ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Δίκαιος χρονοπρογραμματισμός μεγίστου-σταθμισμένου-ελαχίστου, για ένα Crossbar με ουρές αποθήκευσης στις εισόδους και μικρούς εσωτερικούς ενταμιευτές.

Νικόλαος Ι. Μ. Χρυσός

Μεταπτυχιακή Εργασία

Ηράκλειο, Μάρτιος 2002

Δίκαιος χρονοπρογραμματισμός μεγίστου-σταθμισμένου-ελαχίστου, για ένα Crossbar με ουρές αποθήκευσης στις εισόδους και μικρούς εσωτερικούς ενταμιευτές.

Νικόλαος Ι. Μ. Χρυσός

Μεταπτυχιακή Εργασία

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Περίληψη

Τα συστήματα μεταγωγών πακέτων παίζουν κρίσιμο ρόλο στη σχεδίαση δικτύων υψηλών επιδόσεων και υπό αυτή την έννοια πρέπει να κατασκευαστούν έτσι ώστε να είναι αποτελεσματικά και χαμηλού κόστους. Οι μεταγωγείς με ουρές αποθήκευσης στις εισόδους, απαιτούν μνήμες με ταχύτητα ακριβώς διπλάσια της ταχύτητας των γραμμών μεταφοράς, επιδεικνύοντας έτσι πολύ καλές ιδιότητες κλιμάκωσης. Εν τούτοις, αυτά τα συστήματα επίσης απαιτούν ένα κεντρικό χρονοπρογραμματιστή, που θα ρυθμίζει το εσωτερικό δίκτυο του μεταγωγέα, έτσι ώστε πακέτα παρόντα στις εισόδους να μεταβιβάζονται στις εξόδους προορισμού: η πολυπλοκότητα αυτού του χρονοπρογραμματιστή αυξάνει πολυωνυμικά με το μέγεθος του μεταγωγέα. Επιπλέον, είναι ακόμα αμφισβητούμενο αν αυτή η αρχιτεκτονική μπορεί να προσφέρει σύνθετη ποιότητα υπηρεσίας, χωρίς να χρησιμοποιεί υποβαθμισμένα τη λανθάνουσα χωρητικότητα μετάδοσης του συστήματος. Σε αυτή τη μεταπτυχιακή εργασία, προτείνουμε και αναλύουμε μία αρχιτεκτονική για την τοπολογία crossbar, με εσωτερική μνήμη, που ανήκει στην κατηγορία των παραπάνω μεταγωγών και λύνει το πρόβλημα του εσωτερικού χρόνοπρογραμματισμού με τρόπο απλό και αποδοτικό. Χρησιμοποιούμε εσωτερική αποθήκευση στο crossbar, που επιτρέπει σε ξεχωριστούς σε κάθε είσοδο και έξοδο χρόνοπρογραμματιστές, συλλογικά αλλά ταυτόχρονα και ο καθένας αυτόνομα, να κατασκευάζουν το σύνολο των πακέτων που θα προωθηθούν στις εξόδους. Στο σχήμα που

εξετάζουμε σε αυτή την εργασία, οι χρόνοπρογραμματιστές στις εισόδους και στις εξόδους υλοποιούν την αρχή της σταθμισμένης-κυκλικής δρομολόγησης(WRR) και οι οποίοι ασθενώς συγχρονίζονται, μέσω των σημάτων οπισθοδρομικής πίεσης(backpressure) που διαχειρίζονται την εσωτερική μνήμη. Αυτή η αρχιτεκτονική διακρίνεται από υψηλή, επαναλαμβανόμενη, δομική οργάνωση και απλότητα, που επιτρέπει χαμηλού κόστους, γρήγορη και αποτελεσματική υλοποίηση, κατάλληλη για ένα φάσμα από κλίμακες. Μελετούμε το κατά πόσο η εξυπηρέτηση που κατανέμει το προτεινόμενο σύστημα στις ¨ανταγωνιζόμενες¨ ροές, πλησιάζει την ιδεατή δίκαιη κατανομή, του μεγίστου-σταθμισμένου-ελαχίστου(Weighted Max-Min Fairness). Παρέχουμε εξαντλητικές προσομοιώσεις και προκαταρτικές αναλυτικές μαρτυρίες ότι το σύστημα συγκλίνει πολύ κοντά στον προαναφερθέντα, αντικειμενικό στόχο δικαιοσύνης. Επιπρόσθετα μελετάμε το χρόνο που χρείαζεται το σύστημα για να φτάσει στη κατάσταση ισορροπίας, και τις επιδόσεις παροχής του συστήματος, υπό διαφορετικές υποθέσεις για την φύση της εισερχόμενης κίνησης. Αν και χώρος για ένα πακέτο είναι ικανός για την λειτουργικότητα του χρονοπρογραμματισμού, προσομοιώσεις και αναλυτικές ενδείξεις φανερώνουν ότι για συστήματα μέχρι και 32x32, με το να δεσμεύσουμε χώρο αξίας 4-5 πακέτων για κάθε ροή εσωτερικά στο crossbar, λαμβάνουμε κοντά στο μεγίστο, τα προαναφερθέντα χαρακτηριστικά αποτελεσματικότητας.

Επιβλέπων, Ακαδημαίκος Σύμβουλος: **Μανόλης Κατεβαίνης**, Καθηγητής Πανεπιστημίου Κρήτης, τμήμα Επιστήμης Υπολογιστών

Weighted Max-Min Fair Scheduling, for a Crossbar, with Input Buffering and Small Internal Memory.

Nikolaos I. M. Chrysos

Master of Science Thesis

UNIVERSITY OF CRETE
COMPUTER SCIENCE DEPARTMENT

Abstract

Switching elements, play a pivotal role in the design of high performance networks and by these means, they have to be made both effective and inexpensive. Input buffered switches require the memories to run, at exactly twice the line rate, thus, demonstrating very good scalability properties. However, they also require a centralized scheduler, that configures the interconnect, so that a set of packets resident at the inputs, are forwarded to the switch outputs; the complexity of this scheduler, increases polynomially with the switch size. Moreover, it is ambiguous whether such an architecture, can provide sophisticated QoS guarantees, still without sacrificing the potential switching capacity of the system. In this thesis we propose and analyze a buffered c-rossbar architecture, that belongs to the class of input-buffered switches, and resolves the problem of internal scheduling in a simple and effective manner. We use internal buffering, which enables distinct servers at each input and output collectively but still independently, to create the set of flows that will be forwarded through the interconnect. In the scheme that we consider in this thesis, the input and output servers implements the Weighted Round Robin discipline and are loosely coordinated through the backpressure signals, that manage the internal in the crossbar buffers. This architecture is highly modular and simple, enabling cheap, fast and effective hardware implementations, appropriate for a range of scales. We study how close the service allocated, by the proposed system to the "competing" sessions/flows, approximates the ideal

weighted-max-min fair allocation. We provide extensive simulations and preliminary analytical hints, that indicate that this topology-scheduling discipline, converges very close to the aforementioned fairness, objective. In addition we study the time needed, to enter the ideal state, we quantify the unfairness during the stabilization process, and the saturation throughput of the system, under various assumption on the form of traffic. Although one cell buffer space is adequate for the scheduling operation, simulation results and analytical evidence, show that for switches of size up to 32x32, by allocating buffer size worth 4 to 5 cells for each session internally to the crossbar, we get the most of the previously defined performance characteristics.

Supervisor, Academic Consultant: **Manolis Katevenis**, Professor at University of Crete, Computer Science Department

Ευχαριστίες

Θα ήθελα να ευχαριστήσω όλους όσους βοήθησαν στην περάτωση αυτής της εργασίας. Πρώτον από όλους θελω να ευχαριστήσω τον καθ. Μανόλη Κατεβαίνη που έδωσε την ιδέα για την εργασία, για τις ανεκτίμητες υποδείξεις και ιδέες καθόλη τη διάρκεια της εργασίας.

Θέλω να ευχαριστήσω την κυρία Παρασκευή Φραγκοπούλου για τις συζητήσεις πάνω στο φαινόμενο της σύγκλισης και τον κύριο Βασίλη Σύρρη για τις γνωσείς του που προσφερε απλοχερά πάνω σε θέματα δικτύων και τις υποδείξεις/προτάσεις του πάνω σε ζητήματα προσομοίωσης. Επίσης θα ήθελα να ευχαριστήσω τον κύριο Γεωργακόπουλο για την ώρα που δαπάνησε διαβάζοντας μέρος της Αγγλική/πρώτη έκδοσης και για τις συζητήσεις που κάναμε στη συνέχεια, σε θέματα αποδείξεων.

Θέλω να ευχαριστήσω τον Κώστα Χαρτερό για τις πολύτιμες ώρες συνεργασίας και τις συζητήσεις πάνω σε ζητήματα της εργάσίας, όπως και τον Γιώργο Σαπουντζή για τις συζητήσεις που κάναμε και τη βοηθεία του σε θέματα συγγραφής/παρουσίασης.

Λίγο πιο γενικά, θα ήθελα να ευχαριστήσω το Πανεπιστήμιο Κρήτης/Τμήμα Επιστήμης Υπολογιστών για τις βασικές γνώσεις στα συστήματα και την επιστήμη υπολογιστών, όπως και το University of Manchester/Computer Science Department για την υπερπολύτιμη ευκαιρία που μου έδωσε να αντιμετωπίσω το - εώς τότε, σχετικά δύσκολο για μένα - πρόβλημα των υπολογιστών, ως μέρος του γενικού γνωσιακού αντικειμένου. Τους καθηγητές Διονύση Πνευματικάτο, Δημήτρη Σερπάνο και Πάνο Τραχανιά που μου έδωσαν την ευκαιρία να κάνω μεταπτυχιακά. Ειδικά τον καθ. Διονύση Πνευματικάτο για την συνεργασία κατά την διάρκεια του μεταπτυχιακού σε ένα άλλο θέμα που ασχολήθηκα και για την ενδιαφέρουσα διπλωματική εργασία που κάναμε μαζί.

Σίγουρα θέλω να ευχαριστήσω την οικογένεια μου για τη διαρκή υποστήριξη, και πιο ειδικά, τον πατέρα μου που επέμενε να ασχοληθώ με τους υπολογιστές και όχι με τη φυσική στο κρίσιμο σημείο των γενικών εξετάσεων και την μητέρα μου για την εκπαίδευση που μου έδωσε όταν ήμουν μικρός πάνω σε φιλολογι-

κά/φιλοσοφικά ζητήματα αλλά και για τις - αναλόγου περιεχομένου - διορθώσεις της σε αυτό το κείμενο. Τέλος θα ήθελα να ευχαριστήσω και όλους τους φίλους/φίλες μου για την (όποια)επιρροή τους επάνω μου, που σίγουρα πρέπει συνέβαλαν (και)σε αυτήν την εργασία.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Η εκθετική αύξηση του διαδικτύου συνοδεύεται από συνεχώς αυξανόμενες και αποκλίνουσες απαιτήσεις για υπηρεσίες. Επιπρόσθετα, στη συγγενεύουσα περιοχή των κατανεμημένων/παράλληλων μηχανών, δεν έχει βρεθεί ακόμα το κατάλληλο παράδειγμα/πρότυπο επικοινωνίας για μεγάλα (πολύ)υπολογιστικά συστήματα, που θα προαγάγει τις επιδόσεις και τελικά τις υπολογιστικές λύσεις. Για να ανταποκριθούμε στην αύξηση της ζήτησης και τις διαφοροποιήσεις στην φύση των απαιτήσεων, πρέπει να κατασκευάσουμε έξυπνα και αποδοτικά δίκτυα που να λαμβάνουν υπόψιν τους τη διαφορετικότητα των αιτήσεων. Τα συστήματα μεταγωγέων παίζουν καθοριστικό ρόλο σε αυτήν την κατεύθυνση, αφού αυτά είναι τα σημεία όπου ανταγωνιζόμενη κίνηση πολυπλέκεται πάνω στους κοινούς πόρους. Ο ρόλος των μεταγωγέων(ή και πολυδιακοπτών) είναι ουσιαστικά διττός. Από τη μία ρόλος τους είναι να προωθήσουν τα εισερχόμενα πακέτα στους σωστούς προορισμούς και δεύτερον να αποφασίσουν τη σειρά μεταβίβασης/εξυπηρέτησης. Καλή θεωρείται μία αρχή εξυπηρέτησης, όταν μπορεί να λάβει υπόψην της τη διαφορετικότητα των αιτήσεων, όταν κατανέμει τους πόρους δίκαια προσφέροντας προστασία στις ανταγωνιζόμενες ροές, χρησιμοποιώντας ικανοποιητικά την χωρητικότητα που διαχειρίζεται.

Η δημιουργία φτηνών και αποδοτικών συστημάτων μεταγωγής που θα μπορεί να εξασφαλίσει ποιοτική υπηρεσία σε ροές/εφαρμογές χρηστών, αποτελεί ένα από τα πιο ενδιαφέροντα ζητήματα στην περιοχή των δικτύων-συστημάτων. Η κατασκευασιμότητα προσδιορίζεται συνήθως κάτω από υλικο-οικονομικούς(hardwareeconomic) όρους, ενώ η αποδοτικότητα με όρους πολιτικής/δικαιοσύνης-αποτελεσματικότητας. Τα περισσότερα προβλήματα αποδοτικότητας μπορούν να λυθούν αυξάνοντας το κόστος κατασκευής – χρησιμοποιώντας για παράδειγμα είτε συστήματα με ουρές στις εξόδους(pure output queuing systems), είτε εσωτερική αύξηση της ταχύτητας μετάδοσης(internal speed-up). Εδώ θεωρούμε ότι τίποτα από τα δύο δεν είναι ουσιαστικά εφικτό, όπως ισχύει όντως σε μεγάλες κλίμακες ή εξωτερικές γραμμές πολύ υψηλής ταχύτητας.

Το πρόβλημα που αντιμετωπίζουμε επομένως, είναι σαφώς εξαρτώμενο από το τι θεωρούμε ως αποτελεσματικό και δίκαιο, δεδομένων των περιορισμών που επιβάλλουμε στο σύστημα. Αν σκεφτούμε/συλλάβουμε το μεταγωγέα σαν ενα μαύρο κουτί(black box), τότε το ιδεατό είναι να ανταγωνίζονται μόνο ροές που προορίζονται σε <u>κοινό</u> εξωτερικό σύνδεσμο. Αλλά αυτή η θεώρηση είναι επίσης απλουστευτική, αφού τα συστήματα που την πραγματώνουν είναι εξαιρετικά ακριβά κάτω από υλικό/τεχνικό-οικονομικούς όρους, κυρίως λόγω των απαιτήσεων τους για γρήγορη-μεγάλη μνήμη.

Φαινόμενα συναγωνισμού εμφανίζονται οπουδήποτε διαφορετικές ροές κίνησης πολυπλέκονται. Αυτό συμβαίνει τόσο σε δίκτυα μεγάλης/μεσαίας κλίμακας, αφού δεν μπορούμε να υποστηρίξουμε απευθείας συνδέσεις από (κάθε)άκρο σε άκρο, όσο και εσωτερικά στα περισσότερα συστήματα μεγαγωγέων αφού τα output queuing συστήματα είναι ακριβά. Τα συστήματα με τις ουρές αποθήκευσης στις εισόδους(input queuing systems) που προτιμώνται λόγω των χαμηλών/βέλτιστων απαιτήσεων τους σε μνήμες, μαστίζονται από εσωτερικό ανταγωνισμό. Η κατασκευή αποδοτικών input queuing συστημάτων που θα ¨αντιλαμβάνονται¨ την διαφορετικότητα των ροών/πακέτων, ενώ παράλληλα θα εκμεταλλεύονται την λανθάνουσα χωρητικότητα μετάδοσης παραμένει ανοιχτό ζήτημα, κυρίως λόγω των απαιτήσεων στο υποσύστημα ελέγχου προώθησης(εσω-εξωτερικός χρονοπρογραμματιστής μεταγωγής).

Η έρευνα στο χρονοπρογραμματισμό <u>ενός</u> κοινού πόρου αντιθέτως, έχει πολλά καλά/βέλτιστα αποτελέσματα να επιδείξει, στην κατά τα άλλα κοινή κατεύθυνση, διαφοροποίησης/δικαιοσύνης/αποτελεσματικότητας/υλοποιησιμότητας. Η χρησιμοποίηση GPS-like(Generalized Processor Sharing) αρχών δρομολόγησης σε συστήματα πακέτων, αποδεδειγμένα αποτελεί την βέλτιστη λύση για ζυγισμένη-στατιστική πολυπλεξία, όσον αφορά χαρακτηριστικά δικαιοσύνης, προστασίας και ποιότητας υπηρεσίας.

Το πρόβλημα, ή τα κίνητρά μας ήταν να σχεδιάσουμε μία input queuing αρχιτεκτονική χρονοπρογραμματισμού/μεταγωγής πακέτων για πολλούς πόρους, που να παρουσιάζει ιδιότητες παρόμοιες με αυτές των αρχών δρομολόγησης GPS στην περίπτωση ενός κοινού πόρου. Μία πιθανή/ικανή έκφραση αυτού του προβλήματος είναι λοιπόν: *Δεδομένων των υλικό/κατασκευαστικών-οικονομικών περιορισμών, που μας αναγκάζουν να υιοθετήσουμε την input-queuing πρακτική και η οποία εισάγει το πρόβλημα του συναγωνισμού και σε επιπλέον κοινόχρηστες περιοχές/πόρους, τι είναι το καλύτερο που μπορείς να κάνεις για να προσφέρεις υπηρεσία που να θυμίζει GPS, χωρίς να εισάγεις τραγικές απαιτήσεις για άλλα υποσυστήματα - βασικά στην περιοχή του χρονοπρογραμματισμού.*

Φαίνεται ότι υπάρχει μία αρχιτεκτονική, η οποία διατηρεί το γνώρισμα δικαιοσύνης μεγίστου-ζυγιζόμενου/σταθμισμένου-ελαχίστου στην κατανομή των πόρων(Weighted Max-Min Fairness), που παρουσιάζουν και οι GPS servers. Σε αυτή την εργασία, παρουσιάζουμε και υποστηρίζουμε αυτήν την αρχιτεκτονική που μπορεί εύκολα να υλοποιηθεί, και προσφέρει δίκαιη κατανομή των εσω-εξωτερικών πόρων σε διαφορετικές ροές/ζευγάρια-εισόδου
εξόδου. Αν σε κάθε τέτοια ροή, αντιστοιχηθεί ένα βάρος που αντιπροσωπεύει τη σχετική προτεραιότητα/ανάγκη-για-υπηρεσία αυτής της ροής, τότε η κατανομή που θα προσφέρει το σύστημα θα πλησιάζει πολύ κοντά στον παραπάνω αντικειμενικό στόχο δικαιοσύνης.

Η συνεισφορά αυτής της εργασίας είναι διπλή. Αφ' ενός προτείνουμε μία αρχιτεκτονική που επιλύει το πρόβλημα του χρονοπρογραμματισμού σε συστήματα πολλών πόρων με τις ουρές αποθήκευσης στις εισόδους, λαμβάνοντας υπόψιν την ποσοτικοποιημένη – μέσω μονοδιάστατων βαρών – διαφορετικότητα των ροών. Σε αυτή τη κατεύθυνση παρουσιάζουμε το σκεπτικό και τη συλλογιστική, που μας οδήγησαν σε αυτήν τη λύση. Η άλλη συνεισφορά της εργασίας είναι η ανάλυση της προτεινόμενης αρχιτεκτονικής κάτω από διαφορετικούς/πιθανούς αντικειμενικούς στόχους, και ιδιαίτερα η πειραματική/αναλυτική επιβεβαίωση/ποσοτικοποίηση της καταλληλότητας της, στην πραγμάτωση του ιδεατού στόχου δικαιοσύνης μεγίστου-ζυγισμένου-ελαχίστου. Σε αυτή τη περίπτωση, μελετάμε/αναλύουμε/προτείνουμε την επίδραση κρίσιμων σχεδιαστικών παραμέτρων που σχετίζονται με τα αναδυόμενα προβλήματα επιλογής στο χώρο κόστους-επίδοσης-αποτελεσματικότητας(cost-effectiveness-performance ememrging trade-off). Η εργασία είναι όσο πιο γενική μπορέσαμε: δεν κάναμε επιπλέον υποθέσεις πάνω στη φύση των βαρών, πέραν του ότι σχετίζονται – ευθέως ανάλογα – με την σχετική προτεραιότητα/ανάγκη-υπηρεσίας των ροών.

Το σύστημα που προτείνουμε παρουσιάζει εξαιρετικές ομοιότητες με κάποιες πρόσφατες προτάσεις για κατανεμημένο χρονοπρογραμματισμό [DGPS][ChQoS]. Ενώ οι εργασίες έγιναν αυτόνομα, φτάσαμε/καταλήξαμε στην ίδια βασική αρχιτεκτονική. Αλλά στις δικές τους προτάσεις, εξετάζουν συγκεκριμένες προυποθέσεις για ποιότητα υπηρεσίας(delay bounds[1]), απαίτηση που τους ανάγκασε να κάνουν υποθέσεις πάνω στην φύση των βαρών/ροών και αναγκαστικά σχεδόν, να υποεκτιμήσουν την χωρητικότητα του συστήματος. Εμείς μελετάμε τη δικαιοσύνη του συστήματος όταν οι ροές χρησιμοποιούν δυνητικά ολόκληρη την διαθέσιμη χωρητικότητα μετάδοσης.

Το υπόλοιπο της αναφοράς είναι οργανωμένο ως ακολούθως: Στο κεφάλαιο ένα, μελετάμε/προτείνουμε ένα γενικό σχήμα ορισμού αντικειμενικών στόχων στην περιοχή των δικτύων και συνηγορούμε υπέρ της υιοθέτησης της πολιτικής αρχής-δικαιοσύνης μεγίστου-ζυγισμένου-ελαχίστου για την οποία προσφέρουμε διαφορετικούς διαισθητικούς και μαθηματικούς ορισμούς/αλγοριθμούς. Στο κεφάλαιο δύο, παρουσιάζουμε εν συντομία εργασίες που στηριχτήκαμε/χρησιμοποιήσαμε, πάνω σε αρχές χρονοπρογραμματισμού ενός και πολλών πόρων – επιμένοντας κάπως πάνω στην αρχή GPS και στο πρόβλημα χρονοπρογραμματισμού crossbar. Στο κεφάλαιο τρία περιγράφουμε την αρχιτεκτονική και κάποιες αναλυτικές ενδείξεις για την αναμενόμενη συμπεριφορά του. Ακολουθεί η πειραματική επιβεβαίωση λειτουργίας του συστήματος μέσω προσομοιώσεων στο κεφάλαιο τέσσερα και συζητήσεις πάνω στα αποτελέσματα στο κεφάλαιο πέντε. Τέλος καταλήγουμε με συμπεράσματα και σημειώσεις για μελλοντική εργασία στο κεφάλαιο έξι. Σημειώνουμε ότι ουσιαστικά αυτό το κείμενο μερικώς αποτελεί, μεταφορά/μετάφραση/εκλέπτυνση μέρους αναφοράς στην Αγγλική γλώσσα,

[1] Ένα ακόμα μεταφερόμενο χαρακτηριστικό των GPS servers

όπου το περιεχόμενο/συζητήσεις-επεξηγήσεις-αποδείξεις-θέματα-πειράματα, είναι εκτενέστερα. Οι αναφορές σε αυτό το κείμενο([BufCrossbar]) θα είναι με τη σειρά τους εκτενείς.

Κεφάλαιο 2

Αντικειμενικοί στόχοι στην περιοχή των δικτύων και Weighted Max-Min Fairness

2.0.1 Πολιτική, τεχνικές και αποδοτικότητα.

Η περιοχή των δικτύων ασχολείται με την κατασκευή αποδοτικών δικτύων: υπό αυτήν την έννοια μελετά τα προβλήματα τοπολογίας, κατανομής πόρων και φυσικά της ζήτησης. Τελικός σκοπός φυσικά, δεν μπορεί να είναι άλλος απο την αποδοτική ικανοποίηση των αναγκών των χρηστών του δικτύου, δηλαδή, την ικανοποιητικήν κάλυψη της ζήτησης· κάθε αντικειμένικός στόχος λοιπόν, δεν μπορεί παρά να βασίζεται και να στοχεύει στο να εξυπηρετεί τις αιτήσεις που την πραγματώνουν. Μόνο αν όλα τα επιμέρους προβλήματα αντιμετωπισθούν συγχρόνως και ενιαία/ενικά, μπορούν να κατασκευαστούν τα *πιο αποδοτικά* δίκτυα. Φυσικά λόγοι πολυπλοκότητας και οικονομικό-κοινωνικοί περιορισμοί, αναγκαζούν την ανεξάρτητη εξέταση επιμέρους προβλημάτων. Σέ κάθε περιοχή ή και εργασία, διαφορετικοί τρόποι ανακαλύπτονται/χρησιμοποιούνται για να οριστεί η αποδοτικότητα. Φυσικά πάλι, όλοι οι ορισμοί αυτοί στηρίζονται κατηγορηματικά ή όχι, στην βασική υπόθεση ότι το μέτρο επίδοσης, εξασφαλίζει καλύτερη και οικονομικότερη ικανοποίηση της ζήτησης. Αφού ο τελικός σκοπός κάθε προσπάθειας είναι τελικά πολιτικός(διανομή και αποτελεσματικότητα), και ο βασικότερος ορισμός της αποδόσης δεν μπορεί παρά να είναι πολιτικός. Άν θέλουμε να φτιάξουμε/οργανώσουμε μία δικτυακή κοινωνία, θα πρέπει πρώτα να ορίσουμε τη μονάδα της ζήτησης και στη συνέχεια μπορούμε να μιλήσουμε για αποδοτικότητα. Αυτό που ακολουθεί είναι ένα γενικό μοντέλο (προσδ)ορισμού της αποδοτικότητας.

2.0.2 Ένα γενικό μοντέλο ορισμού δικτυακών αντικειμενικών στόχων — Οι ανάγκες/ικανοποίηση της εφαρμογής

Στα δίκτυα πακέτων, οι εφαρμογές αποστέλλουν/ανταλλάσουν πακέτα. Το δυναμικό περιβάλλον εκτέλεσης του προγράμματος εφαρμογής και πραγματώσης της επικοινωνίας, καθώς και η διαφορετική φύση των εφαρμογών, έχουν ως συνέπεια, κάθε πακέτο να μήν έχει την ίδια *ανάγκη* για ποιότητα εξυπηρέτησης. Δηλαδή αλλιώς θα επηρεάσει την εφαρμογή A, αν το πακέτο p_1^A καθυστερήσει να εξυπηρετηθεί και αλλιώς να καθυστερήσει το ίδιο το p_{1000}^A. Όμοια, διαφορετικά θα επηρεαστεί μία εφαρμογή B άν το χιλιοστό πακέτο της(p_{1000}^B) καθυστερήσει το ίδιο με το χιλιοστό πακέτο της A(p_{1000}^A) – για παράδειγμα η A μπορεί να είναι μία εφαρμογή πραγματικού χρόνου(π.χ. tele-conference) ενώ η B μπορεί να είναι μία εφαρμογή μεταφοράς αρχείων(π.χ ftp). Εισάγεται λοιπόν η ανάγκη να θεωρήσουμε δυναμικά βάρη, διαφορετικά για κάθε εφαρμογή αλλα διαφορετικά και για κάθε πακέτο μίας ροής. Η προσθετική χρησιμότητα/ικανοποίηση(marginal utility/MU) που λαμβάνει μία εφαρμογή από την εξυπηρέτηση(B_p) ενός πακέτου(p) μιας από της ροές επικοινωνίας της(f), θα είναι αντιστρόφως ανάλογη του βάρους αυτού του πακέτου(P_p).

$$\mathcal{MU}_p^f = \mathcal{B}_p^f \;/\; \mathcal{P}_p^f$$

Έτσι μία ροή με μεγάλη τρέχουσα ανάγκη για εξυπηρέτηση και συνεπώς μεγάλο τρέχον βάρος, χρειάζεται επιπλέον εξυπηρέτηση για να λάβει την ίδια προσθετική utility, συγκρινόμενη με μία ροή με μικρότερο τρέχον βάρος. Αν τώρα θεωρήσουμε το χρόνο ως μία ακολουθία από διαστήματα όπου μία ροή είτε έχει πακέτο να στείλει – συνοδευόμενο από κάποιο βάρος/βαθμό-ανάγκης-υπηρεσίας –, είτε όχι – οπότε θεωρούμε μηδενική marginal utility –, τότε η συνολική ϊκανοποίηση" της ροής από το δίκτυο(Net) ύστερα από ένα διάστημα T θα είναι:

$$\mathcal{U}_T^f = \sum_T \mathcal{MU}_{t_i}^f$$

Τώρα μπορούμε να ορίσουμε μία κλάση από αντικειμενικούς στόχους, ως μία, οποιαδήποτε, συνάρτηση πάνω στα utilities που κάθε συγκεκριμένο δίκτυο αποδίδει στις ροές F που το χρησιμοποιούν:

$$\mathcal{EFFECTIVENESS}_{\mathcal{T}}^{network} = \mathcal{OBJ}_{f \in F}(\mathcal{U}_{\mathcal{T}}^{f})$$

Σε αυτό το μοντέλο μπορούμε να διαφοροποιούμε μεταξύ διαφορετικών ροών και μεταξύ διαφορετικών πακέτων της ίδιας ροής(U_f), μεταξύ πολιτικών(OBJ), αλλά και μεταξύ του τί θεωρούμε ως εφαρμογή/ροή. Ένας προφανής αντικειμενικός στόχος, είναι το άθροισμα των utilities να είναι όσο το δυνατόν μεγαλύτερο [ShNetObj]. Αν και αυτός ο σκοπός δείχνει να οδηγεί και να δουλεύει για το συνολικό καλό – μέγιστο net-social-welfare –, είναι πιθανό να δημιουργεί κατανομές όπου ορισμένες ροές παίρνουν όλο το utility/πόρους, ενώ άλλες δεν παίρνουν τίποτα· ύπο αυτή την έννοια δεν είναι πάντα δίκαιος. Η δίκαιη κατανομή μεγίστου-ζυγισμένου-ελαχίστου είναι αυτή η κατανομή, που εκπληρώνει τέλεια και αξιωματικά τον αντικειμενικό στόχο:

$$\mathcal{EFFECTIVENESS}_{\mathcal{T}}^{network} = \mathcal{MIN}_{f \in F}(\mathcal{U}_{\mathcal{T}}^{f})$$

Η διαισθητική ομορφιά αυτού του στόχου, είναι ότι προσπαθεί να εξασφαλίσει, στη ροή που θα πάρει το μικρότερο utility – οποιαδήποτε και αν είναι αυτή – όσο το δυνατόν utility γίνεται· στην περίπτωση αυτή μπορείτε να σκεφτείτε τίποτα πιο δίκαιο; Γενικά οι αντιρρήσεις/ενστάσεις σε αυτή την πολιτική, είναι ότι μπορεί να οδηγήσει στο να πάρουν όλες οι ροές πολύ μικρό service και όλες να υπολειτουργήσουν. Το ίδιο επιχείρημα υποστηρίζει/προτιμάει και το ¨θεσμό¨ του admission control. Η απάντηση είναι ότι αν και το επιχείρημα είναι τεχνικά σωστό, ουσιαστικά δεν είναι πολιτικό. Άλλωστε, αν το δίκτυο δεν μπορεί να εξασφαλίσει κάποια ελάχιστη εξυπηρέτηση για όλες τις ροές στη μεγαλύτερη έκταση του χρόνου τότε τεχνικά πρέπει να επανακατασκευαστεί.

Μία άλλη αρχή δικαιοσύνης, είναι η αναλογική δικαιοσύνη(proportional fairness)[1]. Αποδεικνύεται ότι και αυτή η αρχή ανήκει στην παραπάνω κλάση. Διαισθητικά η αναλογική δικαιοσύνη μας λέει ότι όσο πιο μεγάλη η απόσταση επικοινωνίας – μετρημένη σε δικτυακούς κόμβους(hops) –, τόσο χειρότερη η υπηρεσία που θα πάρει. Αν και αυτό σε πρώτο χρόνο είναι λογικό και δίκαιο – μεγαλύτερη απόσταση $\rightarrow$ ανταγωνίζεσαι περισσότερους $\rightarrow$ δικαιούσαι να πάρεις

[1] Πειραματικά, υπάρχουν ενδείξεις ότι στο Internet, κυριαρχεί αυτή η αρχή, μέσω του πρωτοκόλλου επικοινωνίας TCP/IP.

λιγότερο από τον καθένα –, προσωπικά, δεν μπορούμε να την δεχτούμε, καθώς μας λέει ότι: κοίτα να επικοινωνείς με κοντινούς γείτονες/κόμβους αν θες ποιότητα. Αν είναι να φτιάξουμε μία κοινή δικτυακή πλατφόρμα/κοινωνία όπου οι αποστάσεις θα εκμηδενίζονται – αυτό δεν είναι το όνειρο όλων των cyberman/πολιτικών/επιχειρηματιών, όταν φαντάζονται/μιλάνε-για την επερχόμενη κοινωνία της πληροφορίας; –, δεν μπορούμε να υποκρύψουμε αυτή την αρχή στις βάσεις της[2]. Περισσότερη ανάλυση/συζήτηση σε αυτά τα ζητήματα μπορούμε να βρούμε στα [BufCrossbar][UtMaxMin][ShNetObj]. Τώρα θα συνεχίσουμε επικεντρώνοντας την προσοχή μας στον ιδεατό αντικειμενικό στόχο, μέγιστο-ζυγισμένο-ελάχιστο ή και μέγιστο-μικρότερο-utility(Weighted Max-Min Fairness, or Utility Max-Min Fairness).

2.1 Η Δίκαιη κατανομή Μεγίστου-Σταθμισμένου-Ελαχίστου

2.1.1 Διαισθητική περιγραφή

Είδαμε προηγουμένως ότι η δίκαιη κατανομή μεγίστου-σταθμισμένου-ελαχίστου, που ανήκει στην παραπάνω γενική κλάση από πιθανούς δικτυακούς στόχους, εξασφαλίζει ότι η ροή που θα πάρει τελικά το μικρότερο utility – αξιωματικά ίσο με rate/weights αν το βάρος μίας ροής θεωρηθεί σταθερό στο χρόνο – από το δίκτυο, παίρνει όσο το δυνατόν περισσότερο utility γίνεται. Αυτό διαισθητικά και κυριολεκτικά σημαίνει ότι δεν υπάρχει άλλη εφικτή – υπό τους όρους της τοπολογίας του δικτύου και της συγκεκριμένης/σταθερής ζήτησης – κατανομή που να απέδιδε σε αυτή τη ροή περισσότερο utility, χωρίς να "κλέψει" το παραπάνω utility από μία ροή που έπαιρνε λιγότερο ή ίσο utility στη πρώτη κατανομή. Σημαίνει πάλι, ότι αν μία κατανομή μπορεί να αποδώσει σε μία ροή παραπάνω utility χωρίς να επηρεάσει άλλες ροές, τότε πιθανότατα θα έπρεπε να το κάνει, αλλιώς δεν θα είναι δίκαιη σύμφωνα με την οπτική του μεγίστου-σταθμισμένου-ελαχίστου($WMMF$).

[2]Φυσικά αυτό δεν σημαίνει ότι δεν πρέπει να προσπαθούμε, σε άλλα επιπεδα, να διατηρούμε την επικοινωνία "τοπική" όποτε ειναι δυνατό, αλλά αυτό πάλι είναι τεχνικό και δεν μας ενδιαφέρει εδώ.

Ένας απλός νοητικός τρόπος, για να βρούμε τα ποσοστά των ροών με βάση αυτήν την αρχή, είναι να αρχίσουμε να δίνουμε σε κάθε ροή απειροελάχιστο επιπλέον/προσθετικό ρυθμό υπηρεσίας, ανάλογο με το βάρος της κάθε ροής, μέχρι να συμφορήσουμε κάποιο πόρο, ή μέχρι μία ροή να μήν παρουσιάζει επιπλέον ζήτηση. Όταν συμφορηθεί/εξαντληθεί κάποιος πόρος βγάζουμε από το παιχνίδι όλες τις ροές που τον χρησιμοποιούν, ενώ όταν μία ροή δεν έχει επιπλέον ζήτηση βγάζουμε μόνο αυτή.

2.1.2 Μαθηματικά-αυστηροί ορισμοί

Δεδομένου δικτύου και ζήτησης, μπορούμε να ορίσουμε για κάθε εφικτή κατανομή S ένα διάνυσμα κατανομής(allocation vector) V^S, που δεν είναι τίποτα άλλο από ένα μη-φθείνον διάνυσμα από τα utilitiies που η συγκεκριμένη κατανομή έχει αποδώσει στις υποψήφιες ροές F.

$$V^S = (U^S_{f_1}, U^S_{f_2}, , , U^S_{f_n}) : U^S_{f_1} \leq U^S_{f_2} \cdots \leq U^S_{f_n}$$

Με τους παραπάνω όρους, μία κατανομή είναι $WMMF$ αν και μόνο άν, το διάνυσμα κατανομής είναι το μεγαλύτερο δυνατόν στο σύνολο των εφικτών κατανομών S^*, σύμφωνα με τη λεξικογραφική διάταξη.

$$S \text{ is max-min fair} \equiv \qquad V^S >_{lex} V^x \qquad\qquad \forall x \in S^*$$

Το γεωμετρικό πρόβλημα μεγιστοποίησης που προκύπτει λύνεται με γραμμικό προγραμματισμό([HaydenMaxMin]). Ουσιαστικά η λύση – είναι μοναδική αν η συνάρτηση του utility είναι αύξουσα ως προς το βαθμό υπηρεσίας – καταλήγει σε μία κατάσταση ισορροπίας(equilibrium), όπου η παραμικρή διαταραχή αποκλίνει του τέλεια ορισμένα αρχικού στόχου.

Ένας άλλος αυστηρός και χρήσιμος ορισμός γίνεται εφικτός αν εισάγουμε την έννοια της συμφόρησης(bottleneck). Ένας κόμβος/πόρος λέγεται κόμβος-μέγιστης-συμφόρησης/bottleneck για μία ροή, αν σε αυτόν τον πόρο αυτή η ροή παίρνει το μεγαλύτερο – όχι αυστηρά μεγαλύτερο, αλλά το μέγιστο(=δεν υπάρχει μεγαλύτερο) – utility, και ο συγκεκριμένος κόμβος είναι εξαντλημένος(=δεν έχει αδέσμευτη χωρητικότητα). Μία ροή που δεν έχει πολλή ζήτηση, αρκετή για να απορροφήσει το δίκαιο μερίδιο της, λέγεται συμφορημένη στην πηγή(bottlenecked at source). Τώρα μία κατανομή λέγεται και είναι $WMMF$, αν και μόνο αν, κάθε ροή έχει τουλάχιστον ένα κόμβο συμφόρησης.

2.1.3 Αλγόριθμοι

Αλγορίθμους για τον υπολογισμό της κατάστασης δίκαιης ισορροπίας $WMMF$ που περιγράφουμε μπορεί κανείς να βρεί στα [Keshav97][UtMaxMin]. Στην περίπτωση ενός μόνο κόμβου/πόρου απλά κάθε ροή παίρνει ποσοστό ανάλογο με το βάρος της $\rightarrow$ ίσα utilities για τις ροές που δεν είναι συμφορημένες στην πηγή(γιατί;)· στην περίπτωση πού έχουμε πολλούς κόμβους, επαναληπτικά βρίσκουμε τη ροή που δικαιούται το μικρότερο utility – χρησιμοποιώντας τον αλγόριθμο για τη μονοδιάστατη περίπτωση – της αποδίδουμε το ποσοστό που της εξασφαλίζει αυτό το utility και συνεχίζουμε μέχρι να εξαντλήσουμε τις ροές. Πιο αλγοριθμικά/αυστηρά αυτό γίνεται ως εξής:

Ξεκινάμε μαρκάροντας όλες τις ροές ως μη-συμφορημένες(*unbottlencked*) και τερματίζουμε όταν όλες συμφορηθούν[3].

1. At each link l, determine the available bandwidth for the unbottlenecked flows $i \in U \cap F_l$ crossing l:
   . $A_l = C_l - \sum_{i \in B \cap F_l} r_i$
2. At each link l, allocate temporary bandwidth to each unbottlenecked flow crossing l, by finding the bandwidth allocation that either saturates the link and results in equal utility, or absorbs the demand of a flow f:
   $\forall i \in U \cap F_l$, allocate $t_{i,l}$ such that:
   . $\sum_{i \in U \cap F_l} t_{i,l} = A_l$ and
   . $\forall j \in U \cap F_l, f_i(t_{i,l}) = f_j(t_{j,l})$ [4]

[3] F είναι το σύνολο των ροών, U τό σύνολο των μή συμφορημένων ροών και B το σύνολο των συμφορήμένων. Το F_l προσδιορίζει τις ροές που χρησιμοποιούν τον πόρο/γραμμή l, τό C_l είναι η χωρητικότητα του κόμβου l και $f_i(x)$ είναι η συνάρτηση utility της ροής i για τιμή ποσοστούπαροχής x.

[4] Αυτό είναι απλό: απλώς κανονικοποιούμε τα βάρη των μή συμφορημένων ροών που χρησιμοποιούν αυτό το κόμβο l ($\frac{w_i}{\sum_{k \in U \cap F_l} w_k}$) και αποδίδουμε σε κάθε μία από αυτές παροχή ίση με το γινόμενο της διαθέσιμης χωρητικότητας με το αντίστοιχο, κανονικοποιημένο βάρος($w_i / \sum_{k \in U \cap F_l} w_k \cdot A_l$). Άν μία ροή δεν μπορεί να απορροφήσει αυτή την εξυπηρέτηση επειδή είναι συμφορημένη στην πηγή, τότε υπολογίζουμε ώς μικρότερο utility σε αυτο το κόμβο τό utility της πρώτης συμφορημένης ροής. Διαφορετικά η παραπάνω διαδικασία αποδίδει σε όλες τις μή-συμφορημένες στην πηγή ροές ίσα utility.

3. Find all flows with minimum allocated temporary utility and label them *bottlenecked.*
   if there exists a link p and a flow g such that
   $f_g(t_{g,p}) = min_{l,i \in U}(f_i(t_{i,l}))$ then
   . $\quad B = B \cap \{g\}$
   . $\quad U = \overline{F \cap B}$
   . $\quad r_g = t_{g,p}$
   endif

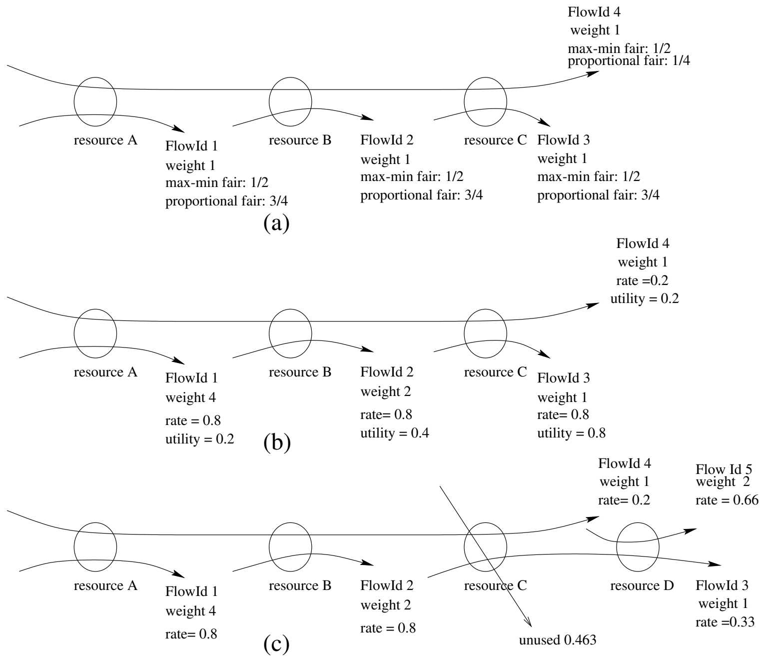
4. If $U = 0$ return
   else goto 1

Ο παραπάνω αλγόριθμός σωστά υπολογίζει τα δίκαια ποσοστά των ροών[UtMaxMin], τερματίζοντας ύστερα από αριθμό επαναλήψεων ίσο με τον αριθμό των κόμβων του δικτύου στη χειρότερη περίπτωση[5]. Μάλιστα, μπορεί να αποδειχθεί(κοίτα για αναφορές το [UtMaxMin]), ότι ο αλγόριθμος τερματίζει ύστερα από το πολύ B επαναλήψεις, όπου B ο αριθμός των εξαντλημένων κόμβων.

2.1.4 Παραδείγματα

Στο Σχ. 2.1 παρουσιάζουμε, τέσσερα απλά παραδείγματα από δίκαιες κατανομές. Βασικά έχουμε τέσσερις ροές$(1, 2, 3, 4)$ και τρεις κόμβους(A, B, C). Όλες οι ροές έχουν απεριόριστη ζήτηση. Στο πρώτο παράδειγμα όλες οι ροές έχουν ίσα βάρη και ίδια είναι και η συμφόρηση που αντιμετωπίζουν, οπότε ίσα είναι και τα ποσοστά που δικαιούνται. Στο δεύτερο η ροή 1 έχει το μεγαλύτερο βάρος και δικαιούται το μικρότερο utility(=0.2) στον κόμβο που χρησιμοποιεί, αναφορικά με όλες τις άλλες ροές και κόμβους – το ίδιο utility(=0.2) δικαιούται και η ροή 4 που χρησιμοποιεί τον ίδιο κόμβο. Πρέπει σίγουρα να αποδώσουμε σε αυτή τη ροή τη δίκαιη παροχή της στον κόμβο A πριν συνεχίσουμε οπότε το κάνουμε πρώτο. Αυτό σημαίνει ότι δίνουμε στην ροή 1 ποσοστό 0.8 και στην ροή 4 ποσοστό 0.2. Συνεχίζοντας με αυτό τον τρόπο βρίσκουμε τα ποσοστά και των υπολοίπων ροών. Στο τρίτο παράδειγμα, παρουσιάζουμε την περίπτωση όπου η δίκαιη κατανομή

[5] Αυτό ισχύει τουλάχιστον στην περιπτωση πού καμία ροή δεν είναι κυριάρχα-συμφορημένη στην πηγή.



Σχήμα 2.1: Παραδείγματα δίκαιων κατανομών.

μεγίστου-σταθμισμένου-ελαχίστου δεν αξιοποιεί πλήρως τη διαθέσιμη χωρητικότητα αν και οι ροές έχουν ανεξάντλητη ζήτηση οπότε αυτό είναι γενικά δυνατό – μία κατανομή που θα πραγμάτωνε το στόχο μεγίστης απασχόλησης θα το πετύχαινε. Τόσο η ροή 4 όσο και η ροή 5 που χρησιμοποιούν το κόμβο C, δεν τα καταφέρνουν να απορροφήσουν όλη τη διαθέσιμη χωρητικότητα αυτού του κόμβου, καθώς και οι δύο αντιμετωπίζουν πιο μεγάλη συμφόρηση στους κόμβους A και D αντίστοιχα. Με το να υπαγορεύουμε σε αυτές τις ροές να είναι δίκαιες με τους γείτονες τους σε αυτούς του κόμβους περιορίζουμε τη μέγιστη παροχή που μπορούν να χρησιμοποιήσουν, με αποτέλεσμα να υπο-χρησιμοποιείται ο κόμβος C. Αν και υπό αυτήν την έννοια, ο αντικειμενικός στόχος που θέσαμε είναι μη αποδοτικός, είναι δίκαιος γιατί διαφορετικά – αν θέλαμε δηλαδή να εκμεταλλευτούμε όλη τη χωρητικότητα του δικτύου – θα δίναμε είτε στην ροή 1, είτε στην ροή 5, μικρότερο utility από αυτό που δικαιούνται.

2.2 Υποθέσεις πάνω στην φύση των βαρών/δεικτών-σχετικής-προτεραιότητας.

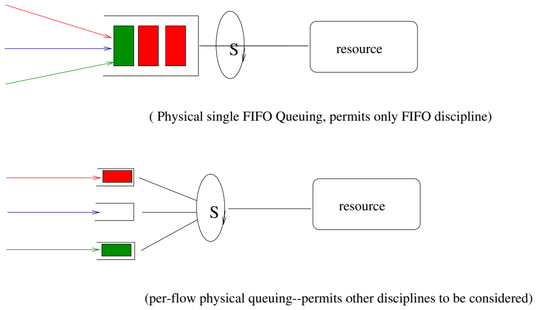
Σε αυτήν τη παράγραφο θέλουμε να τονίσουμε απλώς ότι στην ανάλυση που θα ακολουθήσει, δεν πρόκειται να κάνουμε επιπρόσθετες υποθέσεις πάνω στο σχήμα/τεχνική που αποδίδει βάρη στις ροές, πέραν από το ότι το βάρος εκφράζει/ποσοτικοποιεί τη σχετική προτεραιότητα της κάθε ροής – το βαθμό ανάγκης της για υπηρεσία. Υποψήφια σχήματα αντιστοίχισης βαρών, είτε έχουν τις βάσεις τους σε απόπειρες διαπραγμάτευσης ελάχιστης παροχής ή/και μέγιστης καθυστέρησης – συνήθως/λογικά απαιτούν προσύμφωνα και εγκατάσταση(admission control) –, είτε σε αγορά ποσοστού σταθερής παροχής απο γραμμές στα πλαίσια ενός οργανισμού/θεσμού, είτε προσδιορίζουν ιδιότητες κίνησης στα όρια μέρους του δικτύου, για λόγους λειτουργικότητας και απλότητας. Πάνω στο τελευταίο για παράδειγμα, μπορεί κάποιος να σταθμίσει/ζυγίσει την προτεραιότητα των διαφορετικών γραμμών εισόδου σε ένα μεταγωγέα, με βάση τον αριθμό των ενεργών υπο-ροών που την χρησιμοποιούν· τότε εξυπηρετώντας κάθε γραμμή ανάλογα με το βάρος της, προσπαθεί ουσιαστικά να αποδώσει ίση παροχή σε κάθε υπο-ροή.

Σε περίπτωση πάντως που δεν αναφέρεται ρητά, δεν πρόκειται να κάνουμε υποθέσεις πάνω στην ¨φύση¨ των βαρών και θα μελετήσουμε το γενικό σχήμα ενιαία, θεωρώντας ότι το βάρος εκπέμπει το σχετικό βαθμό προτεραιότητας/βιασύνης, όπως άλλωστε γίνεται και από τα περισσότερα σχήματα· απλώς ορισμένες φορές, αυτά προυποθέτουν επιπλέον ιδιότητες για την εισερχόμενη κίνηση και αναλόγως επιδιώκουν να ικανοποιήσουν αυστηρές απαιτήσεις εξυπηρέτησης.

Κεφάλαιο 3

Υπόβαθρο εργασίας

Σε αυτό το κεφάλαιο περιγράφουμε εν συντομία προηγούμενα αποτελέσματα εργασιών σχετικών με το πρόβλημα του χρονοπρογραμματισμού, σχετικά επομένως και με το δικό μας πρόβλημα. Ορισμένα από τα αποτελεσματα και τις τεχνικες που θα παρουσιάσουμε, τα έχουμε χρησιμοποιήσει για να λύσουμε επιμέρους δικά μας προβλήματα και για να κατασκευάσουμε την προτεινόμενη λύση – ή απλά μας απώθησαν απο συγκεκριμένους δρόμους επιλογής –, οπότε είναι ιδιαίτερα σημαντικά. Η έρευνα στον χρονοπρογραμματισμό που μας ενδιαφέρει εδώ, μπορεί να χωριστεί σε δύο (ανεξάρτητους;)κλάδους/προβλήματα. Το χρονοπρογραμματισμό ενός μόνο πόρου και την δυναμική κατασκευή προγραμμάτων προγραμματισμού crossbar – την εύρεση δηλαδή αυτών των ζευγών εισόδου εξόδου, που θα προωθούνται κάθε χρονική στιγμη μέσα από αυτό. Το δεύτερο πρόβλημα αποτελεί ουσιαστικά παράδειγμα(instance), γενίκευσης του πρώτου για πολλούς πόρους, χωρίς όμως να ανάγεται σε πολλά υποπροβλήματα του πρώτου τύπου, λόγω των εξαρτήσεων στις επιμέρους αποφάσεις[1].



Σχήμα 3.1: Μία μόνο $FIFO$ περιορίζει την δρομολόγηση σε $FIFO(FCFS)$ και δεν εξασφαλίζει προστασία στις ροές. Per flow queuing, δηλαδή η χρήση λογικών ουρών, διακριτών για κάθε ροή, επιτρέπει σε άλλες, πιο δίκαιες, τεχνικές να εφαρμοστούν.

3.1 Τεχνικές κατανομής ενός κοινού πόρου

3.1.1 Κατηγορίες τεχνικών

Οι αρχές χρονοπρογραμματισμού ενός πόρου κατανέμουν την πεπερασμένη χωρητικότητα αυτού, σε 'ἀνταγωνιζόμενες' αιτήσεις χρήσης του. Υπό αυτήν την έννοια διαφορετικές αρχές χρονοπρογραμματισμού διαφοροποιούνται στο να αποδίδουν διαφορετικές καθυστερήσεις σε διαφορετικές αιτήσεις διαφορετικών πηγών. Σε ένα πρωταρχικό επίπεδο μπορούμε να τις διακρίνουμε σε αρχές που δουλέυουν συνεχώς(work conserving) και σε αρχές που δεν δουλεύουν συνεχώς(non-work conserving). Οι work conserving τεχνικές, πάντα διανέμουν τον πόρο όταν υπάρχει ζήτηση, ενώ οι non work conserving όχι. Υπάρχει ο conservation ¨νόμος¨ που μας λέει, ότι ανεξάρτητα από το ποια work conserving τεχνική ακολουθούμε η μέση καθυστέρηση των αιτήσεων παραμένει σταθερή, αφού αν μειώσουμε την καθυστέρηση μίας αίτησης αναγκαστικά θα αυξήσουμε την καθυστέρηση μιας άλλης, οπότε η μέση καθυστέρηση παραμένει σταθερή.

Τώρα οι ¨εργατικές¨ τεχνικές, μπορούν να διακριθούν σε αυτές που δεν διαφοροποιούν τις ροές – εδώ εννοούμε τη διαφοροποίηση ανεξάρτητα της εξυπηρέτησης του παρελθόντος, αλλά διαφοροποίηση πάνω στη φύση των ροών – και σε αυτές που τις διαφοροποιούν. Γνωστές αρχές στην πρώτη κατηγορία είναι η FIFO/FCFS, Round Robin, Longest Queue First, Longest Waiting Time First,

[1]Υπό αυτή την οπτική, η λύση που θα προτείνουμε αργότερα πετυχαίνει ακριβώς αυτό· καθιστά δυνατή την αναγωγή του πολυδιάστατου προβλήματατος, σε πολλά - ανεξάρτητα - μονοδιάστατα.

Least Recently Used κτλ. Η απλή FIFO/FCFS τεχνική δεν προσφέρει προστασία από ροές που στέλνουν συνεχώς(Σχ. 3.1). Οι υπόλοιπες τεχνικές που αναφέραμε εδώ, προσπαθούν να βελτιώσουν αυτήν την κατάσταση και να προσφέρουν προστασία/απομόνωση – ουσιαστικά σε κατασκευαστικό επίπεδο, την απομόνωση την προσφέρει το per flow queuing – στις ροές, δημιουργώντας/κατασκευάζοντας προτεραιότητες που έχουν πεπερασμένη ισχύ/διάρκεια. Έτσι για παράδειγμα η τεχνική RR αφού εξυπηρετήσει μία ροή δίνει μεγαλύτερη προτεραιότητα σε όλες τις άλλες, προσπαθώντας να παράγει δίκαιη κατανομή.

Όπως έχουμε αναφέρει και πιο πριν, είναι απλουστευτικό/απλοικό και μη αποδοτικό να θεωρούμε ότι όλες οι ροές έχουν την ίδια *ανάγκη* για υπηρεσία. Τεχνικές που διαφοροποιούν πάνω στη φύση των ροών μπορούν να χωριστούν με τη σειρά τους σε δύο κατηγορίες. Σε αυτές που χρησιμοποιούν απόλυτες προτεραιότητες – οπότε έχει πακέτο η ροή q θα την εξυπηρετώ – και σε αυτές που χρησιμοποιούν σχετικές προτεραιότητες. Χρησιμοποιώντας απόλυτες προτεραιότητες μπορεί να στερήσουμε την εξυπηρέτηση από μία ροή χαμηλής προτεραιότητας για οσοδήποτε μεγάλο διάστημα(starvation) και για αυτό το λόγο πρέπει να τις αποφεύγουμε και να τις υιοθετούμε μόνο σε ειδικές/κρίσιμες περιπτώσεις.

3.1.2 Τεχνικές ιδεατού χρόνου

Οι σχετικές προτεραιότητες εκφράζονται συνήθως μέσω βαρών. Τα πλέον γνωστά και αποδεχτά σχήματα σε αυτή τη κατηγορία σήμερα, βασίζονται στο ιδεατό Generalized Processor Sharing μοντέλο. Αυτό το μοντέλο είναι αναπτυγμένο πάνω σε ένα κόσμο ρευστών, όπου κάθε ενεργή ροή – ένα διαφορετικό υγρό/ρευστό σε αυτό το μοντέλο – εξυπηρετείται <u>συνεχώς</u> σε ποσοστό ανάλογο του βάρους της. Η μονάδα λοιπόν εξυπηρέτησης σε αυτό το μοντέλο, όντας απεριόριστα μικρή, επιτρέπει σε όλες τις ενεργές ροές να εξυπηρετούνται ταυτόχρονα και σε ποσοστό ανάλογο του βάρους τους. Αυτό το σχήμα έχει ωραίες/χρήσιμες ιδιότητες, επιθυμητές και στα συστήματα πακέτων, όπου η μονάδα εξυπηρέτησης είναι το πακέτο – ένα το πολύ πακέτο εξυπηρετείται κάθε χρονική στιγμή.

- κάθε ροή f με βάρος w_f, έχει σίγουρο ένα ελάχιστο ρυθμό εξυπηρέτησης $r_f^{min} = \frac{w_f}{\sum_{g \in F} w_g}$

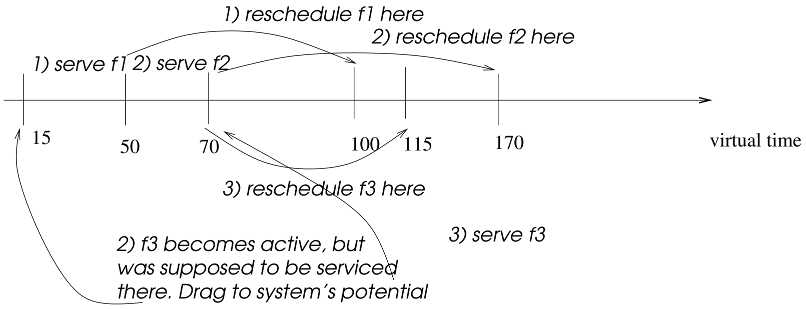
- κάθε ροή παίρνει πραγματικά, σε κάθε χρονικό διαστημα με το σύνολο των ενεργών ροών G σταθερό, ρυθμό $r_f = \frac{w_f}{\sum_{g \in B} w_g} \geq r_f^{min}$.
- συνεπώς η κατανομή που παράγεται είναι <u>απολύτως</u> δίκαιη, υπό τους όρους του μεγίστου-σταθμισμένου-ελαχίστου, αφού όλες οι ενεργές ροές παίρνουν ίση προσθετική/στιγμιαία utility.
- Αν μία ροή είναι leacky-bucket constrained μπορεί της αποδοθεί πάνω όριο, μεγίστης καθυστέρησης[GalPGP].

Τα συστήματα πακέτων που επιδιώκουν παρόμοιες ιδιότητες, προσπαθούν να εξυπηρετούν τα πακέτα των ροών με τη σειρά που θα αναχωρούσαν αυτά στο ιδεατό μοντέλο των ρευστών. Κυρίως χρησιμοποιούν την τεχνική του ιδεατού χρόνου, μία μεταβλητή που σε κάθε χρονική στιγμή προχωρεί με ρυθμό ίσο με το ρυθμό που οι ενεργές ροές λαμβάνουν προσθετική utility στο ιδεατό/ρευστό μοντέλο. Για κάθε ροή, μία ξεχωριστή μεταβλητή διατηρείται που συνήθως ονομάζεται virtual_finish_time· αυτή η μεταβλητή σημειώνει το χρόνο στο ιδεατό σύστημα, που το επόμενο πακέτο αυτής της ροής θα εξυπηρετηθεί πλήρως. Ο χρονοπρογραμματιστής δρομολογεί τις ροές/πακέτα με αύξουσα σειρά virtual_finish_time. Ο ιδεατός χρόνος του συστήματος(virtual time), χρησιμοποιείται για να επανεντάξει δίκαια, μία μη ενεργή ροή στο συναγωνισμό με τις άλλες· αν μία ροή που μόλις έγινε ενεργή ξανά, έχει virtual_finish_time πίσω από τον ιδεατό χρόνο του συστήματος, τότε ο χρονοπρογραμματιστής της αναθέτει ως νέο, κάτι κοντά – εξαρτάται από την υλοποίηση – στο virtual time. Αν την άφηνε εκεί πίσω, τότε αυτή η ροή πιθανότατα θα έπαιρνε ένα burst/ομοστοιχία από εξυπηρετήσεις, πράγμα τό οποίο δεν είναι δίκαιο και ασφαλές για τις υπόλοιπες ροές[2]. Για λεπτομέρειες πάνω στην τεχνική του ιδεατού χρόνου(virtual time) κοιτάξτε στο [BufCrossbar][ParGalPGPS][DemKesShenWFQ].

Διατηρώντας λεπτομερείς/ακριβείς πληροφορίες για το μοντέλο των ρευστών – ουσιαστικά για το σύνολο των ενεργών ροών στο ιδεατό μοντέλο, πράγμα που απαιτεί πράξεις σε διάφορες χρονικές στιγμές και όχι μόνο όταν είναι να παρθεί η επόμενη απόφαση δρομολόγησης –, μπορεί κανείς να αποδώσει σε κάθε ροή εξυπηρέτηση που δεν είναι μικρότερη από αυτή στο ιδεατό μοντέλο για περισσότερο από ένα πακέτο[ParGalPGPS]. Τέτοιες τεχνικές είναι η $PGPS$ και η

[2]Αυτό κάνει πράγματι, η τεχνική ιδεατού ρολογιού/VirtualClock [Virtualclock].

WFQ. Η πολυπλοκότητα μπορεί να μειωθεί σημαντικά, εξάγοντας πληροφορίες για το ιδεατό μοντέλο από το σύστημα πακέτων, όπως κάνει η τεχνική WRR που περιγράφουμε στην επόμενη παράγραφο.



Σχήμα 3.2: One execution example of the WRR ([KaSC91]) scheduling discipline, implemented through the virtual clock method. Flows 1,2 and 3 have service_intervals 50,45 and 100 respectively. Events 1), 2), 3) correspond to continuous scheduling cycles. At time 1 the active flow with the smallest next_service_time is flow 1(virtual time 50), which is served and re-scheduled at virtual time 100. At time 2 the flow with the smallest virtual time is flow 2(VT 70), which is served and re-programmed for virtual time 170. At the end of this cycle flow 3 – which was inactive – becomes active again and is "dragged" to the current virtual time(70), since its previous next_service_time was left behind(15). It is served at step 3 and reprogrammed according to its service interval.

Η τεχνική WRR είναι μία hardware friendly, απλοποιημένη εκδοχή συστήματος που εξομοιώνει το ιδανικό μοντέλο ρευστών. Για κάθε ροή η μεταβλητή next_service_time σημειώνει την επόμενη εξυπηρέτηση αυτής της ροής σε ένα ιδεατό άξονα χρόνου(Σχ. 3.2). Κάθε φορά που πραγματικά εξυπηρετείται μία ροή, προστίθεται στο παλιό της next_service_time το service_interval αυτής της ροής, μία μεταβλητή αντιστρόφως ανάλογη του βάρους της. Τέλος αν το next_service_time μίας ροής που μόλις τώρα έγινε ενεργή, είναι πίσω από το next_service_time όλων των ενεργών ροών, τότε γίνεται ίσο με το μικρότερο από αυτά. Μία παρόμοια τεχνική, προτάθηκε από τον Κατεβαίνη, με διαφορετική υλοποίηση στο [KaSC91].

Τέλος αξίζει να αναφέρουμε την τεχνική WF^2Q που είναι αποδεδειγμένα βέλτιστη στο να προσεγγίζει το ιδεατό μοντέλο. Αν και κάτω από την αρχή χρονοπρογραμματισμού $PGPS$, η εξυπηρέτηση που έχει δεχτεί μία ροή, δεν υστερεί ποτέ περισσότερο από ένα πακέτο από το ιδανικό μπορεί να είναι πολύ πιο μπροστά από αυτό, πράγμα που υπονοεί bursty/όχι-ομοιόμορφες χρονό-κατανομές. Η βελτιωμένη αρχή WF^2Q το λύνει αυτό εξυπηρετώντας μόνο πακέτα που έχουν ξεκινήσει να δέχονται εξυπηρέτηση στο ιδεατό μοντέλο. Επειδή στο σύστημα πακέτων κάθε ροή εξυπηρετείται με μέγιστο ρυθμό, την παροχή της γραμμής, ενώ στο μοντέλο των υγρών με ένα μέρος αυτής που εξαρτάται από το βάρος της ροής και το βάρος όλων των άλλων ενεργών ροών, ένα πακέτο μπορεί να είναι υποψήφιο για προγραμματισμό στο σύστημα πακέτων ενώ στο ιδεατό σύστημα ακόμα εξυπηρετείται το προηγούμενο πακέτο αυτής της ίδιας ροής. Φυσικά η ομοιόμορφη κατανομή που προσφέρει αυτή η τεχνική κοστίζει προσθετικό κόστος πολυπλοκότητας, το οποίο όμως μειώνει αισθητά η παραλλαγή WF^2Q+. Παρακάτω αναφέρουμε συνολικά το βαθμό που προσεγγίζουν οι παραπάνω τεχνικές το ιδεατό μοντέλο:

$$S_f^{GPS}(0,t) - O(1) \leq S_f^{WF^2Q}(0,t) \leq S_f^{GPS}(0,t) + O(1)$$

$$S_f^{GPS}(0,t) - O(1) \leq S_f^{PGPS}(0,t) \leq S_f^{GPS}(0,t) + O(N)$$

$$S_f^{GPS}(0,t) - O(N) \leq S_f^{WRR}(0,t) \leq S_f^{GPS}(0,t) + O(N)$$

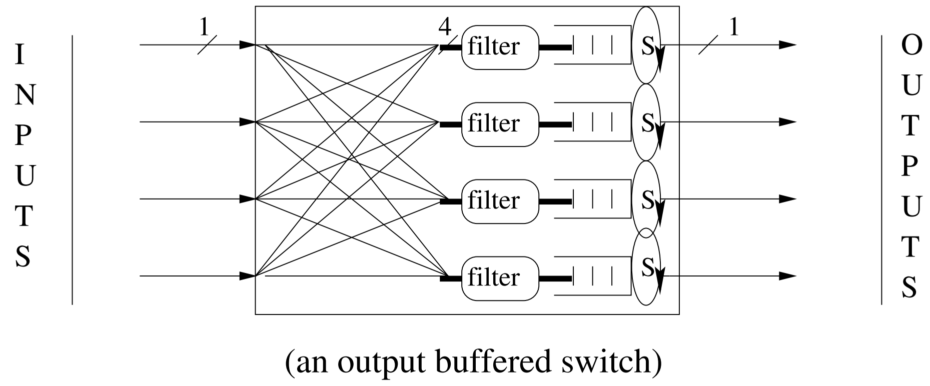
$$S_f^{GPS}(0,t) - O(N) \leq S_f^{Vclock}(0,t) \leq S_f^{GPS}(0,t) + 1/0 \text{ !!}$$

$$S_f^{GPS}(0,t) - O(1) \leq S_f^{RR}(0,t) \leq S_f^{GPS}(0,t) + O(1)$$

$$S_f^{GPS}(0,t) - O(1) \leq W_f^{WF^2Q+}(0,t) \leq S_f^{GPS}(0,t) + O(1) \qquad .$$

3.1.3 Κόστος/πολυπλοκότητα και εφαρμογές.

Οι παραπάνω – δίκαιες – τεχνικές απαιτούνε per flow queuing. Υπάρχουν καλοί/αποδοτικοί και γρήγοροι τρόποι για να υλοποιήσουμε πολλές λογικές ουρές σε μία φυσική μνήμη [Kat534]. Οι περισσότερες απαιτούνε επιπλέον και μία δομή/μηχανή που να βρίσκει τον μικρότερο αριθμό από ένα σύνολο υποψηφίων. Αυτό μπορεί να λυθεί είτε με μία κατάλληλη δομή δεδομένων υλοποιημένη

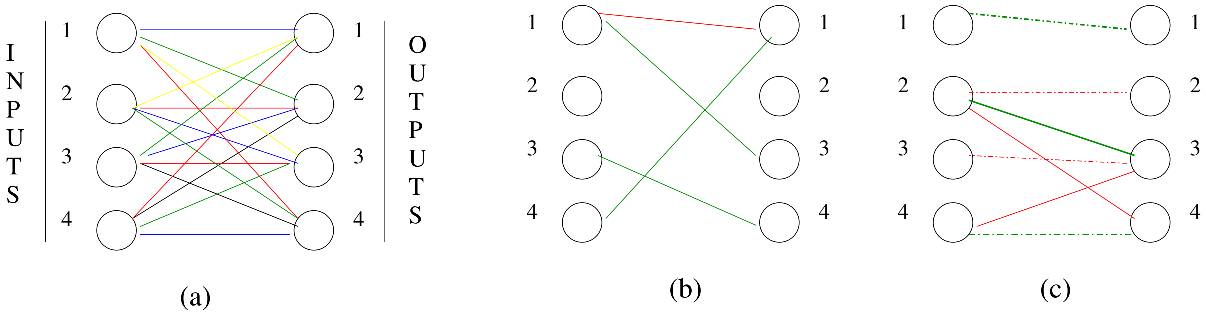


Σχήμα 3.3: The architecture of an output, buffered switch with one resource-servers at the outputs. Although a single buffer is shown, per input or per flow distinct logical queues can be used, at each output. These buffers will need operating throughput $N + 1$ times the throughput of the external links. The filter – that again needs to provide throughput N – forwards to the output buffer only these packets that are destinated for this output.

σε υλικό([IoannouKatHeap]), είτε με παράλληλα/γρήγορα κυκλώματα που δυναμικά(on demand) βρίσκουν τον μικρότερο αριθμό από ένα συνολο υποψηφίων, χωρίς να απαιτούν/διατηρούν κάποια κατάλληλη δομική/μνημονική οργάνωση για το σύνολο αυτό([HartKatBinTree]). Η δεύτερη μέθοδος είναι εξαιρετικά χρήσιμη όταν το σύνολο των υποψηφίων ροών αλλάζει σχετικά γρήγορα.

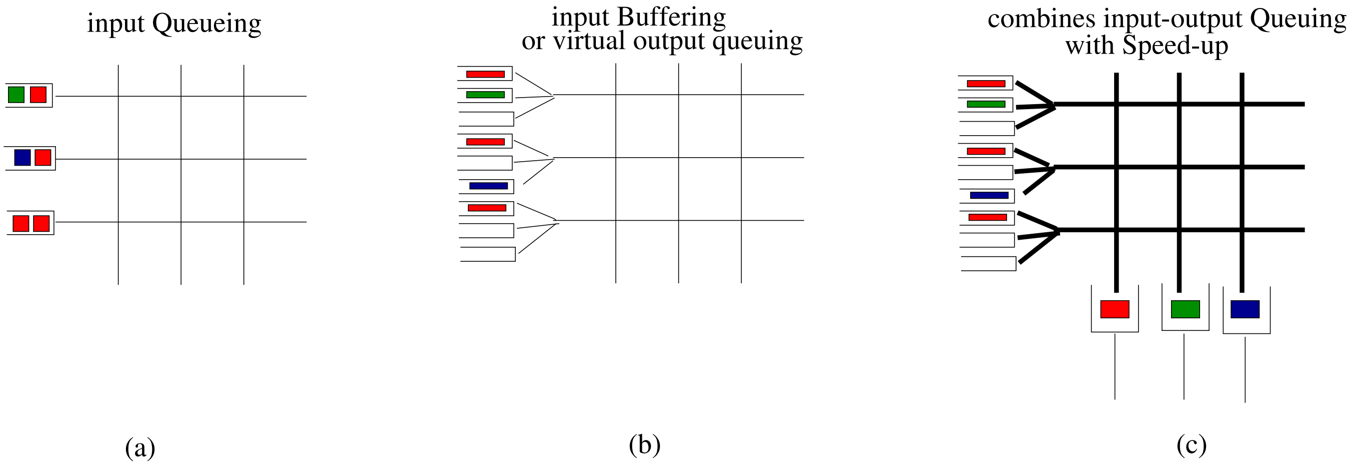
Κλασική εφαρμογή των παραπάνω τεχνικών είναι στις εξόδους ενός μεταγωγέα με τις ουρές αποθήκευσης στις εξόδους(Σχ. 3.3). Αφού οι αποφάσεις των χρονοπρογραμματιστών στις εξόδους είναι ανεξάρτητες, η χρήση WFQ-like τεχνικών ορίζει το ιδανικό σύστημα μεταγωγής/πολύπλεξης πακέτων. Φυσικά αυτό το σύστημα είναι πολύ ακριβό και μάλιστα εφικτό μόνο για ένα περιορισμένο φάσμα από κλίμακες, αφού πρέπει να διαθέτει μνήμες σε κάθε έξοδο με παροχή ανάλογη του αριθμού των εισόδων του συστήματος.

3.2 Το πρόβλημα χρονοπρογραμματισμού ενός crossbar



Σχήμα 3.4: The crossbar scheduling problem defined in graph theory terms. Bipartite is a graph, with edges only from one set of nodes to an other and with no edges among nodes of the same set. Bipartite match is a set of edges in a bipartite graph, where no node participates more than one time. In (a), we can see different matches, assuming all edges are active with different colors. In (b) not all edges are active. The green match has maximum size (3), while any other match that would include the red color would have less size. If a maximum match discipline is followed, then the red colored edge(flow) will not be selected and will be starved. In (c) we see with green think lines, a maximal match under construction. The other green edge, can be included in the maximal match, since no other edge in the match so far needs to change. The red lines cannot be included. The dash-dot lines represent the maximum match, that has size 4, while the maximal match that is being constructed in the figure with the green lines, can only have size 3 – since it is dictated by actions-choices taken previously in the construction of the maximal match.

Είδαμε ότι τα συστήματα με ουρές αποθήκευσης στις εξόδους, άν και ιδανικά, είναι πολύ ακριβά λόγω των απαιτήσεων τους στο μονοπάτι της μνήμης(memory datapath requirements). Αντιθέτως, κάτω από την ίδια οπτική, τα συστήματα με τις ουρές αποθήκευσης στις εισόδους αποτελούν την βέλτιστη λύση· οι μνήμες στις εισόδους χρειάζεται να έχουν ταχύτητα ακριβώς διπλάσια – σταθερή – από αυτή των εξωτερικών γραμμών. Το βασικό πρόβλημα σε αυτά τα συστήματα, βρίσκεται στην επιλογή των πακέτων που θα προωθηθούν κάθε χρονική στιγμή από τις εισόδους στις εξόδους —- μόνο μία είσοδος μπορεί να απασχολήσει μία έξοδο κάθε χρονική στιγμή, διαφορετικά θα χρειαζόμασταν μνήμες και στις εξόδους – για να αποφύγουμε τη σύγκρουση – καί μόνο ένα πακέτο μπορεί να φύγει από μία εισόδο, διαφορετικά, για τον ίδιο λόγο συμμετρικά ορισμένο, θα χρειαζόμασταν εσωτερική επιτάχυνση. Αυτό που προκύπτει είναι ένα κεντρικό(ποιημένο) πρόβλημα ταιριάσματος που πρέπει να λύσει παράλληλα τις απαιτήσεις για ικανοποιητική απασχόληση των εξωτερικών γραμμών, για δικαιοσύνη και τελευταία για ποιότητα υπηρεσίας ή διαφορετικά για αναλογικό χρονοπρογραμματισμό.



Σχήμα 3.5: The alternatives solution for buffers organization on the crossbar topology. We have omitted one last alternative, to be presented under our proposed scheme.

Χρησιμοποιώντας μία μόνο $FIFO$ σε κάθε είσοδο για την αποθήκευση όλων των πακέτων(input queuing) ανεξάρτητα του προορισμού τους, περιορίζονται οι επιλογές του επίδοξου χρονοπρογραμμιστή και προκύπτει το φαινόμενο Head Of Line blocking που μειώνει την μέγιστη δυνατή απασχόληση σε χαμηλά επίπεδα. Η χρήση διαφορετικών ουρών ανά έξοδο σε κάθε είσοδο(virtual output queuing or advanced input queuing) είναι επιτακτική για κάθε αξιόλογο σύστημα, έτσι ώστε να αποκτήσει ο χρονοπρογραμματιστής, που έτσι και αλλιώς έχει δύσκολο έργο, περισσότερες επιλογές. Το πρόβλημα ταιριάσματος εισόδων και εξόδων, θα ήταν λυμμένο αν το ταίριασμα μέγιστου μεγέθους ήταν οικονομικότερο, πιο δίκαιο, αλλά είχε και παραλλαγή που θα ικανοποιούσε και τις απαιτήσεις για αναλογικό χρονοπρογραμματισμό. Ιδιαίτερο ενδιαφέρον έχουν οι προτάσεις για επαναληπτικό-κατανεμημένο-παράλληλο ταίριασμα, που παράγουν maximal matching. Αυτά ορίζονται ως τα ταιριάσματα που δεν μπορούν να μεγαλώσουν σε μέγεθος, χωρίς να αλλάξουν συνδέσεις που έχουν ήδη καταχωρηθεί. Η καλύτερη λύση σε αυτήν την κατηγορία, ανήκει στην κλάση του i-slip - περιγράφεται και στο [SLIPMcKeownAndersComp] - και η οποία βελτιώνει ουσιαστικά, την πρώτη πρόταση για αλγορίθμους ταιριάσματος αυτής της μορφής που προτάθηκε από τον Anderson στο [AnderPIM].

ITERATE

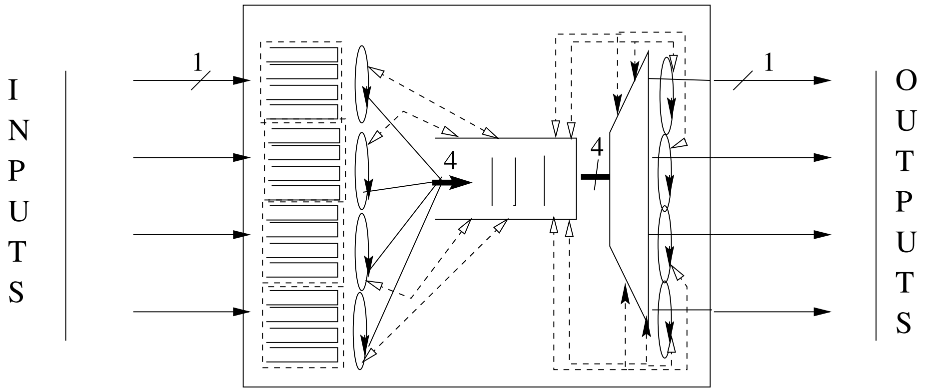
1. *All inputs send request to all outputs for which, they queued cells(input buffering systems have this knowledge)*
2. *All outputs gather the requests sent to them during this iteration, and select one the using round robin, implemented with a next to serve pointer and any, constant enumeration of inputs. The next to serve pointer to, is updated to the next(module arithmetic) input, only if the grant is accepted by the respective input in the next step of the current iteration.*
3. *All inputs that receive at least one grant, accept one using in round robin fashion . Each accept, results into a new connection being added into the crossbar configuration.*

Στο [AnderPIM] αποδεικνύεται ότι οι αλγόριθμοι αυτού του τύπου τερματίζουν ύστερα από log(N) επαναλήψεις κατά μέσο όρο, ενώ υπάρχει και ένα αποτέλεσμα ότι τα maximal match ταιριάσματα έχουν μέγεθος τουλάχιστον το μισό από το μέγιστο δυνατό(για αναφορές κοιτά [Linear]). Το ξεχωριστό στο i-slip είναι ότι τείνει("slips") να παράγει ταιριάσματα μεγάλου μεγέθους(αποδοτικότητα) και διαφορετικά μεταξύ τους(δικαιοσύνη) ακόμα και με μία επανάληψη. Το πρόβλημα είναι ότι για να γενικεύσουμε την τεχνική εισάγοντας προτεραιότητες θα χρειαζόμαστε περισσότερες επαναλήψεις, αφού γενικεύοντας τους Round Robin σε Weighted Round Robin χάνεται το "slip" χαρακτηριστικό(βλ. [BufCrossbar] κεφάλαιο background-crossbar scheduling και κεφάλαιοSimulation Results — Scheduling a Bufferless Crossbar — The difficult Problem).

Άλλες προτάσεις για την εισαγωγή προτεραιοτήτων στο ίδιο σχήμα μπορούν να βρεθούν στα [StilVarmWPIM][FIM][AnderPIM], χωρίς όμως καμία να είναι ουσιαστικά συνολικά αρκετά καλή(κόστος-απόδοση). Μία διαφορετική γενίκευση έχει προταθεί στο [Linear], όπου τα ταιριάσματα είναι πλέον Stable Marriage Matching αντί για maximal. Το σχήμα αν και καλό, δείχνει ακριβό και δεν έχει πείσει ότι μπορεί να διατηρήσει μεγάλη απασχόληση – στο ίδιο paper μάλιστα, προτείνουνε δεύτερους/συμπληρωματικούς αλγορίθμους για να καλύψουν το κενό. Τέλος ενδιαφέρον παρουσιάζουν και οι τεχνικές FIRM, που είναι βελτίωση του i-slip και η Two Dimension Round Robin· αυτές έχουν προταθεί από τον Σερπάνο στα [SerpFIRM] και [Serp2DRR] αντίστοιχα, αλλά δεν προσφέρουν διαφοροποίηση μεταξύ των ροών. Περισσότερες πληροφορίες για αυτές τις τεχνικές αλλά και για άλλες που χρησιμοποιούν εσωτερικό speed up και/ή μικρούς εσωτερικούς buffers μπορείτε να βρείτε στο [BufCrossbar].

Κεφάλαιο 4

Σύστημα



(an input buffered switch, with small internal buffering and distributed scheduling)

Σχήμα 4.1: An input-buffering, internal buffering switch with distributed scheduling. With dashed lines we see the information that must be exchanges between scheduling servers and the internal buffer space "territory", in the view of flow control and scheduling. Each output server, must know which cells in the internal buffer are destinated for the respective output, in order to take a scheduling decision. Each input must now if there exists buffer space in the internal buffer, adequate to receive its candidate packet.

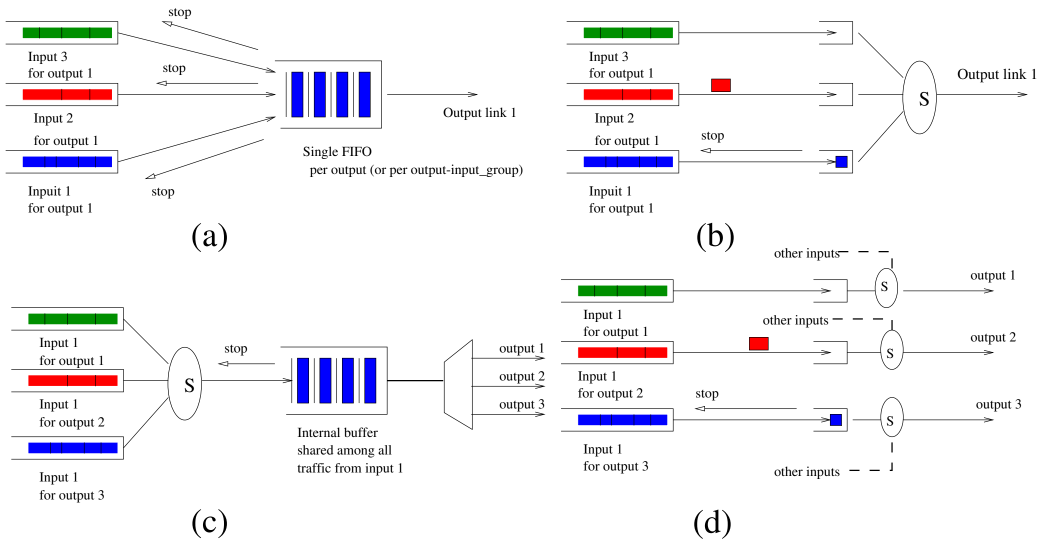
4.1 Γενική Περιγραφή Συστήματος-Λύση προβλήματος εσωτερικής δρομολόγησης.

Αφού τα συστήματα με τις ουρές αποθήκευσης στις εισόδους αποτελούν την βάση για την κατασκευή μεγάλων μεταγωγέων, σίγουρα πρέπει να βασιστούμε σε αυτή τη κλάση συστημάτων για να προτείνουμε μία καλή, συμβατή/εφικτή λύση. Είδαμε όμως ότι τότε εισάγεται η ανάγκη για ένα κεντρικό ¨διαιτητή¨, που συνολικά θα αποφασίζει ποιες ροές θα εξυπηρετηθούν· διαφορετικά μπορεί να προκύψει σύγκρουση και χάσιμο πακέτων. Πέρα από το προσθετικό κόστος, δεν έχει βρεθεί ακόμα η κατάλληλη λύση που θα διαφοροποιεί πάνω στην προτεραιότητα των ροών, χωρίς να υπο-χρησιμοποιεί τη διαθέσιμη χωρητικότητα. Ουσιαστικά σε ένα σύστημα με ουρές αποθήκευσης στις εισόδους, υπάρχει ανταγωνισμός τόσο για τις εισόδους όσο και για τις εξόδους. Ακόμα και για το crossbar, που δεν εισάγει άλλα σημεία ανταγωνισμού, η αλληλεξάρτηση των αιτήσεων πάνω στην τοπολογία, δεν επιτρέπει την χρήση ανεξάρτητων ¨διαιτητών¨ που θα λύνουν συνολικά το πρόβλημα. Αυτό γίνεται εφικτό αν τοποθετήσουμε μικρή εσωτερική μνήμη στο δίκτυο μεταγωγής· τα πακέτα που θα διαλέγουν διακριτοί χρονοπρογραμματιστες στις εισόδους θα αποθηκεύονται εκεί μέχρι να τα επιλέξει ο χρονοπρογραμματιστής της κατάλληλης εξόδου. Επιπλέον απαιτείται ένα σχήμα προστασίας της εσωτερικής μνήμης για να αποτρέπει το χάσιμο πακέτων σε αυτή – αυτή η μνήμη είναι μικρή, αφού δεν υπάρχει για αποθήκευση μεγάλης διάρκειας/ποσότητας, αλλά για να λύσει το πρόβλημα του scheduling.

Το σχήμα που έχουμε προτείνει μέχρι στιγμής, μοιάζει κάπως σαν το Σχ. 4.1 . Απομένουν τα προβλήματα οργάνωσης της εσωτερικής μνήμης, το σύστημα προστασίας αυτής(flow control και το τί χρονοπρογραμματιστές θα βάλουμε στα σημεία ανταγωνισμού.

4.1.1 Τοποθέτηση/διαμερισμός και το σχήμα διαχείρισης, της εσωτερικής μνήμης

Έστω ότι χρησιμοποιούμε κοινή εσωτερική μνήμη, διαθέσιμη πάντα για όλες τις ροές να την απασχολήσουν πλήρως αν θέλουν και μπορούν – αυτό άλλωστε



Σχήμα 4.2: In case (a) three inputs send to the same output, but these inputs share the internal buffer space. In case (b) a distinct per input buffer is used among inputs, that solves the unfairness seen in (a). In (c) all flows from one input, share the same buffer (as in (b)). In (d) a distinct per input-output pair buffer solves the HeadOfLine present at (c).

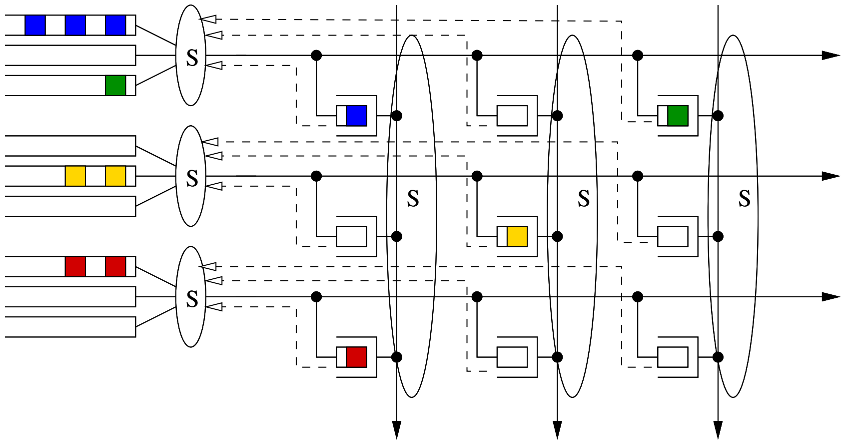
μας συστήνει και η εμπειρίας μας από συστήματα διαχείρισης μνήμης· ενιαί-α μνήμη ίσον μεγαλύτερη αξιοποίηση → μεγαλύτερη αποδοτικότητα. Τότε όμως προκύπτουν δύο προβλήματα για το σύστημα μας. Πρώτον οι χρονοπρογραμματιστές στις εισόδους δεν μπορούν να λειτουργούν ανεξάρτητα όπως θέλαμε εξαρχής, αφού αν έχει μείνει λίγος χώρος στην εσωτερική μνήμη, κάποιος θα πρέπει να αποφασίσει ποία είσοδος θα τον χρησιμοποιήσει – → κεντρικός έλεγχος. Δεύτερον υπάρχει πρόβλημα προστασίας των ροών αν χρησιμοποιήσουμε κοινή μνήμη ανά είσοδο ή ανά έξοδο(Σχ. 4.2 , αλλά και αποδοτικότητας – κάτι σαν HOL blocking. Ίσως κάτι τελευταίο είναι ότι θα χρειαζόμασταν αναγκαστικά ένα προγραμματιστή για το χρονοπρογραμματισμό των αιτήσεων στην κοινή εσωτερική μνήμη, πρόβλημα που δεν είναι καθόλου ευκαταφρόνητο. Τελικά καταλήγουμε να διαχωρίσουμε/τμηματίσουμε την εσωτερική μνήμη άνα είσοδο και άνα έξοδο, πράγμα που συνεπάγεται N^2 διακριτούς/ανεξάρτητους buffers στο crossbar – όσες και οι ροές(=ζευγάρια είσόδων
εξόδων).

Τώρα κάθε ροή θα είναι ΰπεύθυνη̋ για να μήν υπερχειλίζει την μνήμη που της αντιστοιχεί – μέσω του προγραμματιστή εισόδου που την διαχειρίζεται –, άρα και για την συμφόρηση που έχει προκαλέσει στο παρελθόν, άρα δεν θα μπορεί να επηρεάσει τις υπόλοιπες ροές. Αλλά και κάθε χρονοπρογραμματιστής θα μπορεί ανεξάρτητα, ¨βλέποντας¨ την κατάσταση μόνο των εσωτερικών buffer των ροών που του αντιστοιχούν, να κρίνει ποια ροή θα στείλει και σε αυτή του την επιλογή μπορεί να έχει ό,τι κριτήρια θέλει – χωρισμός/decoupling απαιτήσεων για εξεζητημένο χρονοπρογραμματισμό από την αίτηση για διατήρηση μεγάλης συνολικής παροχής.

Αυτό που χρειάζόμαστε είναι ένα καλό σχήμα διαχείρισης της εσωτερικής μνήμης, που να εξασφαλίζει ότι δεν θα χάνονται πακέτα. Το backpressure credit based flow control είναι εδώ κατάλληλο. Λέγεται selective, επιλεκτική, γιατί κάθε είσοδος ¨βλέπει διαφορετικά¨ την συμφόρηση σε κάθε έξοδο από τις άλλες είσόδους και όπως είδαμε αυτό λύνει πολλά υποψήφια προβλήματα δικαιοσύνης/προστασίας/αποτελεσματικότητας. Άρα η λειτουργία του συστήματος μέχρι στιγμής έχει ως εξής:

- *Κ*άθε είσοδος επιλέγει μία ροή, θεωρώντας ως ενεργές τις ροές που έχουν πακέτο στην είσοδο και για τις οποίες υπάρχει διαθέσιμος χώρος εσωτερικά.
- *Κ*άθε έξοδος, διαλέγει για προώθηση μία ροή που έχει πακέτο εσωτερικά.

Στην τοπολογία μας, όλα τα παραπάνω ταιριάζουν απόλυτα, αφου τους εσωτερικούς ενταμιευτές μπορούμε να τους τοποθετήσουμε στο ίδιο chip με τον επιλογέα διασταύρωσης(crosspoint), όπως δείχνει και το Σχ. 4.3 . Αν και πολλοί αυτοί οι ενταμιευτές, καθένας χρειάζεται να έχει παροχή διπλάσια μόνο από αυτή των εξωτερικών γραμμών, ενώ και η λογική που τους συντηρεί/διαχειρίζεται μπορεί να τοποθετηθεί στο ίδιο chip – με τους ρυθμούς που αυξάνεται η χωρητικότητα των ολοκληρωμένων κυκλωμάτων, οι δύο προσθήκες που κάναμε δεν είναι ουσιαστικά πρόβλημα/κόστος.



The input buffered crosspoint buffered crossbar architecture
with selective backpressure

Σχήμα 4.3: The input buffered crossbar with internal buffering & crosspoint buffering. Note that actually we did not have to make all the arguments in support of selective backpressure, since the crosspoint buffer naturally imply it.

4.1.2 Αρχιτεκτονική Δρομολόγησης — Επιλογή χρονοπρογραμματιστών

Μέχρι στιγμής έχουμε περιγράψει τον κορμό της αρχιτεκτονικής. Μας μένει να προσδιορίσουμε τον τύπο των διαιτητών που θα τοποθετήσουμε στα σημεία ανταγωνισμού. Αυτά είναι $2 \cdot N$ στον αριθμό – βασικά κάθε είσοδος και κάθε έξοδος. Φυσικά δεν μπορούμε να πετύχουμε ακριβώς τις επιδόσεις που θα πετύχαινε ένα σύστημα με τις ουρές αποθήκευσης στις εξόδους, αφού έχουμε διαφορετικά προβλήματα, με διαφορετικό πεδίο εφικτών λύσεων. Έχουμε όμως πει ότι θέλουμε αναλογικές κατανομές, να μπορούμε να διαφοροποιούμε τις ροές. Ας βάλουμε λοιπόν αναλογικούς χρονοπρογραμματιστές σε κάθε είσοδο και σε κάθε έξοδο και ό,τι ήθελε προκύψει[1]

Φυσικά τότε, κάθε ροή μπορεί γενικά να δικαιούται διαφορετικό ποσοστό στην είσοδο και διαφορετικό στην έξοδο. Σε μεσοπρόθεσμη διάρκεια όμως, περιμένου-

[1] Βασικά δεν είναι τόσο ¨ανοήτη¨ η επιλογή όσο την παρουσιάζουμε εδώ. Ένας μεταγωγέας δεν είναι παρά ένα μικρό δικτύο. Η Hahne είχε ήδη ασχοληθεί με το επιχείρημα ότι Round Robin Scheduling μαζί με backpressure flow control, μπορούν να υποστηρίξουν max-min fair allocations σε δίκτυα γενικής τοπολογίας. Η γενικεύση με την εισαγωγή βαρών, είναι σχετικά απλή.

με κάθε ροή να περιορίζεται στο μικρότερο από αυτά τα ποσοστά – λόγω back-pressure ή συμμετικά λόγω απουσίας πακέτων στους εσωτερικούς ενταμιευτές – καί αν οι δρομολογητές που χρησιμοποιούμε είναι work conserving, το ποσοστό που περισσεύει να διανέμεται στις υπολοιπές ροές, άν μπορούνε αυτές με τη σειρά τους να το απορροφήσουν. Αν δεν σας θυμίζει η περιγραφή της αναμενόμενης συμπεριφοράς του συστήματος, την κατανομή μεγίστου-σταθμισμένου-ελαχίστου, μάλλον θα έπρεπε να ξαναδιαβάσετε το σχετικό κεφάλαιο.

Είναι όμως αυτό το καλύτερο που μπορούμε να κάνουμε; Εδώ μπορούν να υπάρξουν ενστάσεις όπως θα φανεί στη συνέχεια. Αρχικά ομώς θα αναλύσουμε το σύστημα με δρομολογητές τύπου WRR.

4.2 Ανάλυση συστήματος

Διαισθητικά δημιουργείται η εντύπωση ότι το σύστημα με αναλογικούς χρονο-δρομολογητές, θα πλησιάσει κάπως στο στόχο δικαιοσύνης που έχουμε περιγράψει. Βασικά περιμένουμε πως οι ροές που αντιμετωπίζουν μέγιστη-συμφόρηση σε ένα κόμβο, νομοτελειακά ότι θα κάνουν δειγματοληψία – θα είναι ενεργές σε αυτόν τον κόμβο – συχνότερα από ότι εξυπηρετούνται από αυτόν, και συνεπώς δύσκολα θα χάσουν τη σειρά εξυπηρέτησης από το δρομολογητή αυτού του κόμβου· συνεπώς θα προσαρμοστούν στο δίκαιο ποσοστό τους σε αυτό τον κόμβο, που λογικά[2] θα είναι κοντά και στο συνολικά δίκαιο. Αυτό που ακολουθεί είναι μία απόδειξη ότι αυτό όντως θα ίσχυε – απόλυτα – σε ένα κόσμο ρευστών.

4.2.1 Χαρακτηριστικά αρχιτεκτονικής υπό όρους δρομολόγησης.

Σε ένα κόσμο ρευστών...

Θεώρημα 4.2.1 *Σε ένα κόσμο ρευστών, με ροές σταδερής κατάστασης(ενεργής/ανενεργής), που χρησιμοποιούν το σύστημα, εξοπλισμένο με GPS δρομολογητές και "αυτόματη"*

[2]Ίσως όχι και τόσο λογικά δεδομένης της πολυπλοκότητας του αλγορίθμου μεγίστου-σταθμισμένου-ελαχίστου, αφού σημασία παίζει και η σείρα με την οποία συναντούν συμφόρηση οι ροές.

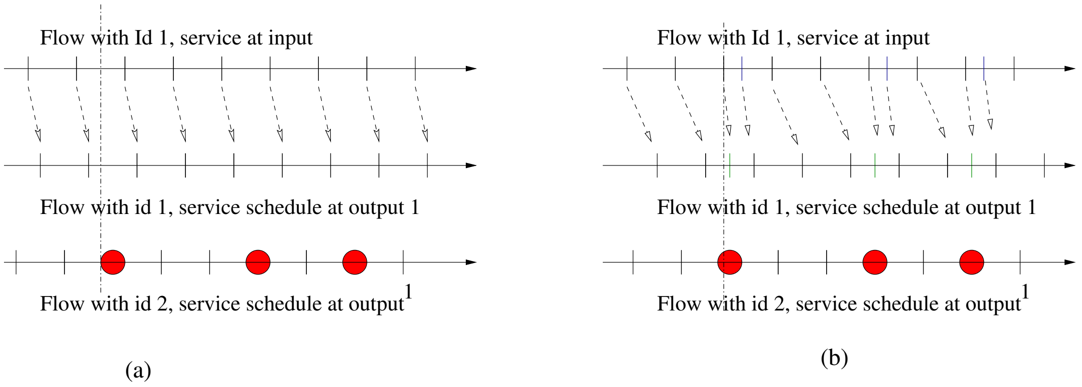
προσαρμογή ρυθμών παροχής - βασικά backpressure με απειροελάχιστο παράθυρο - οι ρόες θα πάρουν τελικά τα Weighted Max-Min fair ποσοστά τους.

Περίγραμμα Απόδειξης

Βασικά η ιδέα είναι να κάνουμε "trace" τις ροές που είναι συνεχώς ενεργές σε κάποιο κόμβο. Για αυτές τις ροές μπορούμε με σιγουριά να βρούμε τι ποσοστό θα πάρουν και επαναληπτικά, μέσω της ίδιας διαδικασίας, να βρούμε τα ποσοστά όλων των ροών-ρευστών. Τότε βασικά θα ξανά-ανακαλύψουμε τον αλγόριθμο εύρεσης ποσοστών για την κατανομή μεγίστου-σταθμισμένου-ελαχίστου. Πιο περιγραφικά-αναλυτική απόδειξη μπορείτε να βρείτε στο [BufCrossbar]. ♠

Άρα στο ιδανικό μοντέλο, το εξιδανικευμένο/προσαρμοσμένο σύστημα, συστήνεται για βέλτιστη κατανομή. Φυσικά σε ένα πραγματικό σύστημα πακέτων υπάρχουν αποκλίσεις τόσο στην ισχύ των υποθέσεων, όσο και στην ακρίβεια της πιθανολογούμενης σύγκλισης. Πρώτον, οι δρομολογητές θα είναι μία προσέγγιση των ιδεατών GPS. Και δεύτερον η αυτόματη προσαρμογή των ρυθμών παροχής μεταξύ δρομολογητή εισόδου και εξόδου, θα είναι τόσο αυτόματη όσο θα την κάνει το πραγματικό σχήμα διαχείρισης των εσωτερικών μνημών.

Σίγουρα κάθε ροή θα πρέπει να μπορεί να πάρει το μικρότερο από τα ποσοστά που στατικά δικαιούται στην είσοδο και στην έξοδο. Με αυτό το πρόβλημα έχουν ασχοληθεί στα [DGPS][ChQoS], βγάζοντας λίγο διαφορετικά αποτελέσματα. Πάντως φαίνεται ότι χρησιμοποιώντας μνήμες αξίας 2 cells σε κάθε crosspoint, εξασφαλίζουμε ότι κάθε ροή θα μπορεί να πάρει αυτό το ελάχιστο σε κάθε περίπτωση, αν τοποθετήσουμε WF^2Q/βέλτιστους δρομολογητές. Γενικά το μέγεθος για τους ενταμιευτές που το εξασφαλίζει αυτό για την χειρότερη περίπτωση, θα εξαρτάται από το πόσο καλή/ακριβής είναι η τεχνική προσέγγισης του μοντέλου ρευστών και αναλόγως ίσως και από το μέγεθος του μεταγωγέα· αν για παράδειγμα χρησιμοποιούμε δρομολογητές WRR στα σημεία ανταγωνισμού, τότε ο ελάχιστος-ικανός χώρος εσωτερικής μνήμης που θα προκύψει από την ανάλυση θα αυξάνει αναλογικά με το μέγεθος του μεταγωγέα, ενώ αν επιλέξουμε εξυπηρετητές WF^2Q τότε ο ικανός και αναγκαίος χώρος εσωτερικής αποταμίευσης είναι ανεξάρτητος από το μέγεθος αυτό.

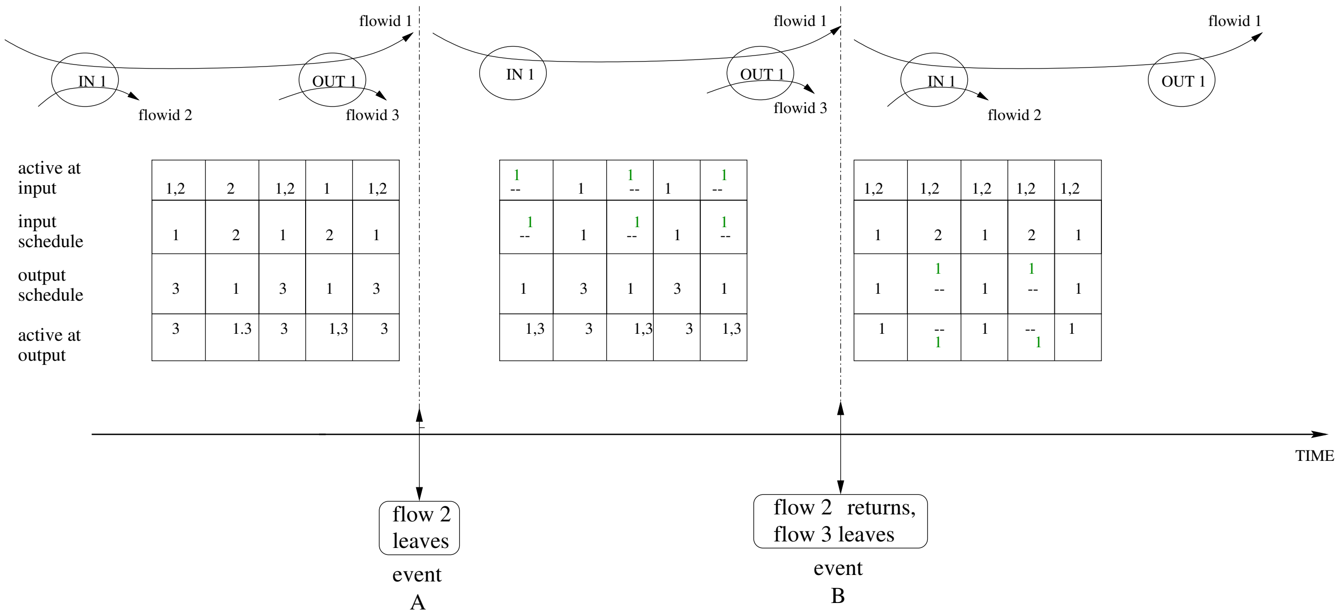


Σχήμα 4.4: Flow with id 2 and flow with id 1 have equal priority for the output link. But flow with id 2 suffers from a more strive bottleneck in the input, which results in empty slots(red circles). With one buffer space, the flow with id 1 – which can accommodate more bandwidth from the input –, can happen to be able to get advantage of these empty slots(as happens in case b), or may not(as happens in case a). If we use additional buffer space, the flow will always be able to get advantage of the excess bandwidth

Μεγάλοι εσωτερικοί ενταμιεύτες και ο ρόλος τους.

Γενικά μεγαλύτεροι εσωτερικοί ενταμιευτές, δημιουργούν καλύτερες προυποθέσεις για την εγκαθίδρυση του στόχου δικαιοσύνης. Μεγαλύτεροι buffers δίνουν την δυνατότητα στους δρομολογητές εισόδου και εξόδου να λειτουργούν πιο ανεξάρτητα, να παίρνουν υπόψιν τους μεγαλύτερο μέρος της πρόσφατης ῒστορίας῞ του συμπληρωματικού κατανεμητή μίας ροής – δεν χρειάζεται να συγχρονίζονται – και συνεπώς η συμπεριφορά να πλησιάζει περισσότερο σε αυτή του ρευστού κόσμου. Το πρόβλημα γενικά είναι αν στο διακριτό σύστημα πακέτων, φαινόμενα κακού συγχρονισμού μεταξύ συμπληρωματικών ως προς μία ροή εξυπηρετητών, δημιουργήσουν επαναλαμβανόμενα φαινόμενα αδικίας. Μεγαλύτεροι ενταμιευτές, περιορίζουν τα φαινόμενα αυτά, αφού ουσιαστικά αυξάνουν την πιθανότητα επιτυχούς δειγματοληψίας(Σχ. 4.4).

Επίσης μεγαλύτεροι ενταμιευτές θέτουν τις βάσεις για μεγαλύτερη απασχόληση των εξωτερικών γραμμών σε μεταβατικές καταστάσεις· οι ενταμιευτές διευκολύνουν την προώθηση πακέτων πιο κοντά στα σημεία προσωρινής συμφόρησης, έτσι ώστε αυτά να μπορούν να εκμεταλλευτούν την ῞πρώτη῞ ευκαιρία δρομολόγησης(



Σχήμα 4.5: Tables illustrates the argument that bigger crosspoint buffer can provide better utilization. These figures shows one input and one output server. There are three flows in the system, flow 1 which uses both the input and the output, flow 2 that uses only the input-and some other output-while flow 3 uses only the output-and some other input. In the case of a single buffer, this output will remain under-utilized after event B, because flow with id 1 cannot use the excess bandwidth since this flow confronts a bottleneck at the input. If this flow could use a buffer to put the excess bandwidth between events A and B, when flow 2 was inactive, then this flow would be able to occupy the output link after event B(as is shown with the green color), even for a while.

Σχ. 4.5). Άλλωστε αν αυξήσουμε υπερβολικά τους ενταμιευτές, θα καταλήξουμε σε ένα σύστημα με ουρές αποθήκευσης στις εξόδους πού είναι όπως έχουμε δεί και πεί, βέλτιστο στην αύξηση του βαθμού απασχόλησης των εξωτερικών γραμμών.

Στα αρνητικά των μεγάλων εσωτερικών ενταμιευτών πέρα από το αυξητικό κόστος κατασκευής που εισάγουν, είναι και η αύξηση της αδυναμίας/αδράνειας του συστήματος να αντιδράσει σε ¨περιβαλλοντολογικές αλλαγές¨ – όπως θα δούμε παρακάτω – και επίσης η καθυστέρηση ¨απεγκλωβισμού¨ απο κακά σενάρια συντονισμού πού περιορίζουν την συνολική παροχή του συστήματος· αν και σχετικά απίθανα αυτά τα σενάρια, μπορούν να συμβούν όπως θα δούμε.

Έχουμε καταλήξει σε μία απόδειξη, η οποία βρίσκεται σε ¨εφηβική μορφή-¨, ότι για το σύστημα πακέτων που έχουμε προτείνει όντως υπάρχει ένα ικανό μέγεθος ενταμιευτών που θα εξασφαλίζει την προσέγγιση του ιδεατού σκοπού μεγιστου-σταθμισμένου-ελαχίστου με απεριόριστη ακρίβεια, όταν οι ροές έχουν σταθερή κατάσταση(βλ. [BufCrossbar]). Για να την περιγράψουμε όμως, πρέπει πρώτα να περιγράψουμε τις αλληλεπιδράσεις μεταξύ των ροών και την γενικά συμπεριφορά του συστήματος στις μεταβατικές/μη-ισορροπημένες καταστάσεις, που ισχύουν και κατά τη διάρκεια της σύγκλισης. Αυτός είναι ο σκοπός των επόμενων παραγράφων.

4.2.2 Χαρακτηριστικά δρομολόγησης σε μεταβατικά φαινόμενα.

Αντικείμενο και κίνητρα παραγράφου

Λέμε ότι μία ροή αλλάζει κατάσταση όταν είτε το βάρος της αλλάζει, είτε όταν γίνεται από ενεργή/ανενεργή και το αντίστροφο. Το γιατί μας ενδιαφέρουν αυτές οι αλλαγές είναι προφανές· δεν μπορούμε να απαιτούμε από τις ροές να διατηρούν την κατάσταση τους μέχρι το σύστημα να βρεί το δίκαιο ποσοστό τους, γιατί αυτό γενικά δεν θα ισχύει σε ένα πραγματικό περιβάλλον. Αλλά και ανεξάρτητα από αυτό, γιατί η ανάλυση που προκύπτει μας αποκαλύπτει την δυναμική του συστήματος κατά τη διάρκεια της σύγκλισης.

Όταν είμαστε σε μία κατάσταση ισορροπίας A και η κατάσταση μίας ροής αλλάζει, περιμένουμε να βρεθούμε σε μία νέα κατάσταση ισορροπίας B. Η ροή που άλλαξε κατάσταση μπορεί να επηρεάσει δυνητικά όλες τις ¨γείτονες ροές¨ – αυτές που μοιράζονται μαζί της ένα κοινό κόμβο/πόρο – και η αλλαγή αυτών, με τη σειρά τους, όλες τις δικές τους γείτονες κτλ. Ουσιαστικά έχουμε ένα αναμενόμενο κύμα αλλαγών στη ζήτηση τών ροών, που μετατρέπει ένα αρχικό equilibrium A σε ένα νέο $\rightarrow B$. Η πρώτη μας ερώτηση είναι ποιές ροές μπορούν να επηρεαστούν από αυτήν την αρχική αλλαγή. Η (πρωτα)αρχική αλλαγή – βασικά μία αλλαγή ζήτησης είτε πρόκειται για αλλαγή βάρους/βαθμού-βιασύνης, είτε για αλλαγή διαθεσιμότητας(ενεργή $\longleftrightarrow$ ανενεργή) –, προκαλεί αλλαγή στο ρυθμό εξυπηρέτησης γειτονικών ροών, συνεπώς δευτερογενή αλλαγή στη ζήτηση αυτών των ροών στις συμπληρωματικές/υπόλοιπες γραμμές που χρησιμοποιούνε, άρα νέα γεγονότα αλλαγής κατάστασης. Σίγουρα οι ροές που θα επηρεαστούν από την πρωταρχική αλλαγή θα πρέπει να ανήκουν σε ένα κοινό υπογράφο αλληλεπίδράσεων(interaction graph για ορισμό βλ. [BufCrossbar]) με την αρχική ροή που άλλαξε κατάσταση. Η κυριαρχούσα όμως αρχή του μεγίστου-σταθμισμένου-ελαχίστου, περιορίζει το σύνολο αυτό, σε αυτό που εδώ ονομάζουμε trajectory of flow f, under equilibrium A ή πεδίο/περιοχή/βολή επιρροής/επιβολής/κυριαρχίας της ροής f κάτω από την κατάσταση ισορροπίας A.

Περιορισμός πεδίου επιρροής/κυριαρχίας μίας ροής στο μοντέλο των ρευστών.

Η ανάλυση που ακολουθεί στηρίζεται περισσότερο στον ορισμό της δίκαιης κατανομής μεγίστου-σταθμισμένου-ελαχίστου, αλλά μπορεί να επεκταθεί όπως είδαμε και για την ιδεατή περίπτωση ρευστών ροών, που χρησιμοποιούν το σύστημα μας εξοπλισμένο με GPS χρονοδρομολογητές.

Αυτό που επιδιώκουν τα παρακάτω θεωρήματα, αποδείξεις των οποίων μπορείτε να βρείτε στο [BufCrossbar], είναι να περιγράψουν το σύνολο των ροών που θα αλλάξουν κατάσταση ύστερα από την αλλαγή της κατάστασης μίας άλλης ροής. Θέλουμε μία περιγραφή όσο πιο μηχανηκιστική γίνεται – υπό όρους αιτίας-αποτελέσματος –, για να βάλουμε σε τάξη το χάος των αλληλεπιδράσεων τόσο στην κατάσταση της μικρής αλλαγής κατάστασης, όσο και στο φαινόμενο, της υποστηριζόμενης αλλά ακόμα υποθετικής για το σύστημα πακέτων, σύγκλισης.

Το πρώτο θεώρημα το κάνει αυτό τελείως ¨καταστασιακά¨, χωρίς να βγάζει/εξάγει καθόλου τα στοιχεία της αλληλεπίδρασης.

Θεώρημα 4.2.1 *Ύστερα από μία ουσιαστική αλλαγή στη ζήτηση μίας ροής f, που ωθεί το δίκτυο από μία κατάσταση ισορροπίας A σε μία νέα B, μία ροή g θα αλλάξει ρυθμό μόνο αν αυτή έπαιρνε μεγαλύτερη-ίση utility από την f, είτε στην πρώτη κατάσταση A, είτε στην τελική B:*

$$U_g^A \geq U_f^A \text{ ορ } U_g^B \geq U_f^B \qquad .$$

Δεν παρουσιάζουμε την απόδειξη αυτή – καθώς και τις υπόλοιπες – αν και είναι σχετικά απλή. Το θεώρημα αυτό μας λέει ουσιαστικά, δεδομένης της αλλαγής, ποιές ροές *μπορεί* να επηρεάστηκαν, δηλαδή όχι και τόσο πολλά. Μας λέει όμως και ποιά αλλαγή υπήρξε η αιτία και ποιά το αποτέλεσμα/συνέπεια. Έτσι με βάση αυτό το θεώρημα μία ροή f μπορεί να επηρεάσει όλες τις άλλες ροές g που βρίσκονται στο γράφο αλληλεπιδράσεων που την περιέχει, ανεξάρτητα της αρχικής της κατάστασης, αρκεί η f να αντιμετωπίζει μεγαλύτερη-ίση συμφόρηση από την g, είτε πρίν είτε μετά την αλλαγή. Στην συνέχεια θα προσπαθήσουμε να κάνουμε αυτό το σύνολο ακόμα πιο ¨στενό¨.

Θεώρημα 4.2.2 *Η βολή επιρροής μίας ροής f, είναι περιορισμένη στους γείτονες αυτής της ροής και σε όλες τις άλλες ροές, που μπορούν να προσεγγιστούν από την*

f ή από κάποιο γείτονα της, μέσω ενός μονοπατιού από αλληλεπιδρώμενες ροές, με όχι-αυστηρά αυξανόμενο utility κάτω από την κατάσταση ισορροπίας A.

Ουσιαστικά η απόδειξη στηρίζεται στην διαισθητική, ότι για να αλλάξει μία ροή παροχή σε δεύτερο χρόνο – δηλαδή χωρίς να αλλάξει το βάρος της ή η κατάσταση διαθεσιμότητας της –, θα πρέπει πρώτα να αλλάξει ένας γείτονας της, και με βάση το προήγουμενο θεώρημα για να επηρεαστεί από την αλλαγή του γείτονα, θα πρέπει αυτός ο γείτονας να έπαιρνε μικρότερο-ίσο utility[3]. Οι γείτονες μπαίνουν στο παιχνίδι, ουσιαστικά γιατί μία ροή με μεγάλο utility μπορεί αυξάνοντας το βάρος της, να επηρεάσει τους γειτόνους της ακόμα και αν αυτοί έχουν αρχικά μικρότερο utility και στην συνέχεια η προκύπτουσα αλλαγή αυτών, να επηρεάσει άλλες ροές. Ουσιαστικά τα κύματα της αλλαγής αποτελούνται από γειτονικές ροές πάνω στο γράφο αλληλεπιδράσεων. Κάθε κύμα/ακολουθία-αλλαγών αποτελείται απο αλυσιδωτά ζευγάρια· σε κάθε ζευγάρι ένας επηρεάζεται θετικά και ένας αρνητικά, θετικά, αρνητικά κτλπ. Το επόμενο θεώρημα, συνέπεια του παραπάνω, μας λέει τι μπορεί να γίνει αν τα βάρη θεωρηθούν σταθερά.

Θεώρημα-συνέπεια 4.2.1 *Μία ροή f που αλλάζει κατάσταση από ενεργή σε ανενεργή ή το αντίστροφο, μπορεί να επηρεάσει μόνο ροές με μεγαλύτερο utility από την f, θεωρούμενο στην ενεργή κατάσταση της.*

4.2.3 Συνέπειες θεωρημάτων

Μηχανική περιγραφή φαινομένων σύγκλισης

Τα θεωρήματα μας λένε ουσιαστικά ότι για κάθε επηρεασμένη ροή, υπάρχει ένα μονοπάτι αλλαγών που την προκάλεσε. Άρα και ο χρόνος της μεταβατικής φάσης είναι περιορισμένος στη μέγιστη διάρκεια κάποιου τέτοιου μονοπατιού. Πολλά τέτοια μονοπάτια/κύματα μπορεί να υπάρχουν παράλληλα, αλλά αυτό της μεγαλύτερης διάρκειας μας ορίζει τη χειρότερη περίπτωση και το χρόνο τελικής σύγκλισης συνολικά. Επίσης τα ίδια θεωρήματα μας λένε, ότι πρώτα σταθεροποιούνται ροές μικρής utility, και στην συνέχεια οι υπόλοιπες μπορούν με

[3] Η απόδειξη στο [BufCrossbar] χρησιμοποιεί ένα σύνολο από προκείμενες/λήμματα, που αποκαλύπτουν επιπλέον προυποθέσεις πραγμάτωσης ενός τέτοιου μονοπατιού μεταφοράς/μετάδοσης του κύμματος αλλαγών.

ασφάλεια να ¨ψάξουν¨ τα δίκαια ποσοστά τους – όπως άλλωστε συμβαίνει και με τον αλγόριθμο που βρίσκει τη δίκαιη κατανομή. Κάτι παρόμοιο συμβαίνει και στην περίπτωση της σύγκλισης του συστήματος από την αρχή.

Αρχικά δημιουργούνται μικρές/προσωρινές θέσεις ισορροπίας(sub-equilibrium, που τις θεσμοθετούν οι δρομολογητές εισόδου πάνω σε γειτονικές ροές με κοινή είσοδο, μέχρι να γεμίσουν κάποιοι εσωτερικοί ενταμιευτές οπότε το backpressure αλλάζει την ζήτηση ορισμένων ροών(αυτών που στο νέο – αλλά και παλι προσωρινό/τοπικό equlibrium – έχουν σημείο συμφόρησης την εξωτερική γραμμή), και συνεπώς αλλάζει την παλιά/αρχική κατάσταση ισορροπίας. Στην συνέχεια πρέπει να φτάσουν τα ολικά φαινόμενα/γεγονότα αλλαγής κατάστασης από ροές με μικρότερο utility, μέσω ενός μονοπατιού όπως μόλις πριν περιγράψαμε, για να καταλήξουν σωστά οι ροές στα ολικώς δίκαια rates τους. Στο [BufCrossbar] παρουσιάζουμε μία πρώτη απόδείξη ότι το σύστημα πακέτων όντως συγκλίνει αν χρησιμοποιήσουμε ένα ελάχιστο μέγεθος ενταμιευτών εσωτερικά.

Παράμετροι χρόνου μετάβασης/σύγκλισης στο σύστημα πακέτων

Ο χρόνος μετάβασης/σύγκλισης επηρεάζεται από την ύπαρξη τέτοιων μονοπατιών, όπως τα περιγράψαμε. Το μήκος αυτών είναι σίγουρα μία πρώτη παράμετρος. Σε ένα σύστημα πακέτων, οι αλλαγές μπορεί να προκαλέσουν αλλαγή του σημείου συμφόρησης από την είσοδο στην έξοδο ή και το αντίστροφο. Άρα για να προχωρήσει μία αλλαγή θα πρέπει πρώτα να αλλάξει κατάσταση ο εσωτερικος ενταμιευτής και αυτό χρειάζεται χρόνο[4].

Βασικά αν μία ροή αυξάνει την παροχή της σε ένα κόμβο, θα μεταβιβάσει το μήνυμα της αλλαγής αυτής χωρίς καθυστέρηση στο συμπληρωματικό κόμβο που χρησιμοποιεί, ενώ μία ροή που μειώνει την παροχή της, θα χρειαστεί χρόνο για να μεταβιβάσει την αλλαγή αυτή – ως μείωση ζήτησης/δειγματοληψίας – στους γείτονες της στο συμπληρωματικό κόμβο, όσο χρόνο θα κάνει να γεμίσει ή να αδειάσει ο δικός της ενταμιευτής· ο ρυθμός αλλαγής κατάστασης του ενταμιευ-

[4]Δεδομένου ότι οι ροές έχουν σταθερή κατάσταση(ενεργή/ανενεργή), σε κατάσταση ισορροπίας, ο εσωτερικός ενταμιευτής μίας ροής, είτε θα είναι συνεχώς σχεδόν γεμάτος, είτε σχεδόν άδειος, αναλόγως του αν αυτή αντιμετωπίζει μέγιστη-συμφόρηση/bottleneck στην έξοδο ή στην είσοδο αντίστοιχα.

τή είναι ίσος με την παλιά της παροχή μείον την καινούργια. Αυτός ο χρόνος, προφανώς εξαρτάται από την παλιά παροχή της ροής και από το μέγεθος του ενταμιευτή που χρησιμοποιούσε αυτή. Άρα μία πρώτη εκτίμηση για το χρόνο σύγκλισης στη χειρότερη περίπτωση είναι $O(N \cdot B/min(old\_rate - new\_rate))$. Όμως αν μετρήσουμε την ¨αδικία¨ κατα τη διάρκεια της προσαρμογής σε μονάδες εξυπηρέτησης bits, τότε αυτή είναι ανεξάρτητη από την προηγούμενη παροχή min(old_rate-new_rate), αφού ο ρυθμός αδικίας(min(old_rate-new_rate)) και η διάρκεια αυτής είναι αντιστρόφως ανάλογα – συνεπώς το γινόμενο τους(=¨αδικα¨ bits) δεν εξαρτάται από το min(old_rate-new_rate). Στο [BufCrossbar] παρουσιάζουμε αναλυτικότερα αυτες τις παραμέτρους, τις οποίες θα επανεξετάσουμε και εδώ στο κεφάλαιο με τα σχετικά πειραματικά αποτελέσματα.

4.2.4 Πιθανοκρατική κίνηση — μέση καθυστέρηση — σύγκριση με το ιδεατό σύστημα με τις κύριες ουρές αποθήκευσης στις εξόδους

Πρόβλημα αρχιτεκτονικής

Μέχρι στιγμής έχουμε μελετήσει το σύστημα κάτω από απλουστευτικές υποθέσεις, γεγονός που διευκόλυνε την εξαγωγή χρήσιμων αποτελεσμάτων. Για παράδειγμα μία απόπειρα απόδειξης ότι το σύστημα αποδίδει δίκαιη εξυπηρέτηση στην περίπτωση πιθανοκρατικών αφίξεων, παρουσιάζει προφανώς μεγάλα προβλήματα – βλέπε [HahRRMaxMin]. Στο πιθανοκρατικό μοντέλο, θα μελέτησουμε πιο απλά πράγματα και πιο συγκεκριμένα, το χρόνο απόκρισης του συστήματος στο να προωθήσει ένα σύνολο αφίξεων. Το ιδανικό σύστημα υπό αυτό το πρίσμα είναι ένα σύστημα με τις ουρές αποθήκευσης στις εξόδους – ανέξαρτητα της αρχής χρονοπρογραμματισμού που χρησιμοποιεί, αρκεί να είναι work conserving.

Διατηρώντας τα πακέτα στις εξόδους εξασφαλίζουμε ότι πάντα μία έξοδος θα είναι απασχολημένη, αν υπάρχει τουλάχιστον μία ροή που προορίζεται σε αυτή. Αυτό αποτελεί γενίκευση της έννοιας ¨συντήρηση/διατήρηση της εργασίας¨/work conserving για την περίπτωση ενός πολυπορικού συστήματος όπως ο μεταγωγέας. Η χρησιμοποίηση work conserving δρομολογητών στο σύστημα μας, δεν μας εξασφαλίζει ότι αυτό συνολικά θα διατηρεί την εργασία· μπορεί <u>ένα</u> πακέτο που έφτασε στο μεταγωγεα και είναι το μοναδικό προς <u>αυτή</u> την έξοδο, να μήν προωθηθεί τελικά, λόγω ύπαρξης συναγωνισμού στην είσοδο με άλλες ροές – που μπορεί να συναντούν συναγωνισμό και στην έξοδο και στην οποία περίπτωση, ώς προς το κριτήριο της ελάχιστης μέσης καθυστέρησης ή αλλιώς της μέγιστης παροχής, θα έπρεπε να επιλεγεί για προώθηση το πακέτο <u>ένα</u>.

Βασικά ο χωρισμός/διαίρεση του προβλήματος σε $2 \cdot N$ ανεξάρτητα υποπροβλήματα, πέραν από τα καλά του, εισάγει και κάποια αρνητικά. Σε μικρό-μεσοπρόθεσμο διάστημα, τίποτα δεν μπορεί να μας εξασφαλίσει ότι ¨συνωμοτικά¨ οι δρομολογητές στις εισόδους δεν θα επιλέγουν συνεχώς ροές που προορίζονται για την ίδια εξωτερική γραμμή τόσο σε σχέση με την προηγούμενη απόφαση τους, οσο και σε σχέση με τις υπόλοιπες εισόδους, περιορίζοντας έτσι την παροχή του συστήματος σε κάτι, από αρκετά ώς πολύ μικρότερο του βέλτιστου[5]. Φυσικά μία έξοδος δεν θα μπορεί να απορροφήσει τις ταυτόχρονες αιτήσεις πολλών εισόδων μαζί, και σε ένα εύλογο χρονικό διάστημα το σύστημα διαχείρισης της μνήμης(backpressure) θα αναγκάσει τελικά τους δρομολογητές στις εισόδους να διαλέξουν διαφορετικές εξόδους και συνεπώς – αυξάνοντας τον αριθμό των ενεργών εξόδων –, έμμεσα να αυξήσουν την παροχή συνολικά και να μειώσουν τη μέση καθυστέρηση των πακέτων.

Μία γρήγορη ανάλυση χειρίστης περιπτώσεως, θα μας πείσει ότι τελικά σε ένα χρονικό διάστημα ίσο με το μέγεθος των ενταμιευτών στα crosspoints B, τό σύστημα μας θα έχει στείλει μόνο B πακέτα, ενώ το βέλτιστο σύστημα θα είχε στείλει $N \cdot B - N$ φορές περισσότερα. Αυτό ισχύει για αυτό το χρονικό διάστημα και όχι μεγαλύτερο. Τελικά συνεχίζοντας την ίδια ανάλυση και μετρώντας το σύνολο των κατανεμητών εξόδου που έχουν έναν τουλάχιστον υποψήφιο, βγαίνει ότι σε ένα χρονικό διάστημα $N \cdot B$ το σύστημα μας μπορεί να στείλει μόνο $1 \cdot B + 2 \cdot B + \ldots + N \cdot B$ πακέτα, ένω το βέλτιστο θα ήταν $N \cdot N \cdot B - \simeq 1/2$. Φυσικά, όσο πιθανό είναι αυτό το σενάριο, άλλο τόσο ακριβώς είναι να πετύχουμε ακριβώς το βέλτιστο, με την συνήθη περίπτωση να βρίσκεται κάπου στη μέση – αλλά εξαρτάται από τα βάρη τών ροών, τις αφίξεις, τους δρομολογητές και τις αποφάσεις τους και ίσως και από άλλα πράγματα, που δυσκολεύουν πολύ την ανάλυση.

Γενικά περιμένουμε, κάτω από κάποιες (προ)υποθέσεις, το σύστημα να δίνει περίπου την ίδια μέση καθυστέρηση με ένα σύστημα με τις ουρές αποθήκευσης στις εξόδους. Αυτές είναι πάνω-κάτω οι εξής: Μία είναι, η συμφόρηση/συνδυασμός-ενεργών-ροών-και-του-βάρους-αυτών στους $2 \cdot N$ πόρους, να είναι περίπου η ίδια. Τότε κάθε ροή θα αντιμετωπίζει κατά μέσο όρο τον ίδιο ανταγωνισμό τόσο στην είσοδο όσο και στην έξοδο, οπότε διαισθητικά, κάποια ροή το πολύ να έχει διπλάσια μέση καθυστέρηση. Αν η ζήτηση είναι ανομοιόμορφα κατανεμημένη – ξανά ώς ζήτηση θεωρούμε τόσο την ύπαρξη πακέτων όσο και το βάρος/σχετική-προτεραιότητα-αυτών –, τότε ένα πακέτο μίας ροής μπορεί να αντιμετωπίζει μεγάλη καθυστέρηση στην είσοδο ενώ ή έξοδος που του αντιστοιχεί

[5]Εμφανίζεται ξανά η πιθανολογούμενη αντινομία ανάμεσα στην απαίτηση για ποιότητα υπηρεσίας και μεγάλης συνολικής παροχής – αν όλες οι ροές ήταν ϊσότιμες" τότε το φράγμα για την απόκλιση από το βελτιστο που θα βγάλουμε παρακάτω, θα μπορούσε να ήταν καλύτερο.

είναι ελεύθερη – στο ιδανικό σύστημα θα έβλεπε μηδαμινή καθυστέρηση. Η δεύτερη υπόθεση είναι, ας είναι η ζήτηση όπως θέλει κατανεμημένη, αλλά ο ρυθμός των αφίξεων να είναι κάποια ¨λογική¨/γραμμικά-ευθέως-ανάλογη συνάρτηση των σχετικών βαρών στις εισόδους. Για προφανείς λόγους και πάλι το σύστημα θα παρουσιάζει συμπεριφορά κοντά στο ιδεατό.

Περιγραφή αναμενόνης συμπεριφοράς, με πειραματική — μερική — επιβεβαίωση.

Δε θα παρουσιάσουμε πειραματικά αποτελέσματα με πιθανοκρατικές αφίξεις σε αυτό το έγγραφο· μπορείτε να βρείτε αρκετές στο [BufCrossbar]. Βασικά αυτές δείχνουν ότι η σχετική συμπεριφορά των δύο συστημάτων κυριαρχείται απο τέτοιες αρχές και επιπλέον ότι η εκδήλωση της απόστασης από το βέλτιστο στο σύστημα μας, εμφανίζεται σε πολύ μεγάλο φόρτο-ζήτησης – το όριο αυτό προφανώς αυξάνει όσο αυξάνει το μέγεθος των εσωτερικών μνημών[6]. Αυτό, όπως είπαμε και παραπάνω, εξαρτάται από τις υποθέσεις-παραμέτρους της εισερχόμενης κίνησης – γενικά είναι από εκεί που αρχίζει να μειώνεται/χαλάει η απόδοση τής output queuing αρχιτεκτονικής –, απλώς το σύστημα μας παρουσιάζει σχετικά χειρότερη επίδοση, μπορεί και πολλαπλασιαστικά στην αύξηση του φόρτου από ένα σημείο και μετά, αποτέλεσμα των μεγάλων – άπειρων στις προσομοιώσεις μας – buffer της γνήσιας output queuing αρχιτεκτονικής σε αντίθεση με τους μικρούς/($[1...10] \cdot N^2$ για μεταγωγείς $[4x4...32x32]$), ¨εξωτερικούς¨ ενταμιευτές στο [BufCrossbar]. Τέλος πρεπει να αναφέρουμε εδώ, ότι το σοβαρό πρόβλημα συντονισμού των εισόδων δεν θα ήταν τόσο τραγικό αν δεν χρησιμοποιούσαμε αναλογικούς δρομολογητές στις εισόδους αλλά κάποιες περισσότερο προσαρμοστικές στις ιδιότητες της εισερχόμενης κίνησης.

Αυτές οι αρχές, διατηρώντας περίπου τη σειρά αφίξεων στις προωθήσεις που κάνουν, θα εξασφαλίζουν ότι το σύνολο των ενεργών εξόδων – αυτές που έχουν πακέτα – δεν θα αποκλίνει πολύ του ίδιου συνόλου στην output queuing αρχιτεκτονική. Φυσικά δεν θα πετυχαίναμε την κατανομή μεγίστου-σταθμισμένου-ελαχίστου, αλλά χρησιμοποιώντας κάποια ¨προσαρμοστική¨(adaptive) μέθοδο

[6] Σχετικά είναι εδώ, διαγραμμάτα συμπεριφοράς ενός output queuing συστήματος, σε συνάρτηση με το μέγεθος των εξωτερικών ενταμιευτών, που μπορούν να βρεθούν στη βιβλιογραφία.

στις εισόδους, μπορεί να πετυχαίναμε πολύ μεγαλύτερη παροχή, ανεξάρτητα υποθέσεων όπως οι παραπάνω. Υποψήφιες είναι το Round Robin, η Longest Queue First η Least Recently Used και ίσως και αλλές[PackSwitDes]. Για την Longest Queue First υπάρχει απόδειξη ότι το σύστημα μπορεί να προσφέρει 100% απασχόληση στις εξόδους [Magil]. Αλλά με αυτή τη μέθοδο δεν προστατεύονται οι ροές από κάποιες που στέλνουνε συνέχεια, οπότε και συνεχώς έχουν τις μεγαλύτερες ουρές με αποτέλεσμα να εξυπηρετούνται συνεχώς. Στο [PackSwitDes] προτείνονται και αναλύονται άλλες εναλλακτικές, καθαρά προσαρμοστικές μέθοδοι(Least Recently Used, Longest Waiting Time), οι οποίες είναι πιο δίκαιες. Αλλά χρησιμοποιώντας τεχνικές τέτοιου τύπου, δεν μπορούμε να εξασφαλίσουμε οσηδήποτε μικρή καθυστέρηση σε κάποια ροή της αρεσκείας μας [DGPS], καθώς η ανάλυση πάντα θα περιέχει ένα παράγοντα της τάξης N και επιπλέον δεν είναι εφικτή η κατανομή μεγίστου-σταθμισμένου-ελαχίστου για προφανείς λόγους.

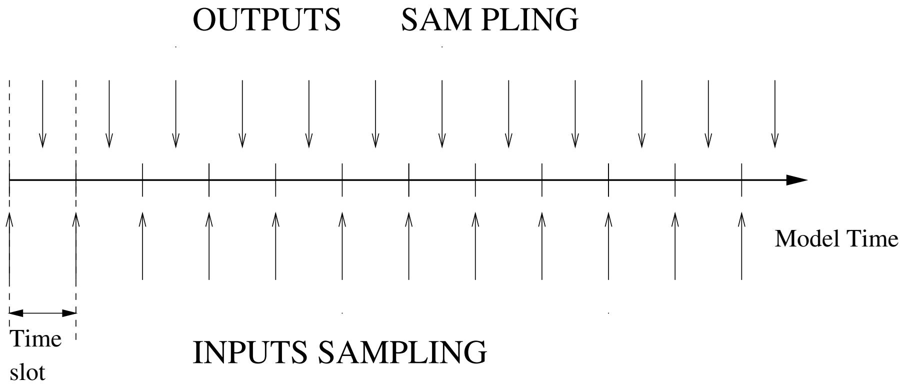
Στο τέλος της εργασίας, στο παράρτημα 1 – και μοναδικό –, παρουσιάζουμε μία novel, ίσως λίγο "adhoc" τεχνική – προς το παρόν διαισθητικά και πειραματικά μόνο δικαιολογημένη και επιβεβαιωμένη –, που ονομάζουμε προσαρμοστική-αναλογικά-κυκλική δρομολόγηση(Adaptive-WRR), ορισμένη κάτω από συμφραζόμενα παρόμοια με αυτά που ορίζει το σύστημα μας, και η οποία φαίνεται ότι μπορεί να ικανοποιήσει ταυτόχρονα και τις δύο απαιτήσεις. Μάλιστα έχει και κάποια άλλα θετικά που ακόμα ερευνούμε.

Κεφάλαιο 5

Πειραματικά αποτελέσματα προσομοιώσεων

5.1 Περιγραφή μοντέλου — υποθέσεις-στόχοι.

Σχέση αιτίας/αιτιατού — υπόθεση χρόνου/καθυστερήσεων



Σχήμα 5.1: The simulation event ordering.

Στό μοντέλο που βασίσαμε την προσομοίωση κάναμε κάποιες απλουστευτικές υποθέσεις, που πιστεύουμε ότι δεν επηρεάζουν την συμπεριφορά από την οπτική που αντιμετωπίζουμε το πρόβλημα. Υποθέσαμε μία αλγοριθμική συμπεριφορά – εννοιολογικά και πραγματικά ταυτόσημη με υλοποιήσεις σε υλικό(βλ. [Hart-BinTree]) –, για τους input και output WRR δρομολογητές που τοποθετήσαμε στο crossbar στις περισσότερες προσομοιώσεις(βλ. [BufCrossbar]) καί μία on/off θεώρηση για το backpressure, με ακαριαία μετάδοση των ανταλλασσόμενων μηνυμάτων.

Η απόδοση της αρχιτεκτονικής του συστήματος σε όρους αλγοριθμικούς, ακολούθησε το μοντέλο αιτιοκρατίας που παρουσιάζεται γραφικά στο Σχ. 5.1 και πιο προγραμματιστικά ακολούθως:

Repeat
. 1 $\forall$ flows $i \in [1 \dots N^2]$ traffic_arrives(i)
. 2 $\forall$ inputs $i \in [1 \dots N]$ input_server_take_decision(i)
. 3 $\forall$ crosspoints $i \in [1 \dots N^2]$ backpressure_propagate_state(i)
. 4 $\forall$ outputs $i \in [1 \dots N]$ output_server_take_decision(i)
. 5 $\forall$ crosspoints $i \in [1 \dots N^2]$ backpressure_propagate_state(i)
. 6 modelTime++

Το σχήμα υποθέτει πακέτα μοναδιαίου/σταθερού μεγέθους που καταφτάνουν, ολόκληρα στην αρχή του κάθε time_slot. Ένα πακέρο p που έρχεται στο time_slot t_i, αν δεν συναντήσει ανταγωνισμό θα εξυπηρετηθεί στιγμιαία από τον δρομολογητή εισόδου, θα μεταφερθεί στιγμιαία στους εσωτερικούς ενταμιευτές και σιτγμιαία θα εξυπηρετηθεί και από τον δρομολογητή εξόδου, αναχωρώντας από το σύστημα στό τέλος του time_slot t_i, ολόκληρο. Η κατάσταση backpressure του ενταμιευτή αυτής της ροής στο επομένο time_slot, t_{i+1}, θα είναι όπως ήταν και πρίν την άφιξη του πακέτου p, στο τέλος δηλαδή του time_slot t_{i-1}. Η υπόθεση αυτή αν και δεν είναι τελείως θεωρητική – μπορούν να κατασκευάστουν τέτοια συστημάτα μικρού μεγέθους –, μειώνει κατά ένα πακέτο τις απαιτήσεις στο μέγεθος των εσωτερικών ενταμιευτών· αυτό για την περίπτωση που υπάρχει καθυστέρηση μίας περιόδου/κύκλου-δρομολόγησης στην μετάδοση των σημάτων backpressure ή γενικότερα αναλογικά με τον αριθμό των κύκλων καθυστέρησης μετάδοσης των σημάτων backpressure, βλέπε [Κατ534].

Γλώσσα προγραμματισμού μοντέλου — περιγραφή πειράματων

Το μοντέλο των προσομοιώσεων, υλοποιήθηκε στη γλώσσα προγραμματισμού $C{+}{+}$ αλλά ακολούθησε μάλλον περισσότερο την δομική τεχνική παρά τήν αντικειμενοστραφή. Βασικά υπήρξαν τρεις εκδόσεις, μία για κάθε σκοπό/θέμα του πειράματος. Μία για την υπόθεση σύγκλισης στη κατανομή μεγίστου-σταθμισμένου-ελαχίστου, μία για τα μεταβατικά φαινόμενα και την εκτίμηση χρόνου σύγκλισης και μία για την ανάλυση προσφερόμενης/χρήσιμης παροχής, ένω υπήρξαν και άλλες εκδόσεις με διαφορετικές αρχές χρονο-προγραμματισμού(WRR, WF^2Q+, VC, $Adapt-WRR$ $LWTF$, LRU, $FIFO$, LQF). Έιναι ενδιαφέρον ότι οι περισσότερες αρχές από αυτές, μπορούν να υλοποιηθούν πάνω στον ίδιο πυρήνα – εντοπισμός μεγίστου/ελαχίστου –, μέ αλλαγές των παραμέτρων. Βάζοντας άπειρους ενταμιέυτες στα crosspoints και προωθώντας τα είσερχόμενα πακέτα αμέσως εκεί, φτιάξαμε ένα πρότυπο της ιδανικής μορφής του output queuing συστήματος. Τέλος προσομοιώσαμε κάποιες ακριβές υπό υλικό-οικονομικούς όρους μεθόδους, που βρίσκουν ταιριάσματα για το παραδοσιακό crossbar χωρίς εσωτερικούς μικρό-ενταμιευτές(βλ. [BufCrossbar]).

5.2 Τελική/μακροπρόθεσμη σύγκλιση στη δίκαιη κατανομή μεγίστου-σταθμισμένου-ελαχίστου.

5.2.1 Περιγραφή περιβάλλοντος πειράματος — Μέτρο βαθμού σύγκλισης.

Για να μελετήσουμε σε πρωτη φάση το βαθμό σύγκλισης στο στόχο δικαιοσύνης που έχουμε περιγράψει, θεωρήσαμε ότι η κατάσταση των ροών στο σύστημα είναι σταθερή – είτε έχουν συνεχώς πακέτα για δρομολόγηση στις ουρές εισόδου, είτε καθόλου και ποτέ. Αυτό έκανε την εξαγωγή και την ανάλυση των αποτελεσμάτων πιο εύκολη/γρήγορη και δεν μπορούμε να πούμε ότι η υπόθεση αυτή είναι τελείως μη-ρεαλιστική· όταν η κίνηση είναι πολύ απαιτητική, συμβαίνει σε κάθε μεταγωγέα με τις ουρές στις εισόδους να υπάρχει μεγάλη και συνεχής ζήτηση, ιδιαίτερα-προφανώς σε μεταγωγείς του backbone δικτύου. Έτσι οι παράμετροι των πειραμάτων ήταν κυρίως η κατανομή των βαρών και της κάτάστασης ζήτησης(on/off) στίς N^2 ροές και το μέγεθος των εσωτερικών ενταμιευτών στα crosspoints[1].

Εδώ μελετάμε το κατά πόσο συγκλίνει το σύστημα, ψάχνοντας βασικά το κα-

[1]Αλλη παράμετρος που δεν παρουσιάζουμε εδώ, είναι ο τύπος των δρομολογητών· αυτή τη παράμετρο την μελετάμε στο [BufCrossbar].

τάλληλο αντιπαράδειγμα που θα καταρρίψει την υπόθεση μας, ενώ παράλληλα μελετάμε την επίδρασης της κατανομής βαρών στις απαιτήσεις σε εσωτερικούς ενταμιευτές και την επιρροή των τελευταίων στο βαθμό ακρίβειας της σύγκλισης· η ανάλυση που παρουσίαζεται στο BufCrossbar], ισχυρίζεται/καταλήγει ότι με μέγεθος/χωρητικότητα ανάλογη του $2 \cdot N$ σε κάθε crosspoint, μακροπρόθεσμα μπορούμε να πετύχουμε σύγκλιση με απεριόριστη ακρίβεια για κάθε περίπτωση ακόμα και για την χειρότερη.

Για να μετρήσουμε το βαθμό ανακρίβειας χρησιμοποιήσαμε την παρακάτω φόρμουλα/μεταβλητή, που προσαρμόζεται στην ιδεατή δίκαιη παροχή κάθε ροής:

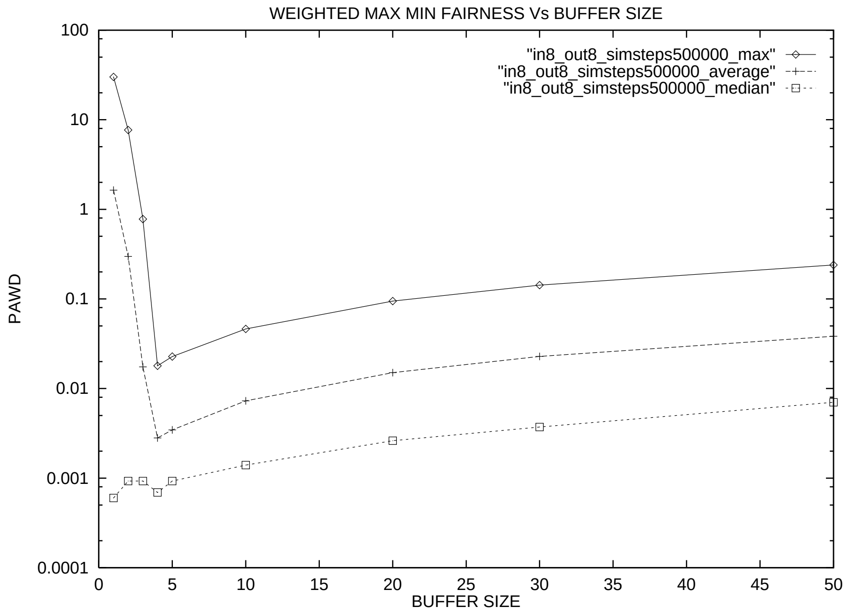
$$PAWD_f^{system}(t_1, t_2) = 100 \cdot \frac{|fair\_service_f(t_1, t_2) - simulate\_service_f^{system}(t_1, t_2)|}{fair\_service_f(t_1, t_2)}$$

Αυτή μας λέει, πόσο λιγότερη ή περισσότερη εξυπηρέτηση, θα πάρει αυτή η ροή από το σύστημά μας, για κάθε 100 μοναδιαιές/στοιχειώδεις εξυπηρέτησεις, που θα έπαιρνε μαθηματικά – ή στο μοντέλο των ρευστών. Στα σχήματα που ακολουθούν, παρουσιάζουμε συνολικά/αθροιστικά αποτελέσματα πάνω σε όλες τις ενεργές ροές(max, average, median).

Τέλος κάτι ενδιαφέρον είναι το εξής: επειδή είχαμε αρχικοποιήσει με τον ίδιο τρόπο όλους τους WRR δρομολογητές στις εισόδους – να διαλέγουν την ίδια ροή/έξοδο στην περιπτώση που η αρχή δρομολόγησης τις θεωρούσε ισότιμες –, στην αρχή των προσομοιώσεων παρουσιαζόταν το φαινόμενο της ¨ομαδικής συνομωσίας¨ που είχε ώς αποτέλεσμα την μικροπρόθεσμη – για τους σκοπούς του πειράματος – επιλογή τών ίδιων εξόδων και συνεπώς/τελικά το φαινόμενο υποχρησιμοποίησης των τελευταίων, όπως περιγράψαμε στο Κεφ. 4.2.4 .

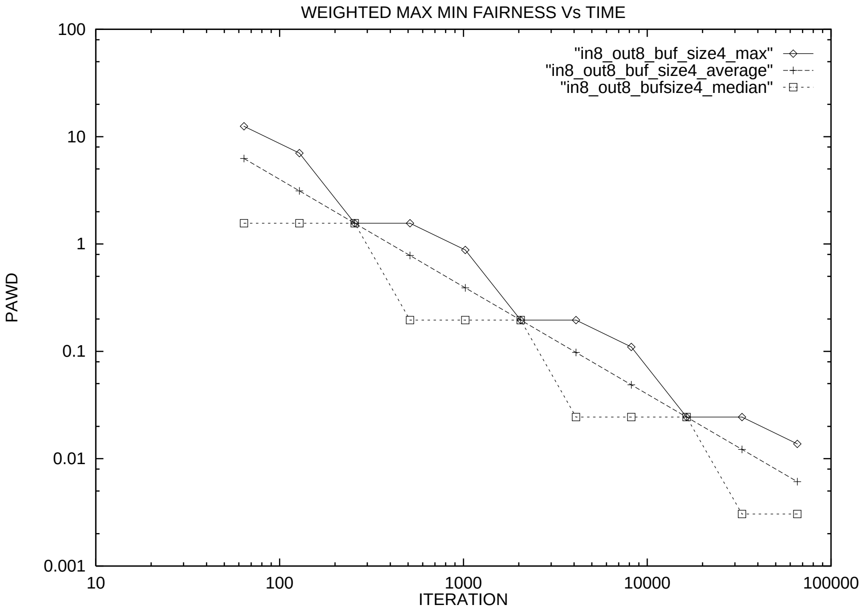
Ίδια βάρη-μέτρο σύγκλισης-μέτρο βαθμού σύγκλισης

Στην περίπτωση που τα βάρη όλων των ροών είναι ίδια και όλες είναι σε ενεργή κατάσταση, τότε περάν από την αρχική ανωμαλία που περιγράψαμε παραπάνω και κυριαρχεί για N κύκλους, το σύστημα αμέσως μπαίνει στην πλήρη κατάσταση δικαιοσύνης, εξυπηρετώντας ισότημα όλες τις ροές(κάθε N κύκλους) και διατηρώντας συνεχώς τις εξωτερικές γραμμές απασχολημένες. Αυτή η ιδανική Round Robin συμπεριφορά στις δύο διαστάσεις δεν επηρεαζόταν από το μέγεθος των ενταμιευτών εσωτερικά· αυτό που την έκανε τόσο σταθερή ήταν ότι το άθροισμα των βαρών των ενεργών ροών, σε κάθε είσοδο και κάθε έξοδο, ήταν το ίδιο και άρα κάθε ροή δικαιούτο ταυτόσημη εξυπηρέτηση τόσο στην είσοδο όσο και στην έξοδο, όποτε δεν εμφανίστηκε η ανάγκη επαναδιανομής αχρησιμοποίητης/excess χωρητικότητας. Το ίδιο αποτέλεσμα – στην αναλογική του έκδοση – έχουμε και όταν υπάρχουν διαφορετικά βάρη αλλά το άθροισμα αυτών, σε κάθε είσοδο και έξοδο, είναι το ίδιο.



Σχήμα 5.2: Average and Max discrepancy from weighted max-min fairness over all flows. Configuration2.

Στο σχήμα Σχ. 5.2 παρουσιάζουμε ένα τρέξιμο για ένα σύστημα με μέγεθος $8x8$, με το 1/4 των ροών ενενεργό, ομοιόμορφα κατανεμημένο στους πόρους. Αφήσαμε το σύστημα να τρέξει για 100000 κύκλους-δρομολόγησης προτού πάρουμε τα συνολικά/τελικά αποτελέσματα. Ο οριζόντιος άξονας είναι το μέγεθος των ενταμιευτών στα crosspoints και ο κάθετος οι συνολικές τιμές της μεταβλητής $PAWD$. Βλέπουμε ότι με χωρητικότητα ενταμίευσης σε κάθε crosspoint ίση με 3-4 μοναδιαία-πακέτα/cells, το σύστημα συγκλινεί και αριθμητικά πάρα πολύ κοντά στον ιδεατό στόχο. Η ροή που παρουσιάζει την μεγαλύτερη απόκλιση θα χάνει ή θα κερδίζει μία εξυπηρέτηση για κάθε $\simeq$ 10000 εξυπηρετήσεις που ιδανικά δικαιούτο, ενώ οι περισσότερες/¨πιθανότερες¨ ροές($median$), θα αποκλίνουν μία εξυπηρέτηση σε κάθε 100000 δίκαιες.

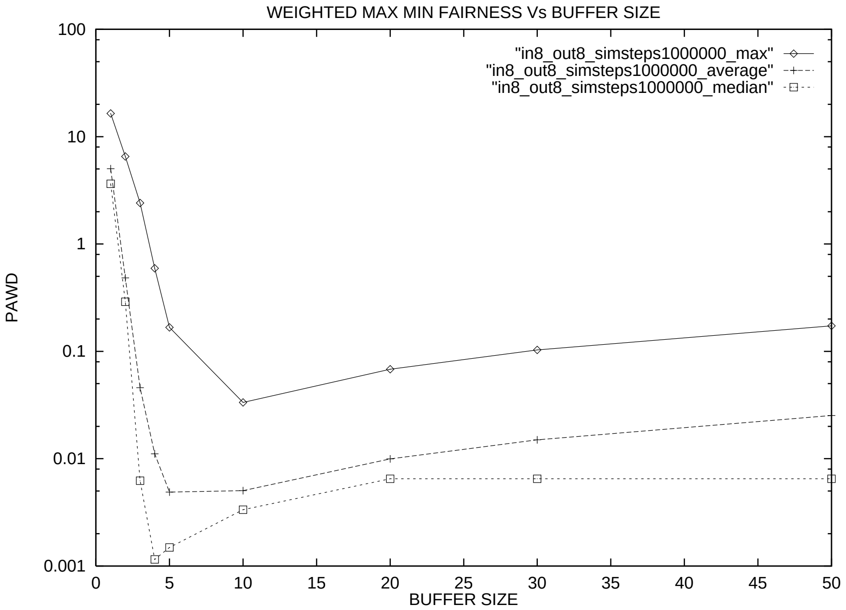


Σχήμα 5.3: Α εραγε ανδ Μαξ δισςρεπανςψ φρομ ωειγητεδ μαξ-μιν φαιρνεσσ ο ερ αλλ φλοωσ ο ερ τιμε. ὂνφιγυρατιον2. ῾ροσσποιντ βυφφερσ σιζε 4.

Στο σχήμα Σχ. 5.3 , παρουσιάζουμε την εξέλιξη των (συν)ολικών μετρικών $PAWD$ στο χρόνο προσομοίωσης. Το περιβάλλον και οι παράμετροι προσομοίωσης είναι ίδιες με αυτές στο Σχ. 5.2 , αλλά φυσικά το μέγεθος των buffer στα crosspoints είναι σταθερό και ίσο με 4 cells. Παρατηρούμε σχεδόν γραμμική μείωση της απόκλισης σε λογαριθμικούς άξονες, πράγμα που υπονοεί χρονική μειώση/εξάλειψη της αρχικής αστάθειας/αδικίας, ενώ η κατανομή των αποκλίσεων στις ροές ακολουθεί περιόδους μεταβάσεων από μορφή long tail σε μορφή πίδακα(impulse) κυκλικά – με γενικά φθίνουσες αποκλίσεις(ορισμένες με την όχι-memoryless μεταβλητή $PAWD$). Τέλος η αύξηση της απόκλισης σε πολύ μεγάλα μεγέθη ενταμιευτών, οφείλεται στο ότι δεν εξυπηρετούμε, όταν τελειώσει ο χρόνος προσομοίωσης, τα πακέτα που βρίσκονται στους εσωτερικούς ενταμιευτές· οφείλεται δηλαδή σε υπολογιστικό λάθος και όχι σε κάποια μυστηριώδη ιδιότητα του συστήματος.

Διαφορετικά βάρη

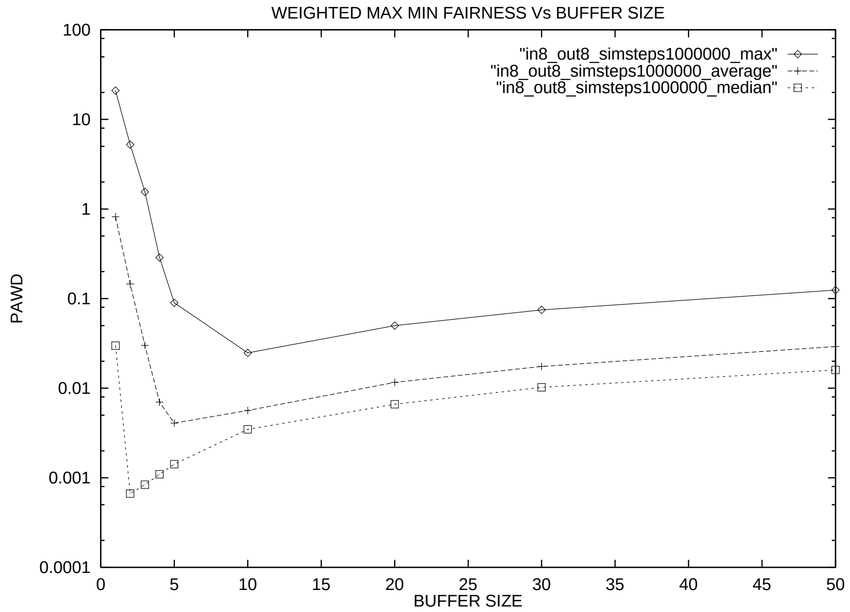
Εδώ διατηρούμε την ίδια ομοιόμορφη κατανομή για τις ανενεργές ροές αλλά αντιστοιχούμε σε αυτές διαφορετικά βάρη. Αρχικά χρησιμοποιήσαμε ομοιόμορφη



Σχήμα 5.4: Average and Max and Median discrepancy from weighted max-min fairness over all flows. Uniform distribution of weights. Configuration3.

κατανομή και για τα βάρη των ροών. Αν και σε αυτήν την περίπτωση περιμένουμε το άθροισμα-των-βαρών/ζήτηση/συμφόρηση σε κάθε είσοδο και έξοδο να είναι ομοιόμορφα κατανεμημένο σε όλες τις εισόδους και εξόδους, τελικά το Σχ. 5.4 δείχνει ότι και η παραμικρή απόκλιση από την ιδανική ομοιομορφία, δημιουργεί σημαντικές απαιτήσεις σε μνήμη. Στο Σχ. 5.4 κάθε ροή έχει ένα τυχαίο βάρος στην περιοχή $[1...1000]$. Και πάλι με το ίδιο περίπου μέγεθος ενταμιευτών, το σύστημα πετυχαίνει εξίσου καλές επιδόσεις σύγκλισης με πρίν – την περίπτωση ίδιων βαρών.

Μία ιδέα για να κάνουμε πιο δύσκολα λίγο τα πράγματα, είναι να δημιουργήσουμε αυτό που λένε hot spots, σημεία δηλαδή μεγάλης συμφόρησης – ιδιάζουσα κατανομή ζήτησης(irregular weighted-demand distribution). Τέτοια σημεία είναι γνωστό ότι υπάρχουν στο Ιντερνετ, και επιπλέον εισάγουν ακόμα μεγαλύτερες απαιτήσεις αναδιακατανομής παραπανήσιας χωρητικότητας. Οι ροές που χρησιμοποιούν ένα κοινό σημείο μεγάλης συμφόρησης(hot-spot) θα χρειάζεται να είναι δίκαιες στο μεταξύ τους ανταγωνισμό – όλες έχουν μεγάλο/παρόμοιο βάρος, οπότε όλες δικαιούνται A –, οπότε θα αφήνουν μέρος του τοπικά-δίκαιου μερίδιου τους($B > A$) στο συμπληρωματικό κόμβο που χρησιμοποιούν, για χρήση από τις άλλες γείτονες ροές. Ο συπληρωματικός κόμβος θα είναι λιγότερο συμφορημένος, αφού θα εξασφαλίσουμε ότι θα έχει γενικά ροές με σημαντικά μικρότερο βάρος.

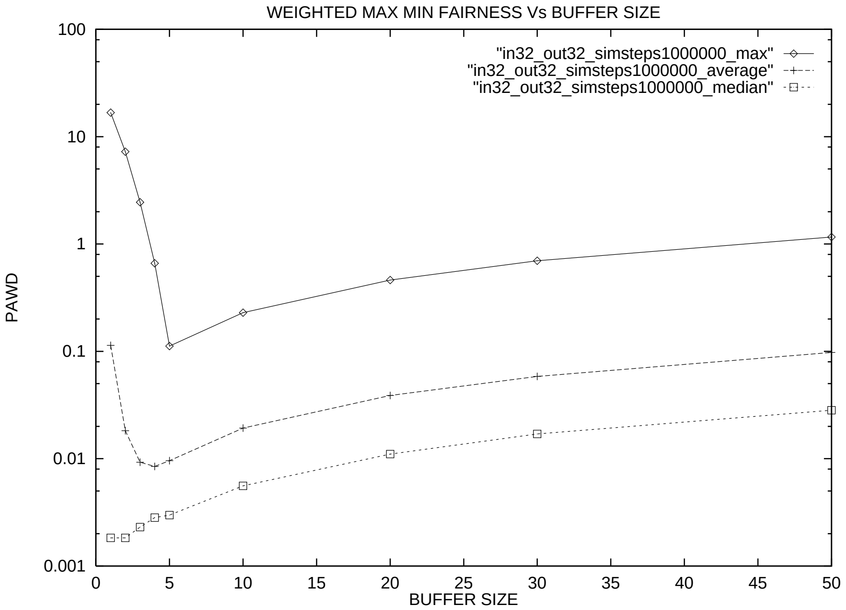


Σχήμα 5.5: Average and Max and Median discrepancy from weighted max-min fairness over all flows. Hot spots case. Configuration5.

Στο πείραμα το αποτέλεσμα του οποίου δείχνουμε στο Σχ. 5.5 , οι ροές έχουν ένα τυχαίο βάρος, ανάλογο του τετραγώνου, της σταθερά αριθμημένης εξόδου που χρησιμοποιούν. Και πάλι με περίπου τις ίδιες απαιτήσεις σε μνήμες, ο ιδεατός στόχος επιτεύχθηκε. Δείχνει ότι αν και προσπαθήσαμε, τελικά δεν επηρεάσαμε την συμπεριφορά του συστήματος.

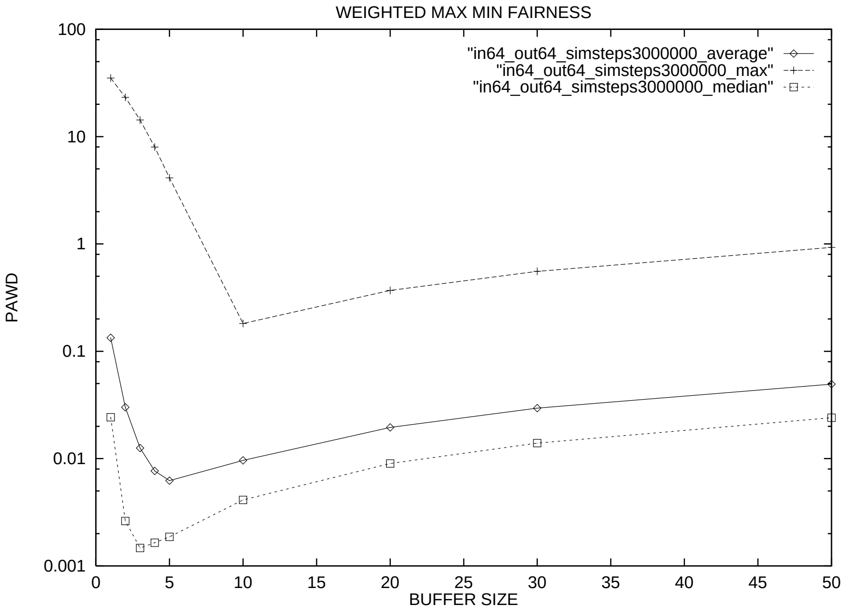
Μεγαλύτεροι μεταγωγείς — Κλιμάκωση(·) επιδόσεων — Αυξημένες(·) απαιτήσεις.

Τα αποτέλεσματα στα Σχ. 5.6 και Σχ. 5.7 , αντιστοιχούν σε μεταγωγείς 32x32 και 64x64 αντίστοιχα. Οι παράμετροι είναι ανάλογοι με αυτές που παρουσιάσαμε στην παράγραφο με την ιδιάζουσα κατανομή βαρών(Hot Spot), μόνο που τα πειράματα έτρεξαν για περισσοτερο χρόνο για λόγους ευστάθειας των αποτελεσμάτων. Μπορούμε να δούμε ότι το σύστημα και πάλι προσεγγίζει το τέλειο αλλά κάτω από αυξημένες απαιτήσεις σε εσωτερικές μνήμες. Για το τρέξιμο στο Σχ. 5.6 (32x32) 4-5 δείχνουν να είναι η βέλτιστη λύση, ενώ για το τρέξιμο στο Σχ. 5.7 (64x64) χρειάζονται τουλάχιστον 10 cells χωρητικότητα σε κάθε crosspoints, για



Σχήμα 5.6: Average and Max and Median discrepancy from weighted max-min fairness over all flows for a 32x32 switch. Hot-spot(irregular) weights. Configuration6.

να φράξουμε αρκετά την πιο αποκλίνουσα ροή. Ακόμα όμως και με μικρότερες προυποθέσεις μνήμης, η μέση απόκλιση των ροών είναι ικανοποιητική. Πάντως δεν εμφανίστηκαν – ή δεν αναγνωρίσαμε – σε αυτά και σε όσα άλλα τρεξίματα δοκιμάσαμε, τίς αναλογικές με το μέγεθος του διακόπτη απαιτήσεις σε μνήμη που ισχυρίζεται η ανάλυση στο [BufCrossbar] ότι εγγυώνται την βέλτιστη, ώς προς την οπτική που εξετάζουμε εδώ, συμπεριφορά.



Σχήμα 5.7: A hot-spot weights assignment, for a 64x64 switch.

5.3 Χρόνος/περιγραφή (επανα)σύγκλισης συστήματος — Αδράνεια μετάβασης(Βαθμός Επιρροής) σε(των) στοιχειώδεις(ων) αλλαγές(ων).

Μέχρι στιγμής ελέγξαμε πειραματικά την υπόθεση μακροπρόθεσμης σύγκλισης του συστήματος, στο στόχο δικαιοσύνης που μας ενδιαφέρει. Φαίνεται ότι το σύστημα όντως συγκλίνει – ή τουλάχιστον έχει τα δομικά στοιχεία και την κατάλληλη οργάνωση, που επιτρέπουν την σύγκλιση αυτή κάτω από ένα σύνολο εξωτερικών παραμέτρων. Αν μάλιστα δεχτούμε και το αναλυτικό αποτέλεσμα που επιβεβαιώνει την ιδιότητα αυτή, τότε πρέπει να δεχτούμε ότι αν χρησιμοποιήσουμε ικανοποιητική ενταμίευση και αν περιμένουμε για ένα x χρονικό διάστημα που οι ροές διατηρούν την κατάσταση ζήτησης τους, τότε τελικά θα τους αποδοθούν δίκαια ποσοστά χωρητικότητας/παροχής. Εδώ θα δεχτούμε αυτήν την μισοεπιβεβαιωμένη υπόθεση και θα προχωρήσουμε την ανάλυση που παρουσιάσαμε στο Κεφ. 4.2.2 με πειραματική επαλήθευση/επιβεβαίωση των εκεί αποτελεσμάτων. Θα προσπαθήσουμε να παρατηρήσουμε δηλαδή το χρόνο σύγκλισης, και το χρόνο εύρεσης της νέας ισορροπίας, ύστερα από μικρές αλλαγές στην κατάσταση των ροών(αλλαγή βάρους ή/και ζήτησης-υπηρεσίας). Σημαντικό είναι γενικά το γεγονός ότι η αύξηση της χωρητικότητας της εσωτερικής μνήμης, αν και παρουσιάζεται να διευκολύνει την σύγκλιση μεγάλης ακριβείας, επιπλέον σίγουρα την καθυστερεί, αφού μεγάλοι ενταμιευτές αλλάζουν πιο αργά κατάσταση και φυσικά αυξάνουν την αδράνεια του συστήματος.

5.3.1 Περιγραφή περιβάλλοντος πειράματος — Ειδικές μεθόδοι

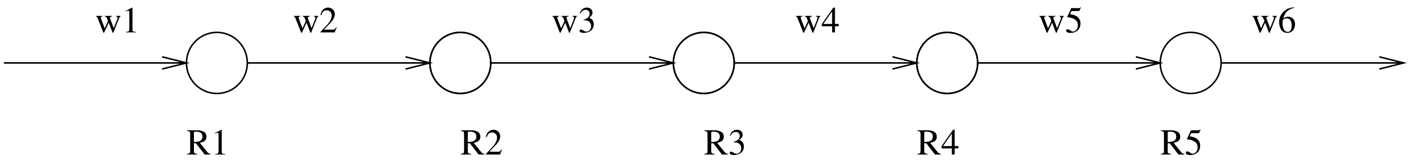
Η μεταβλητή $PAWD$ ήταν κατάλληλη για να μελετήσουμε την υπόθεση της σύγκλισης στην κατάσταση ισορροπίας. Δεν μας φανερώνει όμως το πραγματικό χρόνο – τη στιγμή – σύγκλισης, αφού κουβαλάει/θυμάται συνεχώς τις αρχικές/μεταβατικές/"άδικες" φάσεις-κατάστασης του συστήματος. Εδώ δείχνει να θέλουμε μεταβλητές όπως την χρονική συνάρτηση στιγμιαίας παροχής(marginal service, έτσι ώστε να μπορούμε να πούμε "<u>τώρα</u> έχουμε δικαιοσύνη, ή <u>τώρα</u> δεν έχουμε¨. Δυστυχώς η συνάρτηση αυτή δεν είναι καλώς ορισμένη σε ένα σύστημα πακέτων διακριτού χρόνου και παρόλες τις προσπάθειες που κάναμε(βλ. [BufCrossbar]), δεν βρήκαμε ένα ενιαίο-κατάλληλο σχήμα έκφρασης αυτής της μεταβλητής, κοινό για όλες τις ροές και για όλα τα τρεξίματα – γεγονός που δυσκόλευε την εξαγωγή και ερμηνεία των αποτελεσμάτων.

Ευτυχώς, βρέθηκε ένας απλός τρόπος, κατάλληλος για τους σκοπούς των πειραμάτων. Αυτός δεν είναι άλλος από τη μέτρηση της διαφορά της συνολικής εξυπηρέτησης κάτω από το ιδανικό-μαθηματικό μοντέλο δικαιοσύνης και της εξυπηρέτησης που προσφέρει το σύστημα μας. Όταν αυτή η καμπύλη, για μία συγκεκριμένη ροή, κάτω από ένα συγκεκριμένο περιβάλλον προσομοίωσης, γίνεται οριζόντια, τότε προφανώς η παράγωγος της μηδενίζεται, οπότε η παροχή του συστήματος σε αυτήν την ροή ισούται με την παροχή που δικαιούται και μαθηματικά, ή ισοδύναμα στο μοντέλο των ρευστών.

Τώρα η γενική μέθοδος εξαγωγής των αποτελεσμάτων είναι η αποτύπωση αυτής της μεταβλητής και η παρατήρηση της χρονικής στιγμής που αυτή γίνεται περίπου οριζόντια – άν γίνεται καθόλου. Εφαρμόζοντας αυτήν την τεχνική για να παρατηρούμε τις ροές, είτε ξεκινάμε το σύστημα ψάχνοντας το χρόνο σύγκλισης, είτε αφήνουμε το σύστημα να σταθεροποιηθεί, δημιουργούμε μία μικρή αλλάγη στην κατάσταση μίας(ή περισσοτέρων) ροών και ελέγχουμε με τον ίδιο τρόπο το χρόνο σταθεροποίησης στην νέα ιδεατή/δίκαια κατανομή.

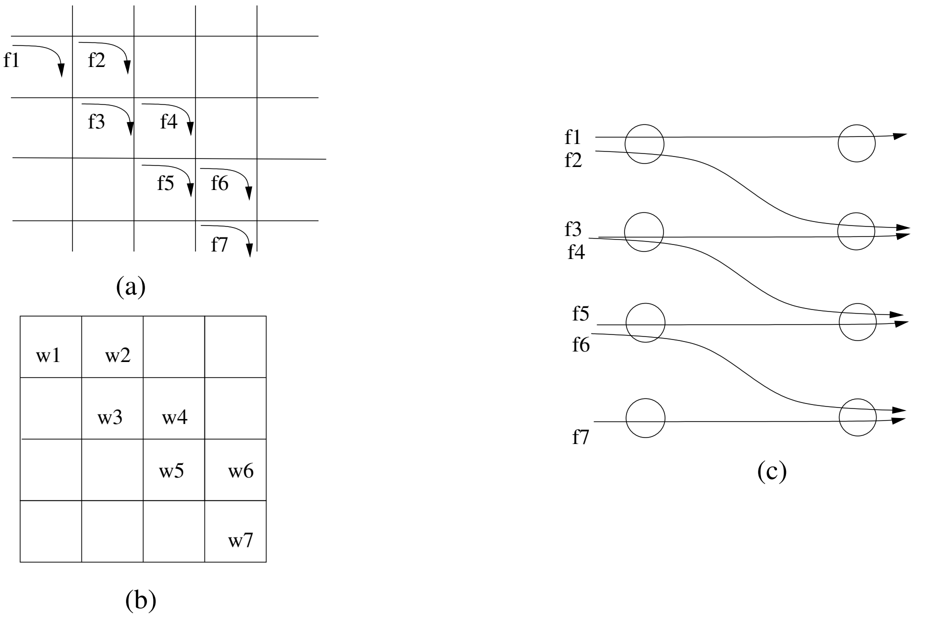
Εκτίμηση βαθμού Επιρροής Στοιχειοδών Αλλαγών — Flows trajectories under equilibirum — και το βασικό σχήμα χειρίστης περιπτώσεως.

Στο Κεφ. 4.2.2 , δώσαμε κάποιες αποδείξεις για το βαθμό επιρροής μίας ροής κάτω απο μία συγκεκριμένη κατάσταση ισορροπίας —- μία περιγραφή του συνόλου των ροών που θα επηρεαστούν ύστερα από μία αλλαγή στην κατάσταση αυτής της ροής – διαφορετικά αν η ροή άλλαξε παροχή, τότε ποιες άλλες σίγουρα δεν άλλαξαν. Το βασικό θεώρημα εκεί ισχυρίζεται, ότι μία ροή g βρίσκεται υπό την επιρροή της f στην κατάσταση ισορροπίας A, μόνο αν μπορούμε να προσεγγίσουμε την g ξεκινώντας από την f ή ένα γείτονα της f, ακολουθώντας ένα μονοπάτι από γειτονικές/αλληλεπιδρώμενες ροές με αυξανόμενο utility – μειωμένο βαθμό συμφόρησης –, στην ισορροπία A. Επιπλέον – αυτό το απαιτούν τα λήμματα που ολοκληρώνουν την απόδειξη στο [BufCrossbar], αλλά δεν γράφεται στο θεώρημα το ίδιο –, επειδή η έκφραση της επιρροής σε αυτό το μονοπάτι ακολουθεί ζευγάρια από θετικά και αρνητικά επηρεασμένες ροές – αυτά τα ζευγάρια αποτελούν τα subequilibria –, και κάθε επιρροή θετικού τύπου – αύξηση παροχής – απαιτεί η ροή που την αποδέχτηκε να αντιμετώπιζε(στην αρχική ισορροπία) ώς κόμβο συμφόρησης(bottleneck) τον κοινό της κόμβο με την ροή-αιτία αυτής της επιρροής/γεγονότος, χρειάζονται κάποιες επιπλέον συνθήκες που να εξασφαλίζουν αυτήν την ιδιότητα – θα φανεί παρακάτω. Εδώ θα απομονώσουμε ένα τέτοιο μονοπάτι, χωρίς να χάσουμε την γενικότητα σύμφωνα με τα θεωρήματα, και θα μελετήσουμε την βαθμό επιρροής μίας ροής.



Σχήμα 5.8: Five resources are being shared by six flows. Each resource is being utilized by each these flows "touching" them. $w1 \geq w2 \geq w3 \geq w4 \geq w5 \geq w6$.

Το βασικό σχήμα που χρησιμοποιούμε φαίνεται στο Σχ. 5.8 . Στην κατάσταση ισορροπίας, ή ροή ένα έχει την μικρότερη utility, ενώ η ποσότητα αυτή αυξάνεται όσο προχωράμε στην αλυσίδα προς ροές με μεγαλύτερο δείκτη – πιο μακριά από την ένα. Η ιδέα είναι ότι παρούσης της ροής ένα, η ροή δύο έχει μειωμένη παροχή/παρουσία την οποία εκμεταλλεύεται η ροή τρία. Για να μπορεί όμως η τρία να χρησιμοποιήσει όλο το excess που αφήνει η δύο, πρέπει να ισχύει η παρακάτω σχέση ανάμεσα στις αναλογίες βαρών:$\frac{w_1}{w_2} \leq \frac{w_3}{w_4}$. Διαφορετικά κάποιοι πόροι θα υποχρησιμοποιούνται. Αν $\frac{w_i}{w_{i+1}} = constant$, τότε αυτή η σχέση ικανοποιείται. Δημιουργήσαμε/κατασκευάσαμε λοιπόν μία αλυσίδα εξαρτήσεων, που σύμφωνα με το θεώρημα στο Κεφ. 4.2.2 , πραγματώνει το χειρότερο – από άποψη σειριακής καθυστέρησης – σενάριο (επανά)σύγκλισης. Αυτό μπορεί να αντιστοιχιθεί στη τοπολογία του crossbar σε ένα σύνολο από ροές $i \in [1 \ldots 2 \cdot N]$, που χρησιμοποιούν την είσοδο $i/2 \in [0 \ldots N-1]$ και την έξοδο $i/2 + i$ μοδ $2 \in [0 \ldots N-1]$, όπως παρουσιάζεται στο Σχ. 5.9 .



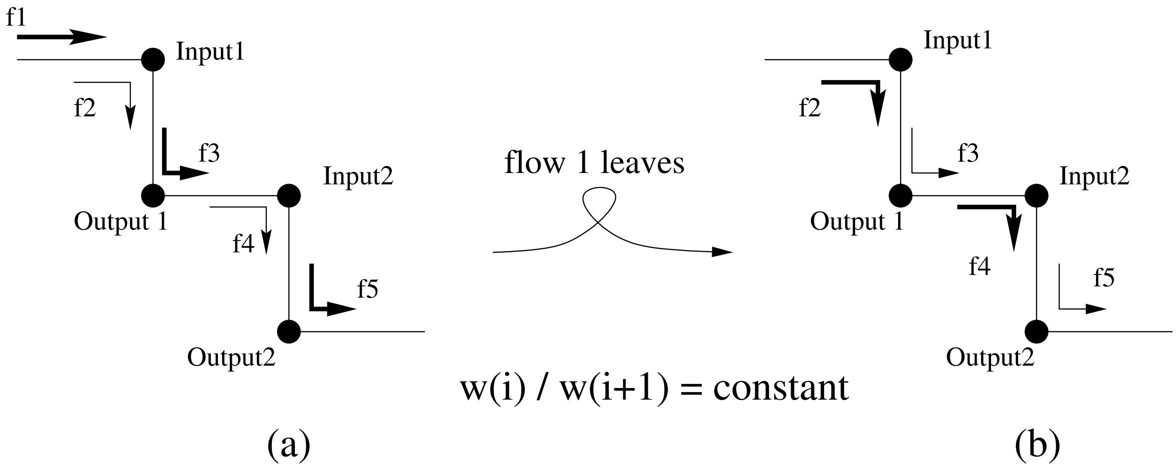
Σχήμα 5.9: One possible map of Sq. 5.8 , into the crossbar topology(a). The weights configuration(b) and the bipartite graph of resources-requests(c)

Εκτίμηση του χρόνου εγκαθίδρυσης ισορροπίας.

Η εκτίμηση που πήραμε για το χρόνο σύγκλισης μίας ροής στο Κεφ. 4.2.2 , ήταν $O(N \cdot B/min(old\_rate - new\_rate))$.

Στο αποτέλεσμα που θα παρουσιάσουμε πρώτο, αλλάζουμε την κατάσταση της πρώτης ροής και παρατηρούμε την μεταβλητή δίκαιης μείον μετρήσιμης εξυπηρέτησης – του συστήματος –, που ορίσαμε στην προηγούμενη παράγραφο. Ο μεταγωγέας του πειράματος έχει μέγεθος 8x8 και εσωτερικούς ενταμιεύτες χωρητικότητας 4 cells σε κάθε crosspoint, ενώ η αναλογία μεταξύ των βαρών των ροών είναι 2. Το αναμενόμενο φαινόμενο/αποτέλεσμα της διαδικασίας μετάβασης, παρουσιάζεται γραφικά στο Σχ. 5.10 καί αντιστοιχεί βασικά σε επέκταση της περιγραφής που δώσαμε προηγουμένως.

Η ροή 2 ¨εκμεταλλεύεται¨ την απουσία της 1 – μπορεί να χρησιμοποιήσει μεγάλο μέρος της παροχής της 1, λόγω της αναλογίας του βάρους της σε σχέση με τη 3 –, η τρία θα σταματήσει να εκμεταλλεύεται την απουσία της 2 και θα μειώσει την παροχή της, η οποία αλλαγή-μείωσης θα προωθηθεί στην 4 μόνο όταν αδειάσουν οι εσωτερικοί ενταμιευτές της(3) – αυτό είναι ένα φαινόμενο-καθυστέρησης που δεν εμφανίζεται στο μοντέλο ρευστών με μηδενικούς/απειροελάχιστους εσω-

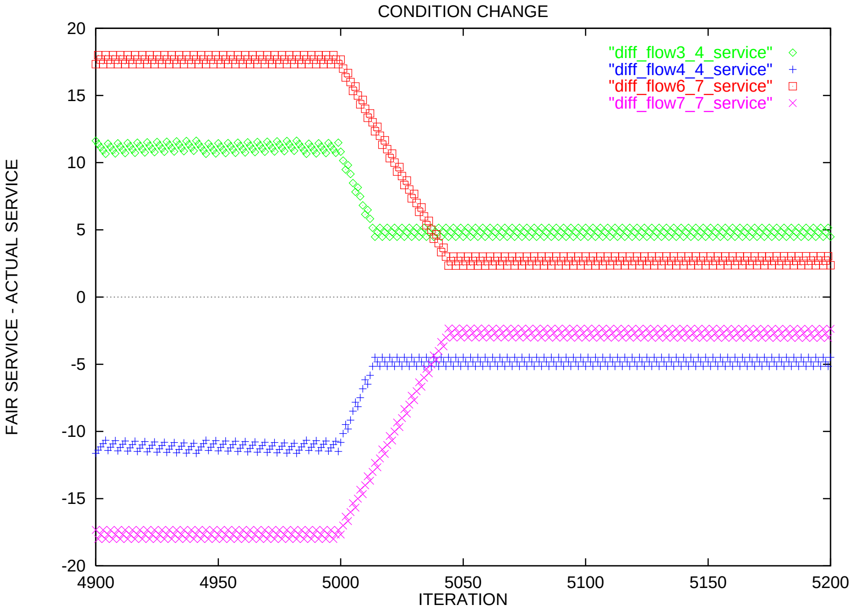


Σχήμα 5.10: Five resources are being shared by six flows. Each resource is being utilized by these flows "touching" it. $w1 \geq w2 \geq w3 \geq w4 \geq w5 \geq w6$.

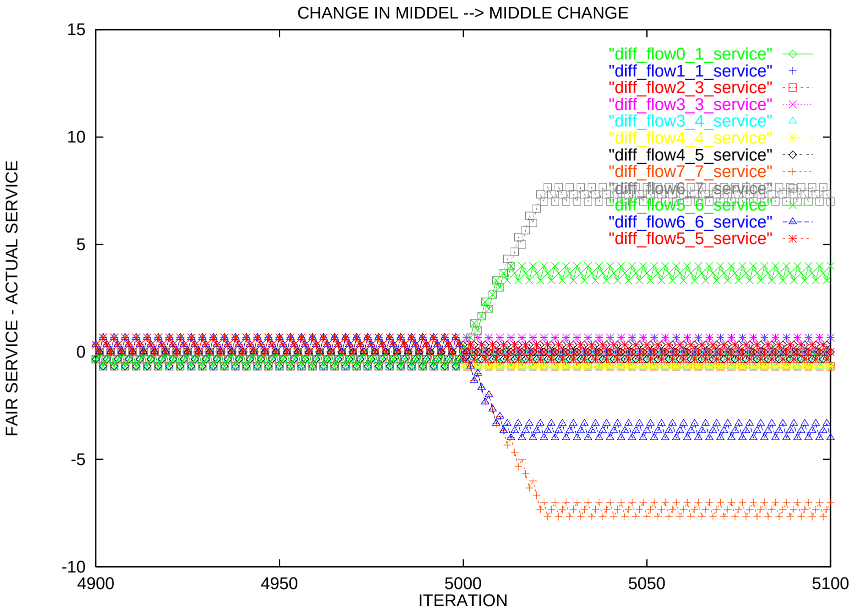
τερικούς buffers –, ή οποία 4 θα αυξήσει την παροχή της και στους δύο κόμβους που χρησιμοποιεί. Όταν γίνει αυτό, αμέσως η 5 θα μειώσει την την παροχή της στο κοινό της κόμβο με την 4, αλλά και πάλι θα μεσολαβήσει χρόνος μέχρι να μεταφερθεί το γεγονός-μείωσης στον συμπληρωματικό κόμβο που χρησιμοποιεί. Αυτή η αλύσωση/αλληλουχία γεγονότων περιγράφει τον τρόπο που το κύμα της αλλαγής μετακινείται μέσα στο δίκτυο.

Στο Σχ. 5.11 βλέπουμε ότι ροές σε μεγαλύτερη απόσταση από το αρχικό γεγονός, χρειάζονται περίπου διπλάσιο χρόνο για να βρουν τη θέση ισορροπίας τους. Επίσης βλέπουμε ότι οι ροές αλλάζουν μαζί σε ζευγάρια γειτονικών ροών(α/β) – θετικά/αρνητικά επηρεασμένες – που ορίζουν/αντιστοιχούν σε προσωρινά/τοπικά/sub-equilibria. Πρέπει να παρατηρήσουμε ότι ο τύπος για το χρόνο σύγκλισης έχει πολύ μεγάλη ακρίβεια(2 time_slot) σε αυτό το παράδειγμα, αν σωστά υπολογίσουμε τις ροές f που μειώνεται η παροχή τους και άρα καθυστερούν την μεταβίβαση/προώθηση της αλλαγής για χρόνο $\frac{buffer\_occupancy_f}{old\_rate_f - new\_rate_f}$.

Για λόγους επαλήθευσης, στο Σχ. 5.12 αλλάζουμε την κατάσταση μίας ροής στη μέση ενός παρόμοιου μονοπατιού και παρατηρούμε ποιές ροές επηρεάστηκαν. Από ό,τι βλέπουμε μόνο ροές πιο μπροστά στο μονοπάτι επηρεάστηκαν, όπως περιμέναμε. Άρα και πειραματικά επαληθεύσαμε ότι δεν μπορούμε να ¨-μεταθέσουμε¨/αποδώσουμε την αλλαγή μίας ροής, στην αλλαγή μίας ροής με μικρότερο βαθμό συμφόρησης.

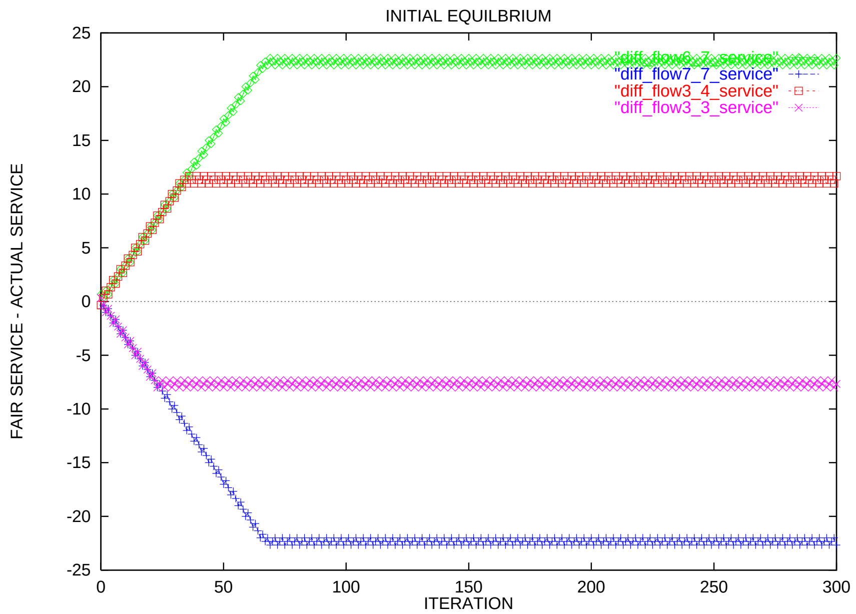


Σχήμα 5.11: This plot present the results for the scenario in Sq. 5.10 , for 8x8 switch with crosspoint buffers size 4 and constant analogy $\frac{w_i}{w_{i+1}} = 1/2$.



Σχήμα 5.12: In this plot we have a scenario similar to Sq. 5.10 but with event difference that the flow that uses input 4 and output 4 becomes inactive(instead of flow 1), affecting only flows with lower congestion index.

Επίδραση μεγέθους χειρίστου μονοπατιού.

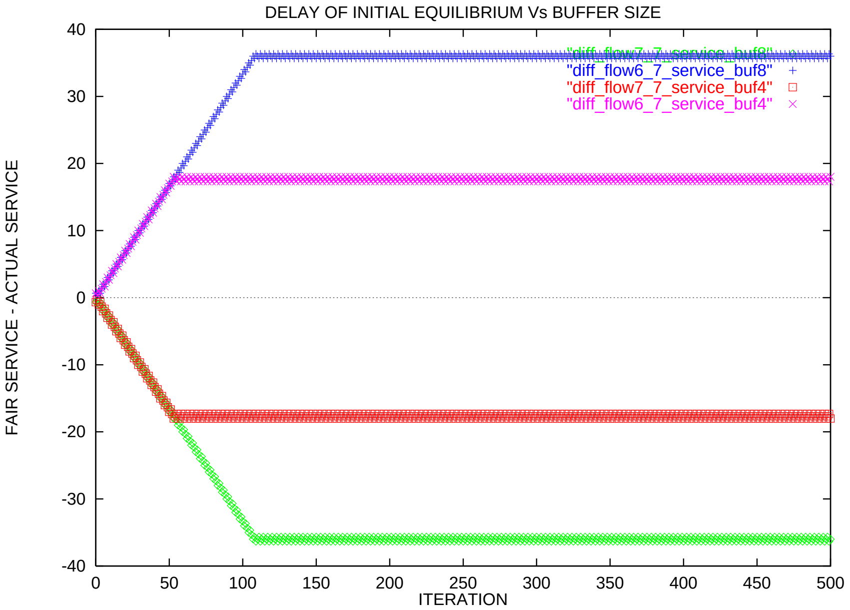


Σχήμα 5.13: The evolution of weighted max-min fairness through time 8x8 switch. Flows have have the same configuration like in Sq. 5.8 (except that flow 0_0(flow 1) is inactive) and in addition $\forall \quad i \quad (w_i/w_{i+1} = 2)$.

Στο Σχ. 5.13 , βλέπουμε πιο καθαρά την αναλογική επίδραση της απόστασης του μονοπατιού αλληλεπίδρασης στο χρόνο σταθεροποίησης, αλλά και στην αδικία(μετρούμενη σε cells), τών ροών. Βασικά το σχήμα αναφέρεται στο χρόνο σύγκλισης από την αρχή και όχι σε κάποιο γεγονός αλλάγής κατάστασης, αλλά επειδή κάθε ροή θα πρέπει να περιμένει να σταθεροποιηθούν οι ροές μικρότερης utility πριν ψάξει στα σίγουρα το δίκαιο ποσοστό της, το φαινόμενο είναι ακριβώς ανάλογο.

Επίδραση μεγέθους ενταμιευτών.

Στο Σχ. 5.14 , βλέπουμε καθαρά την αναλογική επίδραση του μεγέθους των ενταμιευτών στη ταχύτητα σύγκλισης, τόσο σε όρους χρόνου όσο και ποσότητας αδικίας. Βασικά, αν και δεν το έχουμε αναφέρει ρητά εδώ – φαίνεται στο [BufCrossbar] στην απόδειξη σύγκλισης αλλά και από λεπτομερή παρατήρηση των προσομοιώσεων –, στην κατάσταση ισορροπίας οι ροές είτε χρησιμοποιούν ολό-



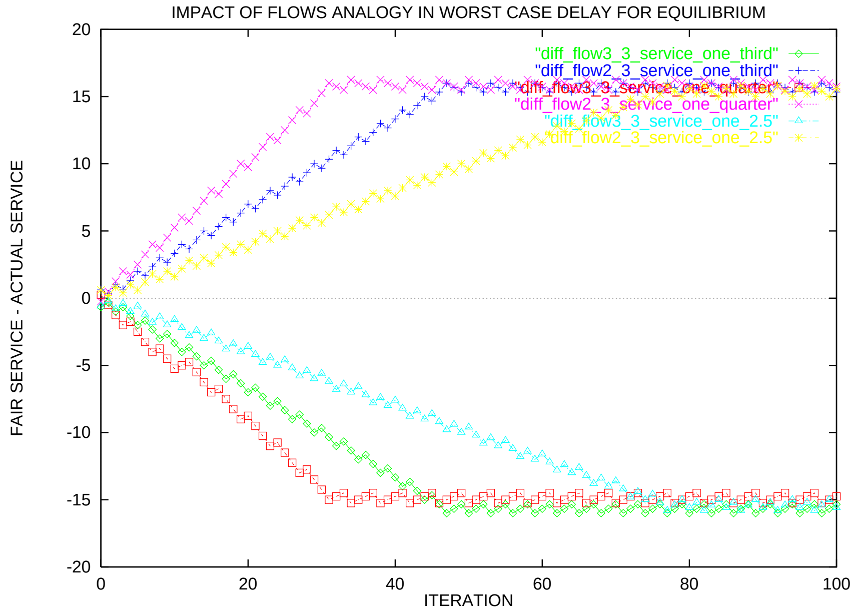
Σχήμα 5.14: The impact of the buffers size in the delay of fairness enforcement.

κληρη τη χωρητικότητα των ενταμιευτών είτε ελάχιστη, αναλόγως του αν συναντάνε bottleneck στο κόμβο εξόδου ή εισόδου αντίστοιχα. Οπότε προκύπτει και η αναλογική επίδραση αυτού του μεγέθους. Αυτό είναι ουσιαστικά το ¨κακό¨/αρνητικό των μεγάλων ενταμιευτών στην συμπεριφορά του μεταγωγέα. Η προτεινόμενη μέθοδος στο πρώτο – και μοναδικό – παράρτημα, φαίνεται να μπορεί να το αντιμετωπίσει και αυτό.

Επίδραση παλιάς αναλογίας βαρών.

Στο Σχ. 5.15 φαίνεται η επίδραση της παλιάς κατάστασης στο χρόνο επιβολής της νέας. Ουσιαστικά αλλάζουμε την αναλογία μεταξύ των βαρών και παρατηρούμε ροές σε ίδια απόσταση από τον τόπο-γεγονός της αρχικής αλλαγής. Γενικά η εκτίμηση του χρόνου σύγκλισης είναι N – ίσος με τον αριθμό των ροών που μειώνουν την παροχή –, επί B – ίσο με την αλλαγή χωρητικότητας σε μνήμη αυτών των ροών –, διά τον ρυθμό αυτής της αλλαγής[2]. Βλέπουμε επίσης ότι ο αριθμός των άδικων bits που εξυπηρετήθηκαν δεν εξαρτάται από αυτήν την αναλογία, αφού ο ρυθμός της άδικης εξυπηρέτησης είναι αντιστρόφως ανάλογος του χρόνου που

[2] Ο ρυθμός αυτός είναι στο παράδειγμα μας: $|r\_in(=\frac{w_i}{w_i+w_{i-1}}) - r\_out(=\frac{w_i}{w_i+w_{i+1}})|$. Γενικά με αναλογία c ο ρυθμός αλλαγής κατάστασης είναι: $\frac{2\cdot c}{(c+1)\cdot(c+2)}$.



Σχήμα 5.15: The impact of the analogy of weights to maximum delay.

αυτή διαρκεί.

Συνολικά

Είδαμε και πειραματικά τις παραμέτρους που επηρεάζουν το χρόνο σύγκλισης. Βασικά αυτός εξαρτάται από την ύπαρξη τέτοιων μονοπατιών αλληλεπιδράσεων. Αν υπάρχουν πολλά τέτοια, το μεγαλύτερο/χειρότερο από αυτά θα ορίσει την διάρκεια και την ποσότητα της αδικίας. Επίσης είδαμε ότι όσο πιο μακριά είναι το σύστημα από την κατάσταση ισορροπίας, τόσο πιο βιαστικό θα είναι να την βρεί ξανά(βλ. προηγούμενη παράγραφο). Αυτό θυμίζει φυσικά-συστήματα-ελατηρίων, και είναι μία απρόσμενη, θετική ιδιότητα για το σύστημα και της δίκαιης αρχής δρομολόγησης που αυτό συνολικά ¨θεσμοθετεί¨.

5.4 (Υπό)Λοιπα Πειράματα.

Για λόγους χρόνου/-χώρου δεν συμπεραλάβαμε σε αυτήν την αναφορά τα υπόλοιπα πειράματα που κάναμε κατά τη διάρκεια – και στα πλαίσια – της εργασίας. Αυτά ουσιαστικά ασχολούνται κυρίως με την προσφερόμενη παροχή του συστήματος κάτω από πιθανοκρατικές αφίξεις και με την χρησιμοποίηση διαφορετικών δρομολογητών στα σημεία συναγωνισμού. Εκεί μελετάμε αν προσαρμοστικές μέθοδοι δρομολόγησης στις εισόδους, μπορούν να προσφέρουν καλύτερη συνολική παροχή – διατηρώντας κατά την προώθηση, πάνω κάτω την σειρά των αφίξεων – και το tradeoff ανάμεσα σε πολυπλοκότητα και απόδοσης, κάτω από διαφορετικές θεωρήσεις για την τελευταία, χρησιμοποιώντας διαφορετικούς δρομολογητές που (καθ)ορίζουν την πρώτη(= πολυπλοκότητα). Επίσης δημιουργήθηκαν γραφήματα με πιο λεπτομερείς παρατηρήσεις των προσομοιώσεων, άλλοτε για λόγους επεξήγησης ή επαλήθευσης υποθέσεων, άλλοτε για να δείξουν/παρουσιάσουν την συλλογιστική επιχειρημάτων και άλλοτε για "καλλωπιστικούς" λόγους. Τέλος υπάρχουν προσομοιώσεις δύο "ακριβών" τεχνικών δρομολόγησης/ταιριάσματος για crossbar χωρίς εσωτερικούς ενταμιευτές, που εξετάζουν ταυτόχρονα την απαίτηση για μεγάλη συνολική παροχή και ποιότητα υπηρεσίας. Τα αποτελέσματα αυτά, μπορείτε να τα βρείτε στο [BufCrossbar].

Κεφάλαιο 6

Συζητήσεις

6.1 Φαινόμενο/υπόθεση σύγκλισης

6.1.1 Σύγκλιση

Γενικά τρέξαμε μερικές εκατονταδες πειραμάτα και δεν βρήκαμε κανένα αντιπαράδειγμα, που να αντικρούεται του επιχειρήματος σύγκλισης. Μάλιστα και οι απαιτήσεις σε μνήμες δεν μας έδειξαν να εξαρτώνται από τη συγκεκριμένη κατανομή των βαρών. Αν δεν υπάρχει μάλιστα ανάγκη αναδιανομής και οι επιμέρουςτοπικές καταστάσεις ισορροπίας(sub-equilibria, ορίζουν/ταυτίζονται με το συνολικό/επιθυμητό, τότε και με αξία εσωτερικών ενταμιευτών ένα cell, πετυχαίνουμε το σκοπό μας. Αυτό συμφωνεί με την ανάλυση του Chiussi([ChQoS]), εφαρμοσμένη στο σύστημα μας(βλέπε [BufCrossbar] section Providing the minimum fair share.). Σύμφωνα με τις προσομοιώσεις, μόνο από το μέγεθος του διακόπτη δείχνουν να εξαρτώνται οι επιδόσεις του συστήματος και πάλι ούτε κάν γραμμικά – όπως ισχυρίζεται/χρησιμοποιεί η ανάλυση στο [BufCrossbar] για να αποτρέψει/αποφύγει την χειρότερη περίπτωση με εύκολα-αναλύσιμο τρόπο –, αλλά κάτι σίγουρα μικρότερης τάξης· θέλει περαιτέρω πειραματική έρευνα και ίσως ανάλυση μέσης περίπτωσης ή καλύτερη ανάλυση της χειρίστης. Φυσικά πρέπει να θυμόμαστε ότι τόσο τα πειράματα, όσο και η ανάλυση βασίστηκαν σε ροές σταθερής κατάστασης(on/off).

6.1.2 Μεταβατικά φαινόμενα — Χρόνος σύγκλισης

Η καταστατική-μηχανιστική περιγραφή του φαινομένου που παραθέσαμε στο Κεφ. 4.2.2 φαίνεται να επιβεβαιώθηκε από τα σχετικά πειράματα. Φαίνεται ότι δεδομένων των παραμέτρων, μπορούμε να φράξουμε το χρόνο – και την ¨αδικία˝ – των μεταβάσεων, αν κάνουμε κάποιες υποθέσεις για την φύση αυτών· η σταθερή κατάσταση των ροών πριν και μετά από ένα γεγονός, είναι μία τέτοια απλουστευτική υπόθεση.

Το θετικό είναι ότι το σύστημα συγκλίνει αρκετά γρήγορα και ανάλογα με το μεγεθος του συστήματος – γενικά μεγάλο σύστημα σημαίνει αυξημένη πιθανότητα-δυνατότητας ύπαρξης μεγάλων μονοπατιών αλληλεπίδρασης –, του πλήθουςς των ροών, όσο και την ¨πολυπλόκοτητα˝ της αλληλεπίδρασης αυτών – πόσο απέχουν τα άμεσα τοπικά/¨προσωρινά˝ equilibria από το συνολικά επιθυμιτό. Στη συνηθισμένη περίπτωση πάντως, δεν περιμένουμε να υπάρχουν μεγάλα μονοπάτια επιρροής, όπως αυτά τα περιγράφει το θεώρημα στο Κεφ. 4.2.2 – τα χαρακτηριστικά προσδιορισμού της χειρότερης περίπτωσης είναι αρκετά ακραία. Ταυτόχρονα είδαμε ότι όσο πιο άδικο εμφανίζεται το σύστημα σε μία χρονική στιγμή, τόσο πιο πολύ προσπαθεί να γίνει γρήγορα δίκαιο. Και πάλι δεδομένου σταθερού πλήθους ροών, όσο πιο μεγάλο το σύστημα, τόσο πιο διάσπαρτες θα είναι αυτές, άρα τόσο θα αυξάνει η παραλληλία($\rightarrow$ ταχύτητα) στην εύρεση της θέσης ισορροπίας. Το μέγεθος των εσωτερικών ενταμιευτών που επηρεάζει αρνητικά την ταχύτητα σύγκλισης, με βάση τα αποτελέσματα στις προυποθέσεις σύγκλισης ακριβείας, δεν χρείαζεται να είναι και τόσο μεγάλο.

Κεφάλαιο 7

Συμπεράσματα/συνεισφορά μελλοντική δουλειά

Συμπεράσματα

Το σύστημα που περιγράψαμε παρουσιάζει πολύ καλα χαρακτηριστικά. Πρώτον και ίσως σημαντικότερο, έχει λογικό/προσιτό κόστος αφού έχει τις κύριες μνήμες αποθήκευσης στις εισόδους και ταυτόχρονα υπερπηδά το πρόβλημα δρομολόγησης της εσωτερικής τοπολογίας χρησιμοποιώντας τεχνικές που γνωρίζουμε να κατασκευάζουμε(διανεμητές ενός κοινού πόρου). Το προσθετικό κόστος των εσωτερικών ενταμιευτών, δείχνει να μην είναι τραγικό δεδομένης της τωρινής αύξησης σε χωρητικότητα on chip και μπορούν εύκολα να κατασκευαστούν στο ίδιο chip με το crosspoint, μαζί μάλιστα και με τα – και πάλι εύκολα – κυκλώματα προσπέλασης τους. Το σύστημα επιτρέπει τη χρήση/επεξεργασία πακέτων μεταβλητού μεγέθους εσωτερικά στο σύστημα – αποτέλεσμα της ανεξαρτησίας των δρομολογητών –, το οποίο είναι θετικό αφού μειώνει το κόστος διαμερισμού των πακέτων στις εισόδους και επανασύνθεσης τους στις εξόδους, εξαλείφει την ανάγκη για εξωτερικούς ενταμιευτές που θα επιτρέπουν την τελευταία διαδικασία, ενώ μειώνει και τις απαιτήσεις για εσωτερική επιτάγχυνση(βλ. [DGPS]).

Παράλληλα η παραγόμενη απο την αρχιτεκτονική, αρχή δρομολόγησης, είναι πολύ αποδοτική. Πλησιάζει τον αντικειμενικό στόχο για κατανομή μεγίστου-σταθμισμένου-ελαχίστου, που από την οπτική δίκαιης-αποτελεσματικότητας είναι ό,τι καλύτερο μπορούμε να φανταστούμε για ένα σύστημα με πολλά σημεία ανταγωνισμού, όπως είναι ένας πολυ-διακόπτης με τις ουρές αποθήκευσης στις εισόδους. Δεύτερον αποδέχεται αναλυτική ερμηνεία αυτής της συμπεριφοράς, γεγονός που δείχνει ότι δεν είναι μία adhoc αρχιτεκτονική. Το μέγεθος των ενταμιευτών που πειραματικά φάνηκε να παράγουν την βέλτιστη απόδοση, δεν είναι παρά μία μικρή σταθερά για ένα εύρος κλίμακας, πράγμα που σημαίνει ότι αυξάνεται η προσαρμοστικότητα του συστήματος.

Η χρηστική-καταλληλότητα/λειτουργικότητα του συστήματος, ουσιαστικά αποδείχτηκε με την ανάλυση πάνω στα μεταβατικά φαινόμενα και τις παρατηρήσεις για την ταχύτητα σύγκλισης ή οποία στην χειρότερη περίπτωση είναι γραμμική σε ένα σύνολο συζευκτικών παραμέτρων – μέγεθος ενεργού-μονοπατιού αλληλεπίδρασης/trajectory –, αλλά πάντα αντιστρόφως ανάλογη της προσωρινής παρατηρούμενης αδικίας κατα τη μετάβαση. Τέλος η ανάλυση/αποτελέσματα πάνω στις πιθανοκρατικές αφίξεις δείχνουν ότι κάτω από ένα μεγάλο φάσμα υποθέσεων, η κατανομή που παράγει το σύστημα πλησιάζει και πάλι το ιδανικό output queuing. Στα αρνητικά πρέπει να συμπεριλάβουμε την έλλειψη προσαρμοστικότητας των δρομολογητών εισόδου και την συνεπαγόμενη μείωση αποτελεσματικότητας όταν οι αφίξεις ακολουθούν κατανομή ανεξάρτητη/διαφορετική των βαρών· το καλό είναι οτι η ¨αχίλλειος πτέρνα¨, εμφανίζεται μόνο υπό αρκετά υψηλό βαθμό χρησιμοποιήσης. Φυσικά θέλουμε το σύστημα να είναι όσο το δυνατόν πιο work conserving γίνεται, αλλά και πάλι εμφανίζεται το κενό ανάμεσα στην απαίτηση για διαφοροποίηση και υψηλό βαθμό απασχόλησης. Με αλλαγή των διανεμητών εισόδου σε κάτι πιο προσαρμοστικό ως προς τις αφίξεις πραγματικού χρόνου και την ιστορία εξυπηρέτησης, πετυχαίνουμε μεγαλύτερη χρησιμοποίηση, άρα και μικρότερη μέση καθυστέρηση, αλλά ταυτόχρονα απομακρυνόμαστε από την ικανοποιήση της ανάγκης για ποιότητα υπηρεσίας. Τέλος το θετικό είναι ότι δεν χρειάζεται να χρησιμοποιήσουμε τους πλέον ακριβείς και ακριβούς δρομολογητές για να πετύχουμε τα παραπάνω ορισμένα χαρακτηριστικά.

Συνεισφορά

Το σύστημα, είχε ανεξάρτητα προταθεί στο [DGPS]. Εκεί προτάθηκε για να μεταφέρει την ιδιότητα φραγμού μέγιστης καθυστέρησης από συστήματα ενός πόρου – όπως προσφέρεται από GPS διανεμητές σε δίκτυα με τις ουρές στις εξόδους –, σε συστήματα με πολλούς με τις κύριες ουρές αποθήκευσης στις εισόδους. Εμείς ασχοληθήκαμε με τη δικαιοσύνη της παραγόμενης κατανομής και αποδείξαμε πειραματικά και εν μέρει αναλυτικά, την σύγκλιση σε κάτι ιδανικό. Αυτό είναι, πιστεύουμε, ιδιαίτερα σημαντικό και για την εργασία αυτών, αφού για να πετύχουνε τον στόχο φραγμού της μέγιστης καθυστέρησης, υποχρησιμοποιούν τη διαθέσιμη παροχή του συστήματος· εξασφαλίζοντας σε κάθε ροή την ελάχιστη των παροχών που αυτή δικαιούται στην είσοδο και στην έξοδο, δεν μπορούν να αναθέσουν/διαθέσουν όλη τη χωρητικότητα, παρά μόνο σε μία τελείως συμμετρική κατανομή των βαρών. Το πρόβλημα που ασχολήθηκαν αποτελεί ουσιαστικά υποπερίπτωση του δικού μας και επιπλέον εμείς τους ¨εγγυώμαστε¨ ότι η παραπανήσια χωρητικότητα θα μοιραστεί δίκαια.

Επιπλέον περιγράψαμε την σύγκλιση με τρόπο που μας έδωσε τη δυνατότητα να προσδιορίσουμε το χρόνο που αυτή κάνει στη χειρότερη περίπτωση, αλλά και πόσο χρόνο θα κάνει δεδομένων των παραμέτρων. Υπό αυτήν την έννοια, μερικώς επεκτείναμε την ανάλυση της Hahne [HahRRMaxMin], σε μέρος φυσικά του γενικού πεδιου εφαρμογής αυτής. Περιορισμένοι στην τοπολογία του crossbar, φέρουμε ενδείξεις σύγκλισης στο στόχο δικαιοσύνης Weighted Max-Min Fair Scheduling – αντι max-min που έλεγξε η Hahne – για το σύστημα με WRR δρομολογητές και backpressure flow control και επιπλέον μία περιγραφή του φαινομένου, που από όσο γνωρίζουμε δεν εχει ξαναπαρουσιαστεί στη βιβλιογραφία.

Ελέγξαμε πειραματικά την απόδοση του συστήματος για πιθανοκρατικές αφίξεις, κάτω από διαφορετικές υποθέσεις για τα βάρη και τη μορφή των εμφανίσεων κίνησης(burtsty, uniform, weights-dependent, weights independent βλ. [BufCrossbar]). Σε όλα τα παραπάνω εξετάσαμε την επίδραση που έχει στην παραγόμενη – σχετικά με το στόχο/ανάλυση – απόδοση, το είδος των δρομολογητών(βασικά WRR Vs WF^2Q+ βλ. [BufCrossbar]), άρα και το κόστος της υλοποίησης σε συνάρτηση με την συμπεριφορά. Τέλος δώσαμε περιγραφή της κρυφής συλλογιστικής που οδήγησε σε αυτή την αρχιτεκτονική και επιπλέον διευκρινίσαμε ορισμένες σχεδιαστικές παγίδες(βλ. selective backpressure, κοινόχρηστη μνήμη).

7.0.3 Μελλοντική εργασία/ ανοιχτά θέματα.

Ένα ανοιχτό θέμα αποτελεί ο έλεγχος προσέγγισης του στόχου δικαιοσύνης υπό πιθανοκρατικες αφίξεις. Το πρόβλημα σε αυτή την περίπτωση γίνεται ακόμα πιο πολύπλοκο, αφού για τους κατανεμητές εισόδου η πραγμάτωση μίας δίκαιης εξυπηρέτησης θα είναι η σύμπτωση τριών γεγονότων, αντί για δύο που ήταν για εμάς μέχρι τώρα: η επιλογή του δρομολογητή εισόδου, η ύπαρξη χώρου ενταμίευσης εσωτερικά <u>και</u> η άφιξη/ύπαρξη πακέτου της ροής στην είσοδο. Ένα δεύτερο ανοιχτό θέμα είναι ο προσδιορισμός του ελάχιστου μεγέθους για τους εσωτερικούς ενταμιευτές που επιτρέπουν σύγκλιση, σε συνδυασμό με την εύρεση αρχών χρονοπρογραμματισμού – ίσως ορισμένων στο πλαίσιο που εξετάζουμε – που θα διευκολύνουν την σύγκλιση με λιγότερες απαιτήσεις σε μνήμη. Όπως και στην περίπτωση της δρομολόγησης ενός πόρου, σίγουρα πολλές μακροπρόθεσμα max-min κατανομές θα υπάρχουν· η περιγραφή της βραχυπρόθεσμα βέλτιστης και ίσως η σχεδίαση της αρχής που την παράγει, έχουν σίγουρα θεωρητικό και πρακτικό ενδιαφέρον.

Ισως κάποιες λεπτομέρειες που λείπουν ή δεν έχουν παρουσιαστεί πλήρως σωστά, πάνω στην περιγραφή του φαινομένου σύγκλισης και του χρόνου που αυτή διαρκεί να χρειάζονται λίγη μελέτη, ενώ σίγουρα χρειάζεται να καλυφθεί ένα κενό στην απόδειξη της τελικής σύγκλισης, όπως αυτή παρουσιάζεται στο [BufCrossbar].

Τέλος η εύρεση αρχών που προσφέρουν μεγαλύτερη χρησιμοποίηση των πόρων – αυτές μας φαίνεται ότι μόνο "προσαρμοστικές" μπορούν να είναι – κάτω από μεγαλύτερο εύρος υποθέσεων, ενώ παράλληλα δε θα χάνουν το στόχο διαφοροποίησης πάνω στη φύση και την κρισιμότητα των αιτήσεων είναι αρκετά σημαντικό, αφού η τάση αύξησης της ζήτησης και της διαφορετικότητας των εφαρμογών ουσιαστικά επιτάσσουν την εξέταση και των δύο παραμέτρων. Φυσικά και πάντα η εύρεση λιγότερο ακριβών τοπολογιών από αυτήν του crossbar(N^2), που σε συνδυασμό με κατάλληλη αρχιτεκτονική δρομολόγησης, μπορούν να προσφέρουν ικανοποιητικές επιδόσεις(π.χ. Weighted Max -Min Fair Schedule) και η μελέτη των ιδιοτήτων που θα πρέπει σίγουρα να διαθέτουν, παρουσιάζουν πάντα ενδιαφέρον(βλ. [SapunjisKatBufBenes]).

Bibliograf'ia

[DGPS] Donpaul C. Stephens Master Thesis, *"Implementing Distributed Packet Fair Queueing in a scalale switch architecture"*, In INFOCOM'98, pages 282–290, San Francisco, CA, March 1998.

[ChQoS] Fabio M. Chiussi and Andrea Fancini, Data Networking Systems Research Department Bell Laboratoris, Lucent Technologies, *"Providing QoS Guarantees in Packet Switches"*, Proceedings of IEEE GLOBECOM'99.

[ShNetObj] Sandeep Bajaj, Lee Breslau Scott Shenker *"Is service Priority Useful in Networks"*, In Proceedings of the ACM Sigmetrics '98, Madison, Wisconsin USA, June 1998.

[UtMaxMin] Zhiruo Cao, Ellen W. Zegura, *"Utility Max-Min: An application-Oriented Bandwidth Allocation Scheme"*, Proceedings of IEEE INFOCOM 99, New York, NY, March, 1999. http://citeseer.nj.nec.com/cao99utility.html

[FaCongCont] Jeonghoon Mo and Jean Warland, *"Fair End-to-End Window-based Congestion Control"*, Available as. http://citeseer.nj.nec.com/337047.html

[HahRRMaxMin] Ellen L. Hahne, AT&T Bell Laboratories, *"Round-Robin Scheduling for Max-Min Fairness in Data Networks"*, IEEE Journal on Selected Areas in Communications, 9(7), September 1991. http://citeseer.nj.nec.com/hahne91roundrobin.htm

[HaydenMaxMin] H. P. Hayden, *"Voice Flow Control in Integrated Packet NetWorks"*, Technical Report LIDS-TH-1152, MIT Laboratory for Information and Decision Systems, 1981.

[JaffeMaxMin] J. M. Jaffe, *"Bottleneck Flow Control"*, IEEE Trans. Commun ,vol. COM-29, pp. 954–962, July 1981.

[Kate87] M. Katevenis: *"Fast Switching and Fair Control of Congested Flow in Broad-Band Networks"*, IEEE Journal on Selected Areas in Communications, Vol. SAC-5, No. 8, October 1987, pp. 1315-1326.

[KaSC91] M. Katevenis, S. Sidiropoulos, C. Courcoubetis: *"Weighted Round-Robin Cell Multiplexing in a General-Purpose ATM Switch Chip"*, IEEE Journal on Selected Areas in Communications, vol. 9, no. 8, October 1991, pp. 1265-1279.

[ParGalPGPS] Abhay K. Parekh and Robert G. Gallager, *"A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The single Node Case"*, IEEE/ACM Transactions on Networking, Vol. 1, No. 3, June 19.

[DemKesShenWFQ] A. Demers, S. Keshav And S. Shenkar, *"Analysis and simulation of a fair Queuing Algorithm"*, Proc. of ACM SIGCOMM, 1989, pp. 3-12.

[ZhangVC] L. Zhang, *"A new architecture for packet switching network protocols"*, MIT PhD Thesis. 1989.

[BenhZhangWFWF] Jon C.R, Bennett, Hui Zhang, *"WF^2Q: Worst-case Fair Weighted Fair Queuing"*, In Proceedings of INFOCOM '96, San Francisco, CA, March 24-28, 1996 http://citeseer.nj.nec.com/zhang96wfq.html.

[Keshav97] S. Keshav: *"An Engineering Approach to Computer Networking"*..

[HierFairQ] Jon C.R. Bennett and H. Zhang, *"Hierarchical Packet Fair Queuing Algorithms"*, In Proceedings of SIGCOMM'96, pages 143–156, August 1996.

[AnderPIM] Thomas E. Anderson, *"High Speed Switch Scheduling for Local Area Network"*, ACM Transactions on Computer Systems, November 1993.

[SLIPMcKeownAndersComp] N. McKeown and T.E. Anderson, *"A quantitative comparison of scheduling algorithms for input queued switches"*, published 1997. http://citeseer.nj.nec.com/mckeown97quantitative.html

[SerpFIRM] D. Serpanos, P. Antoniadis *"FIRM: a Class of Distributed Scheduling Algorithms for High-Speed ATM Switches with Multiple Input Queues"*, In Proceedings of IEEE Infocom 2000.

[Serp2DRR] R. LaMaire, D. Serpanos, *"Two-Dimensional Round-Robin Schedulers for Packet Switches with Multiple Input Queues"*, IEEE/ACM Transactions on Networking Vol.2, No. 5 (Oct. 1994), pp. 471-482.

[StilVarmWPIM] D. Stiliadis and A. Varma, *"Providing bandwidth guarantess in an input-buffered crossbar switch"*, n Proc. IEEE INFOCOM '95, April 1995.

[FIM] Nan Ni, Lax N. Bhuyan, *"Fair scheduling for Input Buffered Switches"*, htpp://citeseer.nj.nec.com/482342.html.

[Kumar98] V. Kumar, T. Lakshman, D. Stiliadis, *"Beyond Best Effort: Router Architectures for the Differentiated Service of Tomorrow's Internet"*, 1998. http://citeseer.nj.nec.com/kumar98beyond.html.

[Linear] Anthony C. Kam, Kai-Yeung Siu, *"Linear complexity algorithms for QoS support in input-queued switches with no speed-up"*, Technical report, d'Arbeloff Laboratory for Information Systems and Technology, Massachusetts Institute of Technology, Cambridge MA, 1998, http://citeseer.nj.nec.com/kam98linear.html.

[LOOFA] P. Krishna, N. Patel, A. Charny, R. Simcoe, *"On the Speed-up Required for Work-Conserving Crossbar Switches"*, In IWQoS 98, 1998.

[In2OutSpeedup] S. Chuang, A. Goel, N. McKeown, B. Prabhakar, *"Matching Output Queuing with a Combined Input Output Queued Switch"*, IEEE Journal on Selected Areas in Communications Volume 17, Number 6, June 1999.

[HarteKatevCompTree] Kostas G.I. Harteros *"Fast Parallel comparison circuits for scheduling"*, Technical Report ICS/FORTH ID Number 304, ftp://ftp.ics.forth.gr/tech-reports/2002/2002.TR304.Fast_Comparison_Circuits_for_Scheduling

[DGPS1] D. Stephens, Hui Zhang: *"Implementing Distributed Packet Fair Queueing in a Scalable Switch Architecture"*, In INFOCOM'98, pages 282–290, San Francisco, CA, March 1998. http://citeseer.nj.nec.com/stephens98implementing.html.

[MagBufFullCap] Tara Javidi, Robert Magill and Terry Hrabik, *"A High-Throughput Scheduling Algorithm for a Buffered Crossbar Switch Fabric"*, Proceedings of IEEE.

[Kat-534] Manolis katevenis *Lectures on Packet Networks Architecture.* http://archvlsi.ics.forth.gr/ kateveni/534/

[PackSwitDes] Cyriel Johan Agnes Minkenberg Thesis, *"On Packet Switch Design"* http://alexandria.tue.nl/extra2/200113175.pdf

[Stephens99] D. Stephens, J. Bennett, H. Zhang, *"Implementing Scheduling Algorithms in High-Speed Networks"*, To Appear in IEEE JSAC, 1999. http://citeseer.nj.nec.com/stephens99implementing.html

[IoanKateHeap] A. Ioannou, M. Katevenis. *"Pipelined Heap (Priority Queue) Management for Advanced Scheduling in High-Speed Networks"*, Proc. IEEE Int. Conf. on Communications (ICC'2001), Helsinki, Finland, June 2001,pp. 2043-2047.

[SapunjisKatBufBenes] G. Sapountzis, M. Katevenis. *"Benes Fabrics with Internal Backpressure:First Work-in-Progress Report"*, Technical Report FORTH-ICS/TR-303,Institute of Computer Science, FORTH, Heraklio, C-rete, Greece, March 2002; http://archvlsi.ics.forth.gr/bpbenes

[BufCrossbar] N. Chrysos *"Weighted Max-Min Fairness in a Buffered C-rossbar Switch with Distributed WFQ Schedulers"*, Master Thesis http://www.csd.uch.gr/ñchrysos/studying, (soonly will appear as Technical Report at ICS/FORTH, with id number 309.)

Κεφάλαιο 8

Παράρτημα Α

8.1 Adaptive WRR — Μία πολλά υποσχόμενη βελτίωση.

Ο αναλογικός προγραμματισμός στις εισόδους και στις εξόδους στο σύστημα μας κάνει εφικτή τη προσέγγιση του στόχου της δίκαιης κατανομής που έχουμε περιγράψει εδώ. Ο στόχος αυτός, αν και αφού απόλυτα δίκαιος, μπορεί να ορίσει το ιδανικό σύστημα, δεν είναι μονοσήμαντα ορισμένος με βάση τη σειρά εξυπηρετήσεων. Από αυτή την άποψη πολλές κατανομές μπορεί να θεωρηθούν δίκαιες στη μικροκλίμακα· φυσικά κάποια μπορεί να θεωρηθεί βέλτιστη. Υπάρχει λοιπόν αυτό το ζήτημα/ερώτημα. Επιπλέον υπάρχει το ερώτημα, αν κάποια από αυτές τις υποψήφιες αρχές, έχει μικρότερες απαιτήσεις σε εσωτερικούς ενταμιευτές. Μάλιστα, όπως φάνηκε και από τα αποτελέσματα για πιθανοκρατικές αφίξεις, μικρές, χρονικά και τοπικά, ¨ανωμαλίες¨ στην εξυπηρέτηση μπορεί να δημιουργήσουν μεγάλες ανωμαλίες για το σύστημα μας σε σχέση με το ιδανικό. Αυτό φαίνεται ξεκάθαρα στην περίπτωση που έχουμε N^2 ροές, με ιδιο ρυθμό αφίξεων $1/N$, οπότε ανεξάρτητα από τα βάρη τους δικαιούνται ίσο ρυθμό εξυπηρέτησης με βάση το μαθηματικό ορισμό δίκαιης κατανομής μεγίστου-σταθμισμένου-ελαχίστου. Όμως το σύστημα δίνει μεγάλη μέση καθύστερηση, αρκετά μεγαλύτερη από αυτή που θα έδινε ένα σύστημα με ουρές στις εξόδους, γεγονός που υπονοεί μεγάλη απόκλιση από την άριστη δίκαιη κατανομή. Είδαμε ότι χρησιμοποιώντας προσαρμοστικές μεθόδους στις εισόδους, βελτιώνουμε κάπως τη συμπεριφορά υπό αυτό το σενάριο – αφού διατηρούμε μεγαλύτερό αριθμό ενεργών εξόδων, θεωρώντας ανυποψίαστα όλες τις ουρές ισοδύναμες και τελικά καταλήγουμε πιο κοντά στη δίκαιη κατανομή(πάνω σε μέσο ρυθμό μετάδοσης) αφού η τελική συνολική ζήτηση είναι τέτοια –, δεν προσσεγγίζουμε όμως το στόχο δικαιοσύνης για την περίπτωση που έχουμε διαφορετικά βάρη και γενικά μή ομοιόμορφη – ή απεριόριστη – ζήτηση. Επιπλέον όπως είπαμε και πιο πάνω, χρησιμοποιώντας προσαρμοστικές μεθόδους δεν μπορούμε να αποδώσουμε όσο μικρή καθυστέρηση θέλουμε σε ροές της προτίμησης μας.

Εδώ προτείνουμε μία παραλλαγή για τούς κατανεμητές εισόδου ή/και εξόδου στο buffered crossbar, που χρησιμοποιεί στοιχεία τόσο του αναλογικού όσο και προσαρμοστικού χρονοπρογραμματισμού. Η μέθοδος αυτή, που γενικά ονομάζουμε, προσαρμοστική-αναλογικά-κυκλική δρομολόγηση, δείχνει, από κάποια πρώτα αποτελέσματα, να παρουσιάζει αρκετά ενδιαφέροντα χαρακτηριστικά. Ουσιαστικά λειτουργεί σε 2(ή και περισσότερες λογικές φάσεις. Στην πρώτη αποκλείει ροές, θεωρώντας τις ως ανενεργές – ενώ κατά τα αλλά κριτήρια είναι ενεργές –, με κριτήρια ¨προσαρμογής¨, καί εκτελεί τον αλγόριθμο αναλογικής δρομολόγησης πάνω στις ροές που πέρασαν αυτό το φίλτρο. Στην δεύτερη, που εκτελείται λογικά αν οι πρώτες δεν βρήκανε ενεργό υποψήφιο, εκτελείται ουσιαστικά ο αλγόριθμος αναλογικής δρομολόγησης, πάνω σε όλες τις ενεργές ροές. Από αυτή την άποψη είναι work conserving. Μάλιστα αν το κριτήριο αποκλεισμού στις πρώτες φάσεις είναι σχετικά απλό, λογικά κυκλώματα μπορούν εξαρχής να μας πούνε ποιά φάση θα βρεί ενεργό υποψήφιο, οπότε το ένα τρέξιμο της μηχανής αναλογικής δρομολόγησης θα είναι αρκετό.

Περιγράφουμε εδώ τη μορφή της, συγκεκριμένα για τους κατανεμητές εισόδου, χρησιμοποιώντας την ορολογία του WRR. Βασικά ο στόχος είναι κάθε δρομολογητής εισόδου, να απασχολήσει μεσοπρόθεσμα όσο το δυνατόν περισσότερες εξόδους – γεμίζοντας τους αντίστοιχους ενταμιευτές –, οπότε να βρεί το σύστημα κάτι κοντά στο συνολικά μέγιστο βαθμό απασχόλησης των εξόδων – ουσιαστικά δεν μπορεί να είναι παρά τοπικό μέγιστο, αφού οι δρομολογητές εισόδου δεν παίρνουν κοινές(=βέλτιστες) αποφάσεις. Στο ψευδοκώδικα που παρουσιάζουμε παρακάτω, αποκλείουμε στον πρώτο γύρο τις ροές που έχουν ένα τουλάχιστον πακέτο στους εσωτερικούς ενταμιευτές – η προσαρμοστική διάσταση της αρχής.

1. *D*rag flows that their next_service_time is left behind virtual_time, to virtual_time only if these are inactive due to backpressure or cell absence – not if these are left behind due to first step selection.

2. Select the active flow with the minimum next_service_time, that has no cell in the crosspoints.

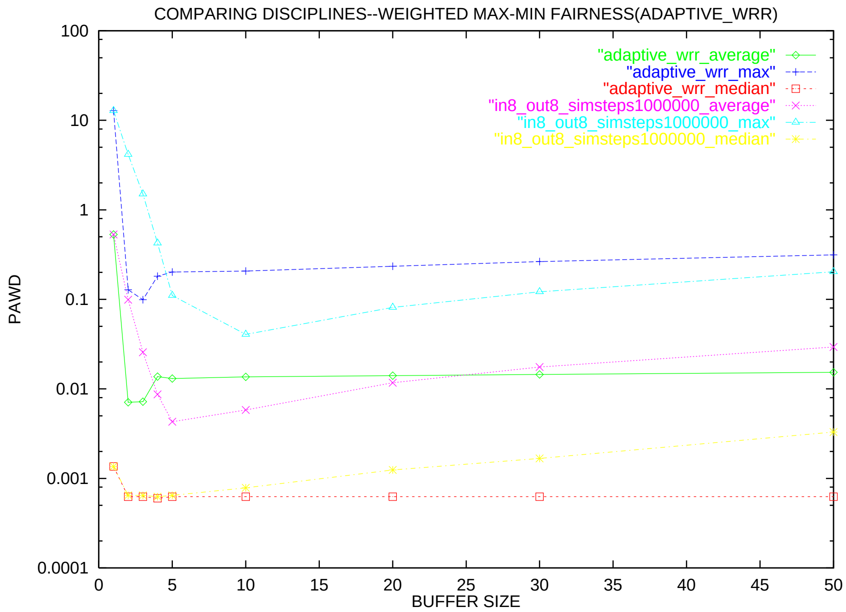
3. *I*f nothing is found in the first step, select the active flow with minimum next_service_time.

4. *U*pdate virtual time to the time of the selected flow and update the next_service_time of the selected flow, by adding the respective service_interval.

Υπάρχουν κάποιες παράμετροι που αξίζει μελετηθούν. Μία είναι αν θα τραβάμε το χρόνο εξυπηρέτησης μίας ροής στον τρέχοντα χρόνο του συστήματος(βλ. WRR), αν αυτές μείνανε πίσω λόγω αποκλεισμού πρώτης φάσης. Διαισθητικά πιο σωστό, δείχνει να είναι καλύτερο να μην το κάνουμε, έτσι ώστε αυτές οι ροές να έχουν μεγαλύτερη προτεραιότητα. Δεύτερο είναι το κριτήριο αποκλεισμού. Αυτό μπορεί να είναι κάποιο κατώφλι της απασχόλησης του εσωτερικού ενταμιευτή. Συμμετρικά ορίζεται και η τεχνική για τις εξόδους. Σημειώνουμε ότι με αυτή τη τεχνική, μπορεί να αποδοθεί όσο μικρή καθύστερηση θέλουμε σε μία ροή – αρκεί να βρεθεί η κατάλληλη ανάλυση για να το αποδείξει.

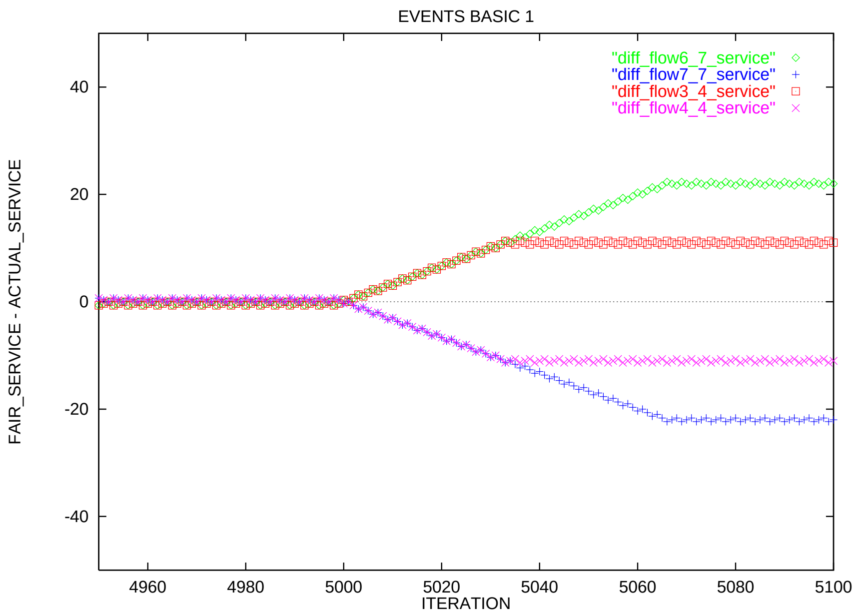
Τα αποτελέσματα παρακάτω, χρησιμοποιούν την μέθοδο $AdaptWRR1$ που περιγράψαμε με τον ψευδο-κώδικα, μόνο στις εισόδους.

Το πρώτο συγκρίνει τις τεχνικές $AdaptWRR1$ και WRR, υπό το κριτήριο σύγκλισης στην ιδεατή κατανομή δικαιοσύνης. Για αυτό το σενάριο η $AdaptWRR$ που περιγράψαμε λειτουργεί πολύ καλύτερα και για μικρότερα μεγέθη εσωτερικής ενταμίευσης. Όπως παρατηρήσαμε, το ποσοστό χρήσης των ενταμιευτών είναι πολύ μικρότερο – κοντά στο ένα – από ότι με την WRR. Στην τελευταία, όλες οι ροές με bottleneck στο εξωτερικό link, χρησιμοποιούν όλο σχεδόν το διαθέσιμο χρόνο, για το περισσότερο του χρόνου – βλ. στοιχεία αναλυτικής επιβεβαίωσης σύγκλισης στο [BufCrossbar] –, αλλά φαίνεται και από την επιδείνωση της συμπεριφοράς για μεγάλους ενταμιευτές, που οφείλεται μεν σε λάθος



Σχήμα 8.1: Aggregate $PAWD$ metrics, under the two schemes(WRR at both input, outputs servers, Vs $AdWRR$ at inputs, WRR at outputs). This configuration is for a hot-spot like distribution of weights. The other parameters are shown in the figure, through the labels that correspond to the traditional WRR system(in8_out8_simsteps1000000*). Y-axis logscale.

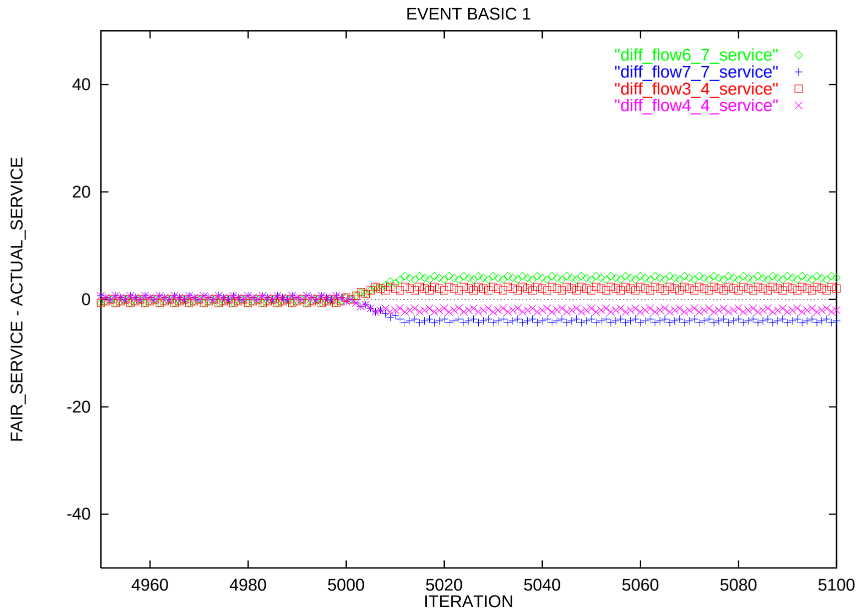
της μέτρησης αλλά αποκαλύπτει τη χρήση του εσωτερικού χώρου· με τη προσαρμοστική μέθοδο δεν υπάρχει αυτή η επιδείνωση. Αυτό σημαίνει και πολύ μικρότερη αδράνεια/καθυστέρηση προσαρμογής σε μεταβατικά φαινόμενα, και αυτό επιβεβαιώνουμε παρακάτω, όπου ο χρόνος ισορρόπησης είναι κοντά στο 4 φορές μικρότερος, αποτέλεσμα της σχεδόν 4 φορές μικρότερη χρησιμοποίηση των εσωτερικών ενταμιευτών(βλ. ανάλυση χρόνου συγκλισης στο Κεφ. 4.2.2).



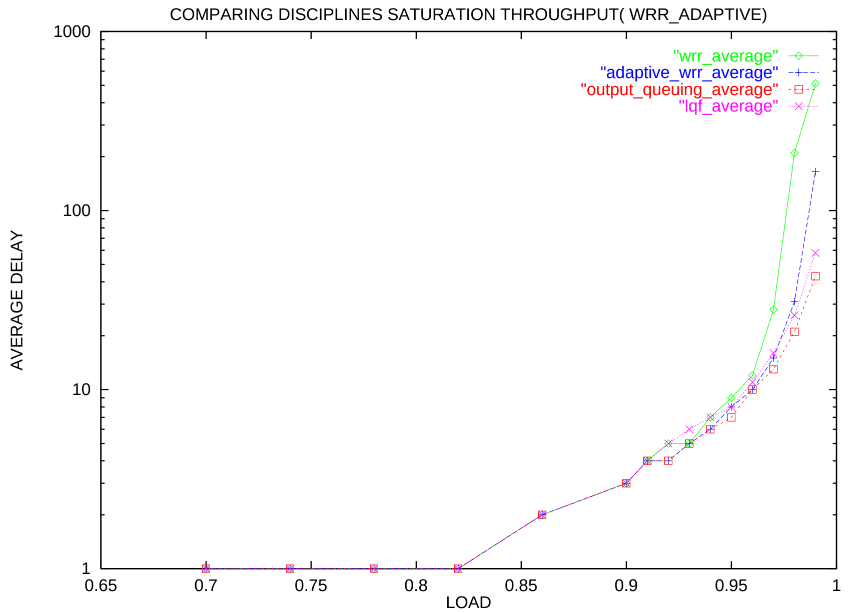
Σχήμα 8.2: Time to converge, after a marginal change in the condition of flow 1, for the scheme presented in Kef. 5.3 (Sq. 5.10), with WRR discipline at both inputs and outputs. Crosspoint buffers size 4.

Τέλος συγκρίνουμε την μέση καθύστερηση, στην περίπτωση ομοιόμορφων και ομοιογενών, πιθανοκρατικών αφίξεων, υπό την ύπαρξη όμως διαφορετικών βαρών που μάλιστα ακολουθούν την hot-spot κατανομή. Το σύστημα παρουσιάζει βελτιωμένη απόδοση περίπου κοντά στο 600%.

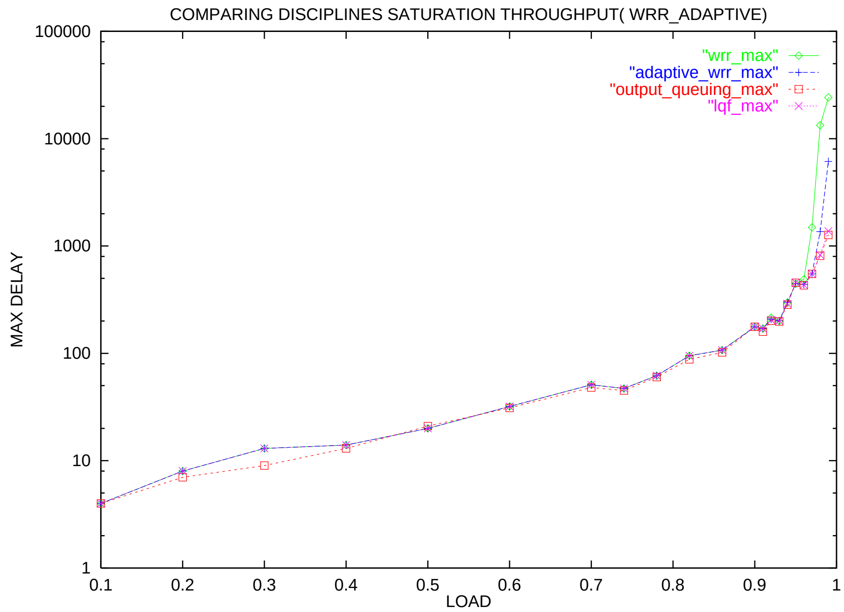
Τα προβλήματα της μεθόδου/εναλλακτηκές/μελλοντική εργασία Κλείνοντας το παράρτημα πρέπει να σημειώσουμε ότι τα πράγματα δυστυχώς δεν είναι τόσο ευνοικά/θετικά. Για το κριτήριο σύγκλισης, βρήκαμε αντιπαραδείγματα, οπου η απόσταση της προτεινόμενης τεχνικής αποκλίνει αρκετά του ιδεατού($\simeq$ 10% σε μέσο όρο). Αν και αυτό πρέπει να κριθεί πόσο κακό/αρνητικό είναι(που



Σχήμα 8.3: Time to converge, after a marginal change in the condition of flow 1, for the scheme presented in Kef. 5.3 (Sq. 5.10), with $AdWRR$ discipline at inputs and WRR at the outputs. Crosspoint buffers size 4.



Σχήμα 8.4: Average delay, for a 8x8 system, hot-spot(irregular) distribution of weights, Bernoulli id.d arrivals, independent of weights. Y-axis in logscsale.



Σχήμα 8.5: Max delay, for a 8x8 system, hot-spot(irregular) distribution of weights, Bernoulli id.d arrivals, independent of weights. Y-axis in logscale.

είναι), μπορεί να διορθωθεί αλλάζοντας το προσαρμοστικό κριτήριο αποκλεισμού – το επιβεβαιώσαμε με επιπλέον πειράματα εφαρμόζοντας τον αποκλεισμό όταν έχουμε απασχόληση 2 πακέτων ή απασχόληση στο μισό του μεγέθους του εσωτερικού ενταμιευτή αν και γενικά το βέλτιστο δείχνει/μπορεί να είναι κάτι ανάλογο του μεγέθους του μεταγωγέα. Τέλος προσομοίωσαμε ένα σύστημα με την προσαρμοστικά-αναλογική τεχνική και στις εξόδους(συμμετρικά ο αποκλεισμός θα είναι όταν η χρησιμοποίηση του ενταμιευτή είναι μικρή) και το σύστημα αν και δεν απέδωσε καλύτερη σύγκλιση, έδωσε μεγαλύτερη απασχόληση, ακόμα πιο κοντά στο ιδανικό output queuing system – αναμενόμενο. Στο [Βυφ῾ροσσβαρ] στην σχετική/αντίστοιχη – με αυτή εδώ – παράγραφο, αναφέρουμε ορισμένες σκέψεις σχετικές με τις κατευθύνσεις/στόχους που μπορεί να ακολουθήσει/κυνηγήσει η έρευνα για την κατασκευή εναλλακτηκών και ίσως καλύτερων αρχών-δρομολόγησης στο buffered crossbar.