2}\log(4\pi\gamma)/(1+\frac{1}{2\gamma})$. Άξιο παρατήρησης είναι το γεγονός ότι οι εξισώσεις (2.4) και (2.7) χρησιμοποιούν τον ορισμό (2.1) του ισοδυνάμου εύρους ζώνης. Η βελτίωση Bahadur-Rao επηρεάζει (αυξάνει) μόνο την ισοδύναμη χωρητικότητα $C^*_{B-R} > C^*$, ενώ οι παράμετροι s, t διατηρούνται γιατί υπολογίζονται βάσει της ίδιας εξίσωσης (2.2).

Στις προηγούμενες παραγράφους εξετάσαμε τις ασυμπτωτικές ιδιότητες της πιθανότητας υπερχείλισης ενός συνδέσμου (εξισώσεις (2.2) και (2.5)). Ένα σχετικό με αυτήν μέγεθος που χρησιμοποιείται ως χαρακτηριστικό της ποιότητας υπηρεσίας σε ευρυζώνια δίκτυα είναι η πιθανότητα απώλειας κελιού σε σύνδεσμο με πεπερασμένο χώρο ενταμιευτή (Cell Loss Probability - CLP), για την οποία ισχύει το ακόλουθο ασυμπτωτικό αποτέλεσμα [26] ως προς N:

$$\mathrm{CLP} = \frac{e^{-NI}}{s^* Nc\rho\sqrt{2\pi\sigma N^3}}\left(1 + O\left(\frac{1}{N}\right)\right) , \tag{2.8}$$

όπου ρ είναι ο βαθμός χρησιμοποίησης του συνδέσμου. Βάσει της εξίσωσης αυτής, λαμβάνουμε την εξής προσέγγιση για πεπερασμένο N:

$$\mathrm{CLP} \approx \frac{1}{s^* Nc\rho\sqrt{2\pi N\sigma^2 s^{*2}}} e^{-NI} = e^{-NI-\frac{1}{2}\log\left(2\pi N\sigma^2 s^2\right)-\log(s^* Nc\rho)} . \tag{2.9}$$

Όπως αναφέραμε, ο όρος $\frac{1}{2}\log\left(2\pi N\sigma^2 s^2\right)$ μπορεί να προσεγγισθεί από τον $\frac{1}{2}\log\left(4\pi NI\right)$ [29], δηλαδή οι εξισώσεις (2.5) και (2.9) εμπλέκουν μόνο τις ποσότητες που περιέχονται στην εξίσωση (2.2) και κατά συνέπεια δεν απαιτούν περαιτέρω υπολογισμούς.

Η ύπαρξη των ισοδυνάμων ευρών ζώνης και η ισχύς των ασυμπτωτικών αποτελεσμάτων έχουν σαν προϋπόθεση μόνο τη στασιμότητα των πηγών. Ενδεικτικά παραδείγματα που αναφέρονται στο [22] και [33] περιλαμβάνουν πηγές εισόδου όπως κλασματική κίνηση Brown, πηγές Markov, πηγές των οποίων η κυκλοφορία έχει αλλάξει μορφή κατά την είσοδό

τους στο δίκτυο και πηγές οι οποίες υπακούν σε μηχανισμούς ελέγχου του δικτύου. Ένα σημαντικό χαρακτηριστικό του προβλήματος βελτιστοποίησης στην εξίσωση (2.3) είναι ότι μπορεί να λυθεί βάσει πληροφορίας η οποία προέρχεται από πραγματικές πηγές κυκλοφορίας. Στην επόμενη παράγραφο περιγράφουμε με ποιό τρόπο είναι δυνατόν να γίνει αυτό.

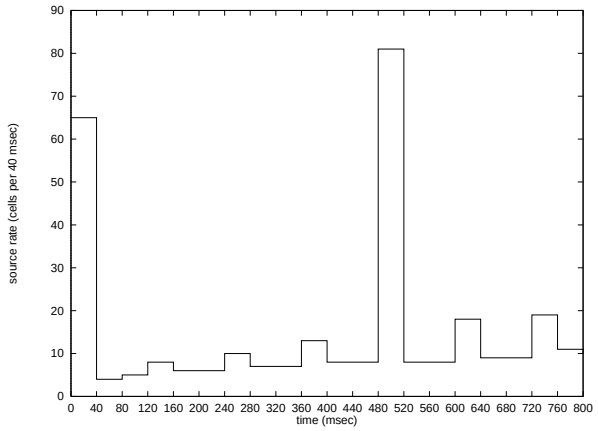
2.3 Ισοδύναμο Εύρος Ζώνης Πραγματικής Κυκλοφορίας

Η κυκλοφορία στα σημερινά ευρυζώνια δίκτυα είναι εύκολο να μετρηθεί λόγω του ευέλικτου σχεδιασμού των μεταγωγέων. Ήδη υπάρχει ένας μεγάλος όγκος από τέτοιες μετρήσεις οι οποίες αποθηκεύονται σε αρχεία επονομαζόμενα ως *αρχεία-ίχνη* (trace files). Τα αρχεία αυτά αναπαριστούν τη διαδικασία άφιξης της κυκλοφορίας μιας συγκεκριμένης πηγής (ή μιας ομάδας πηγών) σε ένα μεταγωγέα. Στη βιβλιογραφία έχουν εκτενώς μελετηθεί αρχεία-ίχνη που αναπαριστούν διαδικασίες αφίξεων σε τμήματα δικτύων Ethernet [25, 36, 18] και πηγών MPEG, JPEG [13, 32].

Αν και η δομή των αρχείων-ιχνών ποικίλει ανάλογα με τον τρόπο που έχουν ληφθεί οι μετρήσεις της πραγματικής κυκλοφορίας και τα χαρακτηριστικά της διαδικασίας άφιξης που μελετώνται, στην παρούσα εργασία θα χρησιμοποιήσουμε μια απλοποιημένη μορφή αρχείου-ίχνους έτσι ώστε να επιδέχεται μαθηματικής ανάλυσης. Στην παρούσα εργασία ένα αρχείο-ίχνος περιγράφεται από το σύνολο $\mathcal{X}_0 = \{x_i; i = 1, \ldots, K\}$ το οποίο υποδηλώνει κυκλοφορία σταθερού ρυθμού x_i κατά τη χρονική περίοδο i. Οι χρονικές περίοδοι $i = 1, \ldots, K$ έχουν την ίδια χρονική διάρκεια, ενώ K είναι η συνολική χρονική διάρκεια μέτρησης. Για παράδειγμα η κυκλοφορία μιας πηγής MPEG-1 θα μπορούσε να περιγραφεί από το σύνολο $\mathcal{X}_0$, όπου x_i είναι το μέγεθος σε κελιά των 53 bytes (συμπεριλαμβάνοντας το μέγεθος της επικεφαλίδας κάθε ATM κελιού) του πλαισίου i, ενώ οι χρονικές περίοδοι έχουν διάρκεια 40 msec (για κωδικοποίηση MPEG-1 με ρυθμό 25 πλαίσια ανά δευτερόλεπτο). Στο Σχήμα 2.2 απεικονίζεται ο ρυθμός μετάδοσης ως συνάρτηση του χρόνου, βάσει ενός αρχείου-ίχνους το οποίο προέρχεται από ένα μικρό τμήμα της ταινίας video Star Wars η οποία έχει κωδικοποιηθεί σύμφωνα με το πρότυπο MPEG-1.

Στο παράδειγμα αυτό, αν και η πραγματική διαδικασία αφίξεων διαφέρει από την περιγραφή που δίδει το σύνολο $\mathcal{X}_0$ λόγω του ότι ο ρυθμός μετάδοσης δεν είναι σταθερός κατά τη διάρκεια μετάδοσης ενός πλαισίου, εντούτοις σκοπός μας είναι η περιγραφή μιας πραγματικής πηγής να είναι αρκετά απλή ώστε η μαθηματική της ανάλυση να είναι εφικτή.

Το επόμενο βήμα είναι να δώσουμε μια στοχαστική περιγραφή του φόρτου $X[\tau, \tau + t]$ που εισάγει η πραγματική πηγή σε ένα σύνδεσμο στο χρονικό διάστημα $(\tau, \tau + t]$, έχοντας την περιγραφή $\mathcal{X}_0$. Για να μπορέσουμε να χρησιμοποιήσουμε τα αποτελέσματα της



Σχήμα 2.2: Ο ρυθμός μετάδοσης πληροφορίας ως συνάρτηση του χρόνου. Το μοντέλο αυτό έχει ανακατασκευαστεί από ένα αρχείο-ίχνος το οποίο προέρχεται από ένα μικρό τμήμα της ταινίας video Star Wars η οποία έχει κωδικοποιηθεί σύμφωνα με το πρότυπο MPEG-1.

Παραγράφου 2.2, θα πρέπει να υποθέσουμε τα εξής:

1. $\tau \in R$, δηλαδή έχουμε μοντέλο συνεχούς χρόνου και

2. η στοχαστική διαδικασία $X[0,t], t > 0$ έχει στάσιμες αυξήσεις, δηλαδή οι τυχαίες μεταβλητές $\{X[\tau, \tau+t], \tau \in R\}$ έχουν την ίδια κατανομή για την ίδια τιμή του t.

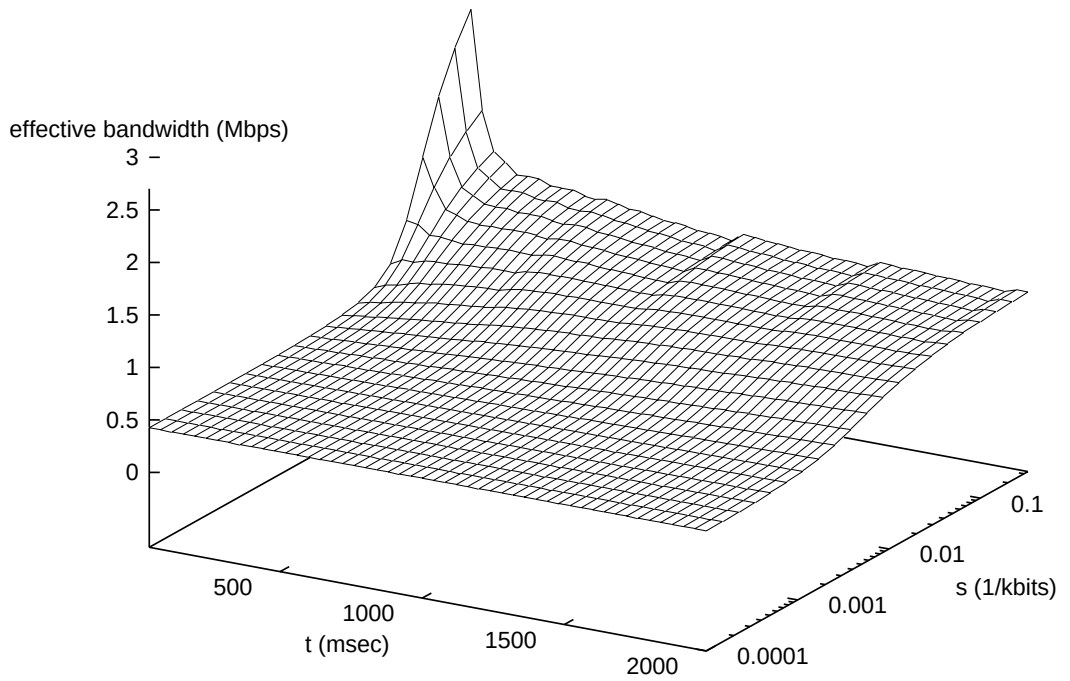
Λόγω της υπόθεσης 1, ο χρονικός ορίζοντας είναι μη φραγμένος σε αντίθεση με τη χρονική διάρκεια των μετρήσεων $\mathcal{X}_0$ η οποία είναι φραγμένη, οπότε πρέπει να επεκτείνουμε περιοδικά το $\mathcal{X}_0$ έτσι ώστε η ακολουθία των ρυθμών μετάδοσης για $x_i, i = 1, \ldots, K$, να επαναλαμβάνεται συνεχώς. Για το σκοπό αυτόν, ορίζουμε $\mathcal{X}_j \stackrel{\Delta}{=} \{x_i : i = jK+1, \ldots, jK+K\}$ όπου $x_{jK+i} = x_i$, για κάθε j στο Z. Τώρα, το σύνολο $\mathcal{X} \stackrel{\Delta}{=} \cup_{j \in Z} \mathcal{X}_j$ αντιπροσωπεύει μια διαδικασία αφίξεων σε συνεχή χρόνο, μη φραγμένης διάρκειας. Έστω δ η διάρκεια των χρονικών περιόδων του αρχείου-ίχνους $\mathcal{X}_0$, το οποίο αρχίζει τη χρονική στιγμή 0, τότε ο φόρτος $X[0,t]$ της διαδικασίας αφίξεων στο χρονικό διάστημα $(\tau, \tau+t]$ δίδεται ως εξής:

$$X[\tau, \tau+t] = \sum_{i \in Z} x_i \mathbf{1}(\tau \leq i\delta + \delta \wedge i\delta \leq \tau + t) \cdot (\delta - (\tau - i\delta)^+ - (i\delta + \delta - \tau - t)^+) \ ,$$

όπου $\mathbf{1}(A) = 0$ εάν η πρόταση A είναι ψευδής, ενώ είναι 1 σε διαφορετική περίπτωση. Επίσης ορίζουμε $x^+ = x \cdot \mathbf{1}(x > 0)$. Τώρα θεωρώντας ότι η διαδικασία που τροφοδοτεί το σύνδεσμο περιγράφεται από το $\mathcal{X}$, αν παρατηρήσουμε σε μια τυχαία χρονική στιγμή το φόρτο που δημιουργεί η πηγή σε χρονικό διάστημα t, θα λάβουμε τη μέτρηση $X[U, U+t]$, όπου U είναι τυχαία μεταβλητή ομοιόμορφα κατανεμημένη στο $(0, \mathrm{K}\delta)$. Οπότε το ισοδύναμο εύρος ζώνης είναι

$$\alpha(s,t) = \frac{1}{st} \log \mathrm{E}_U \left[e^{sX[U,U+t]} \right] = \frac{1}{st} \log \frac{1}{K\delta} \int_0^{K\delta} e^{sX[\tau,\tau+t]} d\tau \ .$$

Η συνάρτηση μέσα στο ολοκλήρωμα είναι κατά τμήματα εκθετική, όπου ο συντελεστής του εκθέτη αλλάζει όταν συμβαίνει μια αλλαγή του ρυθμού μετάδοσης της πηγής μέσα στο διάστημα παρατήρησης $(\tau, \tau + t]$. Κατά συνέπεια το παραπάνω ολοκλήρωμα μπορεί εύκολα να υπολογισθεί ακριβώς κάνοντας μόνο λάθη στρογγύλευσης λόγω πεπερασμένης αριθμητικής ακρίβειας. Στο Σχήμα 2.3 απεικονίζεται το ισοδύναμο εύρος ζώνης $\alpha(s,t)$ για ένα αρχείο-ίχνος που αντιστοιχεί σε μικρό τμήμα της ταινίας video Star Wars η οποία είναι κωδικοποιημένη σύμφωνα με το πρότυπο MPEG-1.



Σχήμα 2.3: Η συνάρτηση ισοδυνάμου εύρους ζώνης $\alpha(s,t)$ της ταινίας video Star Wars η οποία έχει κωδικοποιηθεί σύμφωνα με το πρότυπο MPEG-1.

Ο τρόπος υπολογισμού του ισοδυνάμου εύρους ζώνης βάσει ενός αρχείου-ίχνους που περιγράψαμε, βασίζεται σε αυτόν στο [23], όπου όμως τα αρχεία-ίχνη και η ερμηνεία τους εκεί έχουν διαφορετική μορφή. Συγκεκριμένα ο συγγραφέας στο [23] θεωρεί ότι οι αφίξεις στον πολυπλέκτη είναι σημειακές σε αντίθεση με εμάς όπου υιοθετούμε ένα μοντέλο ρευστού (fluid model) για την κυκλοφορία. Επιπλέον δεν κατασκευάζει ρητά ένα στάσιμο μοντέλο κυκλοφορίας βασισμένο σε ένα αρχείο-ίχνος με αποτέλεσμα να μη μπορεί να χρησιμοποιηθεί το μοντέλο του σε προσομοίωση. Στο Κεφάλαιο 4 θα δούμε ότι είναι σημαντικό το ίδιο μοντέλο να μπορεί να χρησιμοποιηθεί σε προσομοίωση έτσι ώστε να υπάρχει η δυνατότητα ελέγχου της ακρίβειας διαφόρων αποτελεσμάτων που βασίζονται στο

ισοδύναμο εύρος ζώνης κυκλοφοριών.

Κεφάλαιο 3

Ισοδυναμία και Αντικατάσταση Κυκλοφορίας

Στο κεφάλαιο αυτό προτείνουμε συνθήκες οι οποίες μας επιτρέπουν να ορίσουμε μια σχέση ισοδυναμίας μεταξύ της κυκλοφορίας που παράγουν πηγές. Επιπλέον αναλύουμε τις επιπτώσεις που έχει η αντικατάσταση μεταξύ ισοδυνάμων πηγών σε έναν πολυπλέκτη. Η βασική ιδέα ώστε να εξάγουμε συνθήκες ισοδυναμίες μεταξύ κυκλοφορίας με τις κατάλληλες ιδιότητες αντικατάστασης, είναι η εξίσωση (2.2). Οι συνθήκες τις οποίες προτείνουμε, εξασφαλίζουν ότι η παράσταση που περικλείεται μέσα στις αγκύλες παραμένει σχεδόν η ίδια σε μια γειτονιά του σημείου λειτουργίας. Αυτό μας εγγυάται ότι η πιθανότητα υπερχείλισης παραμένει η ίδια και οι υπερχειλίσεις θα συμβαίνουν με τον ίδιο τρόπο (όσον αφορά τη χρονική κλίμακα στην οποία κυρίως υπερχειλίζει ο ενταμιευτής) εκτός και αν δημιουργηθεί ένα άλλο τοπικά βέλτιστο σημείο με μικρότερη τιμή.

3.1 Η Συνθήκη Ισοδυναμίας

Στην παράγραφο αυτή, ορίζουμε τη συνθήκη την οποία πρέπει να ικανοποιεί η κυκλοφορία ενός μοντέλου ώστε να αντικαθιστά την πραγματική κυκλοφορία και αποδεικνύουμε ορισμένες σημαντικές ιδιότητες της αντικατάστασης.

Πρώτα θα ορίσουμε σύμβολα τα οποία θα χρησιμοποιούμε έως το τέλος της εργασίας. Όπως έχει ήδη αναφερθεί, ο πολυπλέκτης ATM αποτελείται από ενταμιευτή μεγέθους B και εξυπηρετεί με ρυθμό C. Αρχικά δέχεται σαν είσοδο την κυκλοφορία $R + V$, όπου R είναι η κυκλοφορία την οποία θα αντικαταστήσουμε και V είναι η υπόλοιπη κυκλοφορία. Μετά την αντικατάσταση, ο σύνδεσμος τροφοδοτείται από την κυκλοφορία $M + V$, όπου M είναι το ‘‘μοντέλο’’ που αντικαθιστά την πραγματική κυκλοφορία R. Εάν αντικαταστήσουμε τη

συνολικά εισερχόμενη κυκλοφορία του συνδέσμου, τότε καταχρηστικά γράφουμε $V = 0$. Ορίζουμε τη συνάρτηση $G_W(s,t) = s(B + Ct) - st\alpha_W(s,t)$, όπου $\alpha_W(s,t)$ είναι το ισοδύναμο εύρος ζώνης της κυκλοφορίας $W \in \{R, V, M, R+V, M+V\}$ και χρησιμοποιούμε (s^*, t^*) ως το σημείο λειτουργίας του συνδέσμου όταν αυτός τροφοδοτείται από την κυκλοφορία $R+V$, δηλ. το σημείο το οποίο βελτιστοποιεί την παράσταση $\inf_t \sup_s G_{R+V}(s,t)$. Θα λέμε ότι η συνάρτηση $\alpha(s,t)$ είναι παραγωγίσιμη σε ένα δοθέν σημείο εάν και οι δύο μερικές παράγωγοί της υπάρχουν και είναι πεπερασμένες.

Ο ακόλουθος ορισμός, ορίζει την ισοδυναμία κυκλοφορίας σε κάποιο αυθαίρετο σημείο (s_0, t_0).

Ορισμός 1 *Έστω M και R κυκλοφορίες με συναρτήσεις ισοδυνάμου εύρους ζώνης $\alpha_M(s,t)$ και $\alpha_R(s,t)$ αντίστοιχα οι οποίες είναι παραγωγίσιμες στο (s_0, t_0). Η κυκλοφορία M είναι ισοδύναμη με την κυκλοφορία R στο σημείο (s_0, t_0) αν και μόνο αν*

$$\begin{cases} \alpha_R(s_0,t_0) = \alpha_M(s_0,t_0) \\ \\ \dfrac{\partial \alpha_R}{\partial s}(s,t_0)\Big|_{s=s_0} = \dfrac{\partial \alpha_M}{\partial s}(s,t_0)\Big|_{s=s_0} \\ \\ \dfrac{\partial \alpha_R}{\partial t}(s_0,t)\Big|_{t=t_0} = \dfrac{\partial \alpha_M}{\partial t}(s_0,t)\Big|_{t=t_0} \end{cases} . \tag{3.1}$$

Είναι εύκολο να δούμε ότι οι παραπάνω συνθήκες ορίζουν μια ανακλαστική, συμμετρική και μεταβατική σχέση, δηλ. μια σχέση ισοδυναμίας τοπικά ορισμένη στο σημείο (s_0, t_0). Μια σημαντική παρατήρηση είναι ότι για να ικανοποιούνται ταυτόχρονα οι παραπάνω συνθήκες μεταξύ της κυκλοφορίας R και την κυκλοφορίας του μοντέλου M πρέπει κάθε υποψήφιο τέτοιο μοντέλο να έχει τουλάχιστον τρεις ελεύθερες παραμέτρους των οποίων η τιμή θα ορίζεται από το παραπάνω σύστημα των τριών εξισώσεων.

Η σημαντική ιδιότητα της ισοδυναμίας κυκλοφορίας είναι ότι αν έχει οριστεί σε ένα σημείο σάγματος, το σημείο αυτό εξακολουθεί να έχει αυτή την ιδιότητα μετά από την αντικατάσταση. Το πρόβλημα που μπορεί να προκύψει είναι ότι η αντικατάσταση μπορεί να προκαλέσει τη δημιουργία και άλλων επιπλέον σημείων σάγματος τα οποία μπορεί να προκαλέσουν αλλαγή του σημείου λειτουργίας. Η παραπάνω ιδιότητα μπορεί να εκφραστεί τυπικά μέσω του θεωρήματος που ακολουθεί.

Θεώρημα 1 *Έστω ότι η συνάρτηση $G_{R+V}(s,t)$ έχει σημείο σάγματος στο (s_0, t_0), σε μια γειτονιά D στην οποία οι συναρτήσεις $\alpha_R, \alpha_M, \alpha_V$ είναι δύο φορές παραγωγίσιμες. Εάν η κυκλοφορία M είναι ισοδύναμη με την κυκλοφορία R στο σημείο (s_0, t_0) και η G_{M+V} έχει μη ημιορισμένη Εσσιανή μήτρα στο σημείο αυτό, τότε η G_{M+V} έχει επίσης σημείο σάγματος στο (s_0, t_0) και επιπλέον $G_{M+V}(s_0, t_0) = G_{R+V}(s_0, t_0)$.*

Απόδειξη: Λόγω των δύο τελευταίων συνθηκών της (3.1), οι μερικές παράγωγοι των G_{R+V} και G_{M+V} στο (s_0, t_0) είναι ίσες και μάλιστα είναι μηδενικές αφού το (s_0, t_0) είναι σημείο σάγματος της G_{R+V}. Λόγω του ότι η Εσσιανή μήτρα της G_{M+V} είναι μη ημιορισμένη, έπεται ότι το παραπάνω σημείο είναι σημείο σάγματος της G_{M+V}. Μένει ακόμα να δείξουμε ότι το σημείο σάγματος αυτό, αντιστοιχεί σε $\inf_t \sup_s$ εν αντιθέσει με $\sup_s \inf_t$. Όμως αυτό όντως ισχύει, αφού η συνάρτηση $G_{M+V}(s,t)$ είναι κοίλη ως προς s για κάθε $t > 0$, γεγονός το οποίο οφείλεται στην κυρτότητα των ροπογεννητριών συναρτήσεων [3]. Πράγματι, λόγω της κοιλότητας της $G_{M+V}(s,t)$ ως προς s, το ακρότατο ως προς s μπορεί μόνο να είναι τοπικό $\sup_s$. Επιπλέον από την πρώτη συνθήκη του συστήματος (3.1), έπεται ότι $G_{M+V}(s_0, t_0) = G_{R+V}(s_0, t_0)$. ■

Η κύρια συνέπεια του Θεωρήματος 1 είναι η εξής: Εάν η κυκλοφορία $R + V$ τροφοδοτεί τον σύνδεσμο του οποίου σημείο λειτουργίας είναι το (s^*, t^*), τότε εάν η M αντικαθιστά την R, το $(s,^*, t^*)$ θα εξακολουθεί να είναι σημείο σάγματος της $G_{M+V}(s,t)$ και συνεπώς *πιθανό* σημείο λειτουργίας. Δεν θα παραμείνει το (s^*, t^*) σαν σημείο λειτουργίας εάν δημιουργήθηκαν και άλλα σημεία σάγματος, κατά την αντικατάσταση, στα οποία η G_{M+V} λαμβάνει μικρότερη τιμή από ότι στο (s^*, t^*). Η φυσική ερμηνεία που μπορούμε να δώσουμε σε αυτή την περίπτωση, είναι ότι η αντικατάσταση της κυκλοφορίας R από το μοντέλο M μπορεί να αλλάξει την πιθανότερη χρονική κλίμακα στην οποία υπερχειλίζει ο ενταμιευτής. Αυτό έχει ως συνέπεια τα δύο συστήματα να έχουν διαφορετικές πιθανότητες υπερχείλισης και οι υπερχειλίσεις των ενταμιευτών να συμβαίνουν με διαφορετικό τρόπο.

Συνοψίζοντας τα παραπάνω, βασική προϋπόθεση έτσι ώστε η προσέγγιση μας στην αντικατάσταση κυκλοφορίας να είναι λειτουργική, είναι το μοντέλο κυκλοφορίας M να *διατηρεί* το σημείο λειτουργίας στο οποίο είναι ισοδύναμο με την πραγματική κυκλοφορία R. Εάν η κυκλοφορία M ικανοποιεί αυτή την προϋπόθεση τότε διατηρούνται τόσο ο τύπος όσο και η συχνότητα των φαινομένων στα οποία οφείλονται κυρίως οι υπερχειλίσεις. Παρακάτω αποδεικνύουμε ότι το παραπάνω ισχύει πάντα στην περίπτωση αντικατάστασης του συνόλου των πηγών με κλασματική κίνηση Brown.

Κάτι που πρέπει να τονισθεί είναι ότι δεδομένου ότι οι προσεγγίσεις της πιθανότητας υπερχείλισης στις εξισώσεις (2.3) και (2.6) ισχύουν για πολλές πηγές (μεγάλο N), θα πρέπει οι κυκλοφορίες $R+V$ και $M+V$ να προέρχονται από την υπέρθεση πολλών μικρότερων πηγών. Μόνο τότε θα ισχύουν τα επιθυμητά αποτελέσματα της αντικατάστασης κυκλοφορίας: διατήρηση του τύπου και της συχνότητας των φαινομένων στα οποία οφείλονται κυρίως οι υπερχειλίσεις.

Όσον αφορά την υπόθεση για μη ημιορισμένη Εσσιανή μήτρα, αυτή μπορεί να ελεγχθεί μόνο πειραματικά αφού μπορεί να έχουμε πραγματικές πηγές (μπορεί να ελεγχθεί και

αναλυτικά σε ειδικές περιπτώσεις όπου ολόκληρη η κυκλοφορία προέρχεται από μοντέλα). Έχουμε εκτενώς πειραματιστεί με συνδυασμούς κυκλοφορίας MPEG (V) και κυκλοφορίας μοντέλων (M) και σε όλες τις περιπτώσεις η Εσσιανή μήτρα ήταν πράγματι μη ημιορισμένη.

Παράδειγμα Αντικατάστασης Κυκλοφορίας

Όπως ήδη έχουμε εξηγήσει στο προηγούμενο κεφάλαιο, οι σχεδιαστές δικτύου θα ήθελαν να αξιολογήσουν την απόδοση του δικτύου σε πραγματικές συνθήκες φόρτου. Χρησιμοποιώντας τη μέθοδο που προτείνουμε, αυτό μπορεί να γίνει αντικαθιστώντας την υπέρθεση πάρα πολλών πραγματικών πηγών με μια ισοδύναμη πηγή η οποία βασίζεται σε απλά μοντέλα κυκλοφορίας (με τουλάχιστον τρεις ελεύθερες παραμέτρους σύμφωνα με τα όσα είπαμε στην προηγούμενη παράγραφο). Για να επιτύχουμε κάτι τέτοιο, πρέπει να αποτιμήσουμε αριθμητικά την τιμή του σημείου λειτουργίας (s^*, t^*) για το πραγματικό σύστημα όπως επίσης και τις τιμές των αριστερών μελών των εξισώσεων του συστήματος (3.1). Με τη χρήση ενός απλού παραδείγματος περιγράφουμε τα βήματα τα οποία πρέπει να ακολουθηθούν.

Έστω ότι θα θέλαμε να αξιολογήσουμε την ποιότητα υπηρεσίας μιας συγκεκριμένης ταινίας video (π.χ. James Bond) όταν αυτή διέρχεται από ένα σύνδεσμο με υψηλό βαθμό πολυπλεξίας, χρησιμοποιώντας τη μέθοδό μας για αντικατάσταση της κυκλοφορίας που δημιουργούν όλες οι υπόλοιπες πηγές. Πιο συγκεκριμένα, έστω ένας σύνδεσμος χωρητικότητας 155 Mbps με ενταμιευτή 2000 κελιών ο οποίος τροφοδοτείται από την υπέρθεση 40 ανεξάρτητων ταινιών video James Bond (κωδικοποιημένες σε MPEG) και 400 ανεξάρτητων ταινιών video Star Wars (σε MPEG επίσης). Ενδιαφερόμαστε να παρακολουθήσουμε οπτικά την επίδοση που απολαμβάνουν οι ταινίες video James Bond. Στη συγκεκριμένη περίπτωση, πρέπει να αντικαταστήσουμε όλες τις ταινίες video Star Wars με μία μόνο πηγή μοντέλο. Για να εφαρμόσουμε την προτεινόμενη προσέγγιση, αρχικά υπολογίζουμε τις συναρτήσεις ισοδυνάμου εύρους ζώνης $\alpha_{SW}(s,t)$, $\alpha_{JB}(s,t)$ των ταινιών video Star Wars και James Bond αντίστοιχα, χρησιμοποιώντας ένα μόνο αρχείο-ίχνος για κάθε μία[1]. Ακολούθως υπολογίζουμε το σημείο λειτουργίας του συστήματος επιλύωντας αριθμητικά το

$$\inf_t \sup_s \left[s(B + Ct) - st(40\alpha_{JB}(s,t) + 400\alpha_{SW}(s,t)) \right] ,$$

όπου λαμβάνονται υπόψιν οι συναρτήσεις ισοδυνάμου εύρους ζώνης των δύο ταινιών, που υπολογίζονται όπως εξηγήθηκε στην Παρ. 2.3. Πρέπει να σημειωθεί ότι εάν προσθέσουμε μία, ή έστω λίγες, πηγές επιπλέον (π.χ. ταινίες Simpsons), το νέο σημείο λειτουργίας θα είναι μια μικρή διαταραχή του αρχικού σημείου, αφού η νέα κυκλοφορία είναι μια μικρή διαταραχή

[1] οι συναρτήσεις περιγράφονται από τις τιμές τους όπως αυτές εκτιμήθηκαν από το αρχείο-ίχνος, πάνω σε ένα πεπερασμένο πλέγμα στο θετικό $s - t$ τεταρτημόριο

της αρχικής. Το γεγονός αυτό μπορεί φανεί χρήσιμο σε περιπτώσεις όπου η κυκλοφορία την οποία παρακολουθούμε είναι μικρή, σε σχέση με τη συνολική κυκλοφορία, και μη παρατηρήσιμη πριν τη διεξαγωγή του πειράματος, π.χ. μια εφαρμογή τηλεσυνδιάσκεψης.

3.2 Αντικατάσταση με πηγές Gaussian

Οι πηγές Gaussian κρίνονται σαν καλές ''υποψήφιες'' για αντικατάσταση, λόγω της απλής αναλυτικής περιγραφής της συνάρτησης του ισοδυνάμου εύρους ζώνης τους και τον άμεσο τρόπο παραγωγής κυκλοφορίας τους που πηγάζει απευθείας από τον ορισμό τους.

Έστω $X[0,t]$ ο φόρτος που παράγεται από μια πηγή στο διάστημα $(0,t]$. Η πηγή αυτή είναι του τύπου Gaussian αν ο φόρτος δίνεται από την

$$X[0,t] = \lambda t + Z(t) \, ,$$

όπου $Z(t)$ είναι μια τυχαία μεταβλητή η οποία ακολουθεί την Κανονική κατανομή και έχει μηδενική μέση τιμή. Το ισοδύναμο εύρος ζώνης είναι [22]

$$\alpha_M(s,t) = \lambda + \frac{s}{2t}\mathrm{Var}Z(t) \, . \tag{3.2}$$

Σε αυτή την περίπτωση το σύστημα εξισώσεων (3.1) αποτιμούμενο στο (s_0,t_0) οδηγεί στο ισοδύναμο σύστημα εξισώσεων

$$\begin{cases} \lambda = \alpha_R(s_0,t_0) - s_0 \left.\dfrac{\partial \alpha_R}{\partial s}(s,t_0)\right|_{s=s_0} \\ \\ \mathrm{Var}Z(t_0) = 2t_0 \left.\dfrac{\partial \alpha_R}{\partial s}(s,t_0)\right|_{s=s_0} \\ \\ \dfrac{s_0}{2}\dfrac{d}{dt}\left(\dfrac{\mathrm{Var}Z(t)}{t}\right)\Big|_{t=t_0} = \left.\dfrac{\partial \alpha_R}{\partial t}(s_0,t)\right|_{t=t_0} \end{cases} . \tag{3.3}$$

Παρατηρούμε ότι για να αντικαταστήσουμε μια δεδομένη πηγή με μια πηγή Gaussian, η μόνη πληροφορία που χρειάζεται να εξάγουμε για τη συνάρτηση $\mathrm{Var}Z(t)$ είναι η τιμή αυτής και της παραγώγου της στο $t = t_0$.

3.3 Υψηλό Κέρδος Στατιστικής Πολυπλεξίας

Στην περίπτωση που το σύστημα παρουσιάζει μεγάλο βαθμό πολυπλεξίας, η παράμετρος χώρου s του συστήματος προσεγγίζει το μηδέν. Τέτοια μεγάλη πολυπλεξία αναμένεται ότι θα παρουσιάζεται σε μεγάλα συστήματα χωρητικότητας 622 Mbps ή 1.2 Gbps. Εάν με $X[0,t]$

συμβολίσουμε τον φόρτο τον οποίο παράγει η πηγή R στο διάστημα $(0,t]$, τότε το ισοδύναμο εύρος ζώνης, για s κοντά στο 0, δίδεται από την [22]

$$\alpha_R(s,t) = \frac{\mathrm{E}X[0,t]}{t} + s\frac{\mathrm{Var}X[0,t]}{2t} + \mathrm{o}(s)\cdot$$

γεγονός που υποδεικνύει ότι μπορούμε να παραλείψουμε τους όρους δευτέρας και μεγαλύτερης τάξεως. Εάν μια πηγή Gaussian M αντικαθιστά την πηγή R στο σημείο λειτουργίας (s^*,t^*), όπου το s^* είναι κοντά στο μηδέν, τότε το σύστημα εξισώσεων (3.1) δίδει

$$\begin{cases} \lambda = \dfrac{\mathrm{E}X[0,t^*]}{t^*} + \mathrm{o}(s^*) \\ \\ \mathrm{Var}Z(t^*) = \mathrm{Var}X[0,t^*] + \mathrm{o}(s^*) \\ \\ \left.\dfrac{d\mathrm{Var}X[0,t]}{dt}\right|_{t=t^*} = \left.\dfrac{d\mathrm{Var}Z(t)}{dt}\right|_{t=t^*} + \mathrm{o}(s^*) \end{cases} .$$

Παρατηρούμε ότι οι παράμετροι του μοντέλου εξαρτώνται από τις πρώτες δύο ροπές της διαδικασίας εισόδου στην πιο πιθανή χρονική κλίμακα όπου υπερχειλίζει ο ενταμιευτής (η χρονική παράμετρος του συστήματος). Σε μεγάλα συστήματα, η μοντελοποίηση κυκλοφορίας μπορεί να γίνεται λαμβάνοντας υπόψιν μόνο τα χαρακτηριστικά πρώτης και δευτέρας τάξεως της διαδικασίας εισόδου. Επίσης βλέπουμε ότι η μέθοδός μας συμφωνεί με γνωστές μεθόδους μοντελοποίησης στην περίπτωση που παρουσιάζεται μεγάλος βαθμός πολυπλεξίας (βλ. [28, 10, 21]). Οι μέθοδοι αυτές αναλύουν την περίπτωση μεγάλης πολυπλεξίας καθώς ο βαθμός χρησιμοποίησης του πολυπλέκτη πλησιάζει την τιμή 1, δηλαδή καθώς το πλήθος των πηγών που πολυπλέκονται αυξάνεται χωρίς να μεταβάλλονται η χωρητικότητα και ο χώρος ενταμίευσης. Τέτοια ασυμπτωτικά σχήματα μπορούν να αναλυθούν με την προσέγγιση που προσφέρει το Κεντρικό Οριακό Θεώρημα όπου η πρώτης και δευτέρας τάξεως ροπές καθορίζουν πλήρως τη συμπεριφορά της κυκλοφορίας.

3.4 Η Περίπτωση $B = 0$

Στην παράγραφο αυτήν, μελετάμε συνδέσμους χωρίς ενταμιευτή όπου αντικαθιστούμε πραγματική κυκλοφορία με πηγές Gaussian. Η πιθανότητα υπερχείλισης του συνδέσμου εξαρτάται μόνο από τα χαρακτηριστικά της στάσιμης κατάστασης της διαδικασίας εισόδου. Η χρονική παράμετρος συστημάτων χωρίς ενταμιευτή είναι πάντα (ανεξαρτήτως της συγκεκριμένης διαδικασίας εισόδου) κοντά στο 0. Για το λόγο αυτόν, μοντελοποιούμε την συμπεριφορά των πηγών στο σημείο λειτουργίας $(s^*,0)$. Η τρίτη εξίσωση του συστήματος (3.1) δεν μπορεί να ληφθεί υπόψιν λόγω της απουσίας ενταμιευτή ($t^* = 0$). Έτσι λοιπόν κάθε μοντέλο

το οποίο ικανοποιεί τις πρώτες δύο εξισώσεις του συστήματος (3.1) τελικά θα έχει τον ίδιο ασυμπτωτικό ρυθμό μείωσης της πιθανότητας υπερχείλισης όπως αυτός δίδεται από το ασυμπτωτικό πολλών πηγών. Οπότε όταν θέλουμε να αντικαταστήσουμε πραγματικές πηγές στο σημείο λειτουργίας $(s^*, 0)$, μπορούμε να λάβουμε υπόψιν και μοντέλα με δύο μόνο ελεύθερες παραμέτρους. Ένα τέτοιο μοντέλο είναι η κίνηση Brown με $\mathrm{Var}Z(t) = \sigma^2 t$ η οποία εμφανίζεται σε μοντέλα υψηλού φόρτου. Χρησιμοποιώντας την (3.2), το ισοδύναμο εύρος ζώνης της κίνησης Brown είναι

$$\alpha_M(s,t) = \lambda + \frac{s\sigma^2}{2} .$$

Λόγω του ότι η κίνηση Brown είναι μια πηγή Gaussian, οι πρώτες δύο εξισώσεις του (3.3) οδηγούν στις εξής σχέσεις:

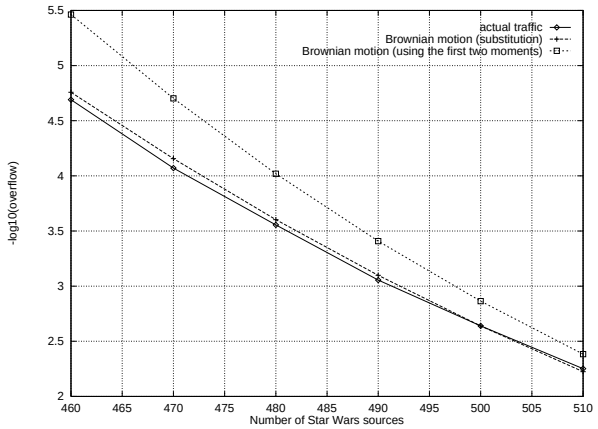
$$\lambda = \lim_{t\to 0} \left[\alpha_R(s^*,t) - s^* \left. \frac{\partial \alpha_R}{\partial s}(s,t) \right|_{s=s^*} \right]$$

και

$$\sigma^2 = 2 \lim_{t\to 0} \left. \frac{\partial \alpha_R}{\partial s}(s,t) \right|_{s=s^*} .$$

Η ακρίβεια της αντικατάστασης πραγματικών πηγών με μια πηγή μοντέλου του τύπου κίνησης Brown προκύπτει από τις καμπύλες του Σχήματος 3.1. Η πραγματική κίνηση αποτελείται από υπέρθεση ανεξάρτητων ταινιών video Star Wars κωδικοποιημένων σύμφωνα με το πρότυπο MPEG-1. Οι κωδικοποιημένες ταινίες video, έχουν γίνει δημόσια διαθέσιμες[2] από τον O. Rose [32] και ακολουθούν την εξής δομή πλαισίου: IBBPBBPBBPBB (12 πλαίσια) με 25 fps (πλαίσια ανά δευτερόλεπτο) ρυθμό μετάδοσης πλαισίων. Οι ταινίες video είναι αμοιβαία ανεξάρτητες με την έννοια ότι κάθε μία αρχίζει από ένα τυχαίο σημείο της ταινίας Star Wars ανεξάρτητο του σημείου που αρχίζουν οι υπόλοιπες ταινίες. Η ακρίβεια της αντικατάστασης συγκρίνεται με αυτήν της προσέγγισης μοντελοποίησης που βασίζεται στην εξίσωση των πρώτων δύο ροπών της περιθωριακής κατανομής της πραγματικής κυκλοφορίας και ενός μοντέλου κίνησης Brown. Είναι φανερό ότι για πιθανότητες υπερχείλισης μικρότερες από 10^{-3} (το οποίο είναι και το εύρος που μας ενδιαφέρει) η προσέγγισή μας στην αντικατάσταση κυκλοφορίας είναι σημαντικά καλύτερη της παραπάνω προσέγγισης μοντελοποίησης, η οποία βασίζεται στο Κεντρικό Οριακό Θεώρημα. Αυτό επιβεβαιώνει το γεγονός ότι οι εκτιμήσεις της πιθανότητας σπάνιων γεγονότων βάσει των φραγμάτων Chernoff είναι πιο ακριβείς από εκτιμήσεις βασιζόμενες στο Κεντρικό Οριακό Θεώρημα.

[2] Διαθέσιμες στο http://ftp-info3.informatik.uni-wuerzburg.de/pub/MPEG/



Σχήμα 3.1: Η πιθανότητα υπερχείλισης συνδέσμου χωρητικότητας 155 Mbps χωρίς ενταμιευτή, ως συνάρτηση του αριθμού των πηγών Star Wars που τροφοδοτούν τον σύνδεσμο.

3.5 Κλασματική Κίνηση Brown

Η κλασματική κίνηση Brown (fractional Brownian motion - fBm) [30, 31] είναι μια πηγή Gaussian με Var$Z(t) = \sigma^2 t^{2H}$, όπου σ και $H \in (0,1)$ σταθερές. Χρησιμοποιώντας επίσης την εξίσωση (3.2), το ισοδύναμο εύρος ζώνης της fBm είναι

$$\alpha_M(s,t) = \lambda + \frac{s}{2}\sigma^2 t^{2H-1} \ . \tag{3.4}$$

Η παράμετρος H ονομάζεται *παράμετρος Hurst* και ανάλογα εάν $H > \frac{1}{2}$ ή $H \leq \frac{1}{2}$ η πηγή έχει μακροπρόθεσμες ή βραχυπρόθεσμες εξαρτήσεις αντίστοιχα. Οι διαδικασίες που παρουσιάζουν μακροπρόθεσμες εξαρτήσεις έχουν μη ολοκληρώσιμη συνάρτηση αυτοσυσχέτισης, ενώ στις διαδικασίες με βραχυπρόθεσμες εξαρτήσεις η συνάρτηση αυτή είναι πάντα ολοκληρώσιμη. Ένα σημαντικό χαρακτηριστικό της fBm είναι ότι μπορεί να οριστεί πλήρως μέσω τριών παραμέτρων, όσες και οι άγνωστες παράμετροι του συστήματος (3.1).

Επιλύωντας το σύστημα (3.3) στην περίπτωση αυτή, οδηγούμαστε στο εξής ισοδύναμο σύστημα:

$$\begin{cases} \lambda = \alpha_R(s_0,t_0) - s_0 \ \dfrac{\partial \alpha_R}{\partial s}(s,t_0) \ |_{s=s_0} \\ \\ 2H - 1 = \dfrac{t_0 \ \dfrac{\partial \alpha_R}{\partial t}(s_0,t)\Big|_{t=t_0}}{s_0 \ \dfrac{\partial \alpha_R}{\partial s}(s,t_0)\Big|_{s=s_0}} \\ \\ \sigma^2 = \dfrac{2 \ \dfrac{\partial \alpha_R}{\partial s}(s,t_0)\Big|_{s=s_0}}{{t_0}^{2H-1}} \end{cases} \ . \tag{3.5}$$

Το ότι ο αριθμός των παραμέτρων είναι ίσος με τον αριθμό των εξισώσεων, μπορεί να δημιουργήσει το ερώτημα εάν υπάρχει αρκετή ελευθερία έτσι ώστε να εξασφαλιστεί ότι

τα μοντέλα fBm των οποίων οι παράμετροι λαμβάνονται μέσω των παραπάνω εξισώσεων είναι πάντα *έγκυρα*, δηλαδή $\lambda \in (0, C)$, $H \in (0, 1)$ και $\sigma^2 > 0$. Πρέπει να σημειώσουμε ότι η συνθήκη που αφορά το σ^2 ικανοποιείται πάντα αφού μια συνάρτηση ισοδυνάμου εύρους ζώνης έχει την ιδιότητα να είναι αύξουσα ως προς s [22]. Από την άλλη, για το $\lambda \in (0, C)$ δεν υπάρχει αλγεβρική απόδειξη για ένα οποιοδήποτε σημείο (s_0, t_0). Παρόλα αυτά, υπάρχουν σημαντικές ενδείξεις ότι η παραπάνω συνθήκη ικανοποιείται εάν το σημείο (s_0, t_0) είναι το σημείο λειτουργίας (s^*, t^*) όπως δείχνουμε στο τέλος αυτής της παραγράφου. Για τον λόγο αυτόν, επικεντρώνουμε τη μελέτη μας στις ιδιότητες της ισοδυναμίας της κυκλοφορίας R με κυκλοφορία μοντέλου fBm στο σημείο λειτουργίας ενώ λαμβάνουμε σαν υπόθεση ότι η συνθήκη $\lambda \in (0, C)$ όντως ισχύει, δηλαδή ότι

$$0 < \alpha_R(s^*, t^*) - s^* \left. \frac{\partial \alpha_R}{\partial s}(s, t^*) \right|_{s=s^*} < C \; . \tag{3.6}$$

Στο λήμμα που ακολουθεί εξάγουμε μια συνθήκη ισοδύναμη με τη δεξιά ανισότητα της εξίσωσης (3.6) η οποία εκφράζεται μέσω της σταθμισμένης συνάρτησης ασυμπτωτικού ρυθμού Φ_R την οποία ορίσαμε στην (2.3). Η συνθήκη αυτή θα χρησιμοποιηθεί παρακάτω στο κύριο αποτέλεσμα της παραγράφου αυτής (Θεώρημα 2).

Λήμμα 1 *Η συνθήκη* $\alpha_R(s^*, t^*) - s^* \, \frac{\partial \alpha_R}{\partial s}(s, t^*) \, |_{s=s^*} < C$ *είναι ισοδύναμη με τη*

$$2\frac{\mathbf{\Phi}_R}{B} > \frac{\partial \mathbf{\Phi}_R}{\partial B} \; . \tag{3.7}$$

Απόδειξη: Αφού το (s^*, t^*) είναι σημείο σάγματος της συνάρτησης $G_R(s, t)$, ισχύει

$$\frac{\partial G_R(s,t)}{\partial s}|_{s=s^*} = B + Ct^* - t^* \alpha_R(s^*, t^*) - s^* t^* \frac{\partial \alpha_R(s^*, t^*)}{\partial s}|_{s=s^*} = 0 \; ,$$

από όπου, βλ. και (2.3), έχουμε

$$s^* \frac{\partial \alpha_R(s^*, t^*)}{\partial s}|_{s=s^*} = C - \alpha_R(s^*, t^*) + \frac{B}{t^*} = \frac{\mathbf{\Phi}_R}{s^* t^*} \; . \tag{3.8}$$

Το γεγονός αυτό υποδεικνύει ότι η ανισότητα

$$\alpha_R(s^*, t^*) - s^* \, \frac{\partial \alpha_R}{\partial s}(s, t^*) \, |_{s=s^*} < C$$

είναι ισοδύναμη με

$$\frac{\mathbf{\Phi}_R}{s^* t^*} > \alpha_R(s^*, t^*) - C \; .$$

Χρησιμοποιώντας πάλι την (2.3), η τελευταία ανισότητα είναι ισοδύναμη με

$$\frac{\mathbf{\Phi}_R}{s^* t^*} > \frac{B}{t^*} - \frac{\mathbf{\Phi}_R}{s^* t^*} \; ,$$

και τελικά με την

$$2\frac{\Phi_R}{B} > s^* .$$

Το λήμμα έπεται λόγω του ότι

$$s^* = \frac{\partial \Phi_R}{\partial B} ,$$

το οποίο αποδεικνύεται παραγωγίζοντας τη Φ_R ως προς B και λαμβάνοντας υπόψιν το ότι το (s^*, t^*) είναι σαγματικό σημείο. ■

Το θεώρημα που ακολουθεί εξασφαλίζει ότι εάν η συνθήκη (3.6) ικανοποιείται στις περιπτώσεις ισοδυναμίας κυκλοφορίας όπου $V = 0$, η παράμετρος H όντως ανήκει στο $(0, 1)$, και συνεπώς το μοντέλο fBm που λαμβάνουμε είναι *έγκυρο*. Επίσης, το θεώρημα εξασφαλίζει ότι το σημείο λειτουργίας *πάντα διατηρείται* μετά από την αντικατάσταση.

Θεώρημα 2 *Έστω κυκλοφορία R με σημείο λειτουργίας (s^*, t^*). Εάν ικανοποιείται η (3.6), τότε υπάρχει έγκυρο μοντέλο fBm M ισοδύναμο με την R στο (s^*, t^*). Ακόμα, εάν το M αντικαταστήσει την R, διατηρείται το ίδιο σημείο λειτουργίας.*

Απόδειξη: Μπορεί να δειχθεί, παρόμοια με την (3.8) ότι

$$t^* \left.\frac{\partial \alpha_R(s^*, t^*)}{\partial t}\right|_{t=t^*} = C - \alpha_R(s^*, t^*) .$$

Σε συνδυασμό με την (3.8) και τη δεύτερη εξίσωση του συστήματος (3.5) λαμβάνουμε ότι

$$2H - 1 = \frac{C - \alpha_R(s^*, t^*)}{C - \alpha_R(s^*, t^*) + B/t^*} = 1 - \frac{B/t^*}{C - \alpha_R(s^*, t^*) + B/t^*} = 1 - \frac{B}{\Phi_R/s^*} , \qquad (3.9)$$

όπου κάναμε και χρήση του ορισμού της Φ_R στην (2.3). Λόγω του ότι η Φ_R είναι πάντα θετική, $H < 1$. Επιπλέον, είναι εύκολο να δειχθεί ότι η συνθήκη $H > 0$ είναι ισοδύναμη με τη $2\Phi/B > s^*$.

Στη συνέχεια, αποδεικνύουμε ότι το σημείο λειτουργίας (s^*, t^*) διατηρείται μετά την αντικατάσταση της κυκλοφορίας R από το μοντέλο M. Ειδικότερα, αποδεικνύουμε ότι η συνάρτηση $G_M(s, t)$ έχει ένα μοναδικό σημείο σάγματος στο (s^*, t^*). Για το σκοπό αυτόν, υπολογίζουμε ευθέως το σημείο σάγματος της $G_M(s, t)$. Σημειώστε πρώτα ότι από την (3.4) έχουμε

$$G_M(s, t) = s(B + (C - \lambda)t) - \frac{s^2}{2} t^{2H} .$$

Είναι προφανές ότι για κάθε σταθερό $t > 0$, το $\sup_s G_M(s, t)$ λαμβάνεται στο

$$\hat{s}_t = \frac{B + (C - \lambda)t}{\sigma^2 t^{2H}} .$$

Επιπλέον,

$$\inf_t \sup_s G_M(s, t) = \inf_t \frac{(B + (C - \lambda)t)^2}{2\sigma^2 t^{2H}} .$$

Για $t > 0$, η παράγωγος ως προς t της συνάρτησης

$$\frac{(B + (C - \lambda)t)^2}{2\sigma^2 t^{2H}} ,$$

ισούται με

$$\frac{(B + (C - \lambda)t)(C - \lambda)(1 - H)}{\sigma^2 t^{2H+1}}(t - \hat{t}) , \tag{3.10}$$

όπου

$$\hat{t} = \frac{H\,B}{(1 - H)(C - \lambda)} , \tag{3.11}$$

έκφραση η οποία είναι θετική λόγω της υπόθεσης (3.6) και της συνέπειάς της, δηλαδή ότι $H \in (0, 1)$. Αφού όλοι οι όροι που πολλαπλασιάζουν το $(t - \hat{t})$ είναι θετικοί και η παραπάνω παράγωγος μηδενίζεται μόνο στο $\hat{t}$, έπεται ότι το $\inf_t \sup_s G_M(s, t)$ λαμβάνεται στο $\hat{t}$ ενώ η συνάρτηση $\sup_s G_M(s, t)$ ως προς t δεν έχει άλλα τοπικά βέλτιστα. Οπότε, η $G_M(s, t)$ έχει ένα μοναδικό σημείο σάγματος στο $(\hat{s}_{\hat{t}}, \hat{t})$. Επειδή λόγω κατασκευής, οι παράγωγοι της $G_M(s, t)$ στο (s^*, t^*) είναι μηδέν, ισχύει $(\hat{s}_{\hat{t}}, \hat{t}) \equiv (s^*, t^*)$. ■

Όσον αφορά το κατά πόσο ισχύει η (3.7), υπάρχουν ενδιαφέρουσες ενδείξεις ότι σε πολλές περιπτώσεις ισχύει μια πιο ισχυρή συνθήκη, δηλαδή ότι

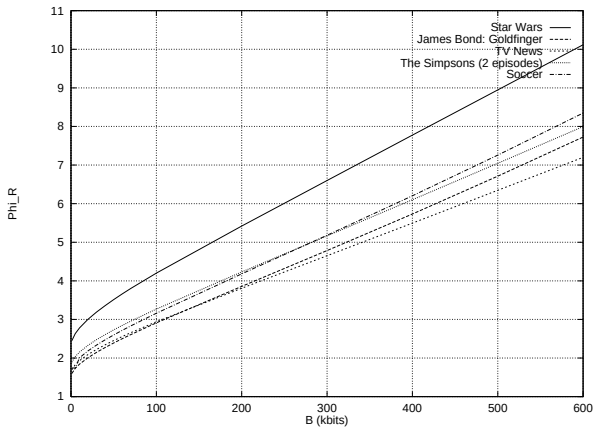
$$\frac{\Phi_R}{B} > \frac{\partial \Phi_R}{\partial B} . \tag{3.12}$$

Αυτό είναι ισοδύναμο με την κοιλότητα της σταθμισμένης συνάρτησης ασυμπτωτικού ρυθμού ως προς B. Πράγματι, στο [35] έχει αποδειχθεί ότι για υπέρθεση πηγών ON/OFF Markov Modulated Fluids και για μικρό B, ισχύει $\Phi_R \approx C_1 + C_2\sqrt{B}$, όπου $C_1, C_2 > 0$. Επίσης, για μεγάλο B η Φ_R συγκλίνει σε μια γραμμική συνάρτηση του B (βλ. [35]). Παρόμοιες ενδείξεις υπάρχουν και στο [10]. Από την άλλη, η κοιλότητα δεν ισχύει πάντοτε, όπως φαίνεται στην περίπτωση των πηγών Markov που εξετάζονται στο [4].

Λόγω του ότι δε μπορούμε να αποδείξουμε ότι οι (3.7) και (3.12) ισχύουν πάντοτε, στραφήκαμε σε πειραματική εξέτασή τους. Πειραματιστήκαμε εκτενώς χρησιμοποιώντας πραγματική κυκλοφορία (MPEG-1) ώστε να μελετήσουμε εάν οι (3.7) και (3.12) ισχύουν στις περιπτώσεις με πρακτικό ενδιαφέρον. Σε όλες τις περιπτώσεις και οι δύο συνθήκες ικανοποιούνταν. Στο Σχήμα 3.2 απεικονίζεται η Φ_R ως συνάρτηση του μεγέθους B του ενταμιευτή. Τα πειράματα έγιναν χρησιμοποιώντας κυκλοφορία που προέρχεται από υπέρθεση πολλών όμοιων και ανεξάρτητων ταινιών video, οι οποίες τροφοδοτούν σύνδεσμο χωρητικότητας 155 Mbps. Σε όλες τις περιπτώσεις ο βαθμός χρησιμοποίησης του συνδέσμου ήταν 88%.

Ενδιαφέρον προκαλεί το γεγονός ότι εάν η (3.12) ισχύει, τότε η (3.9) είναι ισοδύναμη με $H > 1/2$ και το αντίστοιχο μοντέλο fBm παρουσιάζει μακροπρόθεσμες εξαρτήσεις. Επίσης ενδιαφέρον είναι το ότι σε ορισμένες περιπτώσεις ένα μοντέλο fBm με μακροπρόθεσμες

εξαρτήσεις ($H > 1/2$) μπορεί να είναι ισοδύναμο με κυκλοφορία η οποία παρουσιάζει βραχυπρόθεσμες εξαρτήσεις. Για παράδειγμα, όπως ήδη αναφέραμε για μια υπέρθεση πηγών ON/OFF Markov Modulated Fluids η οποία τροφοδοτεί σύνδεσμο με μικρό B, ισχύει $\Phi_R \approx C_1 + C_2\sqrt{B}$ δηλαδή υπάρχει ισοδύναμο μοντέλο fBm με $H > 1/2$.



Σχήμα 3.2: Η Φ_R ως συνάρτηση του μεγέθους B του ενταμιευτή για διάφορες πηγές MPEG-1. Σε κάθε περίπτωση η κυκλοφορία που τροφοδοτεί τον σύνδεσμο (με χωρητικότητα 155 Mbps), προέρχεται από υπέρθεση πολλών όμοιων αλλά ανεξάρτητων ταινιών video. Σε όλα τα πειράματα, ο βαθμός χρησιμοποίησης του συνδέσμου ήταν 88%.

Κεφάλαιο 4

Πειραματική Αξιολόγηση της Μεθόδου Αντικατάστασης

Στο κεφάλαιο αυτό, μελετάμε πειραματικά την ακρίβεια της μεθόδου αντικατάστασης κυκλοφορίας που προτείναμε στο προηγούμενο κεφάλαιο. Διεξήχθησαν πειράματα τριών ειδών:

1. Μελέτη της ακρίβειας της αντικατάστασης, συγκρίνοντας τις πιθανότητες υπερχείλισης (Buffer Overflow Probability - BOP) σε σύνδεσμο, όταν αυτός τροφοδοτείται από την πραγματική κυκλοφορία με την BOP όταν ο ίδιος σύνδεσμος τροφοδοτείται με ένα μοντέλο ισοδύναμο της πραγματικής κυκλοφορίας. Στα πειράματα αυτού του είδους, εξετάζουμε την ακρίβεια της αντικατάστασης τόσο του συνόλου των πραγματικών πηγών που τροφοδοτούν ένα σύνδεσμο, όσο και μέρους των πηγών αυτών. Επίσης συγκρίνουμε την προσέγγισή μας με μια προσέγγιση αντικατάστασης κυκλοφορίας η οποία βασίζεται στην εξίσωση βασικών στατιστικών μεγεθών της πραγματικής κυκλοφορίας και της κυκλοφορίας ενός μοντέλου. Τα αποτελέσματα των πειραμάτων αυτού του είδους περιγράφονται στην Παράγραφο 4.1.

2. Στην Παράγραφο 4.2 γίνεται επιβεβαίωση του ισχυρισμού ότι διατηρώντας το t^* κατά την αντικατάσταση, διατηρούνται και οι κύριες χρονικές κλίμακες υπερχείλισης των ενταμιευτών. Αυτό το είδος πειραμάτων αφορά κυρίως την ακρίβεια του ασυμπτωτικού πολλών πηγών.

3. Σύγκριση της BOP του πραγματικού συστήματος με αυτή της προσέγγισής μας μετά από αντικατάσταση του συνόλου των πηγών, όπως επίσης και με τη BOP όταν χρησιμοποιούμε μόνο ισότητα του ισοδυνάμου εύρους ζώνης και όχι ολόκληρο το σύστημα εξισώσεων του Ορισμού 1. Η σύγκριση αυτή περιγράφεται στην Παράγραφο 4.3

και επιδεικνύει το γεγονός ότι οι συνθήκες πρώτης τάξεως είναι σημαντικές.

Σε όλα τα είδη πειραμάτων, η πιθανότητα υπερχείλισης για οποιαδήποτε κυκλοφορία εισόδου, εκτιμήθηκε χρησιμοποιώντας προσομοιώσεις Monte-Carlo.

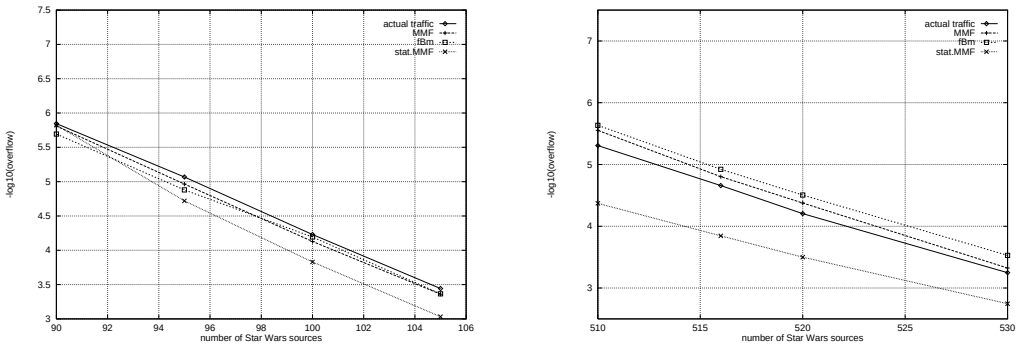
4.1 Ακρίβεια Αντικατάστασης

Στα πειράματα αυτού του είδους, η πραγματική κυκλοφορία που θεωρούμε είναι υπέρθεση ανεξάρτητων ταινιών video Star Wars κωδικοποιημένες σύμφωνα με το πρότυπο MPEG-1. Οι πηγές αυτές, πολυπλέκονται σε ένα σύνδεσμο 34 Mbps με μέγεθος ενταμιευτή 500 κελιά (περίπου 6 msec μέγιστη καθυστέρηση) και σε σύνδεσμο 155 Mbps 2000 κελιών (περίπου 5.5 msec μέγιστη καθυστέρηση). Εξετάζουμε την ακρίβεια δύο μοντέλων που χρησιμοποιήσαμε για αντικατάσταση: μιας πηγής fBm, και υπέρθεσης ανεξάρτητων και ομοίων πηγών ON/OFF Markov Modulated Fluids (ON/OFF MMF). Για τον τελευταίο τύπο πηγών, οι άγνωστες παράμετροι οι οποίες καθορίζονται βάσει του συστήματος (3.1), είναι ο ρυθμός ν μεταβάσεων από την κατάσταση OFF στην ON, ο ρυθμός μ μεταβάσεων από την κατάσταση ON στην OFF, ο ρυθμός μετάδοσης κυκλοφορίας η οποία παράγεται στην κατάσταση ON και ο αριθμός K των πηγών ON/OFF MMF. Το ισοδύναμο εύρος ζώνης μιας πηγής αυτού του τύπου έχει γνωστή αναλυτική μορφή (βλ. [22]). Για να καθοριστούν πλήρως οι παράμετροι (ν, μ, h και K) επιλύωντας το σύστημα (3.1) χρειάζεται ακόμα μια εξίσωση. Εκλέξαμε την εξής: ο μέσος ρυθμός των πηγών ON/OFF MMF να είναι ίσος με τον μέσο ρυθμό των πραγματικών πηγών.

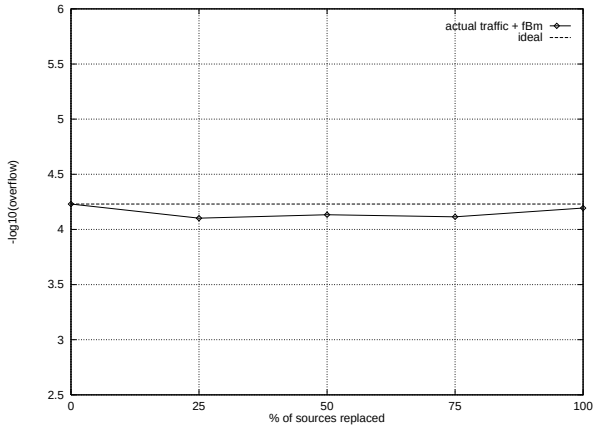
Όπως αναφέραμε παραπάνω, επιδεικνύουμε την πρακτική αξία της προσέγγισής μας συγκρίνοντάς την με ευρέως χρησιμοποιούμενες προσεγγίσεις μοντελοποίησης (βλ Παρ. 1.2). Για το σκοπό αυτό, υπολογίζουμε την BOP ενός συστήματος το οποίο τροφοδοτείται από υπέρθεση ανεξάρτητων και όμοιων πηγών ON/OFF MMF οι οποίες προέρχονται από μια μέθοδο ταύτισης στατιστικών ιδιοτήτων, κατά την οποία οι παράμετροι των παραπάνω πηγών καθορίζονται εξισώνοντας τον μέσο ρυθμό, τον μέγιστο ρυθμό, τη διασπορά του ρυθμού και τον μέσο χρόνο στον οποίο ο ρυθμός μετάδοσης δεν υπερβαίνει το 10% του μέσου ρυθμού, με τα αντίστοιχα στατιστικά μεγέθη της πραγματικής κυκλοφορίας. Στο Σχήμα 4.1 απεικονίζεται η ακρίβεια των μοντέλων που περιγράψαμε παραπάνω, όπου με ‘‘stat.MMF’’ συμβολίζουμε την παραπάνω μέθοδο ταύτισης στατιστικών. Στο σχήμα φαίνεται ότι η BOP της προσέγγισής μας είναι πολύ κοντά σε αυτή των πραγματικών πηγών και για τα δύο μοντέλα που χρησιμοποιήσαμε (ON/OFF MMF και fBm). Αντίθετα η μέθοδος stat.MMF είναι σημαντικά ανακριβής, ιδιαίτερα στο σύνδεσμο των 155 Mbps.

Στο Σχήμα 4.2 απεικονίζεται η BOP όταν αντικαθιστούμε μέρος των 100 πηγών ταινιών

video Star Wars που συνιστούν την πραγματική κυκλοφορία σε αυτή την περίπτωση. Στα πειράματα αυτά χρησιμοποιήσαμε σύνδεσμο των 34 Mbps με μέγεθος ενταμιευτή 500 κελιά. Τα αποτελέσματα της μεθόδου μας συμφωνούν εξαιρετικά καλά με την BOP του συνόλου των πραγματικών πηγών (η οποία αντιστοιχεί στην ‘‘αντικατάσταση’’ 0% των πηγών).



Σχήμα 4.1: Η BOP για διάφορα μοντέλα που χρησιμοποιήθηκαν για αντικατάσταση σε σύγκριση με την BOP των πραγματικών πηγών για σύνδεσμο 34 Mbps (α) και 155 Mbps (β), ως συνάρτηση του αριθμού των πηγών Star Wars των οποίων η υπέρθεση συνιστά την πραγματική κυκλοφορία εισόδου.

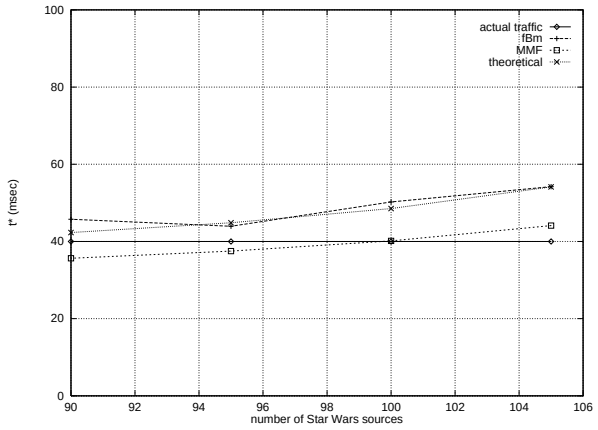


Σχήμα 4.2: Η BOP που επιτυγχάνεται όταν αντικαθίστανται μέρος των πηγών εισόδου από πηγές βασισμένες σε μοντέλα fBm. Η BOP της αρχικής κυκλοφορίας αντιστοιχεί σε ‘‘αντικατάσταση’’ του 0% της κυκλοφορίας.

4.2 Χρονική Κλίμακα Υπερχείλισης

Στο δεύτερο τύπο πειραμάτων, εκτιμάμε τον πιθανότερο χρόνο t_f που απαιτείται για να γεμίσει ο ενταμιευτής στις περιπτώσεις πραγματικής κυκλοφορίας και της ισοδύναμης αυτής. Και για τις δύο περιπτώσεις κυκλοφορίας περιμένουμε οι χρόνοι t_f να είναι κοντά

στους αντίστοιχους t^*. Πράγματι, όπως περιμέναμε, αυτό ισχύει στις περιπτώσεις που μελετήσαμε όπως φαίνεται στο Σχήμα 4.3, όπου πηγές fBm και ON/OFF MMF αντικαθιστούν υπερθέσεις πηγών ταινιών video Star Wars οι οποίες πολυπλέκονται σε σύνδεσμο 34 Mbps με ενταμιευτή μεγέθους 500 κελιών. Η τιμή t^* αντιστοιχεί στην καμπύλη ''theoretical'' του σχήματος. Η συμφωνία των τιμών είναι πολύ καλή, ιδιαίτερα για το μοντέλο fBm, γεγονός το οποίο επιβεβαιώνει ότι πράγματι οι κύριες χρονικές κλίμακες υπερχείλισης διατηρούνται κατά την αντικατάσταση.

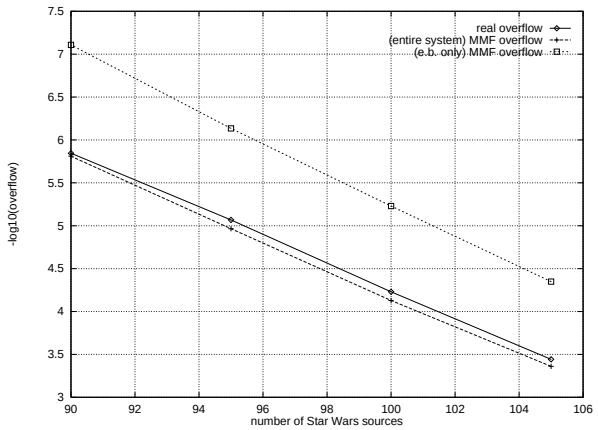


Σχήμα 4.3: Απεικόνιση του πιθανότερου χρόνου στον οποίο γεμίζει ο ενταμιευτής, για την πραγματική κυκλοφορία και πηγές MMF και fBm οι οποίες αντικαθιστούν τις πραγματικές σε σύνδεσμο 34 Mbps με ενταμιευτή μεγέθους 500 κελιά.

4.3 Αντικατάσταση βάσει μόνο του Ισοδυνάμου Εύρους Ζώνης

Στο τελευταίο είδος πειραμάτων, μελετάμε την επίδοση (BOP) του συστήματος όταν η αντικατάσταση κυκλοφορίας γίνεται λαμβάνοντας υπόψιν μόνο ισότητα των ισοδυνάμων ευρών ζώνης, αγνοώντας τις συνθήκες πρώτης τάξης της ισοδυναμίας κυκλοφορίας. Όπως και παραπάνω, θεωρούμε ότι πολλές πηγές Star Wars πολυπλέκονται σε σύνδεσμο 34 Mbps με ενταμιευτή μεγέθους 500 κελιών. Η ισότητα των ισοδυνάμων ευρών ζώνης μπορεί να επιτευχθεί με πολλούς συνδυασμούς των παραμέτρων των πηγών ON/OFF MMF. Αυτό συμβαίνει γιατί οι πηγές αυτές καθορίζονται πλήρως από τέσσερις παραμέτρους, ενώ στην περίπτωση αυτή λαμβάνουμε μία σχέση μόνο υπόψιν (ισότητα των ισοδυνάμων ευρών ζώνης). Το γεγονός αυτό, όπως φαίνεται και στο Σχήμα 4.4, μπορεί να οδηγήσει σε συνδυασμό παραμέτρων των οποίων τα αντίστοιχα μοντέλα αποτυγχάνουν να προσεγγίσουν την απόδοση του πραγματικού συστήματος. Αντίθετα, εάν αντικαταστήσουμε βάσει της ισοδυναμίας κυκλοφορίας πετυχαίνουμε σχεδόν τέλεια συμφωνία της επίδοσης. Η σύγκριση αυτή, επιδεικνύει τη σπουδαιότητα των συνθηκών πρώτης τάξεως κατά την εξαγωγή της

ισοδύναμης κυκλοφορίας.



Σχήμα 4.4: Σύγκριση της BOP όταν δύο μοντέλα ON/OFF MMF αντικαθιστούν την πραγματική κυκλοφορία. Το ένα από αυτά βασίζεται στην έννοια της ισοδυναμίας κυκλοφορίας ενώ το άλλο βασίζεται σε ισότητα μόνο των ισοδυνάμων ευρών ζώνης και δε λαμβάνει υπόψιν τις συνθήκες πρώτης τάξεως.

Κεφάλαιο 5

Υπολογισμός της Πιθανότητας Υπερχείλισης σε Πραγματικό Χρόνο

5.1 Ταχύς Υπολογισμός του Σημείου Λειτουργίας

Η εφαρμογή της μεθόδου μας στην αντικατάσταση κυκλοφορίας προϋποθέτει τη γνώση του σημείου λειτουργίας (s^*, t^*). Έχουμε ήδη αναφέρει ότι το σημείο αυτό μπορεί να βρεθεί, επιλύωντας αριθμητικά το πρόβλημα διπλής βελτιστοποίησης της παράστασης `infsup` στην εξίσωση (2.2). Η λύση αυτού συνιστά σημείο σάγματος της συνάρτησης $G_R(s,t)$. Δεν υπάρχει κάποια γενική μέθοδος λύσης προβλημάτων αυτής της μορφής, ιδιαίτερα όταν η συνάρτηση $\alpha_R(s,t)$ δεν εκφράζεται αναλυτικά αλλά υπολογίζεται βάσει αρχείων-ιχνών πραγματικής κυκλοφορίας. Μια συνήθης μέθοδος λύσης είναι η εύρεση του μεγίστου ως προς s για κάθε t και εν συνεχεία του ελαχίστου ως προς t για κάθε μέγιστο που βρήκαμε στο προηγούμενο βήμα. Η μέθοδος αυτή απαιτεί πολλούς υπολογισμούς γιατί πρέπει να λυθούν *διαδοχικά* δύο προβλήματα βελτιστοποίησης[1]. Έχοντας υπόψιν ότι ακόμα και μία μόνο αποτίμηση (σε ένα συγκεκριμένο σημείο (s,t)) της συνάρτησης ισοδυνάμου εύρους ζώνης για ένα αρχείο-ίχνος πραγματικής κυκλοφορίας μπορεί να απαιτήσει μερικά δευτερόλεπτα υπολογισμών, είναι φανερό ότι η παραπάνω μέθοδος είναι ακατάλληλη για την εύρεση του σημείου λειτουργίας σε πραγματικό χρόνο.

Προτείνουμε μια επαναληπτική μέθοδο για τον υπολογισμό του σημείου λειτουργίας κάνοντας χρήση της ισοδυναμίας και της αντικατάστασης κυκλοφορίας. Η βασική ιδέα είναι η κατασκευή μιας απεικόνισης η οποία έχει σταθερό σημείο στο σημείο λειτουργίας. Η μέθοδος αρχίζει με μια πρώτη (αυθαίρετη) εκτίμηση (s_0, t_0) του σημείου λειτουργίας. Ακολούθως, αντικαθιστούμε *στο σημείο αυτό*, την πραγματική κυκλοφορία με ένα μοντέλο

[1]Μπορούμε να εκμεταλλευτούμε την κυρτότητα της $s(B+Ct) - st\alpha_R(s,t)$ ως προς s ώστε να επιταχύνουμε τη λύση των προβλημάτων μεγιστοποίησης ως προς s.

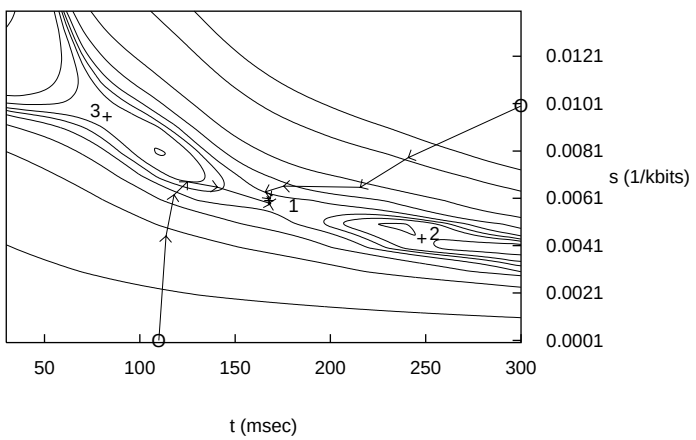
fBm του οποίου οι παράμετροι καθορίζονται λύνοντας το σύστημα (3.5). Έστω fBm_0 το μοντέλο που λάβαμε στο προηγούμενο βήμα, του οποίου οι παράμετροι είναι λ_0, σ_0, H_0. Στη συνέχεια, υπολογίζουμε τη νέα εκτίμηση (s_1, t_1) το οποίο είναι το σημείο λειτουργίας του συνδέσμου ο οποίος τροφοδοτείται από το μοντέλο fBm_0. Το σημείο αυτό μπορεί να υπολογισθεί σε κλειστή μορφή:

$$t_1 = \frac{H_0 B}{(1-H_0)(C-\lambda_0)}$$
$$s_1 = \frac{B + (C-\lambda_0)t_1}{\sigma_0^2 t_1^{2H_0}} .$$

Η μέθοδος επαναλαμβάνει τα παραπάνω βήματα χρησιμοποιώντας το σημείο (s_1, t_1) ως την τρέχουσα εκτίμηση του σημείου λειτουργίας, κατόπιν το (s_2, t_2) κτλ.

Είναι εύκολο να δειχθεί ότι το ζητούμενο σημείο λειτουργίας (s^*, t^*) είναι σταθερό σημείο του παραπάνω επαναληπτικού αλγορίθμου αφού εάν είχαμε το σημείο (s^*, t^*) ως την αρχική εκτίμηση, θα λαμβάναμε ένα μοντέλο fBm το οποίο θα διατηρούσε το σημείο λειτουργίας (βλ. Παρ. 3.5). Δυστυχώς δεν έχουμε αποδείξει ως τώρα ιδιότητες της σύγκλισης, αν και πειραματικά έχουμε εξαιρετικά θετικές ενδείξεις: συνήθως 6 βήματα του αλγορίθμου είναι αρκετά ενώ δεν έχουμε παρατηρήσει περίπτωση στην οποία ο αλγόριθμος αποκλίνει για κανένα σημείο εκκίνησης. Στο Σχήμα 5.1 απεικονίζονται οι ακολουθίες βημάτων του αλγορίθμου για δύο σημεία εκκίνησης.

Πρέπει να προσθέσουμε ότι η πραγματική κυκλοφορία (MPEG) η οποία χρησιμοποιήθηκε στα πειράματα, παρουσίασε πάνω από ένα σαγματικά σημεία. Παρά το γεγονός αυτό, τα σημεία αυτά είχαν σημαντική διαφορά στην τιμή τους σε σύγκριση με αυτή του σημείου λειτουργίας. Είναι πιθανό να κατασκευαστεί παράδειγμα στο οποίο υπάρχει περίπτωση περισσότερες από μία χρονικές κλίμακες να είναι υπεύθυνες για την υπερχείλιση του ενταμιευτή, οπότε και ο αλγόριθμος μπορεί να συγκλίνει σε διαφορετικά σημεία ανάλογα με το σημείο εκκίνησης. Συγκεκριμένα στο παράδειγμα του Σχήματος 5.1, εάν επιλέξουμε ως σημείο εκκίνησης το $s = 0.004, t = 220$ ο αλγόριθμος συγκλίνει σε λάθος σημείο λειτουργίας. Αν και η περίπτωση λάθος σύγκλισης είναι δυνατή, εντούτοις το σημείο εκκίνησης στην περίπτωση αυτή έπρεπε να επιλεχθεί πολύ κοντά στο λάθος σημείο σάγματος. Αντίθετα, ακόμα και εάν επιλέξουμε το σημείο εκκίνησης να απέχει πολύ από το σωστό σημείο σάγματος, ο αλγόριθμος συγκλίνει στο σωστό σημείο.



Σχήμα 5.1: Τα βήματα τα οποία ακολούθησε η μέθοδος γρήγορου υπολογισμού του σημείου λειτουργίας. 510 πηγές Star Wars πολυπλέκονται σε σύνδεσμο ($C =$) 155 Mbps με ενταμιευτή μεγέθους (B $=$) 2000 κελιών. Απεικονίζονται οι ισοσταθμικές καμπύλες της συνάρτησης $G_R(s,t)$ για τη συνολική πραγματική κυκλοφορία. Ενδεικτικά ελήφθησαν δύο σημεία εκκίνησης, το $s = 0.01$, $t = 300$ και το $s = 0.0001$, $t = 110$. Είναι φανερό ότι η σύγκλιση είναι γρήγορη και για τα δύο αρχικά σημεία. Το σημείο λειτουργίας του συστήματος είναι το $s = 0.006$, $t = 168$ (‘1’ στο σχήμα), ενώ υπάρχουν άλλα δύο σταθερά (σαγματικά) σημεία, το $s = 0.0042$, $t = 248$ και το $s = 0.0097$, $t = 82$ (‘2’ και ‘3’ στο σχήμα). Η μέθοδος συνέκλινε στο σωστό σημείο λειτουργίας και για τα δύο σημεία εκκίνησης του αλγορίθμου.

5.2 Υπολογισμός της Πιθανότητας Υπερχείλισης σε ένα Σύνδεσμο ATM

Σε ένα δίκτυο ATM το οποίο εκμεταλλεύεται τη στατιστική πολυπλεξία, σκοπός είναι η υψηλή χρησιμοποίηση του δικτύου και η τήρηση των απαιτήσεων σε QoS με τη βοήθεια των μηχανισμών *Ελέγχου Κυκλοφορίας και Συμφόρησης*. Ο *Έλεγχος Αποδοχής Κλήσης* (Call Admission Control - CAC) είναι μια από τις κύριες λειτουργίες που περιλαμβάνουν οι μηχανισμοί αυτοί, ο οποίος αποφασίζει εάν μια νέα κλήση θα γίνει δεκτή ή όχι, βασιζόμενος στα χαρακτηριστικά κυκλοφορίας της κλήσης αυτής, τις απαιτήσεις σε QoS, όπως επίσης και στο φόρτο του δικτύου τη στιγμή άφιξης της νέας κλήσης. Για αυτόν τον λόγο, είναι σημαντικό το δίκτυο να έχει μηχανισμούς οι οποίοι θα εκτιμούν γρήγορα και με ακρίβεια την πιθανότητα απώλειας κελιού (Cell Loss Probability - CLP) των συνδέσμων ATM.

Το πρόβλημα που υπάρχει είναι ότι τα γεγονότα που μας ενδιαφέρουν είναι σπάνια (π.χ. πιθανότητα απώλειας κελιού της τάξης 10^{-6}-10^{-9}), με αποτέλεσμα κάθε προσπάθεια άμεσης εκτίμησης της συχνότητας των γεγονότων αυτών να είναι ακατάλληλη λόγω του μεγάλου χρόνου που απαιτείται για μια ακριβή εκτίμηση. Για παράδειγμα, έστω ότι η CLP είναι ίση με p και τα γεγονότα απώλειας κελιού είναι ανεξάρτητα, και ισοπίθανα κάθε χρονική στιγμή. Για να εκτιμήσουμε τη CLP με κάποια λογικά διαστήματα εμπιστοσύνης (80%-95%), εκτιμώντας μόνο το πηλίκο των κελιών που χάθηκαν προς το συνολικό αριθμό κελιών που παρήχθησαν στο σύνδεσμο, απαιτείται η παρατήρηση αριθμού κελιών της τάξεως $10^3 p^{-1}$[6]. Έστω ότι έχουμε ένα σύνδεσμο 155 Mbps ο οποίος χρησιμοποιείται σε βαθμό 80% με αναμενόμενη CLP της τάξεως 10^{-6}. Τότε, μια ακριβής εκτίμηση απαιτεί χρόνο περίπου $3 \cdot 10^3$ δευτερόλεπτα. Ένας τέτοιος χρόνος είναι ακατάλληλος για εκτίμηση της CLP σε πραγματικό χρόνο, αφού οι αποφάσεις για αποδοχή νέων κλήσεων στο δίκτυο πρέπει να γίνονται γρήγορα λόγω του ότι η σύνθεση της κυκλοφορίας μπορεί να αλλάξει σε τόσο μεγάλα χρονικά διαστήματα. Ένα σχετικό με τη CLP μέγεθος που χαρακτηρίζει την QoS είναι η πιθανότητα υπερχείλισης (Buffer Overflow Probability - BOP) η οποία είναι ένα άνω φράγμα της CLP.

Στην παράγραφο αυτήν, προτείνουμε μια νέα μέθοδο εκτίμησης της πιθανότητας υπερχείλισης (BOP) που βασίζεται στο ασυμπτωτικό πολλών πηγών (2.2) και στον γρήγορο υπολογισμό του σημείου λειτουργίας που περιγράφηκε στην Παράγραφο 5.1. Η ιδέα είναι η επίλυση του συστήματος (3.5) να γίνεται σε πραγματικό χρόνο. Αυτό μπορεί να γίνει κάνοντας εκτίμηση του ισοδυνάμου εύρους ζώνης και των παραγώγων της εισερχόμενης στο σύνδεσμο κυκλοφορίας.

Έστω ότι ανεξάρτητες πηγές τροφοδοτούν έναν σύνδεσμο όπου η αρχική εκτίμηση

της χρονικής παραμέτρου είναι t_0 και η κυκλοφορία της πηγής j σε κάθε χρονική στιγμή περιγράφεται από την ακολουθία $\{X_1^j, X_2^j, \ldots\}$. Είναι λογικό να θεωρηθεί ένα μοντέλο διακριτού χρόνου δεδομένου ότι στην πράξη η υλοποίηση μοντέλου συνεχούς χρόνου δεν είναι δυνατή. Ως χρονική μονάδα θέτουμε το διάστημα λήψης μιας μέτρησης. Έτσι, για την εκτίμηση του φόρτου σε χρονικά διαστήματα διάρκειας t_0 και $t_0 + 1$ (βλ. (2.1)) σχηματίζουμε τα αθροίσματα

$$\hat{X}_1^j := \sum_{i=1}^{t_0} X_i^j \,, \quad \hat{X}_2^j := \sum_{i=t0+1}^{2t_0+1} X_i^j \,, \quad \ldots \tag{5.1}$$

$$\hat{X'}_1^j := \sum_{i=1}^{t_0+1} X_i^j \,, \quad \hat{X'}_2^j := \sum_{i=t_0}^{2t_0+1} X_i^j \,, \quad \ldots \quad . \tag{5.2}$$

Υποθέτοντας ότι οι αυξήσεις της διαδικασίας εισόδου είναι στάσιμες, εκτιμούμε το ισοδύναμο εύρος ζώνης στις χρονικές κλίμακες t_0 και $t_0 + 1$ αντικαθιστώντας τη μέση τιμή της (2.1) με τον δειγματικό μέσο K αθροισμάτων των εξισώσεων (5.1) και (5.2):

$$\hat{\alpha}_j(s_0, t_0) = \frac{1}{s_0 t_0} \log \frac{1}{K} \sum_{k=1}^{K} e^{s_0 \hat{X}_k^j} \,, \quad \hat{\alpha}_j(s_0, t_0 + 1) = \frac{1}{s_0(t_0 + 1)} \log \frac{1}{K} \sum_{k=1}^{K} e^{s_0 \hat{X'}_k^j} \,. \tag{5.3}$$

Θεωρώντας ότι οι πηγές που εισέρχονται στο σύνδεσμο είναι ανεξάρτητες, μπορούμε να συνδυάσουμε τις παραπάνω εκτιμήσεις για κάθε πηγή, ώστε να εκτιμήσουμε το ισοδύναμο εύρος ζώνης $\alpha_R(s_0, t_0)$ του συνόλου των πηγών:

$$\hat{\alpha}_R(s_0, t_0) = \sum_{j=1}^{J} \hat{\alpha}_j(s_0, t_0) \,,$$

όπου J είναι το πλήθος των πηγών που τροφοδοτούν τον σύνδεσμο. Με παρόμοιο τρόπο εκτιμούμε τα $\alpha_R(s_0, t_0 + 1)$ και $\alpha_R(s_0 + \Delta s, t_0)$, όπου $\Delta s > 0$.

Χρησιμοποιώντας τη διαφορά $\hat{\alpha}_R(s_0, t_0 + 1) - \hat{\alpha}_R(s_0, t_0)$ προσεγγίζουμε την $\frac{\partial \alpha_j(s0,t)}{\partial t}\Big|_{t=t_0}$, ενώ με $\frac{\hat{\alpha}_R(s_0+\Delta s,t_0)-\hat{\alpha}_R(s_0,t_0)}{\Delta s}$ προσεγγίζουμε την $\frac{\partial \alpha_j(s,t_0)}{\partial s}\Big|_{s=s_0}$.

Έχοντας τώρα εκτιμήσεις για τις ποσότητες που εμπλέκονται στα δεξιά μέλη των εξισώσεων του συστήματος (3.5) υπολογίζουμε ένα ισοδύναμο με την κυκλοφορία R, μοντέλο fBm. Εφαρμόζοντας τη μέθοδο που περιγράφηκε στην προηγούμενη παράγραφο, υπολογίζουμε τη νέα εκτίμηση (s_1, t_1) για το σημείο λειτουργίας. Υπολογίζουμε την πιθανότητα υπερχείλισης χρησιμοποιώντας τη σχέση (2.3) και επαναλαμβάνουμε τις παραπάνω εκτιμήσεις για το (s_1, t_1), υπολογίζοντας εκ νέου τα $\hat{\alpha}_R(s_1, t_1)$, $\hat{\alpha}_R(s_1, t_1 + 1)$ και $\hat{\alpha}_R(s_1 + \Delta s, t_1)$.

Το πλεονέκτημα της μεθόδου αυτής είναι ότι δεν παίρνουμε μόνο μια εκτίμηση για την τιμή της σταθμισμένης συνάρτησης ασυμπτωτικού ρυθμού, αλλά και εκτιμήσεις για το σημείο λειτουργίας. Το γεγονός αυτό είναι σημαντικό γιατί μας επιτρέπει να χρησιμοποιήσουμε βελτιώσεις του ασυμπτωτικού πολλών πηγών, όπως η βελτίωση Bahadur-Rao (2.6) και η

προσέγγιση της CLP (2.9), οι οποίες χρησιμοποιούν τον ίδιο ορισμό του σημείου λειτουργίας (βλ. Παρ. 2.2) το οποίο έχει σα συνέπεια το σημείο λειτουργίας να είναι ανεξάρτητο της βελτίωσης του ασυμπτωτικού πολλών πηγών που χρησιμοποιούμε κάθε φορά.

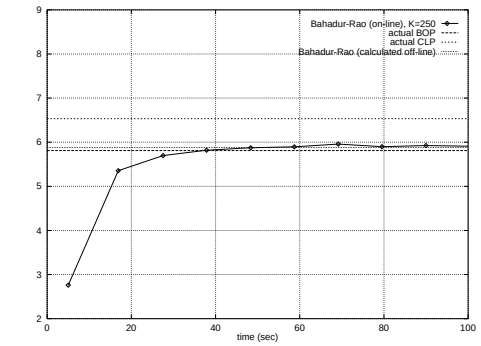
Στην επόμενη παράγραφο εξετάζουμε την επίδοση της μεθόδου που περιγράψαμε χρησιμοποιώντας πηγές κυκλοφορίας Markov καθώς και κυκλοφορία που προέρχεται από πραγματικές πηγές.

5.3 Πειραματικά Αποτελέσματα

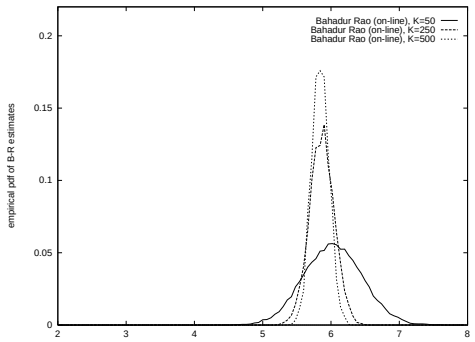
5.3.1 Πηγές ON/OFF Markov Modulated Fluids

Προσομοιώνουμε ένα σύνδεσμο ATM χωρητικότητας 34 Mbps και μεγέθους ενταμιευτή 500 κελιών. Ο σύνδεσμος τροφοδοτείται από 27 ανεξάρτητες πηγές τύπου ON/OFF Markov Modulated Fluids (ON/OFF MMF). Ο ρυθμός μεταβάσεων από την κατάσταση OFF στην ON είναι 0.02244 msec^{-1}, ο ρυθμός μεταβάσεων από την ON στην OFF είναι 0.04008 msec^{-1}, ενώ ο ρυθμός μετάδοσης πληροφορίας στην κατάσταση ON είναι 2.14498 Mbps.

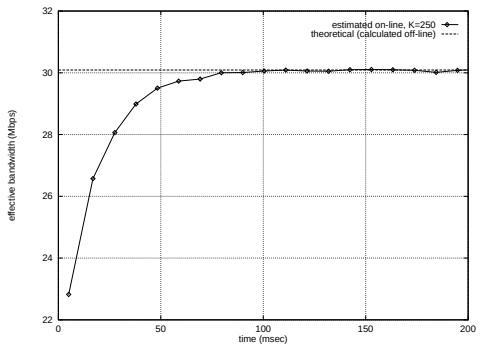
Εκτός από τις εκτιμήσεις του σημείου λειτουργίας και του ισοδυνάμου εύρους ζώνης που χρησιμοποιεί η μέθοδος σε κάθε βήμα, βάσει αυτών δίδονται εκτιμήσεις της σταθμισμένης συνάρτησης ασυμπτωτικού ρυθμού και της βελτίωσης Bahadur-Rao. Η BOP εκτιμάται μέσω της βελτίωσης Bahadur-Rao. Στο Σχήμα 5.2 (a)-(b), απεικονίζεται ο δειγματικός μέσος της της εκτίμησης της BOP και η εμπειρική κατανομή της αντίστοιχα, για 100 ανεξάρτητα πειράματα του συστήματος που περιγράψαμε. Η 'actual BOP' δίνει την πραγματική τιμή της πιθανότητας υπερχείλισης, η 'actual CLP' δίνει την πραγματική τιμή της CLP, ενώ η καμπύλη 'Bahadur-Rao (calculated off-line)' είναι η τιμή της βελτίωσης Bahadur-Rao βάσει της (2.6) επιλύωντας αριθμητικά το πρόβλημα βελτιστοποίησης της παράστασης `infsup`. Παρατηρούμε ότι η μέθοδός μας κάνει γρήγορη και ακριβή (ήδη από το δεύτερο βήμα) εκτίμηση της 'Bahadur-Rao (calculated off-line)' ενώ υπερεκτιμά τη CLP, όπως είναι επιθυμητό. Στο Σχήμα 5.2 (a) χρησιμοποιήθηκαν $K = 250$ όροι στην εξίσωση (5.3), ενώ στο (b) χρησιμοποιήθηκαν 50, 250 και 500. Όπως αναμένουμε υπάρχει ελάττωση της διασποράς των εκτιμήσεων καθώς αυξάνουν οι όροι που χρησιμοποιούμε στην εκτίμηση του ισοδυνάμου εύρους ζώνης. Έτσι, εάν θέλουμε μια πολύ ακριβή εκτίμηση για τη BOP, αρκεί να χρησιμοποιήσουμε αρκετά μεγάλο K, με συνέπεια να μεγαλώνει ανάλογα ο χρόνος που απαιτείται για τη λήψη των K όρων. Παρατηρούμε ότι η επιλογή $K = 250$, είναι ικανοποιητική στη συγκεκριμένη περίπτωση, γεγονός που μας επιτρέπει να εκτιμάμε την BOP σε αρκετά μικρό χρονικό διάστημα (περίπου 40 sec όπως φαίνεται στο Σχήμα 5.2 (a)). Όπως φαίνεται στα Σχήματα 5.2 (c)-(h), η μέθοδός μας παρέχει γρήγορα, αμερόληπτους εκτιμητές με μικρή διασπορά και για



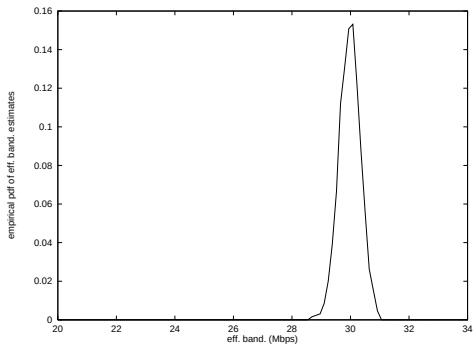
(a)



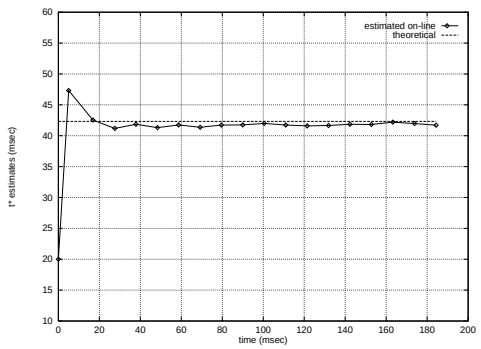
(b)



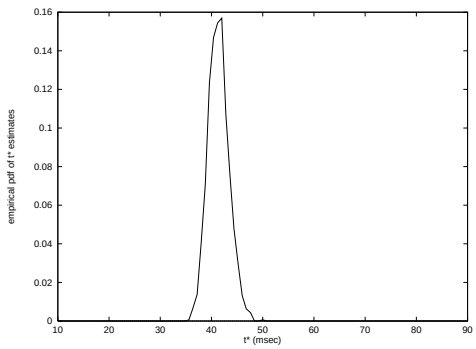
(c)



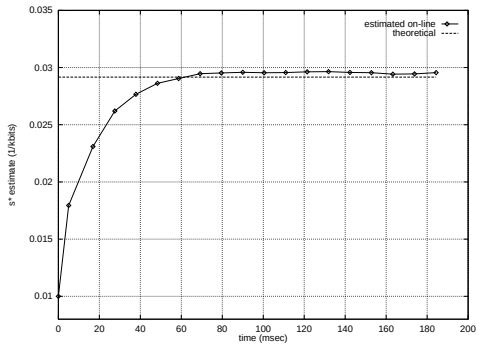
(d)



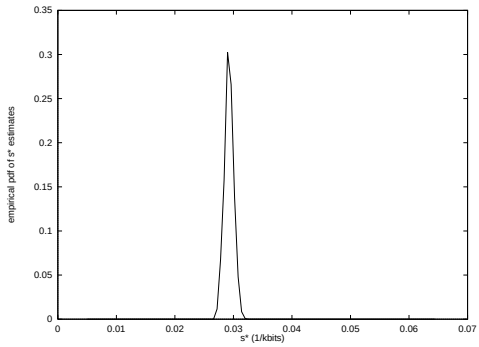
(e)



(f)



(g)



(h)

Σχήμα 5.2: Επίδοση της μεθόδου εκτίμησης σε πραγματικό χρόνο της BOP (a)-(b), του ισοδυνάμου εύρους ζώνης (c)-(d), της παραμέτρου χρόνου του συστήματος (e)-(f) και της παραμέτρου χώρου του συστήματος (g)-(h) για πηγές ON/OFF MMF.

τις υπόλοιπες ποσότητες που μας ενδιαφέρουν όπως, το ισοδύναμο εύρος ζώνης (c)-(d), και τη χρονική και χωρική παράμετρος του συστήματος (Σχήματα 5.2 (e)-(f) και (g)-(h) αντίστοιχα).

5.3.2 Πραγματικές Πηγές Κυκλοφορίας

Στην παράγραφο αυτή χρησιμοποιούμε κυκλοφορία προερχόμενη από την ταινία video Star Wars κωδικοποιημένη σύμφωνα με το πρότυπο MPEG-1, όπως και στο Κεφάλαιο 4. Ο σύνδεσμος που χρησιμοποιήθηκε έχει ρυθμό εξυπηρέτησης 34 Mbps και χώρο ενταμίευσης 500 κελιών. Τα αποτελέσματα αφορούν 250 ανεξάρτητα όμοια πειράματα, όπου πολυπλέκονται 90 ανεξάρτητες πηγές Star Wars.

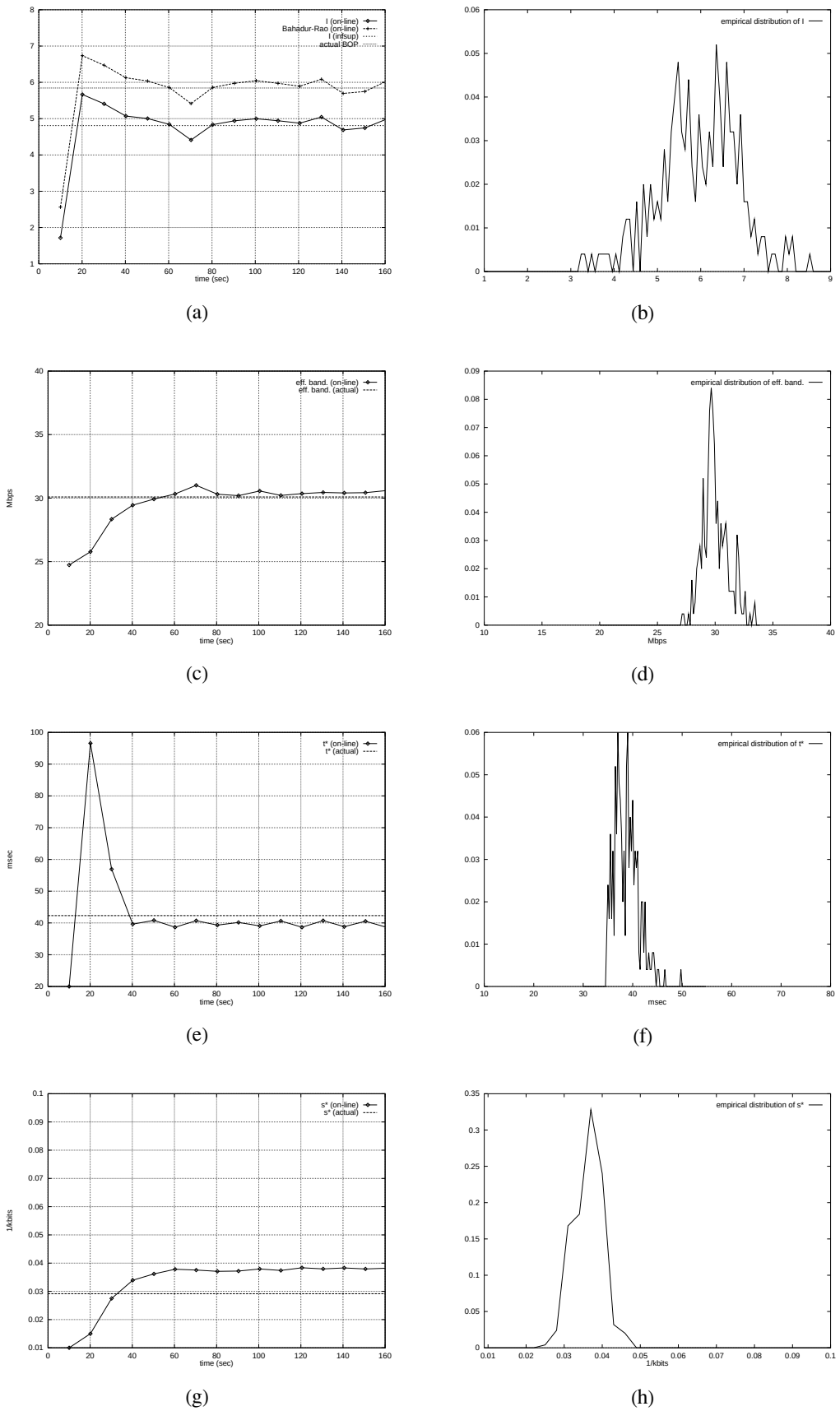
Όπως και στην προηγούμενη παράγραφο, τα διαγράμματα στην αριστερή στήλη του Σχήματος 5.3 αναφέρονται στον δειγματικό μέσο των μετρήσεων για κάθε μέγεθος που λάβαμε στα 250 πειράματα, ως συνάρτηση του χρόνου. Τα μεγέθη που μελετάμε είναι τα ίδια με αυτά της προηγούμενης παραγράφου με την προσθήκη όμως των εκτιμήσεων της τιμής της σταθμισμένης συνάρτησης ασυμπτωτικού ρυθμού Φ ('I (infsup)' στο Σχήμα 5.3). Η BOP όπως και στην προηγούμενη παράγραφο, εκτιμάται μέσω της βελτίωσης Bahadur-Rao.

Στη δεξιά στήλη (διαγράμματα 5.3(b,d,f,h)) απεικονίζεται η εμπειρική κατανομή των μεγεθών που λαμβάνουμε στο 10ο βήμα της επαναληπτικής μεθόδου, δηλ. περίπου 100 sec μετά από την έναρξη της μεθόδου. (Σημείο εκκίνησης είναι το $s = 0.01$ kbits^{-1}, $t = 20$ msec.) Η επιλογή του συγκεκριμένου βήματος έγινε βάσει παρατηρήσεων ότι η κατανομή των εκτιμήσεων όλων των παραπάνω μεγεθών δεν αλλάζει σημαντικά μετά το βήμα αυτό.

Στα διαγράμματα 5.3(e),(g) απεικονίζονται η χρονική και η χωρική παράμετρος του συστήματος όπως εκτιμήθηκαν από τη μέθοδό μας. Παρατηρούμε ότι οι εκτιμήσεις των ποσοτήτων αυτών, συγκλίνουν πολύ γρήγορα (σε λιγότερο από 1 λεπτό) πολύ κοντά στις θεωρητικές τους τιμές. Επίσης (βλ. Σχήμα 5.3(f),(g)), η διασπορά είναι μικρή. Αν και οι εκτιμητές δεν είναι αμερόληπτοι, εντούτοις δεν απέχουν πολύ από τις θεωρητικές τιμές τους. Στην πράξη, το γεγονός αυτό δεν έχει σημαντικές συνέπειες αφού βάσει πειραμάτων στο [7] έχει διαπιστωθεί ότι η συνάρτηση ασυμπτωτικού ρυθμού δεν είναι ευαίσθητη σε μικρές μεταβολές του σημείου λειτουργίας.

Σε αντίθεση με τους εκτιμητές του σημείου λειτουργίας, ο εκτιμητής του ισοδυνάμου εύρους ζώνης έχει μεγάλη διασπορά, με αποτέλεσμα και η διασπορά του εκτιμητή της BOP να είναι μεγάλη (βλ. Σχήμα 5.3(b)). Το γεγονός αυτό έχει ως αποτέλεσμα η πιθανότητα υπερχείλισης ενός ενταμιευτή να μη μπορεί να εκτιμηθεί με ακρίβεια, παρόλο που το σημείο λειτουργίας εκτιμάται με ακρίβεια.

Η διασπορά του εκτιμητή της BOP μπορεί να ελαττωθεί με τον ακόλουθο τρόπο. Αντί να λαμβάνουμε υπόψιν τις μετρήσεις ενός μόνο βήματος (για την εκτίμηση της BOP),



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Σχήμα 5.3: Επίδοση της μεθόδου εκτίμησης σε πραγματικό χρόνο της BOP (a)-(b), του ισοδύναμου εύρους ζώνης (c)-(d), της παραμέτρου χρόνου του συστήματος (e)-(f) και της παραμέτρου χώρου του συστήματος (g)-(h) για πραγματική κυκλοφορία (MPEG-1).

χρησιμοποιούμε και τις εκτιμήσεις που ελήφθησαν στα προηγούμενα βήματα. Αυτό είναι δυνατό, μόνο όταν το εκτιμώμενο σημείο λειτουργίας, δεν αλλάζει σημαντικά από βήμα σε βήμα. Είναι απαισιόδοξο να περιμένουμε το σημείο λειτουργίας να μεταβάλλεται σημαντικά από βήμα σε βήμα, έτσι μπορούμε να χρησιμοποιήσουμε τις εκτιμήσεις των προηγουμένων βημάτων σαν να αφορούν όλες το ίδιο σημείο λειτουργίας.

Εάν συμβολίσουμε με $\hat{s}_n$ και $\hat{t}_n$ τους εκτιμητές του σημείου λειτουργίας στο n-οστό βήμα, με $\hat{\Phi}_n$ τον εκτιμητή του Φ, τότε

$$\hat{\Phi}_n = \hat{s}_n(B + C\hat{t}_n) - \sum_{j=1}^{N} \log \hat{E}_n^j(\hat{s}_n, \hat{t}_n) \, ,$$

όπου $\hat{E}_n^j(s_n, t_n) = \sum_{k=1}^{K} \exp(s_n \hat{X}_k^j)/\mathrm{K}$ και το $\hat{X}_k^j$ ορίζεται στην εξίσωση (5.1).

Παρατηρούμε ότι ακόμα και αν η μέθοδος γνώριζε το σημείο λειτουργίας, ο $\hat{\Phi}_n$ δε θα ήταν *συνεπής* καθώς $n \to \infty$ λόγω του ότι ο $\hat{\Phi}_n$ εξαρτάται μόνο από μετρήσεις φόρτου στη χρονική περίοδο n και ο αριθμός αυτών των μετρήσεων (το K στην εξίσωση (5.3)) είναι πεπερασμένος. Για αυτό, προτείνεται να χρησιμοποιήσουμε τις εκτιμήσεις των προηγούμενων βημάτων $n-1, n-2, \ldots$ κτλ., και σαν εκτιμητή του Φ να χρησιμοποιήσουμε το

$$\hat{\Psi}_n = -\log \frac{1}{n+1} \sum_{i=0}^{n} e^{-\hat{\Phi}_i} \, . \tag{5.4}$$

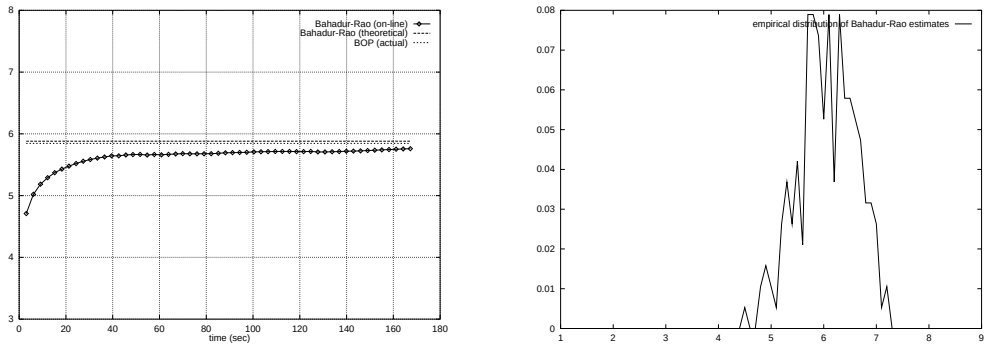
Λόγω της στασιμότητας της διαδικασίας εισόδου, για κάθε j οι τ.μ. $\hat{E}_n^j(s,t)$ έχουν κοινή κατανομή για σταθερά (s,t) και έστω $\hat{E}^j \triangleq \hat{E}_n^j$. Αυτό έχει συνέπεια, οι τ.μ. $\hat{\Phi}_i$ να έχουν κοινή κατανομή επίσης, και έστω $\hat{\Phi} \triangleq \hat{\Phi}_i$. Τότε αν θεωρήσουμε ότι οι εκτιμήσεις του σημείου λειτουργίας (s^*, t^*) δεν μεταβάλλονται σημαντικά, παρ'ότι δεν συγκλίνουν, καθώς το n τείνει στο ∞ έχουμε

$$\hat{\Psi}_n \to -\log \mathrm{E} e^{-\hat{\Phi}} = -\log \mathrm{E} e^{-s^*(B+Ct^*)+\sum_{j=1}^{N} \log \hat{E}^j(s^*,t^*)} = s^*(B+Ct^*) - \sum_{j=1}^{N} \log \mathrm{E}\hat{E}^j(s^*,t^*) = \Phi,$$

με πιθανότητα 1. Δηλαδή το $\hat{\Psi}_n$ είναι συνεπής εκτιμητής του Φ, με την προϋπόθεση ότι οι εκτιμητές του σημείου λειτουργίας έχουν μικρή διασπορά και δεν απέχουν αρκετά από τη θεωρητική τους τιμή, πράγμα που ισχύει όπως φαίνεται στο Σχήμα 5.3 (e-h).

Είναι σημαντικό να αναφέρουμε ότι οι εκτιμητές του σημείου λειτουργίας δεν εξαρτώνται από τις εκτιμήσεις προηγούμενων βημάτων, αφού αυτό γίνεται μόνο για την εκτίμηση του Φ. Έτσι, ακόμα και μετά από μια απότομη αλλαγή στη σύνθεση της κυκλοφορίας η οποία μεταβάλλει κατά πολύ το σημείο λειτουργίας, η μέθοδός μας εξακολουθεί να συγκλίνει στο νέο σημείο λειτουργίας με την ίδια ταχύτητα. Δεν ισχύει το ίδιο για τον εκτιμητή του Φ, αφού επηρεάζεται από παλαιότερες εκτιμήσεις. Μπορούμε να ξεπεράσουμε το μειονέκτημα αυτό, εάν κατά τη μεταβολή της σύνθεσης της κυκλοφορίας (άφιξη ή αναχώρηση κλήσης), η μέθοδος ''ξεχνά'' τις προηγούμενες εκτιμήσεις.

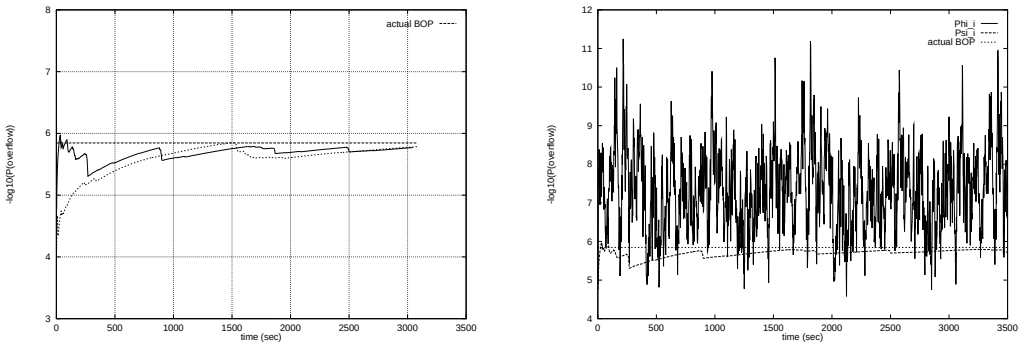
Στο Σχήμα 5.4(a) απεικονίζεται ο δειγματικός μέσος του $\hat{\Psi}_n$ όπως εκτιμήθηκε από 250 ανεξάρτητα πειράματα με $(s_0, t_0) = (0.02, 30)$, ενώ $(s^*, t^*) = (0.029, 42.3)$. Στο Σχήμα 5.4(b) απεικονίζεται η εμπειρική κατανομή του $\hat{\Psi}_n$ τη χρονική στιγμή t=100 sec. Παρατηρούμε ότι υπάρχει σημαντική βελτίωση στη διασπορά του εκτιμητή μεταξύ της περίπτωσης όπου δε λαμβάναμε υπόψιν τις εκτιμήσεις των προηγούμενων βημάτων (βλ. Σχήμα 5.3(b)). Στο Σχήμα 5.5(a) εξετάζουμε το κατά πόσο είναι ακριβής η προσεγγιστική συνέπεια του $\hat{\Psi}_n$, όπου απεικονίζονται οι δειγματικές τιμές της ως συνάρτηση του χρόνου για 2 ανεξάρτητα πειράματα. Η σωστή σύγκλιση των εκτιμήσεων στο Φ, δικαιολογεί την παραπάνω μέθοδο. Η επίδοση του εκτιμητή $\hat{\Psi}_n$ έναντι του $\hat{\Phi}_n$ απεικονίζεται στο Σχήμα 5.5(b).



(a) (b)

Σχήμα 5.4: Επίδοση της μεθόδου εκτίμησης σε πραγματικό χρόνο της BOP. (a) Οι δειγματικοί μέσοι των εκτιμήσεων ως συνάρτηση του χρόνου και (b) η εμπειρική κατανομή των εκτιμήσεων τη χρονική στιγμή t=100 sec.

Συνοψίζοντας, με τη χρήση της μεθόδου που περιγράψαμε μπορούμε να εκτιμήσουμε την πιθανότητα υπερχείλισης σε ένα ενταμιευτή, γρήγορα και με ακρίβεια. Η μέθοδος αυτή μπορεί να χρησιμοποιηθεί ως βάση σε απλούς αλγορίθμους CAC. Έναν τέτοιο απλό αλγόριθμο προτείνουμε και μελετούμε στο επόμενο κεφάλαιο.



(a) (b)

Σχήμα 5.5: (a) Οι δειγματικές τιμές του εκτιμητή της BOP $\hat{\Psi}_n$ ως συνάρτηση του χρόνου, για 2 ανεξάρτητα πειράματα. (b) Οι δειγματικές τιμές των εκτιμητών $\hat{\Psi}_n$ και $\hat{\Phi}_n$ ως συνάρτηση του χρόνου. Με 'actual BOP' απεικονίζεται η πραγματική BOP.

Κεφάλαιο 6

Ένας Απλός Αλγόριθμος CAC

Μέχρι τώρα, έχει κατανοηθεί σχετικά καλά η συμπεριφορά ενός πολυπλέκτη, όπου η κυκλοφορία προέρχεται από πολλές στάσιμες και ανεξάρτητες διαδικασίες των οποίων είναι γνωστή η στοχαστική συμπεριφορά. Αποτελέσματα τα οποία βασίζονται στη θεωρία μεγάλων αποκλίσεων, έχουν διαφωτίσει αρκετά τον τρόπο με τον οποίον ένας πολυπλέκτης μπορεί να εξυπηρετήσει μια πλήρως χαρακτηρισμένη συλλογή από πηγές (βλ. Κεφάλαιο 2). Υπάρχουν όμως πολλές δυσκολίες στην επέκταση των αποτελεσμάτων αυτών έτσι ώστε να κατανοηθεί η συμπεριφορά αλγορίθμων CAC (Ελέγχου Αποδοχής Κλήσεων) που βασίζονται σε μετρήσεις και η κυκλοφορία *δεν* έχει επαρκώς χαρακτηρισθεί.

Ένας λόγος είναι ότι η υπερχείλιση των ενταμιευτών σε ένα τέτοιο δίκτυο, είναι γενικά συνέπεια δύο συνδυασμένων φαινομένων: ακραία λάθη μετρήσεων τα οποία επιτρέπουν την εσφαλμένη αποδοχή πάρα πολλών κλήσεων *ή/και* την ακόλουθη ακραία συμπεριφορά των αποδεχθησών κλήσεων. Το πρώτο φαινόμενο αναλύεται στις χρονικές κλίμακες που αφορούν τις αποφάσεις αποδοχής/απόρριψης και τη διάρκεια των κλήσεων, ενώ η ανάλυση του δευτέρου φαινομένου απαιτεί την κατανόηση της συμπεριφοράς των κλήσεων σε χρονικές κλίμακες συγκρίσιμες με τον χρόνο υπερχείλισης (βλ. Κεφάλαιο 2).

Γενικά, οι απλοί αλγόριθμοι CAC αποτελούνται από δύο βασικά μέρη: τον *εκτιμητή* και τον *ελεγκτή*. Τα δύο αυτά μέρη θεωρείται ότι είναι ανεξάρτητα μεταξύ τους (εδώ οφείλεται και η απλότητα τέτοιων σχημάτων), δηλαδή ο εκτιμητής είναι σχεδιασμένος έτσι ώστε να συγκλίνει γρήγορα στη σωστή πιθανότητα υπερχείλισης, για ένα δεδομένο αριθμό κλήσεων στον πολυπλέκτη, ενώ ο ελεγκτής είναι σχεδιασμένος έτσι ώστε να παίρνει πάντα τη βέλτιστη απόφαση εάν ήξερε *ακριβώς* την πιθανότητα υπερχείλισης για κάθε αριθμό κλήσεων στον πολυπλέκτη.

Στο [14] εξετάζονται αλγόριθμοι CAC σε πολυπλέκτες χωρίς ενταμιευτές, όπου οι κλήσεις είναι πηγές ON/OFF. Παρά την κατανόηση που προσφέρει το απλό αυτό μοντέλο μέσω

της απλής ανάλυσής του, όπως θα δούμε στην Παρ. 6.4 ο βαθμός χρησιμοποίησης που επιτυγχάνουν οι αλγόριθμοι αυτοί είναι πολύ χαμηλή στην περίπτωση πραγματικής κυκλοφορίας (MPEG). Στο [15] μελετώνται οι επιπτώσεις του σφάλματος εκτίμησης απλών αλγορίθμων CAC, όπου μπορεί κάτω από πολύ γενικές συνθήκες να είναι σημαντικές στην επίδοση του πολυπλέκτη.

Σε αυτό το κεφάλαιο εξετάζουμε τη χρήση της μεθόδου εκτίμησης της πιθανότητας υπερχείλισης που μελετήσαμε στην Παράγραφο 5.3.2, σε έναν απλό αλγόριθμο CAC. Η κυκλοφορία που χρησιμοποιούμε, όπως και στις προηγούμενες παραγράφους, προέρχεται από την ταινία video Star Wars κωδικοποιημένη σύμφωνα με το πρότυπο MPEG-1.

Αναφορικά με το [15], δεν εξετάζουμε τις επιπτώσεις των σφαλμάτων εκτίμησης του απλού αλγορίθμου CAC που μελετάμε. Συγκεντρωνόμαστε περισσότερο στη μελέτη της μεθόδου εκτίμησης της πιθανότητας υπερχείλισης που χρησιμοποιούμε και σκοπός μας είναι να εξετάσουμε τη χρήση της στον έλεγχο αποδοχής κλήσης.

6.1 Το Μοντέλο Κλήσεων

Θεωρούμε ότι σε έναν σύνδεσμο χωρητικότητας C=34 Mbps και ενταμιευτή μεγέθους B=500 κελιά, έρχονται αιτήσεις σύνδεσης από κλήσεις. Η διάρκεια κάθε κλήσης είναι εκθετικά κατανεμημένη και οι αιτήσεις σύνδεσης ακολουθούν τη διαδικασία Poisson. Ήδη αναφέραμε ότι το περιεχόμενο των κλήσεων είναι η ταινία video Star Wars. Επιπλέον, η μετάδοση κάθε κλήσης αρχίζει από ένα ομοιόμορφα τυχαίο σημείο μέσα στην ταινία, ενώ όταν η ταινία φτάσει στο τέλος της, συνεχίζει η μετάδοση από τη αρχή.

6.2 Αλγόριθμος CAC

Ο αλγόριθμος χρησιμοποιεί εκτιμήσεις της πιθανότητας υπερχείλισης του ενταμιευτή, βάσει της μεθόδου που περιγράψαμε στην Παράγραφο 5.3.2. Όταν υπάρξει μια αίτηση σύνδεσης από μια νέα κλήση, ο ελεγχτής αποφασίζει την αποδοχή της αίτησης ανάλογα με το εάν η εκτιμώμενη πιθανότητα υπερχείλισης δεν θα υπερβεί, μετά την αποδοχή, ένα επιθυμητό όριο $\mathrm{BOP}_{\mathrm{target}}$ το οποίο καθορίζει και την ποιότητα υπηρεσίας που παρέχει ο πολυπλέκτης.

Πιο αναλυτικά, εάν τη στιγμή άφιξης της αίτησης σύνδεσης εξυπηρετούνται n κλήσεις και $\hat{\Psi}_n$ η τρέχουσα εκτίμηση της σταθμισμένης συνάρτησης ασυμπτωτικού ρυθμού (βλ. εξίσωση (5.4)), τότε η αίτηση γίνεται αποδεκτή εάν

$$\hat{\Psi}_{n+1} + \frac{1}{2}\log(4\pi\hat{\Psi}_{n+1}) > \gamma\,, \tag{6.1}$$

όπου $\gamma = -\log \text{BOP}_{\text{target}}$. Το αριστερό μέλος της (6.1) είναι μια εκτίμηση της Bahadur-Rao βελτίωσης (2.6). Εν γένει τα χαρακτηριστικά της κυκλοφορίας της νέας κλήσης δεν είναι γνωστά, οπότε ο ελεγχτής πρέπει χρησιμοποιώντας την εκτίμηση $\hat{\Psi}_n$ και τις παραμέτρους του συμβολαίου κυκλοφορίας που δόθηκαν, να εκτιμήσει το $\hat{\Psi}_{n+1}$. Μέσω των παραμέτρων αυτών (βλ. [19, 5]), δίδονται ο μέγιστος ρυθμός της κυκλοφορίας καθώς και ένα άνω φράγμα του μέσου ρυθμού. Έστω h ο μέγιστος ρυθμός και m το άνω φράγμα του μέσου ρυθμού (Sustainable Cell Rate - SCR). Τότε ισχύει η εξής ανισότητα [22]:

$$\alpha_{\text{new}}(s,t) \leq \frac{1}{st} \log \left[1 + \frac{m}{h} \left(e^{sht} - 1 \right) \right] \equiv \bar{\alpha}_{\text{new}}(s,t) \,. \tag{6.2}$$

Δηλαδή το ισοδύναμο εύρος ζώνης της νέας κλήσης φράσσεται άνω από το ισοδύναμο εύρος μιας πηγής ON/OFF η οποία έχει μέγιστο και μέσο ρυθμό ίσους με h και m αντίστοιχα. Βασιζόμενοι στην εκτίμηση του σημείου λειτουργίας και στο γεγονός ότι αυτό δεν αλλάζει σημαντικά κατά την προσθήκη μιας ακόμα (‘‘μικρής’’ σχετικά με τη συνολική κυκλοφορία) κλήσης στη σύνθεση της κυκλοφορίας, μπορούμε να προσεγγίσουμε συντηρητικά το $\hat{\Psi}_{n+1}$ ως $\hat{s}(B + C\hat{t}) - \hat{s}\hat{t}\,(\hat{\alpha}(\hat{s},\hat{t}) + \bar{\alpha}_{\text{new}}(\hat{s},\hat{t}))$, όπου $\hat{\alpha}$ είναι η εκτίμηση του ισοδυνάμου εύρους ζώνης των κλήσεων που ήδη εξυπηρετούνται και $(\hat{s},\hat{t})$ είναι η τρέχουσα εκτίμηση του σημείου λειτουργίας.

Ο ρυθμός με τον οποίον έρχονται αιτήσεις σύνδεσης μπορεί να παίξει καθοριστικό ρόλο στην επίδοση ενός αλγορίθμου CAC. Εάν ο ρυθμός αυτός είναι πολύ υψηλός, τότε δεν είναι αρκετό ο αλγόριθμος να λαμβάνει σωστές αποφάσεις κάθε φορά με μεγάλη πιθανότητα επιτυχίας: Εάν ο ρυθμός άφιξης αιτήσεων σύνδεσης είναι πολύ υψηλός, ο ρυθμός με τον οποίο αιτήσεις γίνονται *λανθασμένα* αποδεκτές μπορεί να γίνει πολύ υψηλός. Ένας τρόπος προστασίας από το φαινόμενο αυτό, είναι η ακόλουθη στρατηγική *οπισθοχώρησης* την οποία χρησιμοποιούμε στον αλγόριθμο που προτείνουμε. Όταν μια αίτηση σύνδεσης απορρίπτεται, καμμιά άλλη αίτηση σύνδεσης δε γίνεται δεκτή έως ότου τερματίσει τη σύνδεσή της μια τρέχουσα κλήση. Η στρατηγική αυτή καθώς και άλλες στρατηγικές οπισθοχώρησης περιγράφονται στο [2].

6.3 Βέλτιστος αλγόριθμος CAC

Ο βέλτιστος απλός αλγόριθμος CAC προκύπτει εάν ο ελεγκτής και ο εκτιμητής είχαν τη γνώση για να παίρνουν κάθε φορά τη σωστή απόφαση. Δηλαδή, έχουν πλήρη γνώση για την μελλοντική συμπεριφορά των κλήσεων. Εάν θέσουμε $\text{BOP}_{\text{target}} = 10^{-4}$, τότε κάνοντας προσομοιώσεις βρίσκουμε ότι ο μέγιστος επιτρεπτός αριθμός κλήσεων είναι 101. Συνεπώς ο βέλτιστος βαθμός χρησιμοποίησης του πολυπλέκτη είναι 77.91%. Θα εξετάσουμε κατά πόσο

η επίδοση του αλγορίθμου που περιγράψαμε στην προηγούμενη παράγραφο πλησιάζει την επίδοση του βέλτιστου.

6.4 Απλοί αλγόριθμοι CAC βασιζόμενοι σε πηγές ON/OFF

Στη βιβλιογραφία, για την πρόβλεψη της επίδοσης ενός πολυπλέκτη χωρίς ενταμιευτή κάτω από διάφορους αλγορίθμους CAC, χρησιμοποιούνται πηγές ON/OFF οι οποίες αντιπροσωπεύουν την πραγματική κυκλοφορία. Ο λόγος χρησιμοποίησης του μοντέλου αυτού είναι γιατί είναι απλό έτσι ώστε να οδηγεί σε αναλυτικά αποτελέσματα, κατορθώνει να μοντελοποιήσει τα φαινόμενα της στατιστικής πολυπλεξίας και γιατί παρέχει τη ''χείριστη περίπτωση'' μεταξύ κυκλοφοριών που έχουν τον ίδιο μέσο και μέγιστο ρυθμό. Λέγοντας ''χείριστη περίπτωση'' εννοούμε ότι εάν αντικαταστήσει κυκλοφορία η οποία έχει τον ίδιο μέσο και μέγιστο ρυθμό, η πιθανότητα υπερχείλισης του ενταμιευτή θα αυξηθεί. Το γεγονός αυτό βασίζεται στην ανισότητα (6.2) ενώ παραδείγματα (και αντιπαραδείγματα) δίδονται στο [11].

Πραγματικά, οι εκφράσεις που προκύπτουν στην ανάλυση επιδόσεων με τη χρήση πηγών ON/OFF είναι εξαιρετικά απλές. Για παράδειγμα στα [14, 23], μελετώνται απλοί αλγόριθμοι CAC των οποίων οι ιδιότητες αναλύονται πολύ εύκολα με αποτέλεσμα να κατανοηθούν περισσότερο οι παράμετροι που καθορίζουν την επίδοση τους. Όμως παρά την δυνατότητα κατανόησης των διαφόρων φαινομένων, η ποσοτική τους ανάλυση δεν είναι καθόλου ακριβής με αποτέλεσμα τα μόνα συμπεράσματα τα οποία μπορούμε να εξάγουμε να είναι ποιοτικά. Η αδυναμία ποσοτικής ανάλυσης της επίδοσης πηγάζει από το γεγονός ότι οι πηγές ON/OFF πράγματι αποτελούν μια πολύ ''χείριστη περίπτωση'' και οι αποφάσεις ενός αλγόριθμου CAC είναι εξαιρετικά συντηρητικές.

Για παράδειγμα, έστω ότι χρησιμοποιούμε μοντέλα πηγών ON/OFF για την περιγραφή των κλήσεων. Οι πραγματικές κλήσεις είναι ταινίες video Star Wars και έχουν μέγιστο και μέσο ρυθμό h=3.4556 Mbps και m=0.2623 Mbps αντίστοιχα. Εάν είχαμε N πραγματικές πηγές σε έναν πολυπλέκτη, τότε το σημείο λειτουργίας θα ήταν συνάρτηση του N (έστω (s_N^*, t_N^*)), για σταθερή χωρητικότητα C και χώρο ενταμίευσης B. Εφόσον η χρονική παράμετρος αντιπροσωπεύει τη χρονική κλίμακα στην οποία η υπερχείλιση συμβαίνει με μεγαλύτερη πιθανότητα, μπορούμε να προσεγγίσουμε την πιθανότητα υπερχείλισης του πολυπλέκτη, με την πιθανότητα να συμβεί υπερχείλιση βλέποντας το σύστημα στη χρονική κλίμακα t_N^*. Σε αυτή τη χρονική κλίμακα, το σύστημα είναι ισοδύναμο με ένα σύστημα χωρίς ενταμιευτή και ρυθμό εξυπηρέτησης $C + B/t_N^*$. Μπορούμε τώρα να υπολογίσουμε την πιθανότητα υπερχείλισης του ενταμιευτή, θεωρώντας ότι κάθε πραγματική πηγή είναι πηγή ON/OFF με

τον ίδιο μέγιστο και μέσο ρυθμό με την πραγματική. Δηλαδή:

$$\text{P(overflow)} = 1 - \sum_{k=1}^{N^*} \binom{N}{k} p^k (1-p)^{N-k} \ , \tag{6.3}$$

όπου $p = m/h$ και $N^* = \min\{n : (n+1)h > C + B/t_N^*\}$. Κάνοντας τους απαραίτητους αριθμητικούς υπολογισμούς βρίσκουμε ότι το μεγαλύτερο N έτσι ώστε η παράσταση (6.3) να είναι μικρότερη από 10^{-4} είναι $N = 45$. Δηλαδή ο βαθμός χρησιμοποίησης που θα επιτύγχανε ένας βέλτιστος αλγόριθμος βασισμένος σε πηγές ON/OFF και συστήματα χωρίς ενταμιευτή θα ήταν Nm/C=34.713%.

Από τα παραπάνω συμπεραίνουμε ότι απλοί αλγόριθμοι CAC, δεν μπορούν να εφαρμοστούν στην περίπτωση πραγματικής (MPEG) κυκλοφορίας λόγω του ότι η προσέγγισή τους με πηγές ON/OFF οδηγεί σε εξαιρετικά συντηρητικές αποφάσεις.

6.5 Πειραματικά αποτελέσματα

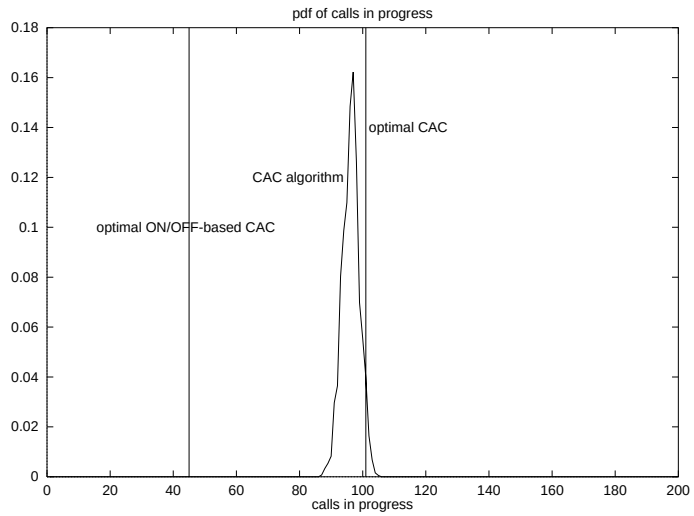
Εκτελέσαμε ένα πείραμα για τη μελέτη της επίδοσης του αλγόριθμου CAC που περιγράψαμε στην Παρ. 6.2, ενώ συγκρίνουμε την επίδοσή του με αυτή του βέλτιστου αλγορίθμου CAC (βλ. Παρ. 6.3) και του βέλτιστου αλγορίθμου CAC που βασίζεται σε πηγές ON/OFF σε σύστημα χωρίς ενταμιευτή (βλ. Παρ. 6.4).

Θέσαμε το ρυθμό άφιξης κλήσεων στις 5 κλήσεις ανά δευτερόλεπτο, έτσι ώστε να είναι πολύ μεγάλος για να πετύχουμε τη μέγιστη δυνατή χρησιμοποίηση του ενταμιευτή. Πρακτικά ένας τέτοιος ρυθμός άφιξης κλήσεων, σημαίνει ότι πάντοτε υπάρχουν αιτήσεις σύνδεσης κλήσεων με αποτέλεσμα το σενάριο αυτό να αποτελεί μια σκληρή δοκιμασία για κάθε αλγόριθμο CAC λόγω του ότι μπορούμε να ελέγξουμε πώς αντιμετωπίζει την κατάσταση υψηλού ρυθμού άφιξης που περιγράψαμε στην αρχή του κεφαλαίου, δηλαδή την πιθανότητα να γίνει δεκτός πολύ μεγάλος αριθμός κλήσεων. Η διάρκεια κάθε κλήσης είναι 1000 sec.

Η επίδοση του αλγορίθμου μελετάται ως προς: (i) τον βαθμό χρησιμοποίησης του πολυπλέκτη που επιτυγχάνει και (ii) την μέση πιθανότητα υπερχείλισης του ενταμιευτή. Επιπλέον κάθε αλγόριθμος CAC, λόγω σφαλμάτων εκτίμησης, όσο μικρά και αν είναι αυτά, υπάρχει η πιθανότητα ο αλγόριθμος να δεχτεί πάρα πολλές κλήσεις, και στιγμιαία να παραβιαστεί η ανώτατη τιμή $\text{BOP}_{\text{target}}$ της πιθανότητας υπερχείλισης. Έτσι, εφόσον δεν μπορούμε να το αποφύγουμε, επιθυμούμε η πιθανότητα να συμβεί το γεγονός αυτό να είναι μικρή.

Λόγω του ότι η διαδικασία άφιξης κλήσεων είναι εργοδική (Poisson) όπως και το σημείο έναρξης κάθε κλήσης μέσα στη ταινία video, η διαδικασία κλήσεων στο σύστημα είναι και αυτή εργοδική. Για τον λόγο αυτό εκτελέσαμε ένα μόνο πείραμα όπου προσομοίωσε

9 ώρες πραγματικής λειτουργίας. Στο Σχήμα 6.1 απεικονίζεται η εμπειρική κατανομή των κλήσεων στο σύστημα κατά το πέρας του πειράματος. Επίσης στο ίδιο σχήμα εμφανίζονται οι κατανομές κλήσεων των δύο βέλτιστων αλγορίθμων των Παρ. 6.3 και 6.4.



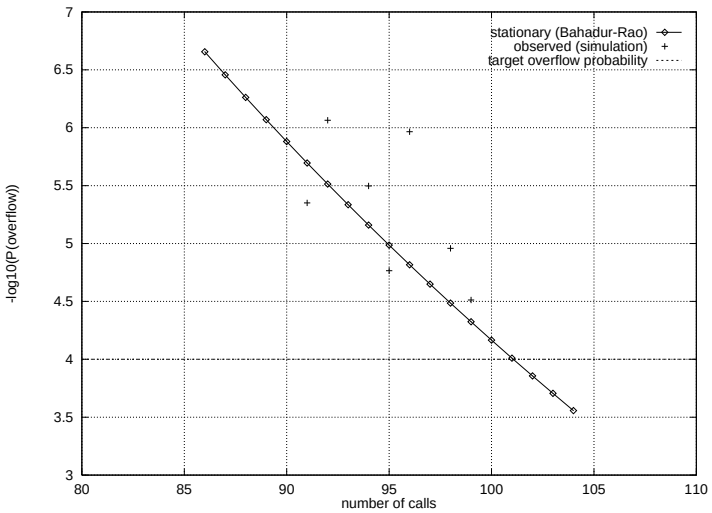
Σχήμα 6.1: Η κατανομή κλήσεων στον πολυπλέκτη: (i) για τον βέλτιστο αλγόριθμο που βασίζεται σε πηγές ON/OFF και σε σύστημα χωρίς ενταμιευτή ('optimal ON/OFF-based CAC'), (ii) για τον αλγόριθμο που προτείνουμε ('CAC algorithm') και (iii) για τον βέλτιστο αλγόριθμο CAC ('optimal CAC').

Ο μέσος βαθμός χρησιμοποίησης του πολυπλέκτη ήταν 74.29%, πολύ κοντά στη βέλτιστη (77.91%), ενώ είναι ασύγκριτα καλύτερος από αυτόν που επιτυγχάνει ο βέλτιστος αλγόριθμος που στηρίζεται σε πηγές ON/OFF. Η μέση πιθανότητα υπερχείλισης του ενταμιευτή ήταν $10^{-5.221}$, μικρότερη από την ανώτερη επιτρεπτή (10^{-4}). Το ποσοστό του χρόνου που ο πολυπλέκτης βρίσκεται σε κατάσταση όπου ο αριθμός των κλήσεων είναι μη επιτρεπτός, δηλ. από 102 κλήσεις μέχρι 104 (ο μέγιστος αριθμός κλήσεων που υπήρξε στο σύστημα) είναι μόλις 1.7% ενώ τα ποσοστά ανά μη επιτρεπτή κατάσταση δίδονται στον παρακάτω πίνακα:

Κατάσταση	Ποσοστό χρόνου
102	1.20%
103	0.42%
104	0.07%

Η εκτίμηση της πιθανότητας υπερχείλισης απαιτεί έναν τεράστιο χρόνο προσομοίωσης λόγω του ότι ορισμένες καταστάσεις, π.χ. αυτή με 104 κλήσεις, είναι πολύ σπάνιες (η πιθανότητα να εξυπηρετεί ο πολυπλέκτης 104 κλήσεις είναι περίπου $10^{-3.01}$). Έτσι, ενώ η εμπειρική κατανομή του αριθμού των κλήσεων στον πολυπλέκτη συγκλίνει γρήγορα, οι πιθανότητες υπερχείλισης σε κάθε κατάσταση δεν μπορούν να εκτιμηθούν με ακρίβεια. Μπορούμε να ξεπεράσουμε αυτή τη δυσκολία εάν υπολογίσουμε τη στάσιμη πιθανότητα υπερχείλισης για κάθε πλήθος πηγών ξεχωριστά. Ο υπολογισμός αυτός μπορεί να γίνει

είτε με ξεχωριστές προσομοιώσεις για κάθε σταθερό αριθμό πηγών, είτε γρήγορα και με επαρκή ακρίβεια χρησιμοποιώντας τη βελτίωση Bahadur-Rao στο ασυμπτωτικό πολλών πηγών. Η μέση πιθανότητα υπερχείλισης εάν χρησιμοποιήσουμε την προσέγγιση αυτή είναι $10^{-4.5823}$. Παρατηρούμε ότι είναι μικρότερη από την ανώτερη επιτρεπτή (10^{-4}), ενώ είναι μεγαλύτερη από την εκτίμηση που λάβαμε στην προσομοίωση ($10^{-5.221}$). Στο Σχήμα 6.2, απεικονίζεται η στάσιμη πιθανότητα υπερχείλισης για κάθε σταθερό πλήθος πηγών, όπως αυτή προσεγγίζεται από τη βελτίωση Bahadur-Rao στο ασυμπτωτικό πολλών πηγών. Επίσης στο ίδιο σχήμα απεικονίζεται η εμπειρική πιθανότητα υπερχείλισης σε κάθε κατάσταση, όπου όπως παρατηρούμε είναι πάντα μικρότερη από 10^{-4} ενώ δεν παρατηρήσαμε υπερχείλιση σε όλες τις καταστάσεις.



Σχήμα 6.2: Η πιθανότητα υπερχείλισης του ενταμιευτή (BOP), για κάθε αριθμό κλήσεων. Απεικονίζεται η στάσιμη BOP μέσω της προσέγγισης Bahadur-Rao καθώς και η BOP που παρατηρήθηκε στην προσομοίωση.

Κεφάλαιο 7

Επίλογος

7.1 Συμπεράσματα

Στην εργασία αυτή, αναλύσαμε μια νέα έννοια: την ισοδυναμία κυκλοφορίας. Η έννοια αυτή βασίζεται στη διατήρηση των κύριων φαινομένων τα οποία οδηγούν στην υπερχείλιση των ενταμιευτών. Είδαμε ότι αντικαθιστώντας την υπέρθεση πραγματικών πηγών με ένα μοντέλο το οποίο διατηρεί αυτά τα φαινόμενα, η επίδοση του συστήματος διατηρείται επιβεβαιώνοντας την πρακτική αξία της έννοιας της ισοδυναμίας κυκλοφορίας. Δείξαμε ότι τέτοια μοντέλα υπάρχουν, κάτω από μη περιοριστικές συνθήκες, για πηγές video κωδικοποιημένες σύμφωνα με το πρότυπο MPEG, το οποίο αναμένεται ότι θα κυριαρχήσει στα ευρυζώνια δίκτυα.

Η αντικατάσταση πραγματικής κυκλοφορίας είναι επιτυχής ακόμα και αν χρησιμοποιήσουμε πολύ απλά μοντέλα, γεγονός που υπογραμμίζει τη σημασία του σημείου λειτουργίας και του ισοδυνάμου εύρους ζώνης σαν περιγραφέων του συστήματος και μιας πηγής αντίστοιχα. Με αυτή την αφαιρετική περιγραφή του συστήματος, τα φαινόμενα της στατιστικής πολυπλεξίας μπορούν να αναλυθούν εύκολα χρησιμοποιώντας απλές αλγεβρικές μεθόδους.

Η ισοδυναμία κυκλοφορίας μπορεί να βρει χρήση κατά τον σχεδιασμό ενός δικτύου και την ανάπτυξη δικτυακών εφαρμογών. Η αντικατάσταση πραγματικών πηγών με ένα απλό μοντέλο μπορεί να χρησιμοποιηθεί στην κατασκευή απλών γεννητριών κυκλοφορίας οι οποίες μπορούν να εξομοιώσουν έναν αυθαίρετα μεγάλο αριθμό πηγών. Οι γεννήτριες αυτές μπορούν να χρησιμοποιούν πολύ απλά μοντέλα των οποίων η κυκλοφορία είναι εύκολο να δημιουργηθεί σε πραγματικό χρόνο. Επίσης, η συνολική κυκλοφορία μπορεί να δημιουργηθεί από σύνθεση πραγματικών πηγών και μοντέλων, επιτρέποντας την πειραματική παρατήρηση και ανάλυση της ποιότητας υπηρεσίας των πραγματικών εφαρμογών όταν μεταδίδονται μέσα

σε μεγάλους πολυπλέκτες.

Παράλληλα, η ισοδυναμία κυκλοφορίας μπορεί να βρει εφαρμογές στον έλεγχο κυκλοφορίας σε πραγματικό χρόνο. Ο υπολογισμός του σημείου λειτουργίας και του ισοδυνάμου εύρους ζώνης μπορεί τόσο να επιταχυνθεί κατά τη φάση της ανάλυσης των συστημάτων του δικτύου όσο και να γίνει εφικτός σε πραγματικό χρόνο. Στο Κεφάλαιο 5 χρησιμοποιήσαμε την έννοια της ισοδυναμίας και της αντικατάστασης για την ανάπτυξη μιας επαναληπτικής μεθόδου για την εύρεση του σημείου λειτουργίας ενός συστήματος. Η μέθοδος αυτή επιταχύνει κατά πολύ την ανάλυση επίδοσης ενός πολυπλέκτη με το ασυμπτωτικό πολλών πηγών, κάνοντας δυνατή την ανάλυση πραγματικών ευρυζωνίων συνδέσμων με πολλά διαφορετικά είδη πηγών. Στο ίδιο κεφάλαιο δείξαμε ότι η επαναληπτική αυτή μέθοδος μπορεί να εφαρμοστεί σε πραγματικό χρόνο, από την οποία λαμβάνουμε ταχείς ακριβείς εκτιμητές μικρής διασποράς για το σημείο λειτουργίας, το ισοδύναμο εύρος ζώνης και την πιθανότητα υπερχείλισης. Δηλαδή, μπορούμε να εκτιμήσουμε όλα τα μεγέθη τα οποία καθορίζουν την επίδοση ενός πολυπλέκτη σε πραγματικό χρόνο. Το ασυμπτωτικό πολλών πηγών έχει αποδειχθεί ότι είναι ένα σημαντικό εργαλείο, και σε πολλές περιπτώσεις το μοναδικό αποδοτικό εργαλείο για την ανάλυση των φαινομένων της πολυπλεξίας σε μεγάλα συστήματα. Εφόσον η λύση του προβλήματος βελτιστοποίησης της παράστασης `infsup` μπορεί να γίνει σε πραγματικό χρόνο για ένα σύνδεσμο, οι μηχανισμοί του δικτύου, όπως ο *Έλεγχος Αποδοχής Κλήσης*, μπορούν να υπολογίσουν γρήγορα το κέρδος της στατιστικής πολυπλεξίας και να οδηγήσουν σε αποδοτική χρήση και απόδοση των πόρων του δικτύου. Πολλά σχήματα χρέωσης υπηρεσιών δικτύων τα οποία βασίζονται στη χρήση των δικτυακών πόρων, χρησιμοποιούν το ισοδύναμο εύρος ζώνης των πηγών υπολογιζόμενο στο τρέχον σημείο λειτουργίας, οπότε η γρήγορη εκτίμησή του σε πραγματικό χρόνο είναι ζωτικής σημασίας.

Η μελέτη των συστημάτων ενός δικτύου ATM κάτω από το πλαίσιο της ισοδυναμίας κυκλοφορίας, και η εφαρμογή της σε μηχανισμούς ελέγχου, εξομοίωσης και χρέωσης, είδαμε ότι μπορεί να βοηθήσει στον καλύτερο σχεδιασμό και λειτουργία του δικτύου χωρίς σημαντική αύξηση της πολυπλοκότητας. Αντίθετα, οι μηχανισμοί που προτάθηκαν στην εργασία αυτή είναι πολύ απλοί στην υλοποίησή τους. Κατά συνέπεια το δίκτυο μπορεί να γίνει πιο ανταγωνιστικό εκμεταλλευόμενο την οικονομία κλίμακας που προσφέρει η στατιστική πολυπλεξία.

7.2 Μελλοντικές Κατευθύνσεις

Παρά τις σημαντικές εφαρμογές της ισοδυναμίας και αντικατάστασης κυκλοφορίας που αναφέραμε στην προηγούμενη παράγραφο, πολλές πτυχές της νέας αυτής προσέγγισης μένουν ακόμα να διερευνηθούν.

Όπως είδαμε στο Κεφάλαιο 5, η μέθοδος για την εκτίμηση του σημείου λειτουργίας, του ισοδυνάμου εύρους ζώνης και την πιθανότητας υπερχείλισης σε πραγματικό χρόνο παρουσίασε πολύ καλά αποτελέσματα στην περίπτωση των πηγών που βασίζονται σε μοντέλα Markov. Παρόλα αυτά για να χρησιμοποιηθεί σε ένα πραγματικό δίκτυο θα πρέπει να εξασφαλιστούν οι συνθήκες σύγκλισης της μεθόδου. Αυτό προϋποθέτει την εξαγωγή και μελέτη των συνθηκών σύγκλισης και πειραματισμό με πραγματική κυκλοφορία. Επίσης η χρήση γνωστών στατιστικών τεχνικών, όπως *bootstraping*, θα μπορούσε να ελαττώσει τη διασπορά του εκτιμητή του ισοδυνάμου εύρους ζώνης και συνεπώς να οδηγήσει σε μικρότερους χρόνους σύγκλισης. Η μέθοδός μας κάνει εκτίμηση σε πραγματικό χρόνο της πιθανότητας υπερχείλισης, ενώ σε ένα δίκτυο ATM το κριτήριο της ποιότητας υπηρεσίας που θα έπρεπε να μελετήσουμε είναι η CLP. Αν και τα δύο αυτά μεγέθη ακολουθούν το ίδιο ασυμπτωτικό σχήμα, στην πράξη διαφέρουν αρκετά. Η μέθοδός μας μπορεί άμεσα να επεκταθεί ώστε να εκτιμήσει την CLP, πολύ απλά χρησιμοποιώντας το ασυμπτωτικό σχήμα των Likhanov και Mazumdar (βλ. εξίσωση (2.9) και [26]) για τη CLP το οποίο βασίζεται στο ασυμπτωτικό πολλών πηγών. Για την εκτίμηση αυτή αρκούν οι μετρήσεις που συλλέγησαν για την εκτίμηση της πιθανότητας υπερχείλισης. Μια άλλη λύση είναι η προσέγγισή μας να συνδυαστεί με μεθόδους εκτίμησης της CLP σε πραγματικό χρόνο όπως αυτή των Courcoubetis, Fouskas και Weber [6] που βασίζεται στην εκτίμηση της CLP σε μικρότερης κλίμακας εικονικά υποσυστήματα. Η μέθοδό μας μπορεί να βελτιώσει την ακρίβεια πρόβλεψης στο [6] στην περίπτωση επιλογής πολύ μικρών εικονικών συστημάτων, όπου η ακρίβεια στην εκτίμηση της τιμής της συνάρτησης ασυμπτωτικού ρυθμού έχει μεγάλη σημασία.

Αν και επιβεβαιώσαμε τη χρησιμότητα του μοντέλου fBm στην αντικατάσταση κυκλοφορίας, πρέπει να μελετηθεί η χρήση μοντέλων με περισσότερες παραμέτρους. Αυτό είναι σημαντικό για την επέκταση του ορισμού της ισοδυναμίας έτσι ώστε να διατηρούνται και άλλα μεγέθη, εκτός του ισοδυνάμου εύρους ζώνης της κυκλοφορίας, όπως το μέσο μέγεθος της ουράς. Η συμπεριφορά τέτοιων μεγεθών έχει εκτενώς μελετηθεί και κατανοηθεί έως τώρα (π.χ. βλ. [28, 24, 21]), οπότε μια προσέγγιση μοντελοποίησης που συνδυάζει όλες αυτές τις μεθόδους με την έννοια της ισοδυναμίας φαίνεται να είναι εφικτή και προβλέπεται αποτελεσματική.

Επίσης, σύμφωνα με νέα αποτελέσματα της θεωρίας πολυπλεξίας στο [37], το ισοδύναμο

εύρος ζώνης μιας πηγής διατηρείται κατά το πέρασμα της κυκλοφορίας της από πολλούς συνδέσμους, ακόμα και αν αλλάζει η σύνθεση των πηγών σε κάθε σύνδεσμο. Τα αποτελέσματα αυτά, ανοίγουν την προοπτική ορισμού της ισοδυναμίας κυκλοφορίας για δίκτυα, όπου μπορούν να μελετηθούν άλλα σημαντικά προβλήματα όπως ο έλεγχος ροής (flow control) και η δρομολόγηση.

Η έννοια της ισοδυναμίας κυκλοφορίας που ορίσαμε στην παρούσα εργασία αποτελεί ένα νέο σημείο εκκίνησης για τη μοντελοποίηση κυκλοφορίας. Η συστηματική μελέτη της μοντελοποίησης κυκλοφορίας σε αυτό το μαθηματικό πλαίσιο μπορεί να οδηγήσει σε αναλυτικά εργαλεία τα οποία με τη σειρά τους να δώσουν το κίνητρο για νέες πρακτικές εφαρμογές.

Βιβλιογραφία

[1] Φ. Χαρμαντζής. *Θέματα σε Χρέωση και Ανάλυση Επίδοσης Κίνησης ATM.* Μεταπτυχιακή Εργασία, Πανεπιστήμιο Κρήτης, Ηράκλειο Κρήτης, 1996.

[2] N. G. Bean. *Statistical multiplexing in broadband communication networks.* PhD thesis, Univ. Cambridge, Cambridge, U.K., 1993.

[3] P. Billingsley. *Probabilty and Measure.* Wiley, New York, 2nd edition, 1986.

[4] D. D. Botvitch and N. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems*, 1995.

[5] T. Committee. *Traffic Management Specification Version 4.0.* The ATM Forum, January 1997.

[6] C. Courcoubetis, G. Fouskas, and R. Weber. An on-line estimation procedure for cell-loss probabilities in ATM links. In *3rd IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, UK, July 1995.

[7] C. Courcoubetis, V. A. Siris, and G. D. Stamoulis. Application and evaluation of large deviation techniques for traffic engineering in broadband networks. In *ACM SIGMETRICS '98/ PERFORMANCE '98 Joint International Conference on Measurement and Modeling of Computer Systems*, Madison, Wisconsin, June 1998.

[8] C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a switch handling many traffic sources. *Journal of Applied Probability*, 1996.

[9] N. Duffield and S. Low. The cost of quality in networks of aggregated traffic. In *Proc. of IEEE INFOCOM'98*, 1998.

[10] A. Elwalid, D. Heyman, T. V. Lakshman, D. Mitra, and A. Weiss. Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing. *IEEE Journal on Selected Areas in Communications*, 13(6), 1995.

[11] A. Elwalid, D. Mitra, and R. H. Wentworth. A new approach for allocating bandwidth and buffer to heterogeneous regulated traffic in an ATM node. *IEEE Journal on Selected Areas in Communications*, 13:1115--1127, August 1995.

[12] V. S. Frost and B. Melamed. Traffic modeling for telecommunications networks. *IEEE Communications Magazine*, 32(3):70--81, March 1994.

[13] M. W. Garrett and W. Willinger. Analysis, modeling and generation of self-similar VBR video traffic. In *ACM SIGCOMM*, pages 269--280, London, UK, Sept. 1994.

[14] R. J. Gibbens, F. P. Kelly, and P. B. Key. A decision-theoretic approach to call admission control in ATM networks. *IEEE Journal on Selected Areas in Communications*, 13(6):1101--1114, August 1995.

[15] M. Grossglauser and D. Tse. A framework for robust measurement-based admission control. submitted to IEEE/ACM Transactions on Networking, February 1998.

[16] B. Hajek and L. He. On variations of queue responses for inputs with identical mean and autocorrelation functions. In *1996 Conference on Information Sciences and Systems*, page 1195, Princeton, NJ, March 1996. Princeton University.

[17] C.-L. Hwang and S.-q. Li. On the convergence of traffic measurement and queueing analysis: A statistical-match queueing (SMAQ) tool. In *Proc. IEEE INFOCOM'95*, pages 602--612, 1995.

[18] Internet Traffic Archive. World wide web resource available at *http://town.hall.org/archives/pub/ita/*.

[19] ITU Recommendation I371. Traffic control and congestion control in B-ISDN.

[20] D. L. Jagerman, B. Melamed, and W. Willinger. Stochastic modeling of traffic processes. In J. H. Dshalalow, editor, *Frontiers in Queueing: Models, Methods and Problems*. CRC Press, 1996.

[21] P. Jelenkovic, A. Lazar, and N. Semret. The effect of multiple time scales and subexponentiality of MPEG video streams on queueing behavior. *IEEE Journal on Selected Areas in Communications*, 15(6):1052--1071, August 1997.

[22] F. P. Kelly. Notes on effective bandwidths. In F. P. Kelly, S. Zachary, and I. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 141--168. Oxford University Press, 1996.

[23] F. P. Kelly and R. J. Gibbens. Measurement-based connection admission control. In *15th International Teletraffic Congress Proceedings*, June 1997.

[24] L. A. Kulkarni and S.-q. Li. Measurement-based traffic modeling: Capturing important statistics. *submitted to Journal of Stochastic Models.*

[25] W. E. Leland, M. S. Taqq, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic. In D. P. Sidhu, editor, *ACM SIGCOMM*, pages 183--193, San Francisco, California, Sept. 1993. ACM. also in *Computer Communication Review* 23 (4), Oct. 1992.

[26] N. Likhanov and R. R. Mazumdar. Cell loss asymptotics for buffer fed with a large number of independent stationary sources. In *Proc of IEEE INFOCOM'98*, April 1998.

[27] N. Likhanov, B. Tsybakov, and N. D. Georganas. Analysis of an ATM buffer with self-similar ("fractal") input traffic. In *Proc. of IEEE INFOCOM'95*, Boston, April 1995.

[28] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins. Performance models of statistical multiplexing in packet video communications. *IEEE Trans. on Comm.*, 36(7):834--844, July 1988.

[29] M. Montgomery and G. de Veciana. On the relevance of time scales in performance oriented traffic characterizations. In *Proc. of IEEE INFOCOM'96*, pages 513--520, April 1996.

[30] I. Norros. A storage model with self - similar input. *Queueing Systems*, 16:387--396, 1994.

[31] I. Norros. On the use of fractional brownian motion in the theory of connectionless networks. *IEEE JSAC*, 13(6):953--962, August 1995.

[32] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. Technical Report No. 101, University of Wuerzburg. Institute of Computer Science Research Report Series, February 1995.

[33] B. K. Ryu and A. Elwalid. The importance of the long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities. In *Proc. of ACM SIGCOMM '96*, pages 3--14, August 1996.

[34] A. Simonian and J. Guibert. Large deviations approximations for fluid queues fed by a large number of on/off sources. *IEEE JSAC*, 13(7):1017--1027, August 1995.

[35] A. Weiss. A new technique for analyzing large traffic systems. *Adv. Appl. Prob.*, 18:506--532, 1986.

[36] W. Willinger. Self-similarity in high speed packet traffic: Analysis of Ethernet traffic measurements. *Statistical Science*, 10(1):67--85, 1995.

[37] D. Wischik. The output of a switch, or, effective bandwidths for networks. To appear in *Queueing Systems.*