

Εθνικό Μετσόβιο Πολυτεχνείο
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Πληροφορικής

**ΕΛΕΓΧΟΣ ΕΙΣΟΔΟΥ ΚΙΝΗΣΗΣ
ΒΑΣΙΖΟΜΕΝΟΣ ΣΕ ΑΝΑΔΡΑΣΗ
ΓΙΑ ΠΑΡΟΧΗ ΠΟΙΟΤΗΤΑΣ
ΣΕ ΔΙΚΤΥΑ ΔΙΑΦΟΡΟΠΟΙΗΜΕΝΩΝ ΥΠΗΡΕΣΙΩΝ**

Διδακτορική Διατριβή
Ελένη Μ. Μυκωνιάτη

Επιβλέπων Καθηγητής
Εμμανουήλ Πρωτονοτάριος

Αθήνα, Νοέμβριος 2003

Εθνικό Μετσόβιο Πολυτεχνείο
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Πληροφορικής

ΕΛΕΓΧΟΣ ΕΙΣΟΔΟΥ ΚΙΝΗΣΗΣ
ΒΑΣΙΖΟΜΕΝΟΣ ΣΕ ΑΝΑΔΡΑΣΗ
ΓΙΑ ΠΑΡΟΧΗ ΠΟΙΟΤΗΤΑΣ
ΣΕ ΔΙΚΤΥΑ ΔΙΑΦΟΡΟΠΟΙΗΜΕΝΩΝ ΥΠΗΡΕΣΙΩΝ

Διδακτορική Διατριβή
Ελένη Μ. Μυκωνιάτη

Συμβουλευτική επιτροπή : Ε. Πρωτονοτάριος, καθηγητής ΕΜΠ
Ν. Μήτρου, καθηγητής ΕΜΠ
Β. Λούμος, αν. καθηγητής ΕΜΠ

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 24^η Νοεμβρίου 2003.

.....
Ε. Πρωτονοτάριος
καθηγητής ΕΜΠ

.....
Ν. Μήτρου
καθηγητής ΕΜΠ

.....
Β. Λούμος
αν. καθηγητής ΕΜΠ

.....
Ε. Συκάς
καθηγητής ΕΜΠ

.....
Μ. Θεολόγου
καθηγητής ΕΜΠ

.....
Γ. Στασινόπουλος
καθηγητής ΕΜΠ

.....
Γ. Σταμούλης
καθηγητής ΟΠΑ

Αθήνα, Νοέμβριος 2003

.....

Ελένη Μ. Μυκωνιάτη

Διδάκτωρ Μηχανικός Πληροφορικής

Copyright © Ελένη Μ. Μυκωνιάτη, 2003.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία ασχολείται με τον έλεγχο εισόδου κίνησης για την παροχή ποιότητας σε δίκτυα Διαφοροποιημένων Υπηρεσιών.

Από την επιτομή των σχετικών μελετών υιοθετούμε το καταναμημένο στους ακραίους κόμβους και βασιζόμενο σε μετρήσεις μοντέλο και προτείνουμε μία νέα προσέγγιση που ονομάζουμε μοντέλο βασιζόμενο σε ανάδραση. Οι πόροι διαμερίζονται εκ των προτέρων σε αποκλειστικά δεσμευμένα και συμπληρωματικά κοινόχρηστα μερίδια. Το ρίσκο παραβίασης της ποιότητας εκτιμάται από τοπικές παθητικές μετρήσεις και το επιτρεπτό του επίπεδο ρυθμίζεται από την πολιτική του παροχέα. Η συμμόρφωση ανιχνεύεται από ενεργές μετρήσεις από άκρο σε άκρο και κατάλληλοι μηχανισμοί αναλαμβάνουν την ομαλή επίλυσή της.

Θεωρούμε τον έλεγχο εισόδου κίνησης ως μέρος ενός ολοκληρωμένου συστήματος για την παροχή ποιότητας υπηρεσίας και διαχωρίζουμε μεταξύ των λειτουργιών διαχείρισης κίνησης που αναλαμβάνουν μεταξύ άλλων το διαμερισμό των πόρων και των λειτουργιών διαχείρισης υπηρεσιών μέρος των οποίων είναι ο έλεγχος εισόδου κίνησης. Θεωρούμε πως της φάσης ενεργοποίησης της υπηρεσίας προηγείται η φάση εγγραφής και κλήσης της υπηρεσίας και ορίζουμε έλεγχο αποδοχής και στα δύο αυτά επίπεδα. Στη φάση εγγραφής ο έλεγχος αποδοχής αποφαίνεται επί της μακροπρόθεσμης δυνατότητας του δικτύου να υποστηρίξει την υπηρεσία και στη φάση κλήσης επί της τρέχουσας διαθεσιμότητας των πόρων.

Έχοντας οριοθετήσει τον έλεγχο εισόδου κίνησης από τις σχετικές και τις βοηθητικές προς αυτόν λειτουργίες αναπτύσσουμε τον αλγόριθμο χρησιμοποιώντας το βασιζόμενο σε ανάδραση μοντέλο. Διακρίνουμε μεταξύ της λογικής που δυναμικά διαμορφώνει την εκάστοτε ενδεδειγμένη στρατηγική αποδοχής και της εφαρμογής της. Η στρατηγική αποδοχής ρυθμίζει την αποδοχή των αιτήσεων κλήσης και τους πόρους που αποδίδονται στις ενεργές ροές, καθορίζοντας αντίστοιχα τη διαθεσιμότητα της υπηρεσίας και την ποιότητα κατά τη χρήση.

Εξετάζονται πρακτικά θέματα που αφορούν στην εξουσιοδότηση βάσει του βαθμού διαθεσιμότητας, στους διαφορετικούς τύπους ενεργοποίησης υπηρεσιών και στη μετάφραση και υλοποίηση των συμβολαίων υπηρεσίας σε ρυθμίσεις του επιπέδου δεδομένων. Ο αλγόριθμος δυναμικής διαχείρισης αποδοχής που εφαρμόζει τις αρχές του βασιζόμενου σε ανάδραση μοντέλου. Ενώ η επίλυση της συμμόρφωσης διασφαλίζεται μηχανιστικά χάρη στο σχεδιασμό του συστήματος, η προληπτική αποφυγή της συμμόρφωσης εξαρτάται από μη ελεγχόμενους παράγοντες όπως η συμπεριφορά των χρηστών σε όλη την επικράτεια του δικτύου. Ο βαθμός στον οποίο είναι αποδεκτή και επιθυμητή η εφαρμογή στρατηγικών αυξημένης αυστηρότητας προληπτικά, ρυθμίζεται από τη λειτουργική πολιτική του παροχέα με την παράμετρο του επιπέδου προφύλαξης.

Η βιωσιμότητα της προτεινόμενης λύσης επαληθεύεται αφενός θεωρητικά, με την ανάλυση της πολυπλοκότητας του συστήματος, και αφετέρου πρακτικά με την ανάπτυξη πρωτοτύπου υλοποίησης και την εφαρμογή του σε εμπορικούς και πειραματικούς δρομολογητές σε διαφορετικά δικτυακά περιβάλλοντα. Επιπλέον, η αξιολόγηση των θεωρητικών συμπερασμάτων και επιτευγμάτων επισφραγίζεται με τη διεξαγωγή εκτενών πειραμάτων χρησιμοποιώντας διαφορετικά σενάρια για την προσφερόμενη κίνηση σε μια ειδικά σχεδιασμένη πλατφόρμα εξομοίωσης δικτύου.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Ποιότητα Υπηρεσίας, Διαφοροποιημένες Υπηρεσίες, Έλεγχος Εισόδου Κίνησης, Μοντέλο Βασιζόμενο σε Ανάδραση, Εγγραφή Υπηρεσίας, Κλήση Υπηρεσίας

ABSTRACT

The scope of this thesis is service admission control for providing QoS in Differentiated Services networks.

From the state of the art review, we adopt the distributed at the edge nodes and measurement-based model and we propose a new approach called the feedback-based model. Network resources are a priori partitioned into hard-reserved and supplementary soft-reserved shares. The risk of QoS deterioration is estimated based on local passive measurements and its acceptable level is adjusted based on the policy of the provider. Congestion is detected using active edge-to-edge measurements and resolved by dedicated mechanisms.

We study service admission control as part of a holistic approach for providing QoS and we distinguish between traffic engineering functions, responsible among others for resource partitioning, and service management functions, part of which is the service admission control. Our view is that using a service is preceded by subscribing to the service and invoking the service and we consider that admission logic applies at both levels. In service subscription phase admission logic deduces the long term capability of the network to sustain the service while in service invocation phase the current availability of the resources.

Having separated service admission control from other relevant or supporting functions we develop the algorithm using the feedback-based model. We distinguish among the logic dynamically formulating the currently optimum admission strategy and the logic to apply it. The admission strategy regulates the admission of invocation requests and the resources admitted to active flows, determining respectively the service availability and the quality granted during its active period.

Applicability technical issues are considered, regarding authorisation based on contractual grade of service, to different service types with respect to activation rights, mapping and implementation of service contracts to traffic conditioning rules in the data plane. The dynamic admission management algorithm is built around the principles of the feedback-based model. While congestion resolution is guaranteed in the engineering level thanks to the design of the system, preventing congestion in anticipation is tied with uncontrollable variables such as network wide user behaviour. The degree up to which, applying strategies of increased severity proactively is considered acceptable and desirable, is tuned by the policy of the provider through the precaution level parameter.

The viability of the approach is verified theoretically, through a system complexity and scalability analysis, and in practice, by implementing a software prototype and applying it upon both commercial and experimental routers in several network environments. Moreover, the evaluation of theoretical conclusions and achievements is validated by undertaking extended tests, under various settings for the offered traffic upon a network emulation platform, designed specifically to fulfil our testing requirements.

KEYWORDS

Quality of Services, Differentiated Services, Service Admission Control, Feedback Based Model, Service Subscription, Service Invocation

ΕΥΧΑΡΙΣΤΙΕΣ

Θα επιθυμούσαμε εκ του παρόντος να ευχαριστήσουμε όλους τους συνεργάτες που συνετέλεσαν στην περάτωση αυτής της εργασίας.

Μεγάλο μέρος της παρούσας εργασίας έλαβε χώρα στα πλαίσια του ευρωπαϊκού ερευνητικού προγράμματος **Traffic Engineering for Quality of Service in the Internet, at Large Scale (IST-TEQUILA)**. Ευχαριστούμε τον Παναγιώτη Γεωργιάτσο, Λεωνίδα Γεωργιάδη και Παναγιώτη Δαμηλάτη για την πολύτιμη καθοδήγηση και συμβολή τους, ιδιαίτερα όσον αφορά στις πολιτικές και τις απαιτήσεις ενός παροχέα υπηρεσιών και στη συμπεριφορά των χρηστών αλλά και στην επιστημονική γνώση της δυναμικής του δικτύου. Οι συνεργάτες Danny Goderis, Sven Van den Bosch και Yves T' Jones συνεισέφεραν με τις προτάσεις τους αναφορικά στον έλεγχο εισόδου κίνησης στα πλαίσια του προαναφερθέντος έργου και την ενεργή συμμετοχή τους στη διαμόρφωση των εναρκτήριων ιδεών της παρούσας εργασίας. Ιδιαίτερες ευχαριστίες απευθύνουμε στους David Griffin, Christian Jacquenet και Steven Van den Berghe για τη συνεισφορά τους στην κατανόηση όλων των θεμάτων που σχετίζονται με την ποιότητα της υπηρεσίας στα δίκτυα διαφοροποιημένων υπηρεσιών και για την κριτική παρακολούθηση της εργασίας μας.

Το μεγαλύτερο μέρος της δραστηριότητας πραγματοποιήθηκε στο εργαστήριο τηλεπικοινωνιών συστημάτων του τομέα πληροφορικής του τμήματος ηλεκτρολόγων μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Ευχαριστούμε τους συνεργάτες Γιώργο Μεμένιο, Δημήτρη Γιαννακόπουλο, Χαράλαμπο Χαραλάμπους, Δημήτρη Καγκλή και Νικόλαο Λιαμπώτη για τη στήριξή τους, τον καθηγητή Ευστάθιο Συκά καθώς και τους επιβλέποντες καθηγητές Εμμανουήλ Πρωτονοτάριο, Νικόλαο Μήτρου και Βασίλειο Λούμο.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	5
ΠΕΡΙΕΧΟΜΕΝΑ	8
ΜΕΡΟΣ Α. ΕΙΣΑΓΩΓΗ.....	A-13
A.1 Ποιότητα Υπηρεσίας.....	A-13
<i>A.1.1 Ιστορικό Πλαίσιο.....</i>	<i>A-13</i>
<i>A.1.2 Ορισμός Ποιότητας Υπηρεσίας</i>	<i>A-13</i>
<i>A.1.3 Ποιότητα Υπηρεσίας για το Χρήστη – Εφαρμογές</i>	<i>A-15</i>
A.1.3.1 Εφαρμογές Αλληλεπίδρασης.....	A-16
A.1.3.2 Εφαρμογές Απόκρισης.....	A-16
A.1.3.3 Εφαρμογές Συνεχούς Ροής	A-17
A.1.3.4 Εφαρμογές Παρασκηνίου.....	A-18
A.1.3.5 Εφαρμογές Ελέγχου Δικτύου	A-18
<i>A.1.4 Ποιότητα Υπηρεσίας για το Δίκτυο</i>	<i>A-18</i>
A.1.4.1 Μηχανισμοί	A-19
A.1.4.2 Αρχιτεκτονικές.....	A-22
A.1.4.3 Λειτουργίες Διαχείρισης.....	A-30
<i>A.1.5 Ανακεφαλαίωση.....</i>	<i>A-36</i>
A.2 Έλεγχος Εισόδου Κίνησης σε Δίκτυα Διαφοροποιημένων Υπηρεσιών	A-39
<i>A.2.1 Ανάλυση και Ορισμός του Προβλήματος.....</i>	<i>A-39</i>
A.2.1.1 Καθορισμός των απαραίτητων πόρων	A-39
A.2.1.2 Εκτίμηση διαθεσιμότητας των απαραίτητων πόρων.....	A-41
A.2.1.3 Πρόσθετες Απαιτήσεις	A-42
<i>A.2.2 Βασικές Αρχές Γνωστών Προσεγγίσεων.....</i>	<i>A-43</i>
A.2.2.1 Μέθοδος ελέγχου εισόδου κίνησης	A-43
A.2.2.2 Τοπολογική και λειτουργική κατανομή	A-47
A.2.2.3 Μέθοδος κατανομής των πόρων	A-49
<i>A.2.3 Σχετικές Μελέτες.....</i>	<i>A-50</i>
A.2.3.1 Συγκεντρωτικές Μελέτες	A-51
A.2.3.2 Κατανεμημένες Parameter-Based Μελέτες.....	A-58
A.2.3.3 Κατανεμημένες Measurement-Based Μελέτες.....	A-63
<i>A.2.4 Ανακεφαλαίωση.....</i>	<i>A-68</i>
ΜΕΡΟΣ Β. ΠΕΡΙΓΡΑΦΗ.....	B-73
B.1 Τοποθέτηση επί του Προβλήματος	B-73
<i>B.1.1 Το Βασικό Μοντέλο.....</i>	<i>B-73</i>
B.1.1.1 Κατανεμημένο Μοντέλο Αμιγώς Ακραίων Κόμβων	B-73
B.1.1.2 Μοντέλο Βασιζόμενο σε Ανάδραση	B-74
B.1.1.3 Ρύθμιση Από Λειτουργική Πολιτική.....	B-77
B.1.1.4 Λογισμός επί της Ικανοποίησης.....	B-78
B.1.1.5 Επίλυση της Συμφόρησης.....	B-79
<i>B.1.2 Περιβάλλον Πλαίσιο</i>	<i>B-79</i>
B.1.2.1 Μοντέλο Υπηρεσιών	B-79
B.1.2.2 Λειτουργικό Μοντέλο	B-87
<i>B.1.3 Ανακεφαλαίωση.....</i>	<i>B-103</i>
B.2 Έλεγχος Εισόδου Κίνησης.....	B-107
<i>B.2.1 Διαχείριση Εγγραφής.....</i>	<i>B-107</i>

B.2.1.1	Πρόβλημα	B-107
B.2.1.2	Παραδοχές και Περιβάλλον.....	B-107
B.2.1.3	Τοποθέτηση	B-107
B.2.1.4	Αλγόριθμος.....	B-109
B.2.2	<i>Διαχείριση Κλήσης</i>	B-112
B.2.2.1	Πρόβλημα	B-112
B.2.2.2	Παραδοχές και Περιβάλλον.....	B-113
B.2.2.3	Τοποθέτηση	B-114
B.2.2.4	Εφαρμογή Στρατηγικής Αποδοχής.....	B-116
B.2.2.5	Δυναμική Διαχείριση Αποδοχής	B-124
B.2.3	<i>Ανακεφαλαίωση</i>	B-143
ΜΕΡΟΣ C.	ΑΞΙΟΛΟΓΗΣΗ	C-147
C.1	Υλοποίηση και Πρακτική Εφαρμογή	C-147
C.1.1	<i>Μοντέλο Οντοτήτων Λογισμικού</i>	C-147
C.1.2	<i>Router Service Invocation Module</i>	C-148
C.1.2.1	Interfaces	C-148
C.1.2.2	Επιμερισμός υπολειτουργιών	C-151
C.1.2.3	Τεχνολογίες.....	C-153
C.1.2.4	Υλοποίηση σε στοιχεία δικτύου.....	C-154
C.2	Θεωρητική Μελέτη Πολυπλοκότητας.....	C-157
C.2.1	<i>Εξυπηρέτηση κλήσης</i>	C-158
C.2.2	<i>Δυναμική προσαρμογή</i>	C-159
C.3	Πειράματα.....	C-161
C.3.1	<i>Πειραματικό Πλαίσιο</i>	C-161
C.3.1.1	Λειτουργίες Πλατφόρμας Δοκιμών	C-161
C.3.1.2	Υλοποίηση Πλατφόρμας Δοκιμών.....	C-164
C.3.1.3	Στήσιμο Πειραμάτων	C-164
C.3.1.4	Διεξαγωγή Πειραμάτων.....	C-164
C.3.2	<i>Πειράματα</i>	C-165
C.3.2.1	Βασικό Στήσιμο Πειραμάτων	C-166
C.3.2.2	Σειρά Δοκιμών TS-R	C-168
C.3.2.3	Σειρά Δοκιμών TS-S.....	C-191
C.4	Συμπεράσματα.....	C-201
C.4.1	<i>Καινοτομία</i>	C-201
C.4.2	<i>Αξιολόγηση</i>	C-202
C.4.3	<i>Περαιτέρω Έρευνα</i>	C-204
ΜΕΡΟΣ D.	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	D-205

ΣΧΗΜΑΤΑ

Σχήμα A-1: ενεργοποίηση υπηρεσίας με RSVP.....	A-25
Σχήμα A-2: λειτουργικό μοντέλο κόμβου IntServ	A-26
Σχήμα A-3: λειτουργίες ακραίου κόμβου DiffServ.....	A-28
Σχήμα A-4: μορφοποιημένη κίνηση υπηρεσίας.....	A-40
Σχήμα A-5: διαμόρφωση λογικής επιφάνειας.....	A-55
Σχήμα A-6: ιεραρχική δομή κόμβων CH	A-55
Σχήμα B-1: το βασικό μοντέλο βήμα προς βήμα	B-73
Σχήμα B-2: παράδειγμα κατανομής της ζήτησης.....	B-75
Σχήμα B-3: εμπειρική ρύθμιση λειτουργίας βασικού μοντέλου	B-78
Σχήμα B-4: Service Subscription Structure.....	B-82
Σχήμα B-5: Service Level Specification	B-84
Σχήμα B-6: ιεραρχικό δομικό μοντέλο υπηρεσιών	B-85
Σχήμα B-7: λειτουργικό μοντέλο ολοκληρωμένου συστήματος παροχής ποιότητας	B-88
Σχήμα B-8: αρχικοποίηση Resource Provisioning Cycle.....	B-92
Σχήμα B-9: λειτουργία πρόβλεψης κίνησης.....	B-94
Σχήμα B-10: περιεχόμενο γραμμής Traffic Matrix	B-94
Σχήμα B-11: λειτουργία διαστασιοποίησης δικτύου.....	B-95
Σχήμα B-12: παράδειγμα βέλτιστης κατανομής.....	B-97
Σχήμα B-13: περιεχόμενο γραμμής Resource Availability Matrix ή Resource Availability Buffer	B-101
Σχήμα B-14: παράδειγμα Resource Availability Matrix	B-101
Σχήμα B-15: ερμηνεία Resource Availability Buffer.....	B-102
Σχήμα B-16: καθορισμός επιτρεπτής περιοχής βάσει του επιπέδου ικανοποίησης	B-110
Σχήμα B-17: λειτουργία διαχείρισης εγγραφής.....	B-111
Σχήμα B-18: λειτουργία διαχείρισης κλήσης	B-115
Σχήμα B-19: επίδραση στρατηγικής αποδοχής στη συνολική διοχετευόμενη κίνηση.....	B-117
Σχήμα B-20: παράδειγμα υλοποίησης υπηρεσίας στο επίπεδο δεδομένων.....	B-120
Σχήμα B-21: πληροφοριακό μοντέλο εφαρμογής στρατηγικής αποδοχής.....	B-122
Σχήμα B-22: διάγραμμα καταστάσεων αλγορίθμου δυναμικής διαχείρισης αποδοχής	B-124
Σχήμα B-23: καθορισμός κρίσιμης περιοχής βάσει του επιπέδου προφύλαξης.....	B-131
Σχήμα B-24: λειτουργία στην ομαλή κατάσταση.....	B-132
Σχήμα B-25: λειτουργία στην κατάσταση αντίδρασης.....	B-134
Σχήμα B-26: λειτουργία στην κατάσταση ομαλοποίησης.....	B-137
Σχήμα B-27: αυστηρότητα σε κατάσταση ομαλοποίησης έναντι ομαλής κατάστασης	B-138
Σχήμα B-28: σενάριο παλινδρομήσεων μεταξύ στρατηγικών.....	B-142
Σχήμα C-1: οντότητες λογισμικού λειτουργικού μοντέλου	C-148
Σχήμα C-2: φυσική αντιστοιχία οντοτήτων λογισμικού	C-148
Σχήμα C-3: RSIM interactions	C-149
Σχήμα C-4: υλοποίηση RSIM.....	C-151
Σχήμα C-5: υλοποίηση GAL	C-154
Σχήμα C-6: δικτυακό περιβάλλον εφαρμογής επί δρομολογητών Cisco	C-155
Σχήμα C-7: οντότητες πλατφόρμας δοκιμών	C-161
Σχήμα C-8: υλοποίηση RSIM και πλατφόρμας δοκιμών	C-164

Σχήμα C-9: τύποι υπηρεσιών και ρυθμίσεις προσομοίωσης εγγραφών	C-167
Σχήμα C-10: φυσική τοπολογία δοκιμών	C-168
Σχήμα C-11: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-1	C-171
Σχήμα C-12: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-2	C-171
Σχήμα C-13: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-3	C-172
Σχήμα C-14: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-4	C-172
Σχήμα C-15: αποτελέσματα πειράματος [TC-R-1].[T-3].....	C-175
Σχήμα C-16: αποτελέσματα πειράματος [TC-R-2].[T-0].....	C-176
Σχήμα C-17: αποτελέσματα πειράματος [TC-R-2].[T-1].....	C-177
Σχήμα C-18: αποτελέσματα πειράματος [TC-R-2].[T-2].....	C-178
Σχήμα C-19: αποτελέσματα πειράματος [TC-R-2].[T-3].....	C-179
Σχήμα C-20: συγκριτικά αποτελέσματα πειραμάτων υπόθεσης δοκιμών TC-R-2	C-180
Σχήμα C-21: αποτελέσματα πειράματος [TC-R-3].[T-0].....	C-181
Σχήμα C-22: αποτελέσματα πειράματος [TC-R-3].[T-1].....	C-182
Σχήμα C-23: αποτελέσματα πειράματος [TC-R-3].[T-2].....	C-183
Σχήμα C-24: αποτελέσματα πειράματος [TC-R-3].[T-3].....	C-184
Σχήμα C-25: συγκριτικά αποτελέσματα πειραμάτων υπόθεσης δοκιμών TC-R-3	C-185
Σχήμα C-26: αποτελέσματα πειράματος [TC-R-4].[T-0].....	C-186
Σχήμα C-27: αποτελέσματα πειράματος [TC-R-4].[T-1].....	C-187
Σχήμα C-28: αποτελέσματα πειράματος [TC-R-4].[T-2].....	C-188
Σχήμα C-29: αποτελέσματα πειράματος [TC-R-4].[T-3].....	C-189
Σχήμα C-30: συγκριτικά αποτελέσματα πειραμάτων υπόθεσης δοκιμών TC-R-4	C-190
Σχήμα C-31: αποτελέσματα πειράματος [TC-S-1].[T-1]	C-193
Σχήμα C-32: αποτελέσματα πειράματος [TC-S-1].[T-2]	C-194
Σχήμα C-33: αποτελέσματα πειράματος [TC-S-2].[T-1]	C-195
Σχήμα C-34: αποτελέσματα πειράματος [TC-S-2].[T-2]	C-196
Σχήμα C-35: αποτελέσματα πειράματος [TC-S-3].[T-1]	C-197
Σχήμα C-36: αποτελέσματα πειράματος [TC-S-3].[T-2]	C-198
Σχήμα C-37: αποτελέσματα πειράματος [TC-S-4].[T-2]	C-199
Σχήμα C-38: συγκριτικά αποτελέσματα πειραμάτων σειράς δοκιμών TS-S	C-200

ΠΙΝΑΚΕΣ

Πίνακας A-1: κατηγορίες εφαρμογών και απαιτήσεις ποιότητας υπηρεσίας.....	A-15
Πίνακας A-2: μελέτες κατανομής ελέγχου εισόδου κίνησης σε δίκτυα DiffServ	A-50
Πίνακας B-1: παράδειγμα παραμέτρων dynamic resource management	B-99
Πίνακας B-2: προδιαγεγραμμένες στρατηγικές	B-126
Πίνακας C-1: μέθοδοι των interfaces της οντότητας RSIM	C-149
Πίνακας C-2: υποθέσεις ακολουθίας δοκιμών TS-R.....	C-169
Πίνακας C-3: πειράματα ανά υπόθεση ακολουθίας δοκιμών TS-R.....	C-170
Πίνακας C-4: πειράματα ανά υπόθεση ακολουθίας δοκιμών TS-S.....	C-191

ΜΕΡΟΣ Α. ΕΙΣΑΓΩΓΗ

A.1 Ποιότητα Υπηρεσίας

A.1.1 Ιστορικό Πλαίσιο

Η γένεση των δικτύων δεδομένων βρήκε άλλα δίκτυα τηλεπικοινωνιών, όπως αυτό της τηλεφωνίας, να προϋπάρχουν. Ο διαχωρισμός μεταξύ της μεταγωγής πακέτων δεδομένων και άλλων υπηρεσιών, όπως της μεταγωγής αναλογικού σήματος φωνής, υπήρξε σε αυτή την πρώιμη εποχή φυσικό και αναμφισβήτητο γεγονός.

Μόλις πριν μερικά χρόνια μία επιχείρηση θα χρησιμοποιούσε TDM (Time Division Multiplexing – TDM) για μετάδοση φωνής, IP για πρόσβαση στο Internet, ISDN (Integrated Services Digital Network – ISDN) για video-conferencing εφαρμογές, SNA (Systems Network Architecture – SNA) και IPX (Internetwork Packet Exchange – IPX) ή άλλα πρωτόκολλα για LAN (Local Area Network – LAN). Αντίστοιχα, ένας παροχέας υπηρεσιών θα χρησιμοποιούσε TDM για φωνή, ATM (Asynchronous Transfer Mode – ATM) ή SONET (Synchronous Optical Network – SONET) δίκτυο κορμού και Frame Relay (Frame Relay) ή ISDN δίκτυο πρόσβασης.

Κάθε τέτοιο δίκτυο εξυπηρετεί κίνηση με ομοιογενή συμπεριφορά και είναι καλά προσαρμοσμένο στις απαιτήσεις των αντίστοιχων εφαρμογών. Ωστόσο, η χρήση ξεχωριστών δικτύων επιφέρει υψηλότερο λειτουργικό κόστος και χαμηλότερο συντελεστή εκμετάλλευσης. Με την ταχύτατη εξέλιξη του τομέα, δημιουργήθηκε η τάση ενοποίησης των διαφορετικών υποδομών, σε μία κοινή υποδομή δικτύου που χρησιμοποιεί μεταγωγή πακέτου και τους κατάλληλους μηχανισμούς ώστε να υποστηρίξει υπηρεσίες διαφορετικής φύσεως.

Η αλματώδης ανάπτυξη του διαδικτύου κατέστησε το πρωτόκολλο IP το de facto standard πρωτόκολλο μεταγωγής πακέτου. Με βάση το IP, πέραν των εφαρμογών μετάδοσης δεδομένων, αναπτύχθηκαν και πληθώρα νέων εφαρμογών πραγματικού χρόνου για μετάδοση φωνής και video, εφαρμογές που βρήκαν γρήγορα τη θέση τους στη νέα αγορά του διαδικτύου.

Οι παροχές υπηρεσιών ανταποκρίθηκαν αναπτύσσοντας ένα μεγάλο και ολοένα επεκτεινόμενο φάσμα υπηρεσιών, κατάλληλων να εξυπηρετήσουν τις νέες εφαρμογές, επιτρέποντας έτσι τον ανταγωνισμό σε ένα ακόμα επίπεδο και τη διαφοροποίηση του κάθε οργανισμού σε αυτή την αγορά σύμφωνα με τις διαφορετικές επιχειρησιακές επιδιώξεις και πολιτικές του.

Δεδομένης μιας τέτοιας εξέλιξης, η έννοια της ποιότητας υπηρεσίας (Quality of Service – QoS) επινοήθηκε και μελετάται ως το απαραίτητο στοιχείο που επιτρέπει σε μία κοινή υποδομή δικτύου την υποστήριξη πολλαπλών υπηρεσιών για κίνηση με ανομοιογενή συμπεριφορά και διαφορετικές απαιτήσεις σε πόρους και εγγυήσεις απόδοσης του δικτύου.

A.1.2 Ορισμός Ποιότητας Υπηρεσίας

Ένας από τους εύστοχους γενικούς ορισμούς που απαντώνται στη βιβλιογραφία αναφέρει πως "Ο όρος Ποιότητα Υπηρεσίας χρησιμοποιείται για να περιγράψει τη συνολική εμπειρία που αποκομίζει ένας χρήστης ή μια εφαρμογή από το δίκτυο" [NortelQoS].

Πέραν αυτής της διαισθητικής περιγραφής του όρου, η επιστημονική κοινότητα καλείται να προσδιορίσει με ακρίβεια εκείνες τις παραμέτρους που διέπουν την Ποιότητα Υπηρεσίας ανεξάρτητα από τα χαρακτηριστικά μιας συγκεκριμένης εφαρμογής ή την υποκειμενική αντίληψη του χρήστη.

Συναίνεση ως προς τον ακριβή ορισμό της έννοιας ποιότητας υπηρεσίας δεν έχει επιτευχθεί ακόμα. Ο κυριότερος λόγος πιστεύεται πως είναι το μεγάλο εύρος θεμάτων που καλύπτει η ίδια η έννοια, και κατά συνέπεια οι κατά κόρον αποσπασματικές προσεγγίσεις που έχουν επιχειρηθεί από μεμονωμένους ερευνητές ή/και οργανισμούς υπό το εκάστοτε ίδιον πρίσμα.

Η *Υπηρεσία* αναφέρεται στο εμπορικό αγαθό που προσφέρεται από τους παροχείς υπηρεσιών δικτύου στους χρήστες της υπηρεσίας, το οποίο περιγράφεται από συγκεκριμένους και με ακρίβεια καθορισμένους όρους και παραμέτρους, ανάλογα με το επίπεδο και τον τύπο προσφοράς του κάθε παροχέα. Διαφορετικά επίπεδα υπηρεσίας μπορεί να είναι το επίπεδο μεταφοράς, το επίπεδο εφαρμογής, το επίπεδο περιεχομένου.

Η παρούσα εργασία εστιάζεται σε υπηρεσίες επιπέδου μεταφοράς.

Η *Ποιότητα* σε ό,τι αφορά τη μεταφορά πακέτων σε ένα δίκτυο για μια δεδομένη ροή κίνησης από άκρο σε άκρο, διακρίνεται από την *επίδοση* (performance), την *αξιοπιστία* (reliability) και την *ένταση* (volume) που χαρακτηρίζουν τη μετάδοση της ροής και προσμετράται από:

- (1) την καθυστέρηση κατά την μεταφορά των πακέτων από την πηγή στον προορισμό τους, τόσο από το μέσο όρο του μεγέθους (delay) όσο και από τη διασπορά του (jitter),
- (2) το ποσοστό απώλειας πακέτων (loss) που σημειώνεται είτε εξαιτίας της υπερχειλίσης των αποταμιευτών μνήμης στις ουρές αναμονής μετάδοσης των δρομολογητών, είτε λόγω αλλοίωσης από το θόρυβο των φυσικών μέσων μετάδοσης, είτε από αναδιάταξη (reordering) λόγω μεταφοράς πακέτων της ίδιας ροής μέσω διαφορετικών μονοπατιών,
- (3) τον ελάχιστο ρυθμό ροής (throughput) της κίνησης για την οποία ισχύουν καθορισμένες τιμές των (1) και (2).

Η *Ποιότητα Υπηρεσίας* μεταφοράς αναφέρεται σε εκείνα τα χαρακτηριστικά της υπηρεσίας που καθορίζουν την ποιότητα μεταφοράς (1-3) καθώς και επιπλέον χαρακτηριστικά που περιγράφουν τις δεσμεύσεις του παροχέα ως προς την παροχή της συμφωνηθείσας ποιότητας αλλά και τις προϋποθέσεις που οφείλει να πληροί ο χρήστης για να κάνει χρήση της προσφερόμενης υπηρεσίας.

Οι δεσμεύσεις του παροχέα εκφράζονται σε σχέση με τη *διαθεσιμότητα* (availability) της προσφερόμενης υπηρεσίας. Χαρακτηριστικά του τύπου μπορεί να αποτελέσουν το ωράριο διάθεσης της υπηρεσίας, ο καθορισμός των σημείων πρόσβασης, το ποσοστό επιτυχούς ενεργοποίησης προς τον συνολικό αριθμό των αποπειρών ενεργοποίησης από το χρήστη, το ποσοστό επιτυχούς ολοκλήρωσης χρήσης της υπηρεσίας δεδομένης επιτυχούς ενεργοποίησης, ο μέγιστος χρόνος ανάνηψης σε περίπτωση βλάβης, ποσοστά επιτρεπτής απόκλισης επί των συμφωνηθέντων χαρακτηριστικών (1-3) κ.ά.

Οι προϋποθέσεις που οφείλει να πληροί ένας χρήστης εκφράζονται σε σχέση με τα χαρακτηριστικά της κίνησης που πρόκειται να αποστείλει ή/και να λάβει ο χρήστης μέσω του δικτύου του παροχέα. Ως τέτοια χρησιμοποιούνται ο μέσος ρυθμός ροής (mean rate), ο μέγιστος ρυθμός ροής (peak rate), το μέγεθος του μέγιστου καταιγισμού (burst size), το μέγιστο μέγεθος πακέτου (packet size), κ.ά.

Από τα παραπάνω διακρίνονται οι δύο άξονες της ποιότητας υπηρεσίας, ο χρήστης και οι εφαρμογές που παράγουν την κίνηση από τη μία πλευρά, το δίκτυο, οι πόροι και οι μηχανισμοί που εγγυώνται τα χαρακτηριστικά ποιότητας μετάδοσης αυτής της κίνησης από την άλλη. Για την υλοποίηση της ποιότητας υπηρεσίας σε ένα τέτοιο σύστημα είναι έτσι απαραίτητο ένα κοινό πλαίσιο μέσα στο οποίο θα λειτουργούν

οι χρήστες και οι παροχείς. Η σύναψη συμβολαίων μεταξύ των δύο μερών, χρήστη και παροχέα ή ακόμα και μεταξύ παροχέων, επιβάλλει την ύπαρξη από κοινού αποδεκτού τρόπου προσδιορισμού της ποιότητας υπηρεσίας, έτσι ώστε ο μεν χρήστης να μπορεί επακριβώς να προσδιορίσει τις απαιτήσεις του και ο δε παροχέας να συναγάγει και υλοποιήσει τις απαραίτητες ρυθμίσεις στο δίκτυό του.

Η σημασία της προτυποποίησης των χαρακτηριστικών που διέπουν τη συμφωνία μεταξύ ενός παροχέα και ενός χρήστη για την ποιότητα μιας προσφερόμενης υπηρεσίας, καθώς και της σύνταξης και σημειολογίας ενός συμβολαίου που επικυρώνει τη συμφωνία, έχει γίνει αντιληπτή από την επιστημονική κοινότητα και διάφορες ερευνητικές προσπάθειες έχουν πραγματοποιηθεί σχετικά.

Στο πλαίσιο της αρχιτεκτονικής δικτύου διαφοροποιημένων υπηρεσιών [DS-ARCH, DS-TERM] γίνεται ήδη ένας διαχωρισμός μεταξύ των τεχνικών χαρακτηριστικών της ποιότητας μεταφοράς και των χαρακτηριστικών της ποιότητας υπηρεσίας στις έννοιες της Συμφωνίας Επιπέδου Υπηρεσίας (Service Level Agreement – SLA) και της Προδιαγραφής Επιπέδου Υπηρεσίας (Service Level Specification – SLS). Μια προσπάθεια αποσαφήνισης και προτυποποίησης αυτών των εννοιών αποτελεί η πρόταση [SLS-T], οι αρχές της οποίας χρησιμοποιήθηκαν κατά την εκπόνηση της παρούσας διατριβής.

A.1.3 Ποιότητα Υπηρεσίας για το Χρήστη – Εφαρμογές

Ανάλογα με το σκοπό που εξυπηρετεί, η κάθε εφαρμογή καλείται να ανταποκριθεί σε διαφορετικές προσδοκίες του χρήστη. Οι προσδοκίες του χρήστη απορρέουν από την ανθρώπινη αντίληψη για την ταχύτητα απόκρισης και την ποιότητα ενός μέσου όπως η φωνή ή το video. Έτσι, η κάθε εφαρμογή παράγει κίνηση με συγκεκριμένη συμπεριφορά, ανάλογα με τις απαιτήσεις της θέτει κάποιες ελάχιστες απαιτήσεις ποιότητας της υπηρεσίας μεταφοράς επί της οποίας λειτουργεί και υλοποιεί μηχανισμούς για την εξομάλυνση παροδικών αποκλίσεων από αυτές τις ελάχιστες απαιτήσεις ποιότητας.

Οι εφαρμογές κατατάσσονται σε διακριτές κατηγορίες ανάλογα με τα χαρακτηριστικά της κίνησης που παράγουν και τις απαιτήσεις σε ποιότητα υπηρεσίας μετάδοσης όπως φαίνεται στον Πίνακα A-1.

Κατηγορία Εφαρμογών	Παραδείγματα Εφαρμογών	Απαιτήσεις Ποιότητας		
		Μέση Καθυστέρηση	Διακύμανση Καθυστέρησης	Ποσοστό Απώλειας
Αλληλεπίδρασης	VoIP, video conferencing, interactive gaming	****	****	***
Απόκρισης	web browsing, server access, e-commerce	***	***	****
Συνεχούς Ροής	VoD, internet radio, news broadcast, teletraining	***	***	***
Παρασκηνίου	e-mail, μεταφορά αρχείων	*	*	****
Ελέγχου Δικτύου	συναγερμοί παρακολούθησης, δρομολόγηση, σηματοδосία	****	*	****

Πίνακας A-1: κατηγορίες εφαρμογών και απαιτήσεις ποιότητας υπηρεσίας

A.1.3.1 Εφαρμογές Αλληλεπίδρασης

Εφαρμογές αλληλεπίδρασης (interactive applications) καλούνται εκείνες οι εφαρμογές δια μέσου των οποίων δύο ή περισσότεροι χρήστες συμμετέχουν ενεργά σε μια αμφίδρομη επικοινωνία. Παραδείγματα τέτοιων εφαρμογών αποτελούν οι: τηλεφωνία μέσω διαδικτύου (Voice over Internet Protocol – VoIP), τηλεδιάσκεψη (video conferencing), interactive gaming κ.ά.

Οι συμμετέχοντες προσδοκούν η αμφίδρομη επικοινωνία μέσω του δικτύου να είναι πραγματικού χρόνου, να ανταποκρίνεται δηλαδή στα χαρακτηριστικά της αντίστοιχης επικοινωνίας χωρίς τη μεσολάβηση του δικτύου. Η ανθρώπινη αντίληψη είναι δεκτική ως ένα βαθμό όσον αφορά καθυστερήσεις στο χρόνο απόκρισης και υποβάθμιση της ποιότητας του μέσου που χρησιμοποιείται, είτε πρόκειται για φωνή είτε για video. Η παραβίαση αυτών των ορίων καθιστά την επικοινωνία ανώφελη. Ως εκ τούτου, οι εφαρμογές αλληλεπίδρασης εγείρουν τις μεγαλύτερες απαιτήσεις για ποιότητα υπηρεσίας.

Καθυστερήσεις στην μεταφορά των πακέτων δημιουργούν την αίσθηση πως η εφαρμογή δεν αποκρίνεται, επιβάλλοντας με αυτό τον τρόπο αυστηρά άνω όρια στη μέση καθυστέρηση.

Αυστηρά μικρή πρέπει να διατηρείται και η διακύμανση της καθυστέρησης. Η ψηφιακή μετάδοση αναλογικών οπτικοακουστικών μέσων πραγματοποιείται με την αποστολή διακριτών στιγμιότυπων του σήματος που έχουν ληφθεί με συγκεκριμένη συχνότητα, τα οποία παρατάσσονται με την ίδια συχνότητα και ανασυνθέτουν το αναλογικό σήμα στον προορισμό. Εάν τα πακέτα που μεταφέρουν αυτά τα στιγμιότυπα φτάνουν με καθυστερήσεις που διαφέρουν κατά πολύ τότε το αναλογικό σήμα στον προορισμό παρουσιάζει παραμόρφωση. Οι εφαρμογές που διαχειρίζονται οπτικοακουστικά μέσα αντιμετωπίζουν τις τυχόν μεγάλες διακυμάνσεις στην καθυστέρηση μεταφοράς των πακέτων καταχωρώντας τα πακέτα που φτάνουν πρώτα και ξεκινώντας την ανασύνθεση του σήματος μετά από κάποιο χρονικό διάστημα (offset) ώστε να επιτρέψουν στα πακέτα με τη μεγαλύτερη καθυστέρηση να φτάσουν εγκαίρως για την ανασύνθεση του σήματος (μηχανισμός buffering). Ενώ σε εφαρμογές άλλων κατηγοριών όπως αυτές της συνεχούς ροής (βλέπε ενότητα A.1.3.3) αυτός ο μηχανισμός λύνει το πρόβλημα, στις εφαρμογές αλληλεπίδρασης δεν μπορεί να χρησιμοποιηθεί καθώς το χρονικό διάστημα που μεσολαβεί μέχρι να ξεκινήσει η ανασύνθεση του σήματος εισάγει καθυστέρηση στον χρόνο απόκρισης. Τα πακέτα που φτάνουν με μεγαλύτερη από την ελάχιστη καθυστέρηση απλά απορρίπτονται προξενώντας έτσι υποβάθμιση της ποιότητας του αναλογικού σήματος.

Αντίστοιχα υποβαθμίζεται η ποιότητα του αναλογικού σήματος λόγω απώλειας πακέτων από την υπηρεσία μεταφοράς. Ακόμα περισσότερο σημαντική είναι η απώλεια πακέτων για εφαρμογές αλληλεπίδρασης που αφορούν τη μετάδοση δεδομένων. Είναι χαρακτηριστικό πως οι εφαρμογές αλληλεπίδρασης κατά την πλειοψηφία τους δεν χρησιμοποιούν το Transport Control Protocol (TCP) που διασφαλίζει την αξιόπιστη μεταφορά των πακέτων, αλλά το Universal Datagram Protocol (UDP) πρωτόκολλο μεταφοράς. Ο μηχανισμός επαναμετάδοσης πακέτου σε περίπτωση απώλειας που παρέχει το TCP δεν χρησιμεύει στις εφαρμογές αλληλεπίδρασης καθώς η εγκυρότητα του κάθε πακέτου συνδέεται άρρηκτα με τον χρόνο που έχει παρέλθει από την αποστολή του.

A.1.3.2 Εφαρμογές Απόκρισης

Στις *εφαρμογές απόκρισης* (responsive applications) η επικοινωνία λαμβάνει χώρα μεταξύ ενός χρήστη και μιας απομακρυσμένης αυτοματοποιημένης διαδικασίας. Τέτοιες εφαρμογές είναι οι: περιήγηση στον ιστό (web browsing), πρόσβαση σε εξυπηρετητή (server access), ηλεκτρονικού εμπορίου (e-commerce), κ.ά.

Οι εφαρμογές απόκρισης συχνά αναφέρονται και ως "σχεδόν πραγματικού χρόνου" καθώς ο χρήστης αναμένει η απόκριση από την απομακρυσμένη διαδικασία να είναι κατά το δυνατόν άμεση, αποδεχόμενος

ωστόσο κάποια καθυστέρηση ανάλογη και με την πολυπλοκότητα της απομακρυσμένης διαδικασίας, όπως αυτή γίνεται αντιληπτή από το χρήστη.

Ο χρήστης καλεί την απομακρυσμένη διαδικασία αποστέλλοντας τα απαραίτητα δεδομένα. Στην περίπτωση του web browsing αυτό μεταφράζεται στην επιλογή του υπερσυνδέσμου (hyperlink) για τη μετάβαση σε νέα ιστοσελίδα (web page), σε e-commerce εφαρμογές τη συμπλήρωση μιας φόρμας παραγγελίας, την εντολή πληρωμής, τη διαμόρφωση της ερώτησης προς μια απομακρυσμένη βάση δεδομένων κ.λπ.

Η εφαρμογή επιβεβαιώνει την αποστολή των δεδομένων από το χρήστη και αναμένει την λήψη των αποτελεσμάτων. Η περάτωση της διαδικασίας σε εύλογο χρονικό διάστημα είναι πολύ σημαντική καθώς μεγάλες καθυστερήσεις μπορεί να ερμηνευθούν από το χρήστη ως σφάλμα του δικτύου και να επαναλάβει τη διαδικασία. Εκτός από κακή διαχείριση του χρόνου του χρήστη και των πόρων του δικτύου, η επανέναρξη μιας διαδικασίας μπορεί να αποσυντονίσει τη διαδικασία ή και να την εκτελέσει για δεύτερη φορά, κάτι που στην περίπτωση π.χ. μιας οικονομικής συναλλαγής είναι ανεπιθύμητο.

Ενώ η άμεση και αξιόπιστη απόκριση προς το χρήστη εγείρει υψηλές απαιτήσεις σε ποιότητα υπηρεσίας μεταφοράς, ωστόσο τα περιθώρια απόκλισης είναι μεγαλύτερα από αυτά των εφαρμογών αλληλεπίδρασης όσον αφορά την μέση καθυστέρηση και τη διακύμανση καθυστέρησης, μικρότερα ωστόσο όσο αφορά την απώλεια πακέτων. Το πρωτόκολλο TCP χρησιμοποιείται κατά κόρον για να διασφαλίσει την αξιόπιστη ανταλλαγή των δεδομένων του χρήστη και των αποτελεσμάτων της διαδικασίας. Εάν η υπηρεσία μεταφοράς του δικτύου παρουσιάζει αυξημένο ποσοστό απώλειας, το TCP θα πραγματοποιήσει πολλές αναμεταδόσεις, αυξάνοντας σημαντικά το χρόνο απόκρισης.

A.1.3.3 Εφαρμογές Συνεχούς Ροής

Οι *εφαρμογές συνεχούς ροής* (streaming applications) είναι ιδιότυπες εφαρμογές απόκρισης, όπου ο χρήστης αιτείται και ελέγχει την μετάδοση ενός οπτικοακουστικού μέσου συνεχόμενης ροής από την απομακρυσμένη διαδικασία. Τέτοιες εφαρμογές χρησιμοποιούνται για ραδιόφωνο και τηλεόραση μέσω διαδικτύου, video κατά αίτηση (Video On Demand – VoD), τηλεεκπαίδευση (teletraining), κ.ά.

Όπως και με τις εφαρμογές αλληλεπίδρασης που διαχειρίζονται οπτικοακουστικά μέσα, έτσι και στις εφαρμογές συνεχούς ροής, η επικοινωνία από την απομακρυσμένη διαδικασία προς το χρήστη πρέπει να διακρίνεται από χαμηλή μέση καθυστέρηση και διακύμανση καθυστέρησης. Ωστόσο, χάρη στον μηχανισμό buffering (βλέπε ενότητα A.1.3.1), οι απαιτήσεις ποιότητας των εφαρμογών συνεχούς ροής είναι κατά πολύ μικρότερες από τις αντίστοιχες των εφαρμογών αλληλεπίδρασης.

Ενώ ο χρόνος απόκρισης είναι αποδεκτό να είναι πολύ μεγαλύτερος από αυτόν που προσδοκά ένας χρήστης από το συνομιλητή του σε μια εφαρμογή VoIP, δεν μπορεί να είναι αυθαίρετα μεγάλος. Σε μια εφαρμογή VoD για παράδειγμα, θα πρέπει οι εντολές παγώματος (freeze), αναστροφής (reverse play), προώθησης (forward) που δίνει ο χρήστης να εκτελούνται σε εύλογο χρονικό διάστημα.

Όπως και με τις αντίστοιχες εφαρμογές αλληλεπίδρασης, η απώλεια επηρεάζει την ποιότητα του μέσου. Ο μηχανισμός buffering συνεισφέρει και ως προς αυτή την παράμετρο καθώς προσφέρει το χρονικό περιθώριο που είναι απαραίτητο μεταξύ της διάγνωσης της απώλειας ενός πακέτου και της επαναμετάδοσής του πριν παρέλθει η σειρά του στην ανασύνθεση του αναλογικού σήματος. Το πρωτόκολλο που χρησιμοποιείται είναι και εδώ κατά κανόνα το UDP και η επαναμετάδοση των απολεσθέντων πακέτων υλοποιείται από την ίδια την εφαρμογή.

A.1.3.4 Εφαρμογές Παρασκηνίου

Εφαρμογές παρασκηνίου (background applications) χαρακτηρίζονται οι εφαρμογές που μετά την ενεργοποίησή τους εκτελούνται στο παρασκήνιο χωρίς να απαιτούν άλλη παρέμβαση από το χρήστη. Εφαρμογές αυτής της κατηγορίας αποτελούν το ηλεκτρονικό ταχυδρομείο (e-mail) και η μεταφορά αρχείων (file transfer).

Ο χρήστης μιας εφαρμογής παρασκηνίου αντιλαμβάνεται μόνο την επιτυχημένη περάτωσή της, όπου επιτυχημένη περάτωση είναι η ορθή μεταφορά όλων των δεδομένων.

Το πρωτόκολλο TCP χρησιμοποιείται, διασφαλίζοντας την ορθή μεταφορά των δεδομένων. Υψηλό ποσοστό απώλειας της υπηρεσίας μεταφοράς έχει ως αποτέλεσμα περισσότερες αναμεταδόσεις από το TCP, αυξάνοντας σημαντικά το χρόνο περάτωσης της διαδικασίας και μειώνοντας την αποτελεσματικότητα των εφαρμογών. Ενώ η διακύμανση της καθυστέρησης έχει αμελητέα επιρροή στην απόδοση των εφαρμογών παρασκηνίου, ο χρόνος περάτωσης της εφαρμογής είναι γραμμικά ανάλογος με τη μέση καθυστέρηση της υπηρεσίας μεταφοράς.

Το άνω όριο του χρόνου περάτωσης και κατά συνέπεια οι απαιτήσεις σε χαρακτηριστικά ποιότητας υπηρεσίας εξαρτώνται κυρίως από το γενικότερο πλαίσιο στο οποίο χρησιμοποιείται η εκάστοτε εφαρμογή. Η αποστολή και λήψη e-mail έχει πολύ διαφορετική σημασία για τους οικιακούς (residential) χρήστες από ό,τι για τα στελέχη κάποιας εταιρείας.

Γενικά, για την αποτελεσματική λειτουργία των εφαρμογών παρασκηνίου τίθενται αυστηρά όρια στην απώλεια πακέτων, ενώ η μέση καθυστέρηση και η διακύμανση της καθυστέρησης παίζουν δευτερεύοντα ρόλο.

A.1.3.5 Εφαρμογές Ελέγχου Δικτύου

Στις *εφαρμογές ελέγχου δικτύου* (network control applications) δεν εμπλέκεται τελικός χρήστης, η επικοινωνία λαμβάνει χώρα μεταξύ των κατανεμημένων διαδικασιών ελέγχου του δικτύου. Τέτοιες εφαρμογές μπορεί να χρησιμοποιούνται για τη δρομολόγηση των πακέτων, για πρωτόκολλα σηματοδότησης, για την προώθηση συναγερμών σε κεντρικά σημεία ελέγχου και διαχείρισης των πόρων του δικτύου, κ.λπ.

Η εξασφάλιση της ποιότητας που απαιτούν οι εφαρμογές ελέγχου δικτύου είναι αποφασιστικής σημασίας καθώς εκεί στηρίζεται ο μηχανισμός ανάνηψης του δικτύου σε περίπτωση δυσλειτουργίας που επηρεάζει και τις εφαρμογές των χρηστών. Πολλοί παροχείς χρησιμοποιούν ξεχωριστό φυσικό δίκτυο για τη μετάδοση δεδομένων ελέγχου δικτύου (out-of-band έλεγχος). Όταν οι φυσικοί πόροι του δικτύου χρησιμοποιούνται από κοινού από εφαρμογές ελέγχου δικτύου και εφαρμογές χρηστών (in-band έλεγχος), οι πρώτες θα πρέπει να μεταδίδονται πάντα με απόλυτη προτεραιότητα σε σχέση με τις δεύτερες.

Οι εφαρμογές ελέγχου δικτύου έχουν υψηλές απαιτήσεις σε μέση καθυστέρηση και απώλεια. Ένας συναγερμός που λαμβάνεται με καθυστέρηση ή μία αλλαγή δρομολόγησης που δεν φτάνει σε κάποιο κόμβο μπορεί να έχουν σοβαρό αντίκτυπο στην απόδοση του δικτύου. Η διακύμανση καθυστέρησης δεν επηρεάζει σημαντικά αυτές τις εφαρμογές.

A.1.4 Ποιότητα Υπηρεσίας για το Δίκτυο

Η TCP/IP στοίβα πρωτοκόλλων, γνωστή απλά και ως IP, είναι αναμφισβήτητα ο κοινός παρονομαστής στο διαδίκτυο. Αντίθετα, οι τεχνολογίες των υποκείμενων επιπέδων διαφέρουν από επικράτεια σε επικράτεια με αποτέλεσμα να δημιουργούνται ασυμβατότητες και ασυνέχειες στην παροχή ποιότητας υπηρεσίας σε αυτά

τα επίπεδα. Έτσι, παρότι υπάρχουν τεχνολογίες φυσικού επιπέδου και επιπέδου σύνδεσης δεδομένων που υποστηρίζουν παροχή ποιότητας υπηρεσίας, η από άκρο σε άκρο ποιότητα υπηρεσίας εξασφαλίζεται στο επίπεδο δικτύου και μεταφοράς από τους μηχανισμούς του IP πρωτοκόλλου. Στη συνέχεια θα αναφερόμαστε αποκλειστικά σε αυτό το επίπεδο.

Το παραδοσιακό IP είναι πρωτόκολλο χωρίς σύνδεση (connectionless) και δεν υποστηρίζει μηχανισμούς για τη διάκριση της κίνησης βάσει του τύπου της εφαρμογής που εξυπηρετεί και κατ' επέκταση της ποιότητας που πρέπει να λάβει. Η κοινή αυτή υπηρεσία που προσφέρει το IP ονομάζεται *βέλτιστης προσπάθειας* (best effort). Η ποιότητα υπηρεσίας που λαμβάνει η κάθε ροή είναι μεταβλητή και εξαρτάται από τα επίπεδα συμφόρησης του μονοπατιού που ακολουθεί τη συγκεκριμένη στιγμή. Είναι συνεπώς αδύνατη η πρόβλεψη και η διασφάλιση συγκεκριμένων και σταθερών επιπέδων ποιότητας όπως απαιτείται από την εκάστοτε εφαρμογή.

Για να ξεπεραστεί ο περιορισμός της μοναδικής κοινής υπηρεσίας που παρέχει το IP, απαιτούνται μηχανισμοί που να καθιστούν εφικτή τη διαφοροποίηση στη μεταχείριση της κάθε ροής κίνησης, δημιουργώντας έτσι τη δυνατότητα παροχής περισσότερων από μία υπηρεσιών με διαφορετικά χαρακτηριστικά ποιότητας υπηρεσίας. Επιπλέον, η διασφάλιση συγκεκριμένων τιμών για τα χαρακτηριστικά ποιότητας των διαφορετικών υπηρεσιών που μπορούν να προκύψουν προϋποθέτει έλεγχο και συγκράτηση των επιπέδων συμφόρησης του δικτύου.

Αρχιτεκτονικές για παροχή ποιότητας υπηρεσίας καθορίζουν την μεθόδευση και ενορχήστρωση των ως άνω μηχανισμών ενώ υπερκείμενες λειτουργίες διαχείρισης είναι υπεύθυνες για τη ρύθμιση της καλής λειτουργίας τους.

A.1.4.1 Μηχανισμοί

Η ποιότητα υπηρεσίας που λαμβάνει ένα πακέτο σε επίπεδο δρομολογητή αφορά στη καθυστέρηση στον καταχωρητή μετάδοσης και την ενδεχόμενη απώλεια του πακέτου λόγω του πεπερασμένου μεγέθους των καταχωρητών μετάδοσης. Οι μηχανισμοί που ρυθμίζουν αυτές τις λειτουργίες είναι αντίστοιχα ο *μηχανισμός χρονοπρογραμματισμού* (scheduling) και ο *μηχανισμός διαχείρισης καταχωρητή* (buffer management).

Σε επίπεδο δικτύου η εξυπηρέτηση ενός πακέτου αφορά στον καθορισμό και την επιλογή της ακολουθίας των δρομολογητών που θα προωθήσουν το πακέτο στον τελικό προορισμό του. Αυτή η διαδικασία είναι δυνατό να συνυπολογίζει κριτήρια ποιότητας υπηρεσίας και να αποφαινεται το μονοπάτι που θα ακολουθήσει μια συγκεκριμένη ροή σύμφωνα με τη αθροιστική συμπεριφορά των δρομολογητών και κατά συνέπεια την από άκρο σε άκρο συμπεριφορά του δικτύου σε κάθε δυνατό μονοπάτι.

Για να είναι σε θέση ένας παροχέας να υποστηρίξει υπηρεσίες ποιότητας πρέπει να μπορεί να εγγυηθεί σε κάποιο βαθμό την ποιότητα των πακέτων που θα εξυπηρετηθούν από την υποδομή δικτύου που διαθέτει. Δεδομένης της ύπαρξης των κατάλληλων μηχανισμών διαφοροποίησης της εξυπηρέτησης ενός πακέτου η ποιότητα εξυπηρέτησης εξαρτάται επιπλέον από τα επίπεδα συμφόρησης στο δίκτυο που δημιουργούνται από τις ροές που το χρησιμοποιούν. Η συνολική προσφερόμενη κίνηση ενδέχεται να υπερβαίνει τις δυνατότητες εξυπηρέτησης του δικτύου, έχοντας ως αποτέλεσμα την υποβάθμιση της ποιότητας των προσφερόμενων υπηρεσιών. Καθώς η φυσική χωρητικότητα της κάθε υποδομής είναι πεπερασμένη, ο παροχέας πρέπει να είναι σε θέση να ελέγχει τον αριθμό, τον τύπο και την ένταση των ενεργών ροών που έχουν πρόσβαση στους πόρους του δικτύου. Ο μηχανισμός που φέρνει εις πέρας αυτές τις λειτουργίες ονομάζεται *έλεγχος εισόδου κίνησης* (admission control).

Στις παρακάτω ενότητες εξετάζονται λεπτομερέστερα κάποιοι από τους πλέον ευρέως διαδεδομένους μηχανισμούς και τεχνολογίες που χρησιμοποιούνται για την παροχή ποιότητας υπηρεσίας σε δίκτυα IP.

A.1.4.1.1 Χρονοπρογραμματισμός

Ο μηχανισμός χρονοπρογραμματισμού καθορίζει τη σειρά με την οποία μεταδίδονται τα πακέτα και κατ' επέκταση την καθυστέρηση που υφίσταται το καθένα. Πακέτα που ανήκουν σε μια εφαρμογή VoIP με ευαισθησία στην καθυστέρηση θα πρέπει να μεταδίδονται με προτεραιότητα έναντι των πακέτων μιας εφαρμογής e-mail. Για να επιτευχθεί παρόμοια διαφοροποίηση στη συμπεριφορά ενός μηχανισμού χρονοπρογραμματισμού χρησιμοποιούνται περισσότεροι του ενός καταχωρητές, ο καθένας από τους οποίους λαμβάνει διαφορετική μεταχείριση από τη διαδικασία επιλογής του επόμενου πακέτου προς μετάδοση. Τέτοιοι μηχανισμοί είναι οι: απόλυτης προτεραιότητας (strict Priority Queuing – PQ), δίκαιος με βάρη (Weighted Fair Queuing – WFQ) [DKS90], κυκλικής σειράς με βάρη (Weighted Round Robin – WRR) και WRR με έλλειμμα (Deficit Weighted Round Robin) [K97], υβριδικό όπως ο βασιζόμενος σε κλάσεις (Class Based Queuing – CBQ) [FJ95] κ.ά.

Με κάθε τέτοιο μηχανισμό σε κάθε διαφορετικό καταχωρητή αποδίδεται τελικά κάποιο μερίδιο του εύρους ζώνης της φυσικής ζεύξης μετάδοσης. Δεδομένων των ρυθμίσεων του κάθε μηχανισμού, το μερίδιο αυτό εξαρτάται από το φόρτο των καταχωρητών. Διακρίνουμε έτσι μεταξύ των μηχανισμών *σταθερού έργου* (work-conserving) και *μη σταθερού έργου* (not work-conserving). Στους μηχανισμούς work-conserving το μερίδιο εύρους ζώνης που αποδίδεται σε κάθε καταχωρητή είναι σταθερό ανεξάρτητα με το αν ο συνολικός φόρτος είναι χαμηλός και αφήνει μέρος του εύρους ζώνης αναξιοποίητο. Αντίθετα, στους μηχανισμούς not work-conserving όταν υπάρχει εναπομένον εύρος ζώνης αυτό μοιράζεται μεταξύ των καταχωρητών που έχουν πακέτα προς μετάδοση αυξάνοντας έτσι το μερίδιο που τους αντιστοιχεί.

A.1.4.1.2 Διαχείριση Καταχωρητή

Ο μηχανισμός διαχείρισης καταχωρητή καθορίζει τα πακέτα που σε περιόδους συμφόρησης δεν μπορούν να εισαχθούν στον καταχωρητή και άρα διαγράφονται. Καθώς ο μηχανισμός αποφυγής συμφόρησης (congestion avoidance) του TCP ρυθμίζει την ένταση με την οποία το πρωτόκολλο διοχετεύει κίνηση στο δίκτυο και ενεργοποιείται από την ανίχνευση απώλειας πακέτων, η απόρριψη πακέτων από τον μηχανισμό διαχείρισης καταχωρητή εμμέσως επηρεάζει τα επίπεδα συμφόρησης στο δίκτυο. Σε αυτή την κατεύθυνση, χρησιμοποιείται ο αλγόριθμος τυχαίας πρότερης διάγνωσης (Random Early Detection – RED) [FJ93]. Όταν ο καταχωρητής τείνει να εξαντληθεί, όταν δηλαδή το μέγεθος της ουράς των πακέτων στον καταχωρητή (queue size) ξεπεράσει κάποιο ελάχιστο κατώφλι (minimum threshold) ενεργοποιείται ο αλγόριθμος και τα καινούρια πακέτα απορρίπτονται σύμφωνα με μια ομοιόμορφη κατανομή πιθανότητας, η τιμή της οποίας διαμορφώνεται ως η γραμμική συνάρτηση του queue size με τιμή μηδέν στο minimum threshold και μέγιστη πιθανότητα απόρριψης (maximum dropping probability) στο μέγιστο κατώφλι (maximum threshold). Οι τιμές των minimum threshold, maximum threshold και maximum dropping probability είναι οι παράμετροι του αλγορίθμου. Όταν το queue size ξεπερνά το maximum threshold τότε τα πακέτα απορρίπτονται με πιθανότητα τη μονάδα. Εφόσον όλα τα πακέτα υπόκεινται στην ίδια κοινή κατανομή, με αυτόν τον μηχανισμό οι ροές δεν διαφοροποιούνται. Ο μηχανισμός RED με βάρη (Weighted RED – WRED) επιτυγχάνει διαφοροποίηση με βάση το πεδίο IP precedence στο Type of Service – ToS byte της IP επικεφαλίδας. Στον WRED, ανάλογα με την τιμή του IP precedence πεδίου χρησιμοποιούνται διαφορετικές τιμές για τις παραμέτρους minimum threshold, maximum threshold και maximum dropping probability, επιτυγχάνοντας έτσι διαφοροποιημένη ευαισθησία σε ενδεχόμενη συμφόρηση και διαφοροποιημένη ένταση

απόρριψης. Ένας παρόμοιος μηχανισμός, ο επαυξημένος RED (enhanced RED) παρουσιάζεται στο [FKSS97].

A.1.4.1.3 Έλεγχος Εισόδου Κίνησης

Ο μηχανισμός ελέγχου εισόδου κίνησης ρυθμίζει τον αριθμό, τον τύπο και την ένταση των ενεργών ροών που έχουν πρόσβαση στους πόρους του δικτύου, επενεργώντας:

- (α) στη διαδικασία λήψης αποφάσεων για την αποδοχή ή απόρριψη αιτήσεων ενεργοποίησης υπηρεσίας,
- (β) στον καθορισμό του επιτρεπτού ρυθμού πρόσβασης (traffic profile) των ροών κίνησης που αντιστοιχούν σε αποδεκτές αιτήσεις υπηρεσίας και
- (γ) στον καθορισμό των διορθωτικών μέτρων επί της κίνησης που υπερβαίνει τον επιτρεπτό ρυθμό πρόσβασης.

Ως διορθωτικά μέτρα μπορεί να χρησιμοποιηθούν ο *μηχανισμός μορφοποίησης* (shaping), ο *μηχανισμός αστυνόμευσης* (policing) ή η υποβάθμιση ποιότητας (quality degradation). Ο μηχανισμός μορφοποίησης καθυστερεί τα πακέτα έτσι ώστε, με την τεχνητή αυτή καθυστέρηση, ο ρυθμός μετάδοσης να μην υπερβαίνει τον επιτρεπτό ρυθμό πρόσβασης. η ροή που προκύπτει να συμμορφώνεται με το traffic profile της. Ο μηχανισμός αστυνόμευσης απλά διαγράφει τα πακέτα που υπερβαίνουν τον επιτρεπτό ρυθμό. Εναλλακτικά, κατά τη διαδικασία υποβάθμισης ποιότητας, τα μη συμμορφούμενα πακέτα στέλνονται χωρίς τεχνητή καθυστέρηση μέσω καναλιού χαμηλότερης ποιότητας υπηρεσίας ή υπηρεσίας best effort χωρίς εγγυήσεις ποιότητας.

Ο μηχανισμός ελέγχου εισόδου κίνησης είναι μεγάλης σημασίας για το δίκτυο καθώς καθορίζει κατά πόσον η συμφωνηθείσα ποιότητα υπηρεσίας πράγματι αποδίδεται στους χρήστες και επιπλέον το βαθμό χρησιμοποίησης των πόρων του δικτύου. Αλγόριθμοι που τείνουν να είναι περισσότερο ανεκτικοί ενδέχεται να οδηγούν το δίκτυο συχνά σε συμφόρηση ενώ αντίθετα υπερβολικά συντηρητικοί αλγόριθμοι ενδέχεται να εμποδίσουν την είσοδο κίνησης ενώ υπάρχουν διαθέσιμοι πόροι στο δίκτυο.

Ο έλεγχος της συμφόρησης στο δίκτυο είναι πολύ σημαντικός ακόμα και για δίκτυα που δεν διαφοροποιούν την ποιότητα των υπηρεσιών που προσφέρουν. Πρωτόκολλα για παράδειγμα που παρέχουν αξιόπιστη επικοινωνία όπως το Transport Control Protocol (TCP), διαθέτουν μηχανισμούς επαναμετάδοσης σε περίπτωση απώλειας πακέτων σε περιόδους συμφόρησης, έχοντας ως αποτέλεσμα την αλληπάλληλη αποστολή των ίδιων πακέτων και κατά συνέπεια την κακή χρήση των πόρων του δικτύου.

Κατάλληλοι αλγόριθμοι έχουν σχεδιαστεί για αποφυγή και διαχείριση συμφόρησης και εφαρμόζονται στα δίκτυα σήμερα. Οι αλγόριθμοι αυτοί λειτουργούν δεδομένης της έντασης της εισερχόμενης κίνησης και άρα η αποτελεσματικότητά τους περιορίζεται από αυτή. Επιπλέον η λειτουργία αυτών των μηχανισμών όταν ανιχνεύεται συμφόρηση επιφέρει κάποιο κόστος στην ευστάθεια του συστήματος και στην επιβάρυνση του δικτύου από την ανταλλαγή των απαραίτητων μηνυμάτων όπου τέτοια εφαρμόζονται. Συνεπώς, η εφαρμογή μηχανισμών ελέγχου εισόδου κίνησης συνεισφέρει στην αποτελεσματικότητα της επίλυσης και της αποφυγής συμφόρησης και κατ' επέκταση στην ευστάθεια του δικτύου.

Τέλος, ο έλεγχος εισόδου κίνησης είθισται να συνδυάζεται με τους απαραίτητους για τους παροχείς υπηρεσιών μηχανισμούς εξακρίβωσης γνησιότητας, εξουσιοδότησης και χρέωσης (Authentication- Authorisation-Accounting – AAA). Αυτή η διάσταση του ελέγχου εισόδου κίνησης είναι έξω από τα πλαίσια της παρούσας διατριβής, αν και σχετικά ζητήματα λήφθηκαν υπόψη κατά τη σχεδίαση και υλοποίηση των αλγορίθμων, για λόγους επαλήθευσης.

A.1.4.2 Αρχιτεκτονικές

Η αρχιτεκτονική οργάνωση των λειτουργιών ενός περιβάλλοντος για την παροχή ποιότητας υπηρεσίας καθορίζει τους μηχανισμούς που χρησιμοποιούνται, τις αλληλεπιδράσεις μεταξύ τους και με το εξωτερικό περιβάλλον, δηλαδή με τους χρήστες, με τα ομότιμα δίκτυα παροχής ποιότητας υπηρεσίας και με τις υπερκείμενες λειτουργίες διαχείρισης. Θέματα που απασχολούν τον αρχιτεκτονικό σχεδιασμό είναι η κατανομή των λειτουργιών σε ακραίους και μη κόμβους ή η συγκέντρωσή τους σε μια κεντρική μονάδα, η διατήρηση και διαχείριση πληροφορίας ανά ροή ή/και ανά κλάση ροών, ο καταμερισμός των πόρων με αποκλειστική ανά ροή δέσμευση πόρων ή με επικάλυψη ανά σύνολο ροών, η μοντελοποίηση του περιβάλλοντος και η μοντελοποίηση των υπηρεσιών, οι παράμετροι πολιτικής του παροχέα που επηρεάζουν τη λειτουργία του συστήματος κ.λπ. Σημειώνεται πως μια αρχιτεκτονική είναι δυνατό να υλοποιείται ένα υποσύνολο των παραπάνω θεμάτων παρέχοντας τη δυνατότητα επίλυσης των υπολοίπων στο ανώτερο επίπεδο των λειτουργιών διαχείρισης και σύμφωνα με τις πολιτικές του παροχέα.

Στις επόμενες ενότητες εξετάζουμε τις αρχιτεκτονικές *Ενοποιημένων Υπηρεσιών* και *Διαφοροποιημένων Υπηρεσιών* που προτάθηκαν στα πλαίσια του οργανισμού προτυποποίησης IETF (Internet Engineering Task Force)¹.

A.1.4.2.1 Ενοποιημένες Υπηρεσίες

Η ομάδα εργασίας *Ενοποιημένων Υπηρεσιών* (Integrated Services – IntServ) [IS-WG] ιδρύθηκε με σκοπό την επαύξηση της βασικής αρχιτεκτονικής του διαδικτύου σε μια νέα IntServ αρχιτεκτονική όπως αρχικά προτάθηκε στο [IS-ARCH], έτσι ώστε να καθίσταται δυνατή η εξυπηρέτηση εφαρμογών πραγματικού χρόνου με απαιτήσεις ποιότητας υπηρεσίας πέραν της best-effort.

Ως αποτέλεσμα αυτής της προσπάθειας δημιουργήθηκε ένα μοντέλο υπηρεσιών που περιγράφει τις υπηρεσίες που υποστηρίζονται από τη νέα αρχιτεκτονική και αποσκοπούν στην εξυπηρέτηση όλων των τύπων εφαρμογών, προδιαγράφηκαν οι επιπρόσθετοι μηχανισμοί *Ελέγχου Κίνησης* (Traffic Control – TC) στους δρομολογητές του δικτύου που είναι απαραίτητοι για την υλοποίηση της αρχιτεκτονικής καθώς και τα σχετικά interfaces με τους υπάρχοντες μηχανισμούς. Ακόμα, αναπτύχθηκε το *Πρωτόκολλο Δέσμευσης Πόρων* (Resource Reservation Protocol – RSVP) [RSVP-FS], το οποίο επιτρέπει στις εφαρμογές να επιλέγουν το επιθυμητό επίπεδο ποιότητας υπηρεσίας για την κίνηση που δημιουργούν και πρόκειται να εξυπηρετηθεί από το δίκτυο. Για τον λεπτομερή σχεδιασμό του πρωτοκόλλου RSVP ιδρύθηκε η RSVP ομάδα εργασίας [RSVP-WG].

A.1.4.2.1.1 Μοντέλο Υπηρεσιών IntServ

Σύμφωνα με το IntServ μοντέλο υπηρεσιών οι εφαρμογές κατατάσσονται σε δύο μεγάλες κατηγορίες, στις εφαρμογές πραγματικού χρόνου (real-time) και στις εφαρμογές μη πραγματικού χρόνου ή ελαστικές (non-real-time/elastic). Στις real-time εφαρμογές ταυτόχρονα με τη λήψη των πακέτων στον προορισμό πραγματοποιείται ασύγχρονα και ανεξάρτητα η επεξεργασία τους για τη σύνθεση του κατακερματισμένου σήματος χρησιμοποιώντας μηχανισμό buffering (βλέπε ενότητα A.1.3.1). Τα πακέτα που φτάνουν αφού έχει παρέλθει το στιγμιότυπό τους στη σύνθεση του σήματος δεν μπορούν να χρησιμοποιηθούν, θεωρούνται άκυρα και απορρίπτονται. Αντίθετα, στις non-real-time εφαρμογές η επεξεργασία των πακέτων είναι

¹ <http://www.ietf.org>

σύγχρονη με τη λήψη τους, μπλοκάρει μέχρι την άφιξη του επόμενου στη σειρά πακέτου η οποία εξασφαλίζεται μέσω μηχανισμών ανίχνευσης λάθους (error detection) και επαναμεταδόσεων από την πηγή.

Οι real-time εφαρμογές διακρίνονται περαιτέρω σε δύο υποκατηγορίες, στις ανεκτικές (tolerant) και μη ανεκτικές (intolerant) real-time εφαρμογές. Tolerant εφαρμογές χαρακτηρίζονται εκείνες οι real-time εφαρμογές οι οποίες δύνανται να αφομοιώσουν διακυμάνσεις σε delay και jitter και να λειτουργήσουν ικανοποιητικά σε αντίθεση με τις intolerant εφαρμογές των οποίων η ποιότητα υποβαθμίζεται σε αντίστοιχες συνθήκες.

Το IntServ μοντέλο υπηρεσιών ορίζει τρεις τύπους υπηρεσίας για την εξυπηρέτηση των ως άνω κατηγοριών εφαρμογών, την υπηρεσία *Εγγυημένης Ποιότητας* (Guaranteed QoS service) για intolerant real-time εφαρμογές, την υπηρεσία *Ελεγχόμενου Φόρτου* (Controlled Load service) για tolerant real-time εφαρμογές και την υπηρεσία *Βέλτιστης Προσπάθειας* (Best Effort service) για elastic εφαρμογές.

A.1.4.2.1.1.1 Υπηρεσία Guaranteed QoS

Η υπηρεσία Guaranteed QoS [IS-GS] παρέχει μηδενική απώλεια πακέτων, εγγυημένο εύρος ζώνης και φραγμένη καθυστέρηση. Ο αποστολέας (sender) ξεκινά τη διαδικασία ενεργοποίησης της υπηρεσίας καθορίζοντας τα χαρακτηριστικά κίνησης της ροής που θα χρησιμοποιήσει την υπηρεσία και στη συνέχεια ο παραλήπτης (receiver) καθορίζει το επιθυμητό επίπεδο ποιότητας υπηρεσίας. Όσο ο αποστολέας στέλνει στο δίκτυο κίνηση που υπακούει στα δοθέντα χαρακτηριστικά τα πακέτα φτάνουν στον αποστολέα χωρίς απώλειες εξαιτίας ενδεχόμενης συμφόρησης και με καθυστέρηση μικρότερη ή το πολύ ίση με το δοθέν άνω όριο καθυστέρησης. Για να είναι δυνατή η πραγματοποίηση της υπηρεσίας Guaranteed QoS, κάθε κόμβος στο από άκρο σε άκρο μονοπάτι θα πρέπει να υλοποιεί τους απαραίτητους μηχανισμούς που την υποστηρίζουν. Σημειώνεται πως εγγυήσεις παρέχονται για την μέγιστη καθυστέρηση και όχι για την ελάχιστη ή κατά μέσο όρο καθυστέρηση, αλλά και κατά κανένα τρόπο για τη διακύμανση της καθυστέρησης.

A.1.4.2.1.1.2 Υπηρεσία Controlled Load

Η υπηρεσίας Controlled Load [IS-CLS] προσεγγίζει τη συμπεριφορά της γνωστής υπηρεσίας best-effort σε ελαφρώς φορτωμένα δίκτυα, ανεξάρτητα με τον πραγματικό φόρτο του δικτύου την κάθε στιγμή. Κατ' ουσία, το ποσοστό απώλειας πακέτων είναι ιδιαίτερα χαμηλό και η καθυστέρηση εξαιτίας του χρόνου αναμονής στις ουρές μετάδοσης των κόμβων του δικτύου ελάχιστη. Η υλοποίηση αυτής της υπηρεσίας προϋποθέτει, όπως και για την Guaranteed QoS υπηρεσία, η εφαρμογή του χρήστη να τροφοδοτήσει το δίκτυο με τα στοιχεία της κίνησης που πρόκειται να παράγει, σε αυτή την περίπτωση μια εκτίμηση των χαρακτηριστικών της κίνησης. Στη συνέχεια, κάθε κόμβος στο μονοπάτι μεταφοράς της συγκεκριμένης υπηρεσίας χρησιμοποιεί αυτή την εκτίμηση για να υπολογίσει αν οι διαθέσιμοι πόροι είναι αρκετοί για την εξυπηρέτησή της, χωρίς όμως να κάνει χρήση αυστηρών μαθηματικών μοντέλων και των παραμέτρων που αφορούν την καθυστέρηση και την απώλεια. Κατ' αυτό τον τρόπο είναι πιθανό, διακυμάνσεις στην ένταση της κίνησης που τελικά παράγεται από τις πηγές να οδηγεί σε περιστασιακή συμφόρηση με αποτέλεσμα καθυστερήσεις ή/και απώλειες πακέτων, οι οποίες όμως, καθώς δεν είναι ιδιαίτερα έντονες, δεν επηρεάζουν την ποιότητα των tolerant εφαρμογών που εξυπηρετούν.

A.1.4.2.1.1.3 Υπηρεσία Best Effort

Η υπηρεσία Best Effort είναι παρόμοια με τη γνωστή υπηρεσία που παρέχεται από τη βασική αρχιτεκτονική του διαδικτύου κάτω από οποιεσδήποτε συνθήκες φόρτου που ποικίλουν από ελαφρώς φορτωμένα δίκτυα

έως και υπερφορτωμένα δίκτυα. Σημειώνεται πως αυτή η υπηρεσία είναι και η υπηρεσία που παρέχεται στις εφαρμογές που δεν υλοποιούν την αρχιτεκτονική IntServ.

A.1.4.2.1.2 Resource Reservation Protocol – RSVP

Οι real-time εφαρμογές και οι αντίστοιχες υπηρεσίες που περιγράφονται από το μοντέλο υπηρεσιών IntServ απαιτούν κάποιες εγγυήσεις ποιότητας υπηρεσίας. Αυτές οι εγγυήσεις δεν μπορούν να πραγματοποιηθούν χωρίς τη δέσμευση πόρων του δικτύου. Το πρωτόκολλο RSVP [RSVP-FS] επινοήθηκε και χρησιμοποιείται για τη διαβίβαση των απαιτήσεων της εφαρμογής που αιτείται την υπηρεσία στο δίκτυο και τη δέσμευση πόρων σε κάθε κόμβο που διατρέχει το μονοπάτι από τον αποστολέα στον παραλήπτη και πρόκειται να εξυπηρετήσει τη ροή κίνησης της εφαρμογής. Το RSVP είναι απλά ένα μονόδρομο (simplex) πρωτόκολλο σηματοδότησης (signalling) που μεταφέρει αιτήσεις δέσμευσης πόρων για ροές μίας κατεύθυνσης και επιστρέφει μια ένδειξη για την επιτυχή ή αποτυχημένη περάτωση της διαδικασίας στην πλευρά που πραγματοποιεί την αίτηση. Δεν είναι το ίδιο πρωτόκολλο δρομολόγησης αλλά χρησιμοποιεί τους πίνακες δρομολόγησης όπως έχουν διαμορφωθεί από άλλους μηχανισμούς.

Η δέσμευση πόρων σε ένα κόμβο για κάθε ροή κίνησης που τον διατρέχει συνεπάγεται την αποθήκευση και διαχείριση πληροφορίας κατάστασης της ροής στον κόμβο, χαρακτηριστικό που αποτελεί θεμελιώδη και ουσιαστική αλλαγή στη βασική αρχιτεκτονική του διαδικτύου. Είναι σημαντικό ακόμα να αναφέρουμε πως το RSVP είναι πρωτόκολλο χαλαρής κατάστασης (soft-state), που σημαίνει πως, ανά τακτά χρονικά διαστήματα, αυτή η πληροφορία κατάστασης πρέπει να ανανεώνεται με την αποστολή και λήψη νέων μηνυμάτων για να συνεχίζει να θεωρείται έγκυρη.

Η διαδικασία δέσμευσης πόρων με το πρωτόκολλο RSVP (βλέπε Σχήμα A-1) πραγματοποιείται σε δύο φάσεις.

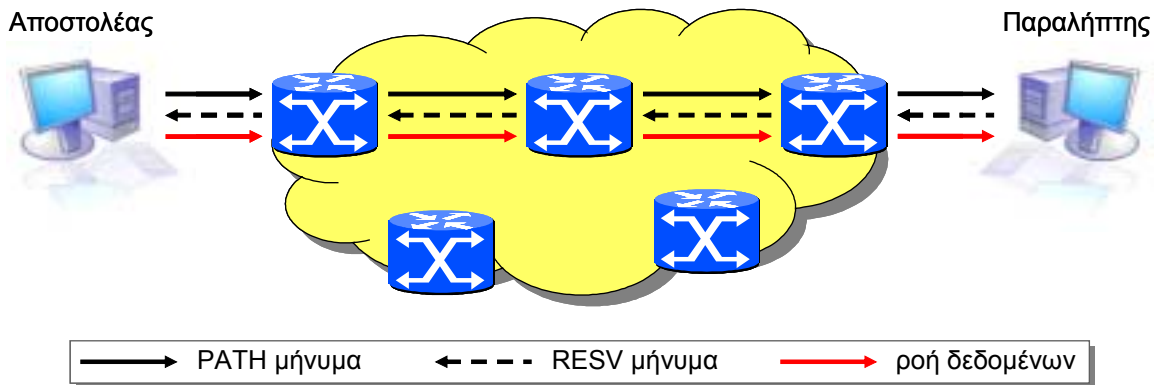
Αρχικά, PATH μηνύματα στέλνονται από τον αποστολέα (αιτούσα πλευρά) προς τον προορισμό. Αυτά τα μηνύματα περιλαμβάνουν πληροφορία που αφορά στα χαρακτηριστικά κίνησης της ροής που θα παράγει ο αποστολέας και τη διεύθυνση IP του προηγούμενου στη σειρά κόμβου ώστε να είναι δυνατή η αντίστροφη δρομολόγηση της απόκρισης από τον παραλήπτη προς τον αποστολέα χρησιμοποιώντας το ίδιο μονοπάτι (reverse path forwarding). Σε αυτή τη φάση, σε κάθε ενδιάμεσο κόμβο αποθηκεύεται η κατάσταση μονοπατιού (path state) για τη συγκεκριμένη ροή με βάση τα PATH μηνύματα αλλά δεν γίνεται καμία απόπειρα δέσμευσης πόρων.

Αφού ολοκληρωθεί αυτή η φάση, ο παραλήπτης αποκρίνεται με RESV μηνύματα προς τον αποστολέα τα οποία περιλαμβάνουν πληροφορία σχετικά με τις απαιτήσεις σε ποιότητα υπηρεσίας και ακολουθούν το αντίστροφο μονοπάτι όπως το path state υποδεικνύει σε κάθε ενδιάμεσο κόμβο. Σε αυτή τη φάση με βάση τα RESV μηνύματα σε κάθε κόμβο επιχειρείται δέσμευση πόρων και αντίστοιχα αποθηκεύεται η κατάσταση δέσμευσης (reservation state) για τη συγκεκριμένη ροή. Στον επόμενο κόμβο προς τον αποστολέα προωθείται το μήνυμα RESV αν επιτύχει η δέσμευση πόρων ή ένα μήνυμα λάθους σε αντίθετη περίπτωση.

Με την επιτυχή περάτωση της διαδικασίας ο αποστολέας ξεκινά την μετάδοση πακέτων που συνιστούν τη ροή δεδομένων που εξυπηρετείται από τους δεσμευμένους πλέον από άκρο σε άκρο πόρους του δικτύου.

Το γεγονός πως το πρωτόκολλο RSVP προσανατολίζεται στον παραλήπτη (receiver-oriented) παρέχει πολλά πλεονεκτήματα σε ένα περιβάλλον εφαρμογών πολλαπλής αποστολής (multicast), π.χ. κάθε φύλλο του δένδρου πολλαπλής αποστολής απαιτεί να έχει το δικό του επίπεδο ποιότητας υπηρεσίας και οι αντίστοιχες δεσμεύσεις πόρων στους ενδιάμεσους κόμβους είναι δυνατό να συγχωνεύονται.

Το πρωτόκολλο RSVP έχει σχεδιαστεί ανεξάρτητα από τις προδιαγραφές των υπηρεσιών του μοντέλου IntServ. Η αντιστοίχιση των λειτουργιών του στις απαιτήσεις και τις ανάγκες των υπηρεσιών Guaranteed QoS και Controlled Load περιγράφονται στο [IS-RSVP].



Σχήμα A-1: ενεργοποίηση υπηρεσίας με RSVP

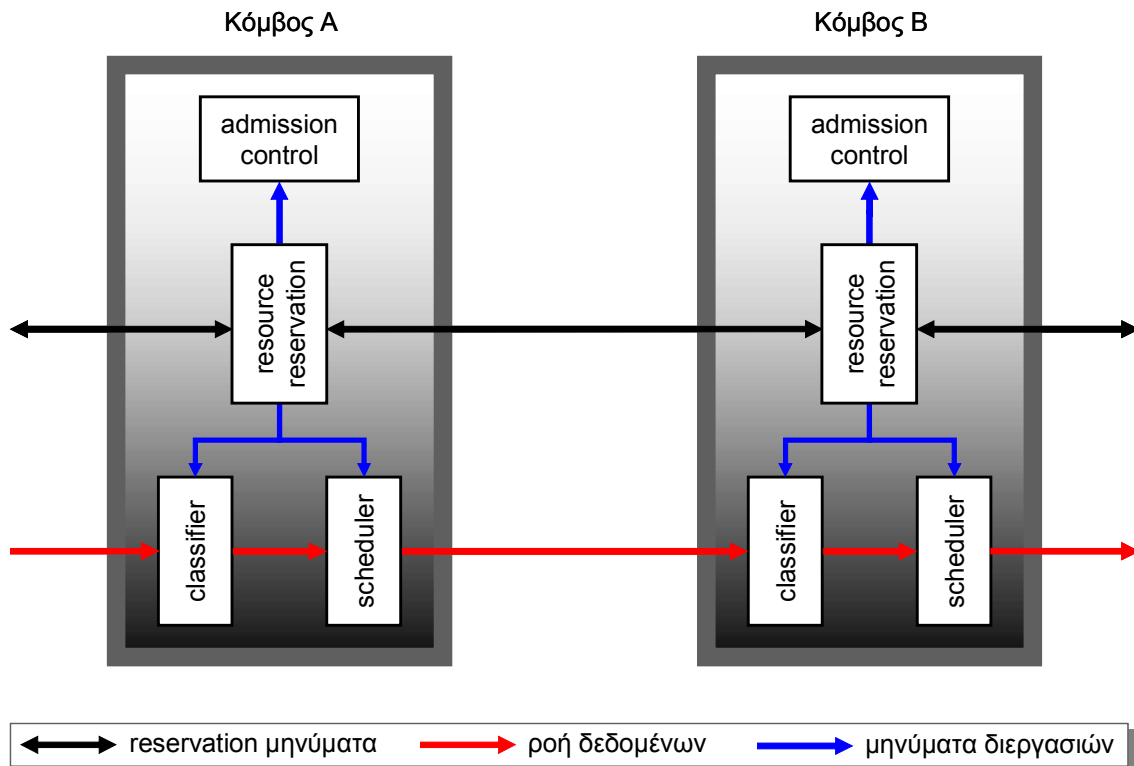
A.1.4.2.1.3 Μηχανισμοί Ελέγχου Κίνησης

Η αρχιτεκτονική IntServ ορίζει τρεις μηχανισμούς που αποτελούν τις λειτουργίες του ελέγχου κίνησης στους δρομολογητές του δικτύου (βλέπε Σχήμα A-2), τον *χρονοπρογραμματισμό πακέτων* (packet scheduler), τον *ταξινομητή πακέτων* (packet classifier) και τον *έλεγχο εισόδου κίνησης* (admission control). Οι μηχανισμοί ελέγχου κίνησης μαζί με το μηχανισμό δέσμευσης πόρων, όπως τελικά υλοποιείται από το πρωτόκολλο RSVP, αποτελούν τα τέσσερα δομικά στοιχεία που είναι απαραίτητα στους κόμβους για την υλοποίηση της αρχιτεκτονικής IntServ [IS-ARCH].

Ο scheduler είναι ο μηχανισμός χρονοπρογραμματισμού του δρομολογητή και μπορεί να είναι οποιοσδήποτε κατάλληλος μη FIFO μηχανισμός υλοποιείται από τον εκάστοτε τύπο του δρομολογητή. Θεωρείται πως ενσωματωμένος σε αυτό τον μηχανισμό είναι και ο μηχανισμός *αστυνόμευσης κίνησης* (traffic policing) καθώς της μετάδοσης των πακέτων προηγείται έλεγχος έναντι του επιτρεπτού ρυθμού όπως έχει δηλωθεί στα χαρακτηριστικά κίνησης της ροής κατά τη δέσμευση πόρων της συγκεκριμένης υπηρεσίας και τα πακέτα που κρίνονται πως ξεπερνούν τον επιτρεπτό ρυθμό απορρίπτονται.

Ο classifier ταξινομεί κάθε εισερχόμενο πακέτο σε μια συγκεκριμένη κλάση. Στις διαμορφωμένες αυτές κλάσεις επενεργεί στη συνέχεια ο scheduler παρέχοντας διαφοροποιημένη ανά κλάση ποιότητα υπηρεσίας. Η κλάση μπορεί να ορίζεται με βάση τα περιεχόμενα της IP επικεφαλίδας του πακέτου ή/και κάποιο επιπλέον χαρακτηριστικό ταξινόμησης που έχει προσδοθεί σε κάθε πακέτο. Οι κανόνες ταξινόμησης μπορεί να διαφέρουν από κόμβο σε κόμβο με την ίδια λογική που ο μηχανισμός scheduler μπορεί είναι διαφορετικός.

Ο έλεγχος εισόδου κίνησης αποφαινεται εάν μία νέα ροή μπορεί να εξυπηρετηθεί στην ποιότητα υπηρεσίας που αιτείται χωρίς να υποβαθμιστεί η ποιότητα υπηρεσίας των ήδη ενεργών ροών. Ο έλεγχος εισόδου κίνησης καλείται κατά την αίτηση για μία νέα υπηρεσία σε κάθε κόμβο στο μονοπάτι της ροής που θα παράγει ο χρήστης. Η απόφαση βασίζεται στη διαθεσιμότητα των τοπικών πόρων του κόμβου και στις πολιτικές του παροχέα που μπορεί να ορίζουν επιπρόσθετους όρους και κανόνες για την αποδοχή μίας αίτησης. Ο αλγόριθμος ελέγχου εισόδου κίνησης πρέπει να είναι συμβατός με το υπό χρήση μοντέλο υπηρεσιών.



Σχήμα Α-2: λειτουργικό μοντέλο κόμβου IntServ

A.1.4.2.1.4 Αξιολόγηση

Το μεγαλύτερο επίτευγμα της αρχιτεκτονικής IntServ είναι ίσως το ισχυρά δομημένο μοντέλο υπηρεσιών, η διάκριση των εφαρμογών σε κατηγορίες και ο καθορισμός των αντίστοιχων υπηρεσιών, καθώς και οι απόλυτες και επικυρωμένες από μαθηματικά μοντέλα εγγυήσεις ποιότητας υπηρεσίας από άκρο σε άκρο.

Η αρχιτεκτονική IntServ ωστόσο παρουσιάζει σημαντικά προβλήματα εφαρμογής σε μεγάλη κλίμακα (scalability) που απορρέουν από τη χρήση του RSVP [RSVP-AS]. Οι πόροι σε ισχύ επεξεργασίας και χώρο μνήμης ενός δρομολογητή που αναλίσκονται για τις ανάγκες του RSVP σε επίπεδο ελέγχου (control plane) και σε επίπεδο δεδομένων (data plane) αυξάνουν αναλογικά με τον αριθμό των ροών που απαιτούν ποιότητα υπηρεσίας, εκείνων δηλαδή που χρησιμοποιούν υπηρεσία άλλης της best effort. Έχει αποδειχθεί πως οι περισσότερες IP συνδέσεις από άκρο σε άκρο διαρκούν πολύ λίγο και πως μερικές δεκάδες χιλιάδες τέτοιες συνδέσεις είναι ενεργές ανά πάσα στιγμή σε ένα μέσο δρομολογητή ενός δικτύου κορμού. Τα RSVP μηνύματα για την εγκατάσταση κάθε τέτοιας σύνδεσης επιβαρύνουν σημαντικά το δίκτυο σε εύρος ζώνης αναλογικά με το εύρος ζώνης που τελικά χρησιμοποιείται από αυτές τις συνδέσεις. Είναι φανερό πως, ιδιαίτερα δίκτυα κορμού με φυσικές συνδέσεις υψηλής ταχύτητας, η απόδοση των δρομολογητών επηρεάζεται σημαντικά από την επιβάρυνση σε επεξεργαστική ισχύ εξαιτίας του RSVP, της διαχείρισης πληροφορίας για τις ενεργές ροές και της ταξινόμησης κάθε εισερχόμενου πακέτου σε μία από αυτές τις ενεργές ροές. Επιπλέον, οι περισσότερες γνωστές υλοποιήσεις scheduling αλγορίθμων δεν μπορούν να διαχειριστούν παρόμοιο πλήθος ταυτόχρονων ενεργών ροών.

A.1.4.2.2 Διαφοροποιημένες Υπηρεσίες

Αναγνωρίζοντας τα προβλήματα εφαρμογής σε μεγάλη κλίμακα της αρχιτεκτονικής IntServ ο οργανισμός IETF δημιούργησε την ομάδα εργασίας *Διαφοροποιημένων Υπηρεσιών* (Differentiated Services – DiffServ) [DS-WG] με σκοπό τη δημιουργία μιας εναλλακτικής πρότασης επαύξησης της βασικής αρχιτεκτονικής του διαδικτύου για παροχή ποιότητας υπηρεσίας που να είναι εφαρμόσιμη σε μεγάλη κλίμακα (scalable).

Η βασική αρχή της αρχιτεκτονικής DiffServ [DS-ARCH] είναι η εξώθηση της διαχείρισης σε επίπεδο ροών κίνησης στα άκρα του δικτύου, η εκεί ανάθεση και συσσωμάτωση (aggregation) των ροών κίνησης σε πεπερασμένες και προδιαγεγραμμένες κλάσεις και η διαχείριση σε επίπεδο αυτών των κλάσεων στον κορμό του δικτύου. Έτσι, η ογκώδης πληροφορία και η πολύπλοκη επεξεργασία (π.χ. αστυνόμευση, μορφοποίηση) των επιμέρους ροών περιορίζεται στους ακραίους κόμβους, ενώ οι κόμβοι κορμού, ανεξάρτητα από το πλήθος των ενεργών ροών, καλούνται να διαχειριστούν ανά πάσα στιγμή τον ίδιο αριθμό κλάσεων, καθιστώντας την αρχιτεκτονική εφαρμόσιμη σε οσοδήποτε μεγάλη κλίμακα.

Απόρροια των παραπάνω είναι και η απαλλαγή των κόμβων κορμού από τη διαδικασία δέσμευσης πόρων ανά ροή και κατά συνέπεια από την από άκρο σε άκρο σηματοδότηση κατά την ενεργοποίηση μιας ροής. Η πληροφορία για την ποιότητα υπηρεσίας που είναι να λάβει μια ροή φέρεται στα ίδια τα πακέτα της ροής με την αναγραφή στην επικεφαλίδα του πακέτου της κλάσης στην οποία έχει ενταχθεί η ροή. Για αυτό το σκοπό χρησιμοποιείται το πεδίο *Differentiated Services Codepoint* (DSCP) [DS-FIELD] που αντικαθιστά τα πεδία Type of Service (ToS) στο IPv4 [IPv4] και Traffic Class στο IPv6 [IPv6].

Αντί της προδιαγραφής των ίδιων των υπηρεσιών η αρχιτεκτονική DiffServ προδιαγράφει τις πεπερασμένες κλάσεις που υλοποιεί το δίκτυο και οι οποίες αποτελούν τα δομικά στοιχεία με βάση τα οποία στη συνέχεια ο κάθε παροχέας μπορεί να δομήσει τις υπηρεσίες που θα παρέχει στην επικράτειά του όπως θα εξυπηρετούν καλύτερα τις ιδιαίτερες ανάγκες του. Συγκεκριμένα κάθε τέτοια κλάση προδιαγράφεται μέσω της αντίστοιχης *Ανά Σταθμό Συμπεριφοράς* (Per Hop Behaviour – PHB), δηλαδή των ανά κλάση χαρακτηριστικών προώθησης (forwarding) σε κάθε δρομολογητή όπως γίνεται αντιληπτός από έναν εξωτερικό παρατηρητή και ανεξαρτήτως της μεθόδου υλοποίησης που διασφαλίζει αυτά τα χαρακτηριστικά. Η ποιότητα υπηρεσίας που λαμβάνει μια ροή δεδομένης κλάσης από άκρο σε άκρο εξαρτάται από τους πόρους που έχουν δεσμευτεί για το αντίστοιχο PHB σε κάθε δρομολογητή του μονοπατιού που διατρέχει και τις άλλες ενεργές ροές της ίδιας κλάσης σε αυτό το μονοπάτι. Έτσι, ο παροχέας ρυθμίζει τα χαρακτηριστικά ποιότητας των υπηρεσιών όπως κρίνεται επιθυμητό, καθορίζοντας τους ανά κλάση δεσμευμένους πόρους σε κάθε δρομολογητή και τις ροές που αποκτούν πρόσβαση σε αυτούς ανά πάσα στιγμή.

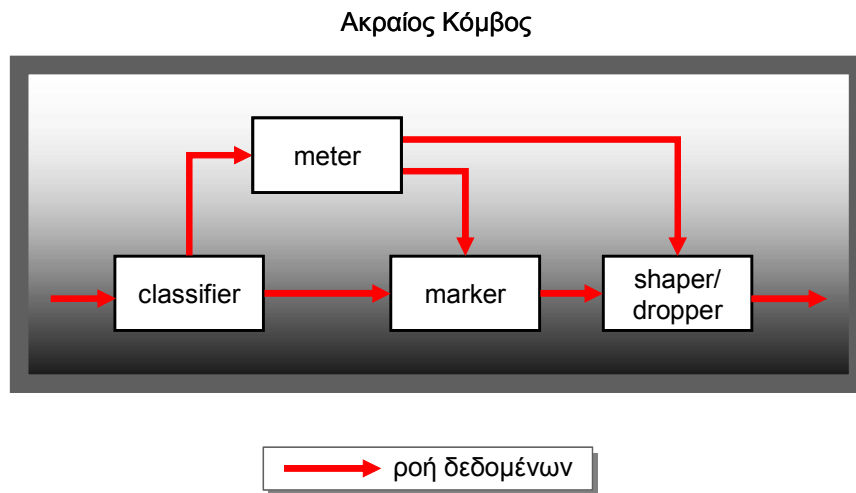
Καθώς δεν υπάρχουν προδιαγεγραμμένες υπηρεσίες, οι υπηρεσίες που προσφέρει ο κάθε παροχέας περιγράφονται από τις *Συμφωνίες Επιπέδου Υπηρεσίας* (Service Level Agreements – SLAs). Στο SLA περιγράφονται οι παράμετροι του συμβολαίου υπηρεσίας που καθορίζουν τις δεσμεύσεις του παροχέα ως προς την παρεχόμενη ποιότητα υπηρεσίας, τη διαθεσιμότητα της υπηρεσίας, τον τρόπο χρέωσης, κ.λπ., και τις υποχρεώσεις του πελάτη ως προς τα χαρακτηριστικά κίνησης της ροής που θα λάβει την υπηρεσία. SLAs δημιουργούνται και μεταξύ ομότιμων παροχέων για την εξυπηρέτηση πελατών με ανάγκες επικοινωνίας πέραν της επικράτειας ενός μόνο παροχέα. Το υποσύνολο των παραμέτρων που αφορούν τα τεχνικά χαρακτηριστικά ποιότητας υπηρεσίας ονομάζεται *Προδιαγραφή Επιπέδου Υπηρεσίας* (Service Level Specification – SLS) [DS-TERM].

Η αρχιτεκτονική DiffServ προδιαγράφει τα PHBs που εφαρμόζουν οι κόμβοι κορμού σε επίπεδο κλάσεων και τις λειτουργίες στους ακραίους κόμβους που εφαρμόζουν τα SLAs στην εισερχόμενη κίνηση σε επίπεδο

ρών σημειώνοντας τελικά στο DSCP πεδίο των πακέτων το κατάλληλο PHB όπως αυτό καθορίζεται στα SLAs και δημιουργώντας έτσι τις κλάσεις στις οποίες επιδρούν οι κόμβοι κορμού.

A.1.4.2.2.1 Λειτουργίες Ακραίων Κόμβων

Οι λειτουργίες στους ακραίους κόμβους αφορούν την *ταξινόμηση της κίνησης* (traffic classification) και την *ρύθμιση της κίνησης* (traffic conditioning) [DS-ARCH, DS-RMOD] (βλέπε Σχήμα A-3). Σε κάθε SLA περιέχονται οι κανόνες ταξινόμησης και ρύθμισης. Στους κανόνες ταξινόμησης περιγράφονται τα στοιχεία του πακέτου που επιτρέπουν την αναγνώριση της ροής κίνησης που πρόκειται να λάβει την υπηρεσία που περιγράφει το SLA. Οι κανόνες ρύθμισης αφορούν την περαιτέρω επεξεργασία και ανάθεση σε PHBs που πρόκειται να υποστεί η συγκεκριμένη ροή πριν εισέλθει στο DiffServ δίκτυο.



Σχήμα A-3: λειτουργίες ακραίου κόμβου DiffServ

Ο *ταξινομητής* (classifier) ελέγχει την IP επικεφαλίδα των εισερχόμενων πακέτων, τα αντιστοιχεί σε SLAs σύμφωνα με τους κανόνες ταξινόμησης και τα προωθεί για επεξεργασία στις λειτουργίες ρύθμισης όπως ορίζεται από τους κανόνες ρύθμισης στο κάθε SLA. Τα πεδία της IP επικεφαλίδας που καθορίζουν την ταυτότητα μιας ροής και χρησιμοποιούνται στους κανόνες ταξινόμησης για ένα δεδομένο φυσικό interface του ακραίου κόμβου μπορεί να είναι οποιαδήποτε από τα DSCP, διεύθυνση πηγής (source address), διεύθυνση προορισμού (destination address), αριθμός πόρτας πηγής (source port number), αριθμός πόρτας προορισμού (destination port number), κωδικός πρωτοκόλλου (protocol identifier). Σε περίπτωση που το υπό εξέταση πακέτο δεν αντιστοιχεί σε κανένα SLA ακολουθεί τους default κανόνες ρύθμισης όπως έχουν καθοριστεί από τον παροχέα.

Οι κανόνες ρύθμισης που περιγράφονται στα SLAs στη γενική περίπτωση καθορίζουν τα χαρακτηριστικά κίνησης της ροής του χρήστη (traffic profile) με βάση τα οποία τα εισερχόμενα πακέτα κρίνονται ως συμμορφούμενα (in-profile) ή μη συμμορφούμενα (out-of-profile). Αυτά τα χαρακτηριστικά αφορούν το ρυθμό άφιξης των πακέτων (arrival rate) και το μέγεθος έκρηξης (burst size). Διαφορετικές ενέργειες ρύθμισης καθορίζονται για τα in-profile και out-of-profile πακέτα. Η λειτουργία που ελέγχει τη ροή των πακέτων έναντι του traffic profile της ονομάζεται *μετρητής* (meter).

Οι ενέργειες ρύθμισης περιλαμβάνουν τη *σημείωση* (marking), την *μορφοποίηση* (shaping) και τη *διαγραφή* (dropping). Ο marker σημειώνει κάποια τιμή στο πεδίο DSCP καθορίζοντας έτσι το PHB που θα χρησιμοποιηθεί από τους κόμβους κορμού για να εξυπηρετήσει το πακέτο και κατά συνέπεια την ποιότητα

που θα λάβει από το δίκτυο. Για τη συμμόρφωση της ροής με το traffic profile της χρησιμοποιούνται ο shaper και ο dropper. Ο shaper καθυστερεί τα πακέτα έτσι ώστε, με την τεχνητή αυτή καθυστέρηση, η ροή που προκύπτει να συμμορφώνεται με το traffic profile της. Ο shaper χρησιμοποιεί καταχωρητές πεπερασμένης χωρητικότητας για την αποθήκευση των πακέτων που έρχονται πιο γρήγορα από όσο τα μεταδίδει. Κατά την υπερχειλίση αυτών των καταχωρητών τα πακέτα διαγράφονται. Ο dropper απλά διαγράφει τα out-of-profile πακέτα. Αυτή η διαδικασία είναι γνωστή και ως *αστυνόμευση* (policing).

A.1.4.2.2.2 Per Hop Behaviours

Τα PHBs υλοποιούνται σε κάθε δρομολογητή μέσω των μηχανισμών χρονοπρογραμματισμού και διαχείρισης καταχωρητή που ρυθμίζουν αντίστοιχα την καθυστέρηση μετάδοσης και την πιθανότητα απώλειας (βλέπε ενότητα A.1.4.1). Το κάθε PHB αντιστοιχεί σε μία μοναδική τιμή του πεδίου DSCP. Εκτός του default PHB που αντιστοιχεί στη γνωστή best-effort συμπεριφορά προώθησης (Best Effort – BE) [DS-FIELD] έχουν προδιαγραφεί δύο κατηγορίες PHBs, η *Εσπευσμένη Προώθηση* (Expedited Forwarding – EF) και η *Κατοχυρωμένη Προώθηση* (Assured Forwarding – AF). Σημειώνεται πως, πέραν των προδιαγεγραμμένων, ο κάθε παροχέας μπορεί στην επικράτειά του να ορίζει ειδικά PHBs σύμφωνα με τις ανάγκες του.

Το BE PHB αντιστοιχεί σε μετάδοση με χαμηλότερη προτεραιότητα μετάδοσης και υψηλότερη προτεραιότητα απόρριψης και χρησιμοποιείται στη γενική περίπτωση για την εξυπηρέτηση εκείνων των ροών που δεν αντιστοιχούν σε κάποιο SLA και κατ' επέκταση των ροών που προκύπτουν από εφαρμογές μη συμβατές με το DiffServ.

Το EF PHB [DS-EF, DS-EFS] ομοιάζει με υπηρεσία *εικονικής μισθωμένης γραμμής* (Virtual Leased Line – VLL) όπου οι τελικοί χρήστες έχουν την εντύπωση πως χρησιμοποιούν σύνδεση πάνω από αποκλειστικά δική τους φυσική ζεύξη. Η υλοποίηση της EF απαιτεί ο ρυθμός μετάδοσης να είναι σε κάθε περίπτωση μεγαλύτερος ή ίσος του ρυθμού άφιξης των πακέτων και να χρησιμοποιεί απόλυτη υψηλότερη προτεραιότητα μετάδοσης επί των άλλων PHBs. Με αυτό τον τρόπο οι χρόνοι αναμονής στις ουρές των δρομολογητών μηδενίζονται καθώς και οι απώλειες πακέτων λόγω υπερχειλίσης των καταχωρητών των ουρών, επιτυγχάνοντας θεωρητικά μετάδοση με μηδενικό delay, loss και jitter, πέραν αυτών που οφείλονται στις ζεύξεις του φυσικού επιπέδου. Εξαιτίας της απόλυτης υψηλότερης προτεραιότητας μετάδοσης, οι ροές που εξυπηρετούνται με EF αστυνομεύονται έτσι ώστε να μην ξεπερνούν ποτέ το traffic profile τους, καθώς κάτι τέτοιο θα υποβάθμιζε τις ροές των άλλων PHBs.

Η AF ομάδα PHBs [DS-AF] δημιουργήθηκε για να εξυπηρετήσει πελάτες και εφαρμογές που επιθυμούν εγγυήσεις για ένα μέρος της κίνησης που παράγουν αλλά και τη δυνατότητα να το ξεπερνούν όταν υπάρχουν διαθέσιμοι πόροι στο δίκτυο ακόμα κι αν αυτό συνεπάγεται μεγαλύτερες καθυστερήσεις. Η ομάδα AF περιλαμβάνει τέσσερις κλάσεις PHBs, AF1x έως AF4x, που προσφέρουν τέσσερα επίπεδα εγγυήσεων μετάδοσης από τα πιο υψηλά (AF1x) έως τα πιο χαμηλά (AF4x) αντίστοιχα. Σε κάθε κλάση εκχωρούνται ξεχωριστά πόροι εύρους ζώνης και μνήμης καταχωρητών ρυθμίζοντας έτσι ξεχωριστά και τα επίπεδα εγγυήσεων ανά κλάση. Σε περίπτωση που κάποιες κλάσεις δεν κάνουν χρήση όλου του εύρους ζώνης που τους έχει ανατεθεί η περίσσεια είναι δυνατό να χρησιμοποιηθεί από τις άλλες κλάσεις. Σε κάθε κλάση διακρίνονται τρία επίπεδα προτεραιότητας απόρριψης, π.χ. AF11, AF12 και AF13 από τη χαμηλότερη προς την υψηλότερη. Καθώς οι ροές που εξυπηρετούνται με AF είναι δυνατό να ξεπερνούν το traffic profile τους, αναμένεται σε κάποιες περιπτώσεις οι ρυθμοί άφιξης να είναι μεγαλύτεροι από το δεσμευμένο εύρος ζώνης και άρα να δημιουργούνται ουρές και καθυστερήσεις. Επιπλέον, εφόσον το πλεόνασμα της κίνησης μπορεί να είναι οσοδήποτε μεγάλο αλλά οι καταχωρητές των ουρών πεπερασμένοι αναμένεται να παρουσιάζεται

συμφόρηση και να υπάρχουν απώλειες. Τότε τα διαφορετικά επίπεδα απόρριψης διασφαλίζουν τα πακέτα που έχουν βρεθεί in-profile και φέρουν DSCP χαμηλότερης προτεραιότητας απόρριψης (AF11, AF21 κ.λ.π.) εις βάρος όσων επιπλέον out-of-profile πακέτων με DSCP υψηλότερων προτεραιοτήτων είναι απαραίτητα να απορριφθούν για την αντιμετώπιση της συμφόρησης.

A.1.4.2.2.3 Αξιολόγηση

Η αρχιτεκτονική DiffServ σχεδιάστηκε με επίκεντρο το δίκτυο και όχι τις υπηρεσίες όπως η αρχιτεκτονική IntServ. Επιτυγχάνεται με αυτό τον τρόπο η απεμπλοκή της λειτουργίας του δικτύου από την πρόβλεψη και παροχή νέων πολύπλοκων και απαιτητικών υπηρεσιών, καθιστώντας εφικτή την εφαρμογή της αρχιτεκτονικής σε οσοδήποτε μεγάλη κλίμακα και αξιοποιώντας παράλληλα τις δυνατότητες των δικτύων υψηλών ταχυτήτων. Επιπλέον, καθώς δεν προαπαιτείται διαδικασία σηματοδότησης από άκρο σε άκρο για κάθε ενεργοποίηση μιας υπηρεσίας όπως στην περίπτωση της αρχιτεκτονικής IntServ μέσω του RSVP, οι χρήστες απαλλάσσονται από τις αντίστοιχες καθυστερήσεις. Το μεγαλύτερο όφελος αποκομίζουν χρήστες και εφαρμογές που λειτουργούν με πολλαπλές κλήσεις μικρής διάρκειας. Ακόμα, η δυνατότητα υπέρβασης του καθορισμένου traffic profile σε συνδυασμό με την ευέλικτη χρήση των διαθέσιμων πόρων από ροές διαφορετικών PHBs ανάλογα με τις πραγματικές ανάγκες όπως διαμορφώνονται από τις εκάστοτε ενεργές ροές συντελούν στην πιο εντατική χρήση των πόρων του δικτύου και κατά συνέπεια στην καλύτερη αξιοποίησή τους, στην εξυπηρέτηση περισσότερων πελατών και σε υψηλότερα οικονομικά οφέλη.

Η αρχιτεκτονική DiffServ από μόνη της, σε αντίθεση με την αρχιτεκτονική IntServ, δεν παρέχει μηχανισμούς που να εξασφαλίζουν συγκεκριμένα ποσοτικά επίπεδα ποιότητας υπηρεσίας. Αντ' αυτού προδιαγράφει τα PHBs ως κλάσεις σχετικής διαφοροποίησης της ποιότητας υπηρεσίας, η τελική τιμή της οποίας ρυθμίζεται από υπερκείμενες λειτουργίες διαχείρισης των πόρων του δικτύου και ελέγχου εισόδου της κίνησης. Συνεπώς, οι υπηρεσίες που προκύπτουν και το επίπεδο των εγγυήσεων των αντίστοιχων επιπέδων ποιότητας υπηρεσίας άπτονται αυτών των υπερκείμενων λειτουργιών και δεν είναι απόρροια της ίδιας της αρχιτεκτονικής. Κατ' επέκταση, οι παροχείς που εφαρμόζουν την αρχιτεκτονική DiffServ έχουν την ελευθερία να καθορίσουν αλλά και την ευθύνη να υλοποιήσουν την καλή σύμφωνα με τις ανάγκες τους λειτουργία του δικτύου τους.

Από τον ορισμό των EF και AF PHBs είναι φανερό πως το πρώτο χρησιμοποιείται για εφαρμογές που απαιτούν απόλυτες εγγυήσεις για συγκεκριμένα ποσοτικά επίπεδα ποιότητας υπηρεσίας και το δεύτερο για εφαρμογές που διαθέτουν την ευελιξία να προσαρμόζονται σε διαφορετικά επίπεδα ποιότητας υπηρεσίας μεταβάλλοντας την ένταση της κίνησης που διοχετεύουν στο δίκτυο ανάλογα με τους εκάστοτε διαθέσιμους πόρους. Τέτοιες εφαρμογές ονομάζονται *προσαρμοστικές* (adaptive) εφαρμογές και χρησιμοποιούν στην πλειοψηφία τους τον μηχανισμό αποφυγής συμφόρησης του TCP. Υπάρχουσες πρακτικές έχουν δείξει πως με κατάλληλο σχεδιασμό μπορούν να επιτευχθούν απόλυτες εγγυήσεις για την κίνηση EF και κατά κανόνα ικανοποιητικά επίπεδα ποιότητας υπηρεσίας για την κίνηση AF, με περιστασιακές περιόδους συμφόρησης όπου όμως οι εφαρμογές μολονότι λαμβάνουν χαμηλότερη ποιότητα υπηρεσίας εξακολουθούν να αντεπεξέρχονται και μάλιστα ανταποκρίνονται μειώνοντας την ένταση της κίνησης και αποσυμφορίζοντας το δίκτυο.

A.1.4.3 Λειτουργίες Διαχείρισης

Το πρόβλημα που καλούνται να λύσουν οι λειτουργίες διαχείρισης ενός δικτύου παροχής ποιότητας υπηρεσίας είναι η μεγιστοποίηση του κέρδους του παροχέα. Το κέρδος του παροχέα είναι συνάρτηση των αποδοχών από την εξυπηρέτηση πελατών και του πάγιου κόστους από την επένδυση σε πόρους δικτύου και

λογισμικά πακέτα για τη διαχείρισή του και του λειτουργικού κόστους του δικτύου από την αμοιβή εξειδικευμένου προσωπικού για το χειρισμό των ως άνω λογισμικών πακέτων.

Στις παρακάτω ενότητες εξετάζεται το πρόβλημα μεγιστοποίησης του κέρδους του παροχέα και παρουσιάζονται οι σχετικές λειτουργίες σε επίπεδο εξοπλισμού σε φυσικούς πόρους και σε επίπεδο διαχείρισης των δεδομένων φυσικών πόρων. Ακόμα, παρουσιάζεται η εναλλακτική προσέγγιση του υπερ-εφοδιασμού (over-provisioning) που υποστηρίζει πως το κέρδος μεγιστοποιείται με τον εξοπλισμό με όσους φυσικούς πόρους είναι απαραίτητοι για την απαλλαγή από κάθε λειτουργικό κόστος και άλλη διαδικασία διαχείρισης.

A.1.4.3.1 Εξοπλισμός σε Φυσικούς Πόρους

Το κέρδος του παροχέα είναι συνάρτηση των αποδοχών από την εξυπηρέτηση πελατών και του πάγιου κόστους από την επένδυση σε πόρους δικτύου και λογισμικά πακέτα για τη διαχείρισή του και του λειτουργικού κόστους του δικτύου από την αμοιβή εξειδικευμένου προσωπικού για το χειρισμό των ως άνω λογισμικών πακέτων. Οι δύο αυτοί παράγοντες που επηρεάζουν τη συνάρτηση εξαρτώνται μεταξύ τους. Περισσότεροι πόροι συνεπάγονται καλύτερη εξυπηρέτηση και μακροπρόθεσμα μεγαλύτερη ζήτηση και άρα περισσότερες αποδοχές και αντίστροφα περιορισμένοι πόροι έχουν ως αποτέλεσμα προβληματική εξυπηρέτηση και απώλειες αποδοχών λόγω δυσαρέσκειας των πελατών. Δεδομένου ότι η ζήτηση των πελατών δεν αυξάνεται στο άπειρο υπάρχει κάποιο σημείο που το επιπλέον κόστος για περισσότερους πόρους είναι μεγαλύτερο από την αντίστοιχη αύξηση των αποδοχών και άρα το κέρδος μειώνεται και η συνάρτηση απομακρύνεται από τη μέγιστη τιμή της.

Η λειτουργία διαχείρισης που, δεδομένων κάποιων εκτιμήσεων για τη ζήτηση υπηρεσιών, καθορίζει τους πόρους του δικτύου που μεγιστοποιούν το κέρδος του παροχέα ονομάζεται *σχεδιασμός του δικτύου* (network planning) και παράγει ως αποτέλεσμα τη φυσική τοπολογία του δικτύου σε δρομολογητές και ζεύξεις μεταξύ τους.

A.1.4.3.2 Αξιοποίηση των Φυσικών Πόρων

Δεδομένων των φυσικών πόρων και της χωρητικότητάς τους όπως προκύπτουν από το network planning και δεδομένου ότι η ζήτηση είναι δυναμικά μεγαλύτερη από τη χωρητικότητα του δικτύου, οι λειτουργίες διαχείρισης καλούνται να επιλύσουν το πρόβλημα μεγιστοποίησης του κέρδους του παροχέα στο επίπεδο εξυπηρέτησης της ζήτησης.

Κάθε χρήστης μιας υπηρεσίας δεν διοχετεύει κίνηση στο δίκτυο αδιάκοπα και σταθερά αλλά μόνο σε κάποια χρονικά διαστήματα και μεταβαλλόμενα με μέγιστη ένταση που ονομάζεται κίνηση αιχμής (peak traffic). Απόλυτες εγγυήσεις ποιότητας υπηρεσίας μπορούν να δοθούν μόνο με το σχήμα *ανάθεσης αιχμής* (peak allocation) όπου, για κάθε συμβόλαιο υπηρεσίας, οι πόροι που είναι απαραίτητοι για να εξυπηρετήσουν την αντίστοιχη κίνηση αιχμής είναι διαθέσιμοι ανά πάσα στιγμή. Έτσι όμως παρατηρείται χαμηλή αξιοποίηση των πόρων και απώλειες αποδοχών από απόρριψη επιπλέον συμβολαίων υπηρεσιών για τα οποία τελικά υπάρχουν διαθέσιμοι πόροι χάρη στη χρονική επικάλυψη και στις αυξομειώσεις της έντασης της κίνησης που παράγουν τα υπάρχοντα συμβόλαια. Οι απώλειες αυτές είναι ακόμα περισσότερες όσο περισσότερο η κίνηση είναι καταγιστική φύσεως (bursty) με μεγάλες ανενεργές περιόδους και μικρές περιόδους υψηλού ρυθμού.

Εναλλακτικά, η διαδικασία εξυπηρέτησης της ζήτησης είναι δυνατό να συνυπολογίζει παρόμοιες επικαλύψεις και να αποδέχεται περισσότερες αιτήσεις υπηρεσίας από όσες η δεδομένη χωρητικότητα του

δικτύου δύναται να εξυπηρετήσει με απόλυτες εγγυήσεις. Αυτό το σχήμα ονομάζεται *επικαλυπτόμενης συνδρομής* (oversubscription) και χρησιμοποιείται κατά κόρον από τους παροχείς υπηρεσιών.

Καθώς δεν ισχύουν πλέον απόλυτες εγγυήσεις είναι στατιστικά βέβαιο πως θα υπάρχουν περιόδοι συμφόρησης και ιδιαίτερα σε ώρες αιχμής όπου η κίνηση των χρηστών θα είναι μεγαλύτερη από τη χωρητικότητα του δικτύου με αποτέλεσμα την υποβάθμιση της ποιότητας υπηρεσίας. Λόγω της ιδιαίτερα απρόβλεπτης φύσης της συμπεριφοράς των χρηστών η απόκλιση από τις απόλυτες εγγυήσεις και η υποβάθμιση της ποιότητας υπηρεσίας δεν είναι δυνατόν να προβλεφθούν με ακρίβεια. Βέβαιο αντίθετα είναι πως, όσο μεγαλύτερη η εντατικοποίηση του oversubscription με αποδοχή περισσότερων αιτήσεων υπηρεσίας, τόσο πιο χαλαρές οι εγγυήσεις παροχής ποιότητας και δριμύτερη η υποβάθμισή της. Παρατηρείται έτσι κάποιο όριο όπου η επιβολή χρηματικών κυρώσεων για παραβίαση συμβολαίων και η συσσωρευμένη δυσαρέσκεια των πελατών είναι περισσότερο επιζήμιες από το κέρδος που προκύπτει από την μεγαλύτερη αξιοποίηση των πόρων.

Έτσι, το πρόβλημα μεγιστοποίησης του κέρδους του παροχέα επαναπροσδιορίζεται ως πρόβλημα μεγιστοποίησης των αρκούντως ικανοποιημένων συμβολαίων [S95]. Η ερμηνεία αυτού του προβλήματος είναι σε μεγάλο βαθμό θέμα της αντίληψης του κάθε παροχέα και της ιδιαιτερότητας της αγοράς στην οποία απευθύνεται. Η παροχή ακριβών υπηρεσιών με υψηλές εγγυήσεις ή φθηνών υπηρεσιών με χαμηλές εγγυήσεις ή ο οποιοσδήποτε συνδυασμός των παραπάνω καθορίζεται από τις επιχειρησιακές επιδιώξεις του παροχέα και είναι γι' αυτό απαραίτητο η επίλυσή του να γίνεται σε επίπεδο πολιτικής του παροχέα που θα ρυθμίζει κατάλληλα τις λειτουργίες διαχείρισης και όχι στο επίπεδο των ίδιων των λειτουργιών.

Όποιες κι αν είναι οι επιλογές του παροχέα, γενικότερα ισχύει πως η εφαρμογή μηχανισμών για την παροχή ποιότητας υπηρεσίας δεν ωφελεί ένα δίκτυο ανεπαρκώς εξοπλισμένο. Εάν η ποιότητα της ομοιόμορφης υπηρεσίας best-effort είναι πολύ χαμηλή τότε η προσθήκη μηχανισμών διαφοροποίησης των υπηρεσιών δεν συνεισφέρει στον περιορισμό της συμφόρησης. Το μόνο που επιτυγχάνει είναι ίσως διαφοροποίηση ως προς το αντίκτυπο της συμφόρησης ευνοώντας τις υπηρεσίες ανώτερων επιπέδων εις βάρος των υπηρεσιών χαμηλού επιπέδου και υποβαθμίζοντας την ποιότητα των τελευταίων σε μη αποδεκτά επίπεδα. Η σημερινή πρακτική έχει δείξει πως ένα μεγάλο μέρος χρηστών επιλέγει υπηρεσίες χαμηλών έως και καθόλου εγγυήσεων όπως η γνωστή υπηρεσία best-effort και άρα ένας παροχέας δεν είναι δυνατό να παραμελεί την ποιότητα που προσφέρει σε αυτό το επίπεδο.

Περαιτέρω, οι λειτουργίες διαχείρισης διακρίνονται σε δύο κατηγορίες, στις *λειτουργίες διαχείρισης κίνησης* (traffic engineering) και στις *λειτουργίες διαχείρισης υπηρεσιών* (service management). Οι λειτουργίες διαχείρισης κίνησης αφορούν στη βέλτιστη κατανομή των πόρων του δικτύου δεδομένης της κίνησης που καλούνται να εξυπηρετήσουν, στοχεύοντας στη μεγιστοποίηση της παρεχόμενης ποιότητας υπηρεσίας. Οι λειτουργίες διαχείρισης υπηρεσιών αφορούν στη βέλτιστη εξυπηρέτηση των αιτήσεων υπηρεσίας δεδομένης της κατανομής των πόρων του δικτύου, στοχεύοντας στη μεγιστοποίηση της κίνησης που αποκτά πρόσβαση στους πόρους του δικτύου σε επίπεδο συμβολαίων αλλά και σε επίπεδο έντασης της κίνησης που παράγει το κάθε συμβόλαιο, διασφαλίζοντας ταυτόχρονα τις δεσμεύσεις του παροχέα σε εγγυήσεις ποιότητας υπηρεσίας.

A.1.4.3.2.1 Traffic Engineering

Το traffic engineering διαχειρίζεται τις διαδικασίες δέσμευσης πόρων ανά δρομολογητή και δημιουργίας μονοπατιών από άκρο σε άκρο. Βέλτιστο αποτέλεσμα επιτυγχάνεται όταν η κατανομή των πόρων αντανakλά κατά το δυνατόν περισσότερο την κατανομή της εισερχόμενης κίνησης. Έτσι, της διαστασιοποίησης του δικτύου από το traffic engineering προηγείται μία περίοδος μετρήσεων που επιτρέπει

την ανάλυση της κίνησης ανά χρονικές περιόδους όπου παρατηρείται ομοιογενής συμπεριφορά. Από αυτή την ανάλυση προκύπτουν στατιστικά δεδομένα για την ένταση της κίνησης με συγκεκριμένες απαιτήσεις ποιότητας υπηρεσίας από άκρο σε άκρο. Αυτά τα δεδομένα τροφοδοτούνται στο traffic engineering που κατανέμει κατάλληλα τους πόρους του δικτύου. Η ικανοποίηση του κριτηρίου βέλτιστης κατανομής επαληθεύεται κατά τη διάρκεια της λειτουργίας του δικτύου μέσω μετρήσεων για τον εντοπισμό των σημείων συνωστισμού, των υπερφορτωμένων και των ανεκμετάλλευτων ζεύξεων, τη συχνότητα, την ένταση, τη διάρκεια φαινομένων συμφόρησης κ.λπ. Τα αποτελέσματα αυτών των μετρήσεων χρησιμοποιούνται για τη βελτίωση των μοντέλων ανάλυσης της κίνησης και ανατροφοδοτούνται στον επόμενο κύκλο διαστασιοποίησης του δικτύου. Οι παραπάνω λειτουργίες είναι μέρος αυτού που ονομάζεται *στατικό traffic engineering*.

Σε κάποια συστήματα όταν διαπιστώνεται δυσλειτουργία ή σημαντική απόκλιση από το κριτήριο βελτιστοποίησης λαμβάνονται διορθωτικά μέτρα. Οι σχετικές λειτουργίες ανήκουν στη σφαίρα του *δυναμικού traffic engineering*. Είναι πολύ σημαντικό, η δυναμική αυτή προσαρμογή της κατανομής των πόρων του δικτύου στις απαιτήσεις της εισερχόμενης κίνησης να μην διαταράσσει την ευστάθεια του συστήματος.

Σε ένα τόσο πολύπλοκο σύστημα είναι συχνά δύσκολο να εκτιμηθεί το αντίκτυπο κάποιας ενέργειας. Ιδιαίτερα σε συστήματα σχεδιασμένα να είναι πιο ευαίσθητα σε μεταβολές, μία ενέργεια προσαρμογής ενός παράγοντα σε ένα σημείο του δικτύου είναι δυνατό να επηρεάσει πολλούς άλλους παράγοντες σε πολύ μεγαλύτερη επικράτεια και να προκαλέσει αλυσιδωτές ενέργειες που κλιμακωτά να αποσταθεροποιήσουν εντελώς τη λειτουργία του δικτύου. Η οριοθέτηση της ανοχής στις στατιστικές αποκλίσεις, η επιβολή φραγμών στη συχνότητα και την ένταση των διορθωτικών μέτρων και κυρίως η κατανόηση των αλληλεπιδράσεων μεταξύ των διαφορετικών μηχανισμών και των παραγόντων που τους ρυθμίζουν είναι ζωτικής σημασίας.

Ιδιαίτερη προσοχή χρήζει η σταθερότητα του συστήματος δρομολόγησης. Μεταβολές των μονοπατιών έχουν ενδεχομένως ως αποτέλεσμα την αναδιάταξη (reordering) των πακέτων μιας ροής και αύξηση της διακύμανσης καθυστέρησης (jitter). Αύξηση του jitter συμβαίνει λόγω της διαφοράς της καθυστέρησης από το παλιό στο νέο μονοπάτι. Αναδιάταξη συμβαίνει όταν πακέτα που έπονται στη διάταξη μιας ροής ακολουθώντας το νέο μονοπάτι φτάνουν στον προορισμό πριν από τα προηγούμενα πακέτα της ίδιας ροής που ακολούθησαν το παλιό μονοπάτι. Εφαρμογές που χρησιμοποιούν μηχανισμούς buffering (βλέπε ενότητα A.1.3.1) αντιμετωπίζουν την αναδιάταξη χωρίς περαιτέρω επιπλοκές ενώ στην αντίθετη περίπτωση ερμηνεύεται ως απώλεια των ενδιάμεσων πακέτων υποβαθμίζοντας έτσι την ποιότητα λόγω αυξημένου ποσοστού απώλειας (loss).

A.1.4.3.2 Service Management

Το service management διαχειρίζεται τις διαδικασίες συνδιαλλαγής με τους πελάτες για τη διαπραγμάτευση και αποδοχή των αιτήσεων υπηρεσιών, καθώς και τις διαδικασίες ενεργοποίησης και ελέγχου των αποδεκτών αιτήσεων. Βέλτιστο αποτέλεσμα επιτυγχάνεται όταν η κίνηση που τελικώς διοχετεύεται στο δίκτυο από τις ενεργές υπηρεσίες όπως προκύπτουν από την εφαρμογή του service management, προσεγγίζει κατά το δυνατό περισσότερο αλλά δεν ξεπερνά τη χωρητικότητα των πόρων του δικτύου όπως έχουν κατανεμηθεί από τις λειτουργίες του traffic engineering.

Έτσι, κατά τη διαδικασία αποδοχής ή απόρριψης των αιτήσεων υπηρεσιών, ο έλεγχος εισόδου κίνησης (βλέπε ενότητα A.1.4.1.3) που είναι η υπεύθυνη διαδικασία, θα πρέπει να είναι σε θέση να εκτιμήσει τη διαθέσιμη χωρητικότητα του δικτύου, τις απαιτήσεις σε πόρους της υπό εξέταση αίτησης υπηρεσίας και εάν

η διαθέσιμη χωρητικότητα επαρκεί για την εξυπηρέτηση της. Η εκάστοτε διαθέσιμη χωρητικότητα είναι δυνατό να εκτιμάται είτε πρακτικά λαμβάνοντας τις απαραίτητες μετρήσεις, είτε θεωρητικά υπολογίζοντας με μαθηματικά μοντέλα τη διαφορά των διαθέσιμων πόρων όπως προκύπτει από την κατανομή του στατικού traffic engineering και των πόρων που διατίθενται ήδη στις υπάρχουσες ενεργές ροές, είτε με συνδυασμό των δύο. Τα θέματα που προκύπτουν σχετικά, όπως ο ρόλος του δυναμικού traffic engineering, οι αποκλίσεις των μαθηματικών μοντέλων, η αξιοπιστία των μετρήσεων κ.λπ. εξετάζονται σε επόμενες ενότητες.

Επιπλέον, το κριτήριο μη υπερφόρτωσης του δικτύου με κίνηση πέραν της χωρητικότητάς του και υποβάθμισης κατ' επέκταση της ποιότητας υπηρεσίας των ενεργών ροών επαληθεύεται μέσω μετρήσεων για την ανίχνευση συμφόρησης τέτοιας που το δυναμικό traffic engineering εάν εφαρμόζεται δεν είναι σε θέση να αποτρέψει. Σε αυτές τις περιπτώσεις το service management πρέπει να είναι σε θέση να επιλύσει τη συμφόρηση είτε διακόπτοντας κάποιες ενεργές ροές (pre-emption) είτε μειώνοντας την ένταση με την οποία διοχετεύουν κίνηση στο δίκτυο, ελαχιστοποιώντας κατά το δυνατό το αντίκτυπο σε ποιότητα υπηρεσίας και κατανέμοντάς το κατά το δυνατό δίκαια στις ενεργές ροές.

Κάποιοι πελάτες μπορεί να επιθυμούν διαβεβαίωση της υπηρεσίας που χρησιμοποιούν (service assurance), επαληθεύοντας την τήρηση των δεσμεύσεων του παροχέα ως προς την ποιότητα υπηρεσίας. Κάτι τέτοιο απαιτεί λεπτομερείς μετρήσεις της ποιότητας υπηρεσίας έως και σε επίπεδο ροής. Τέτοιου είδους μετρήσεις καταναλώνουν σημαντικούς πόρους σε επεξεργασία κατά τη μέτρηση στους δρομολογητές, σε εύρος ζώνης κατά την εξαγωγή και μεταφορά τους από τους δρομολογητές στα σημεία επεξεργασίας και σε χώρο μνήμης στα σημεία αποθήκευσης. Συνηθίζεται έτσι το επίπεδο λεπτομέρειας και συχνότητας λήψης των μετρήσεων για τη διαβεβαίωση της υπηρεσίας να είναι διαπραγματεύσιμο κατά τη σύναψη του συμβολαίου.

A.1.4.3.3 Over-Provisioning

Η εφαρμογή οποιουδήποτε μηχανισμού και λειτουργίας διαχείρισης για την παροχή ποιότητας υπηρεσίας επιφέρει κάποιο κόστος στον παροχέα υπηρεσιών. Το κόστος αυτό προέρχεται από:

- Τη μειωμένη απόδοση του δικτύου. Λόγω της προόδου στον τομέα των μέσων μετάδοσης και της χωρητικότητας σε εύρος ζώνης, ο συνωστισμός πλέον οφείλεται κυρίως στην επεξεργασία που εκτελείται στους κόμβους. Κατά συνέπεια πιο πολύπλοκη επεξεργασία συνεπάγεται μειωμένη απόδοση.
- Τη χρήση και δέσμευση μέρους των πόρων του δικτύου για τις λειτουργίες του μηχανισμού. Πρωτόκολλα καταναλωμένου ελέγχου χρησιμοποιούν μέρος του διαθέσιμου εύρους ζώνης για την ανταλλαγή των μηνυμάτων τους, αλγόριθμοι διαχείρισης επιβαρύνουν σε μνήμη και φόρτο επεξεργασίας τα διαθέσιμα μηχανήματα, κ.λπ.
- Το αυξημένο οικονομικό κόστος αγοράς του δικτυακού εξοπλισμού και του λογισμικού διαχείρισης που υλοποιούν προηγμένους πολύπλοκους μηχανισμούς για την παροχή ποιότητας υπηρεσίας.
- Το κόστος σε εργατοώρες εξειδικευμένου προσωπικού για την εγκατάσταση και το χειρισμό του ως άνω δικτυακού εξοπλισμού και λογισμικού.

Η βασική αρχή του *υπερ-εφοδιασμού* (over-provisioning) είναι πως η ικανοποίηση των απαιτήσεων ποιότητας των προσφερόμενων υπηρεσιών είναι δυνατό να διασφαλιστεί με τον επαρκή εξοπλισμό σε φυσικούς πόρους της δικτυακής υποδομής. Ο εξοπλισμός ενός δικτύου με μεγαλύτερη χωρητικότητα σε εύρος ζώνης από την απαραίτητη για να εξυπηρετηθεί η κίνηση ακόμα και σε ώρες αιχμής και απορροφώντας κάθε είδους διακυμάνσεις είναι ικανή συνθήκη για την εκπλήρωση των απαιτήσεων της κάθε

εφαρμογής, έτσι ώστε η εφαρμογή οποιουδήποτε περίπλοκου μηχανισμού για την παροχή ποιότητας υπηρεσίας καθίσταται περιττή και κατά συνέπεια αποφεύγεται το σχετικό κόστος.

Η επιχείρηση που υιοθετεί το over-provisioning απαλλάσσεται από τη μελέτη της αγοράς ως προς τις τάσεις, τη συμπεριφορά και τις απαιτήσεις των χρηστών, τον σχεδιασμό των σχετικών πολιτικών και το κόστος της εφαρμογής αυτών των πολιτικών με τη χρήση μηχανισμών παροχής ποιότητας και σχετικών λειτουργιών διαχείρισης. Αυτά τα πλεονεκτήματα συνοδεύονται από την επιβάρυνση της επιχείρησης με αυξημένο πάγιο οικονομικό κόστος για τον εξοπλισμό του δικτύου με χωρητικότητα σε εύρος ζώνης τέτοια ώστε να υπερκαλύπτει υπό οποιεσδήποτε συνθήκες την κίνηση των χρηστών. Το κόστος που συνεπάγεται η απόκτηση των επιπλέον φυσικών πόρων αντισταθμίζεται με τις σημαντικές ελαφρύνσεις στο κόστος διαχείρισης και λειτουργίας.

Η αξία αυτής της προσέγγισης έγκειται στην απλότητά της. Το βασικό της μειονέκτημα είναι το υψηλό κόστος των επενδύσεων που απαιτούνται για τη διαμόρφωση της απαιτούμενης υποδομής. Ένα από τα βασικά προβλήματα είναι πως σε μια αγορά η οποία αναπτύσσεται ραγδαία, οι οργανισμοί που εφαρμόζουν αυτή τη μέθοδο, για να εξασφαλίσουν την απόδοση των επενδύσεών τους, τείνουν να είναι συντηρητικοί ως προς τις εκτιμήσεις τους σε ζήτηση και κατά συνέπεια σε απαραίτητους πόρους που πρέπει να διατεθούν.

Το πρόβλημα αυτό οξύνεται όσο μεγαλώνει η πολυπλοκότητα της δικτυακής υποδομής, η ένταση και η ποικιλότητα της κίνησης που αναμένεται να υποδεχθεί το δίκτυο. Λόγω της απρόβλεπτης φύσης της κίνησης, η μέγιστη ροή κίνησης μπορεί να σημειώνεται σε διαφορετικά μονοπάτια κάθε φορά. Όσο αυξάνεται η συνολική ένταση της κίνησης τόσο αυξάνει και η μέγιστη ροή. Επιπλέον, καθώς η μέγιστη ροή είναι η σύνθεση μικρο-ροών (micro-flows) διαφορετικών εφαρμογών με διαφορετικές απαιτήσεις, θα πρέπει να εξυπηρετείται έτσι ώστε να ικανοποιούνται οι υψηλότερες από αυτές, που σε κάποιες εφαρμογές μπορεί να αφορούν το χρόνο μετάβασης με επιστροφή (Round Trip Time – RTT), σε άλλες τη διακύμανση της καθυστέρησης (jitter) κ.λπ. Χωρίς τους κατάλληλους μηχανισμούς διαχωρισμού και ελέγχου της κίνησης αυτό μεταφράζεται σε πολύ υψηλές ανάγκες εφοδιασμού για όλα τα μονοπάτια και κόμβους που πιθανώς να αποτελέσουν σημεία συνωστισμού.

Επιπλέον, οι παροχείς υπηρεσιών που μισθώνουν εύρος ζώνης έχουν πιο περιορισμένες επιλογές. Εάν για παράδειγμα ένας παροχέας παρέχει best-effort υπηρεσίες πάνω από μισθωμένες γραμμές και επιθυμεί να διευρύνει τη δραστηριότητά του με την παροχή υπηρεσιών για εφαρμογές VoIP, τότε η χρήση μηχανισμών για παροχή ποιότητας υπηρεσίας επιτρέπει τη διαφοροποίηση μεταξύ των δύο διαφορετικών τύπων κίνησης χωρίς αναγκαστική επαύξηση του εύρους ζώνης. Έτσι καθίσταται δυνατή η παροχή περισσότερων και πιο ακριβών υπηρεσιών με το ίδιο κόστος μίσθωσης, αυξάνοντας κατ' αυτό τον τρόπο τα περιθώρια κέρδους του παροχέα.

Η εφαρμογή μηχανισμών και λειτουργιών διαχείρισης για την παροχή ποιότητας υπηρεσίας προσφέρει τη δυνατότητα της ρύθμισης ενός δικτύου με πεπερασμένους φυσικούς πόρους έτσι ώστε η κάθε εφαρμογή να λαμβάνει την ποιότητα που είναι απαραίτητη για την ομαλή της λειτουργία. Το over-provisioning μπορεί να εφαρμόζεται σε μικρότερο βαθμό σε συνδυασμό με άλλους μηχανισμούς, επιτυγχάνοντας ελαφρύνσεις στο κόστος διαχείρισης και λειτουργίας του δικτύου και παράλληλα καλύτερη αξιοποίηση των φυσικών του πόρων.

Σήμερα, υπάρχουν πράγματι δίκτυα κορμού τεράστιας χωρητικότητας σε εύρος ζώνης και προσεκτικά διαστασιοποιημένα έτσι ώστε να αποφεύγεται η συμφόρηση. Ωστόσο, τα περισσότερα δίκτυα πρόσβασης διαθέτουν περιορισμένο εύρος ζώνης, αυξάνοντας έτσι τα επίπεδα συμφόρησης που υφίσταται η κίνηση από άκρο σε άκρο (end-to-end) έχοντας ως αποτέλεσμα κυμαινόμενη, μη προβλέψιμη και σε κάποιες περιπτώσεις τελικά ανεπαρκή ποιότητα υπηρεσίας.

A.1.5 Ανακεφαλαίωση

Η *ανάγκη για παροχή ποιότητας υπηρεσίας* πάνω από μια κοινή υποδομή δικτύου με βάση το πρωτόκολλο IP είναι σήμερα μια πραγματικότητα που προέκυψε από την αλματώδη ανάπτυξη του διαδικτύου και την πληθώρα νέων εφαρμογών με διαφορετικές απαιτήσεις σε πόρους και εγγυήσεις απόδοσης του δικτύου.

Η ποιότητα υπηρεσίας περιλαμβάνει όλες εκείνες τις παραμέτρους που συνθέτουν τη συνολική εμπειρία που αποκομίζει ο χρήστης. Τέτοιες είναι στο επίπεδο μεταφοράς η μέση καθυστέρηση (delay), η διακύμανση καθυστέρησης (jitter), το ποσοστό απώλειας (loss), ο ρυθμός ροής (throughput) για τον οποίο ισχύουν καθορισμένες τιμές των παραπάνω, αλλά και παράμετροι σχετικά με τη διαθεσιμότητα της υπηρεσίας ως προς το χρόνο και τα σημεία πρόσβασης, τις εγγυήσεις για μηδενική ή φραγμένη στατιστική απόκλιση από τις παραπάνω τιμές, τον τρόπο χρέωσης, τη μέθοδο επαλήθευσης και τις σχετικές κυρώσεις, τη δυνατότητα επαύξησης σε περίπτωση που υπάρχουν διαθέσιμοι πόροι στο δίκτυο κ.λπ.

Η παροχή ποιότητας υπηρεσίας βασίζεται στο πρωτόκολλο IP που είναι αναμφισβήτητο ο κοινός παρονομαστής στο διαδίκτυο σε αντίθεση με τις τεχνολογίες υποκείμενων επιπέδων που είναι ενδεχομένως πιο κατάλληλες αλλά διαφέρουν από επικράτεια σε επικράτεια με αποτέλεσμα να δημιουργούνται ασυμβατότητες και ασυνέχειες στην ποιότητα της υπηρεσίας από άκρο σε άκρο. **Το παραδοσιακό πρωτόκολλο IP παρέχει μία κοινή υπηρεσία γνωστή ως best-effort, η ποιότητα της οποίας μεταβάλλεται ανεξέλεγκτα και ανάλογα με τους εκάστοτε διαθέσιμους πόρους.**

Για να ξεπεραστεί ο περιορισμός της μοναδικής κοινής υπηρεσίας που παρέχει το IP, απαιτούνται μηχανισμοί που να καθιστούν εφικτή τη διαφοροποίηση στη μεταχείριση της κάθε ροής κίνησης, δημιουργώντας έτσι τη δυνατότητα παροχής περισσότερων από μία υπηρεσιών με διαφορετικά χαρακτηριστικά ποιότητας υπηρεσίας. Σε επίπεδο δρομολογητή έχουν αναπτυχθεί και παγιωθεί κατάλληλοι μηχανισμοί χρονοπρογραμματισμού και διαχείρισης καταχωρητή.

Η διασφάλιση συγκεκριμένων τιμών για τα χαρακτηριστικά ποιότητας των διαφορετικών υπηρεσιών που μπορούν να προκύψουν προϋποθέτει έλεγχο και συγκράτηση των επιπέδων συμφόρησης του δικτύου. Σε αυτή την κατεύθυνση εξετάζονται υπάρχουσες μελέτες και διερευνώνται νέες σχετικά με τους απαραίτητους μηχανισμούς, τις αρχιτεκτονικές που καθορίζουν την μεθόδευση και ενορχήστρωση τους και τις υπερκείμενες λειτουργίες διαχείρισης που ρυθμίζουν την καλή λειτουργία τους.

Η μελέτη της αρχιτεκτονικής IntServ (βλέπε ενότητα A.1.4.2.1) έδειξε πως, αν και πλεονεκτεί σε ακρίβεια και αξιοπιστία ως προς τις εγγυήσεις ποιότητας υπηρεσίας χάρη στο ισχυρά δομημένο μοντέλο υπηρεσιών της, ωστόσο παρουσιάζει προβλήματα εφαρμογής σε μεγάλη κλίμακα καθώς οι πόροι σε ισχύ επεξεργασίας και χώρο μνήμης σε επίπεδο ελέγχου (control plane) και σε επίπεδο δεδομένων (data plane) που αναλίσκονται από το RSVP αυξάνουν αναλογικά με τον αριθμό των ροών που απαιτούν ποιότητα υπηρεσίας. Ιδιαίτερα σε δίκτυα κορμού με φυσικές συνδέσεις υψηλής ταχύτητας η επιβάρυνση εξαιτίας τους RSVP επηρεάζει σημαντικά την απόδοσή τους.

Στοχεύοντας στην αντιμετώπιση των παραπάνω προβλημάτων η αρχιτεκτονική DiffServ (βλέπε ενότητα A.1.4.2.2) σχεδιάστηκε με επίκεντρο το δίκτυο και όχι τις υπηρεσίες, επιτυγχάνοντας έτσι την απεμπλοκή της λειτουργίας του δικτύου από την πρόβλεψη και παροχή πολύπλοκων υπηρεσιών, καθιστώντας εφικτή την εφαρμογή της σε οσοδήποτε μεγάλη κλίμακα και αξιοποιώντας παράλληλα τις δυνατότητες των δικτύων υψηλών ταχυτήτων. Ακόμα, επιτρέποντας τη χρήση των διαθέσιμων πόρων με ευέλικτο τρόπο μεταξύ των διαφορετικών ροών, εντείνεται η αξιοποίησή τους και κατ' επέκταση η αποτελεσματικότητα του σχήματος.

Η αρχιτεκτονική DiffServ προδιαγράφει κλάσεις ποιότητας υπηρεσίας και όχι υπηρεσίες με καθορισμένα ποσοτικά επίπεδα ποιότητας υπηρεσίας. Ελλείπει καθορισμένου μοντέλου υπηρεσιών, οι υπηρεσίες που θα

προκύψουν, τα επίπεδα ποιότητας και οι σχετικές εγγυήσεις άπτονται των υπερκείμενων λειτουργιών διαχείρισης.

Συμπεραίνουμε πως *η αρχιτεκτονική IntServ είναι πιο κατάλληλη για μικρά επιχειρηματικά δίκτυα* ενώ τα εγγενή προβλήματα εφαρμογής της σε μεγαλύτερη κλίμακα, αν και είναι δυνατό να εξομαλυνθούν σε ένα βαθμό όπως προτείνεται στο [IS-RAGGR], ωστόσο δεν απαλείφονται, υποδεικνύοντας την *αρχιτεκτονική DiffServ ως περισσότερο κατάλληλη για μεγάλα δίκτυα κορμού και παροχής υπηρεσιών, δεδομένου ότι οι ελλείψεις της τελευταίας σχετικά με ένα καθορισμένο μοντέλο υπηρεσιών που να προσφέρει υπηρεσίες προβλέψιμης ποιότητας αναπληρώνονται από τις κατάλληλες λειτουργίες διαχείρισης.*

Για τους παραπάνω λόγους η παρούσα διατριβή επικεντρώνεται σε δίκτυα DiffServ.

Το πρόβλημα που καλούνται να λύσουν οι λειτουργίες διαχείρισης ενός δικτύου παροχής ποιότητας υπηρεσίας είναι η μεγιστοποίηση του κέρδους του παροχέα. Η προσέγγιση του over-provisioning (βλέπε ενότητα A.1.4.3.3) επιλύει το πρόβλημα μεγιστοποίησης του κέρδους σε επίπεδο εξοπλισμού σε φυσικούς πόρους. Η ποιότητα υπηρεσίας διασφαλίζεται χάρη στην υπό οποιεσδήποτε συνθήκες διαθεσιμότητα των απαραίτητων πόρων, το πάγιο κόστος εξοπλισμού αυξάνεται αλλά το λειτουργικό κόστος του δικτύου ελαχιστοποιείται καθώς πολύπλοκες λειτουργίες διαχείρισης δεν είναι πλέον αναγκαίες. Πέρα από τα συναφή προβλήματα εφαρμογής και το υψηλό πάγιο κόστος που για κάποιους οργανισμούς είναι απαγορευτικό, *το over-provisioning δεν καταργεί την ανάγκη για λειτουργίες διαχείρισης για την παροχή ποιότητας υπηρεσίας πάνω από ένα δίκτυο με πεπερασμένους φυσικούς πόρους* που είναι και πραγματικότητα για την πλειοψηφία των παροχέων σήμερα.

Δεδομένων πεπερασμένων φυσικών πόρων το πρόβλημα μεγιστοποίησης του κέρδους του παροχέα μετατίθεται στο επίπεδο εξυπηρέτησης της ζήτησης και αξιοποίησης των πόρων (βλέπε ενότητα A.1.4.3.2). Με peak allocation εξασφαλίζονται απόλυτες εγγυήσεις ποιότητας υπηρεσίας αλλά η αξιοποίηση των πόρων είναι ιδιαίτερα χαμηλή. Εναλλακτικά χρησιμοποιείται oversubscription που βελτιώνει την αξιοποίηση των πόρων εις βάρος όμως των εγγυήσεων για ποιότητα υπηρεσίας. Ενώ η πραγματική συμφόρηση και κατ' επέκταση η υποβάθμιση της ποιότητας υπηρεσίας δεν είναι πρακτικά δυνατόν να προβλεφθούν με ακρίβεια είναι αντίθετα βέβαιο πως μεγαλύτερη εντατικοποίηση του oversubscription συμβάλλει σε δριμύτερη υποβάθμιση της ποιότητας. *Το όριο όπου η επιβολή χρηματικών κυρώσεων για παραβίαση συμβολαίων και η συσσωρευμένη δυσαρέσκεια των πελατών είναι περισσότερο επιζήμιες από το κέρδος της μεγαλύτερης αξιοποίησης των πόρων προσδιορίζεται από τις πολιτικές του παροχέα* και διαφοροποιείται ανάλογα με την αγορά στην οποία απευθύνεται και τις επιχειρηματικές επιδιώξεις του.

Οι λειτουργίες διαχείρισης διακρίνονται σε λειτουργίες traffic engineering και service management.

Το traffic engineering διαχειρίζεται τις διαδικασίες δέσμευσης πόρων ανά δρομολογητή και δημιουργίας μονοπατιών από άκρο σε άκρο με σκοπό τη βέλτιστη κατανομή των πόρων του δικτύου δεδομένης της κίνησης που καλούνται να εξυπηρετήσουν. Το στατικό traffic engineering διαστασιοποιεί το δίκτυο χρησιμοποιώντας εκτιμήσεις για την ένταση της κίνησης με συγκεκριμένης απαιτήσεις ποιότητας υπηρεσίας από άκρο σε άκρο όπως προκύπτουν από ανάλυση της κίνησης προηγούμενων χρονικών περιόδων. Το δυναμικό traffic engineering λαμβάνει διορθωτικά μέτρα όταν διαπιστώνεται σημαντική απόκλιση από το κριτήριο βελτιστοποίησης της κατανομής των πόρων κατά τη λειτουργία του δικτύου. Είναι σημαντικό η δυναμική αυτή προσαρμογή να μη διαταράσσει την ευστάθεια του συστήματος.

Το service management διαχειρίζεται τις διαδικασίες συνδιαλλαγής με τους πελάτες για τη διαπραγμάτευση και αποδοχή των αιτήσεων υπηρεσιών, καθώς και τις διαδικασίες ενεργοποίησης και ελέγχου των αποδεκτών αιτήσεων με σκοπό τη μεγιστοποίηση της κίνησης που αποκτά πρόσβαση στους

πόρους του δικτύου σε επίπεδο συμβολαίων αλλά και σε επίπεδο έντασης της κίνησης που παράγει το κάθε συμβόλαιο, διασφαλίζοντας ταυτόχρονα τις δεσμεύσεις του παροχέα σε εγγυήσεις ποιότητας υπηρεσίας. Ο έλεγχος εισόδου κίνησης που διαχειρίζεται τη διαδικασία αποδοχής ή απόρριψης αιτήσεων υπηρεσιών εκτιμά εάν η διαθέσιμη χωρητικότητα του δικτύου επαρκεί για την ικανοποίηση των απαιτήσεων της υπό εξέταση αίτησης. Η εκάστοτε διαθέσιμη χωρητικότητα εκτιμάται είτε λαμβάνοντας τις απαραίτητες μετρήσεις, είτε υπολογίζοντας με μαθηματικά μοντέλα τη διαφορά των πόρων από την κατανομή του στατικού traffic engineering και των πόρων που διατίθενται ήδη στις υπάρχουσες ενεργές ροές, είτε με συνδυασμό των δύο. Σε περίπτωση που διαπιστώνεται υπερφόρτωση του δικτύου ο έλεγχος κίνησης πρέπει να είναι σε θέση να επιλύσει τη συμφόρηση.

Η μεγιστοποίηση του κέρδους μέσω της μεγιστοποίησης των ικανοποιημένων αιτήσεων υπηρεσιών επαφίεται της λειτουργίας του ελέγχου εισόδου κίνησης. Οι λειτουργίες του traffic engineering κατανέμοντας βέλτιστα τους φυσικούς πόρους στην εισερχόμενη κίνηση απλά επαυξάνουν τα περιθώρια αυτής της μεγιστοποίησης.

Η παρούσα διατριβή επικεντρώνεται στον έλεγχο εισόδου κίνησης σε δίκτυα DiffServ.

Στην ενότητα που ακολουθεί ορίζεται το πρόβλημα σχετικά με τον έλεγχο εισόδου κίνησης σε δίκτυα DiffServ, παρουσιάζονται οι υπάρχουσες μελέτες, επιχειρείται αξιολόγηση των προτεινόμενων λύσεων και προσδιορισμός των θεμάτων προς περαιτέρω έρευνα.

A.2 Έλεγχος Εισόδου Κίνησης σε Δίκτυα Διαφοροποιημένων Υπηρεσιών

A.2.1 Ανάλυση και Ορισμός του Προβλήματος

Ο πρωταρχικός σκοπός του ελέγχου εισόδου κίνησης είναι η διασφάλιση της ποιότητας υπηρεσίας των αιτήσεων υπηρεσιών που γίνονται αποδεκτές και αποκτούν πρόσβαση στους πόρους του δικτύου. Ο δευτερεύων σκοπός του ελέγχου εισόδου κίνησης είναι η μεγιστοποίηση των αιτήσεων που γίνονται αποδεκτές ή αλλιώς η μεγιστοποίηση της αξιοποίησης των πόρων του δικτύου και κατ' επέκταση η μεγιστοποίηση του κέρδους του παροχέα.

Η ποιότητα υπηρεσίας διασφαλίζεται όταν οι πόροι που απαιτούνται από την κάθε υπηρεσία είναι διαθέσιμοι στο δίκτυο. Για το σκοπό αυτό απαιτείται καταρχήν να καθοριστούν οι πόροι που είναι απαραίτητοι για την εξυπηρέτηση στη συμφωνηθείσα ποιότητα της κίνησης που θα παράγει η υπό εξέταση υπηρεσία και έπειτα να εκτιμηθεί εάν αυτοί οι πόροι είναι διαθέσιμοι στο δίκτυο.

Στις ακόλουθες ενότητες περιγράφονται με λεπτομέρεια τα δύο αυτά στάδια που συνοψίζουν τη λειτουργία του ελέγχου εισόδου κίνησης και παρουσιάζονται κάποιες πρόσθετες σχετικές απαιτήσεις.

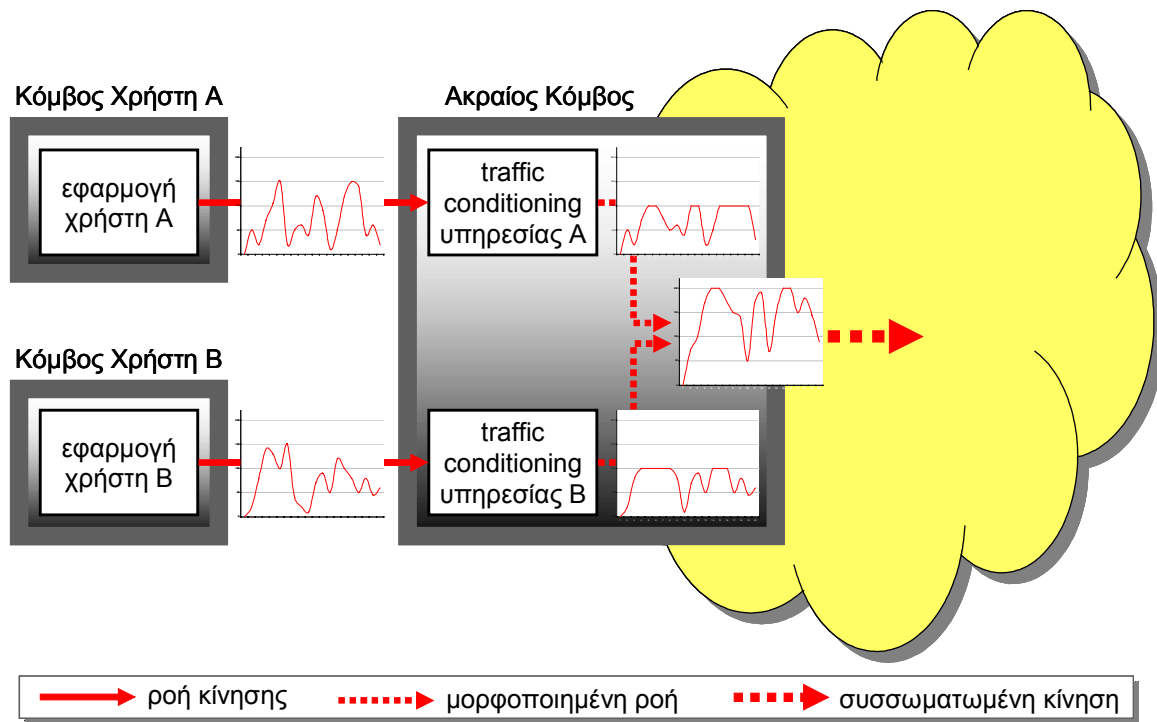
A.2.1.1 Καθορισμός των απαραίτητων πόρων

Οι απαραίτητοι πόροι για την εξυπηρέτηση μιας ροής καθορίζονται συναρτήσει των χαρακτηριστικών της κίνησης της ροής και της ποιότητας υπηρεσίας που απαιτείται.

Η κίνηση που παράγεται από την κάθε υπηρεσία εξαρτάται από τον τύπο της εφαρμογής που τη χρησιμοποιεί και μπορεί να είναι είτε σταθερού είτε μεταβλητού ρυθμού (constant και variable bit rate), με διαφορετική κατανομή, διαφορετικό μέσο ρυθμό (mean rate), ρυθμό αιχμής (peak rate), μέγιστο καταγισμό (maximum burst size) κ.λπ. Κατά κανόνα τα σύμβολα υπηρεσίας που εγγυώνται κάποιο επίπεδο ποιότητας θέτουν και περιορισμούς στην κίνηση του θα εξυπηρετηθεί με αυτή την ποιότητα, συμφωνείται ένα μοντέλο στο οποίο η κίνηση της εφαρμογής οφείλει να συμμορφώνεται ειδάλλως θα αστυνομεύεται. Έτσι, αντί για την ίδια την κίνηση της εφαρμογής αρκεί να μπορεί να καθοριστεί η κίνηση που θα προκύπτει μετά τη μορφοποίηση, αυτή που τελικά θα διοχετεύεται στο δίκτυο (βλέπε Σχήμα A-4).

Η μορφοποιημένη κίνηση παραμένει μεταβλητού ρυθμού με μέσο ρυθμό και ρυθμό αιχμής. Εκτενείς μελέτες που έγιναν σχετικά στα πλαίσια της τεχνολογίας ATM έχουν δείξει πως η κίνηση μεταβλητού ρυθμού είναι δύσκολο να χαρακτηριστεί και η παρεμβολή μορφοποίησης δεν λύνει το πρόβλημα, απλά το μεταθέτει στο χαρακτηρισμό της μορφοποιημένης και προφανώς μικρότερης διακύμανσης κίνησης.

Οι πόροι του δικτύου χρησιμοποιούνται από όλες τις ενεργές ροές και το κριτήριο μεγιστοποίησης του κέρδους του παροχέα υποδεικνύει πως η επικάλυψη μεταξύ τους δεν είναι δυνατόν να παραβλεφθεί (βλέπε ενότητα A.1.4.3.2). Συγκεκριμένα, χάρη στη στατιστική πολυπλεξία η συσσωματωμένη κίνηση (aggregated traffic) (βλέπε Σχήμα A-4) που παράγεται από ένα σύνολο ροών παρουσιάζει κατανομή με ρυθμό αιχμής μικρότερο από το άθροισμα των ρυθμών αιχμής της κίνησης των επιμέρους ροών και άρα συνολικά απαιτούνται λιγότεροι πόροι για την εξασφάλιση της ποιότητας υπηρεσίας κάθε ροής. Το κέρδος αυτό σε απαιτούμενους πόρους ονομάζεται κέρδος πολυπλεξίας (multiplexing gain). Ο έλεγχος διαθεσιμότητας των πόρων γίνεται για τη συσσωματωμένη κίνηση και όχι για τις επιμέρους ροές.



Σχήμα A-4: μορφοποιημένη κίνηση υπηρεσίας

Ανάλογα με τον τύπο της εφαρμογής διαφέρει και η ανεκτικότητα σε περιστασιακή υποβάθμιση της ποιότητας και κατ' επέκταση οι εγγυήσεις για εξασφάλιση της συμφωνηθείσας ποιότητας υπηρεσίας. Σε πρώτο επίπεδο η ποιότητα υπηρεσίας είναι δυνατό να αποτιμάται απόλυτα ή στατιστικά και σε δεύτερο επίπεδο η απόλυτη ή στατιστική ικανοποίησή της μπορεί να είναι είτε εγγυημένη είτε αρκετά πιθανή αλλά χωρίς καθορισμένες εγγυήσεις. Συγκεκριμένα διακρίνουμε τρεις τύπους υπηρεσιών, τις εγγυημένες υπηρεσίες (guaranteed services) με εγγυημένη απόλυτη ικανοποίηση της συμφωνηθείσας ποιότητας, τις υπηρεσίες στατιστικών εγγυήσεων (probabilistic services) με εγγυημένη στατιστική ικανοποίηση και τις προβλέψιμες υπηρεσίες (predictive services) με πιθανή αλλά όχι αυστηρά καθορισμένη ικανοποίηση. Για παράδειγμα έστω η μέγιστη καθυστέρηση συμφωνείται σε μια τιμή d_{max} , σε μια εγγυημένη υπηρεσία κανένα πακέτο δεν θα καθυστερήσει περισσότερο από d_{max} , σε μια υπηρεσία στατιστικών εγγυήσεων η πιθανότητα ένα πακέτο να υπερβεί το d_{max} είναι καθορισμένη έστω p_d και άρα το ποσοστό των πακέτων με καθυστέρηση μεγαλύτερη από d_{max} είναι άνω φραγμένο στο p_d , ενώ σε μια προβλέψιμη υπηρεσία το ποσοστό των πακέτων με καθυστέρηση μεγαλύτερη από d_{max} εκτιμάται πως θα είναι χαμηλό χωρίς όμως εγγυημένο άνω όριο.

Οι εγγυημένες υπηρεσίες εξασφαλίζουν την ποιότητα υπηρεσίας σε κάθε περίπτωση, ακόμα και για το εξαιρετικά απίθανο ενδεχόμενο να συμπέσει η μετάδοση σε ρυθμό αιχμής για όλες οι ενεργές πηγές. Η επικάλυψη μεταξύ των ροών δεν υπολογίζεται κι έτσι το multiplexing gain είναι μηδενικό. Το σχήμα αυτό λέγεται peak allocation (βλέπε ενότητα A.1.4.3.2) και επιτυγχάνει απόλυτες εγγυήσεις με κόστος την ιδιαίτερα χαμηλή αξιοποίηση των πόρων. Αντίθετα, στις υπηρεσίες στατιστικών εγγυήσεων, όσο μεγαλύτερη η πιθανότητα απόκλισης από τη συμφωνηθείσα ποιότητα τόσο μεγαλύτερο το περιθώριο υπολογισμού της επικάλυψης μεταξύ των διαφορετικών ροών και συνεπώς τόσο μεγαλύτερο multiplexing gain. Στις προβλέψιμες υπηρεσίες το γεγονός πως δεν παρέχεται καμία άκαμπτη εγγύηση προσδίδει ελευθερία στον καθορισμό των απαραίτητων πόρων. Έτσι αφενός η διαδικασία μπορεί να απλοποιηθεί σημαντικά και αφετέρου παύει να είναι απαραίτητη η θέσπιση συντηρητικών a priori υποθέσεων για τα χαρακτηριστικά της κίνησης με αποτέλεσμα την ακόμα καλύτερη αξιοποίηση των πόρων.

Ο τύπος των υπηρεσιών, η αντίστοιχη ποιότητα υπηρεσίας και οι σχετικές εγγυήσεις προκύπτουν από τις απαιτήσεις των εφαρμογών και τις σχετικές πολιτικές του παροχέα. Έτσι, μία σημαντική πτυχή του προβλήματος είναι οι σχετικές παραδοχές με βάση τις οποίες οικοδομείται το μοντέλο υπηρεσιών που χρησιμοποιεί ο κάθε αλγόριθμος ελέγχου εισόδου κίνησης, και κατά πόσο είναι δεσμευτικές ως προς τον τύπο και τρόπο εξασφάλισης της ποιότητας υπηρεσίας.

A.2.1.2 Εκτίμηση διαθεσιμότητας των απαραίτητων πόρων

Για τη διασφάλιση ποιότητας υπηρεσίας οι απαραίτητοι πόροι πρέπει να είναι διαθέσιμοι σε κάθε κόμβο που διατρέχει μια ροή. Ενώ στην αρχιτεκτονική IntServ ο έλεγχος εισόδου κίνησης γίνεται με σηματοδοσία μέσω του RSVP σε κάθε κόμβο όπου και εκτιμάται η διαθεσιμότητα με βάση την πληροφορία για όλες τις ροές που διέρχονται από τον κόμβο, στην αρχιτεκτονική DiffServ δεν υπάρχει αντίστοιχη πληροφορία στους κόμβους κορμού και η πολυπλοκότητα της διαχείρισης σε επίπεδο ροής εξωθείται στους ακραίους κόμβους.

Το σημείο που επιφορτίζεται με τον έλεγχο εισόδου κίνησης αντί των κόμβων κορμού πρέπει να είναι σε θέση να εκτιμήσει τους διαθέσιμους πόρους για κάθε κόμβο. Αυτό συνεπάγεται πληροφορία για τις δυνατότητες και τη χωρητικότητα του κάθε κόμβου, πληροφορία δρομολόγησης των ενεργών ροών για τον εντοπισμό αυτών που διατρέχουν τον κάθε κόμβο και πληροφορία για τα χαρακτηριστικά κίνησης της κάθε μίας από αυτές τις ενεργές ροές. Γίνεται εύκολα αντιληπτό ότι η διαχείριση αυτής της πληροφορίας σε μεγάλη κλίμακα είναι δυσχερής.

Το κριτήριο μεγιστοποίησης της αξιοποίησης των πόρων του δικτύου επιτάσσει τον κατά το δυνατό βέλτιστο διαμορισμό των πόρων σε όλη την έκταση του δικτύου σύμφωνα με την πραγματική ζήτηση για πόρους από τις αιτήσεις υπηρεσίας. Η λήψη αποφάσεων γίνεται με βέλτιστο τρόπο μόνο όταν συνεκτιμάται όλη η ως άνω πληροφορία όπως ανά πάσα στιγμή διαμορφώνεται από την πραγματική ζήτηση. Η ιδιαίτερα επίπονη διαδικασία της λήψης αποφάσεων με βέλτιστο τρόπο είναι δυνατόν να αποφορτιστεί σημαντικά εάν η λειτουργία της βασιστεί σε κάποιες παραδοχές, εις βάρος ωστόσο της μεγιστοποίησης της αξιοποίησης των πόρων.

Έτσι, διακρίνουμε τη μέθοδο *κοινόχρηστων πόρων* (sharing) και τη μέθοδο *διαμερισμού πόρων* (partitioning). Σύμφωνα με τη μέθοδο sharing οι πόροι θεωρούνται κοινοί και η εκάστοτε διανομή τους πρέπει να επιλύεται με βέλτιστο ανά πάσα στιγμή τρόπο μεταξύ των ανταγωνιζόμενων αιτήσεων υπηρεσίας. Αντίθετα, στη μέθοδο partitioning οι πόροι διαμοιράζονται a priori μεταξύ των ακραίων κόμβων, ο έλεγχος εισόδου κατανέμεται σε κάθε ακραίο κόμβο για τις αιτήσεις υπηρεσίας που διέρχονται από αυτόν και για το μερίδιο των πόρων κορμού που του αντιστοιχεί, συνεισφέροντας αφενός στην αποφόρτιση της διαδικασίας υστερώντας αφετέρου σε αξιοποίηση των πόρων που μένουν αχρησιμοποίητοι από τα μερίδια άλλων κόμβων εισόδου.

Στην αρχιτεκτονική DiffServ οι πόροι κάθε κόμβου κατανέμονται μεταξύ των διαφορετικών PHBs που υλοποιεί ο κόμβος. Αν χρησιμοποιείται not work-conserving μηχανισμός χρονοπρογραμματισμού (βλέπε ενότητα A.1.4.1.1) επιτρέπεται η χρήση πόρων που έχουν αρχικά κατανεμηθεί σε ένα συγκεκριμένο PHB αλλά τυχαίνει να μην χρησιμοποιούνται από αυτό, για άλλα PHBs που την ίδια στιγμή υπερβαίνουν το δικό τους μερίδιο. Όταν χρησιμοποιείται η μέθοδος partitioning αυτό το πλεονέκτημα χάνεται καθώς ο διαμερισμός των πόρων κορμού στους ακραίους κόμβους γίνεται με την παραδοχή των ανά PHB για κάθε κόμβο μεριδίων και η δυνατότητα υπέρβασης αυτού του μεριδίου δεν λαμβάνεται υπόψη.

Η μεγαλύτερη πρόκληση για τον έλεγχο εισόδου κίνησης σε δίκτυα DiffServ έγκειται στην ανεύρεση εκείνου του διαχειρίσιμου υποσυνόλου της πληροφορίας για την πραγματική κατάσταση του δικτύου και

των κατάλληλων παραδοχών και μηχανισμών που να επαρκούν για τη διασφάλιση της ποιότητας υπηρεσίας με τις απαιτούμενες σε κάθε περίπτωση εγγυήσεις χωρίς αδικαιολόγητα μεγάλο αντίκτυπο στην αξιοποίηση των πόρων.

A.2.1.3 Πρόσθετες Απαιτήσεις

Οι πρόσθετες απαιτήσεις για το μηχανισμό ελέγχου εισόδου αφορούν στην αποτελεσματικότητα, τη δυνατότητα επίλυσης της συμφόρησης, την αμεροληψία, την ευστάθεια και την ευρωστία, και την πρακτική εφαρμογή του.

A.2.1.3.1 Αποτελεσματικότητα

Η αποτελεσματικότητα του μηχανισμού ελέγχου εισόδου κίνησης έχει αντίκτυπο στο δίκτυο αλλά και στους χρήστες. Πέραν των βασικών του στόχων, ο έλεγχος εισόδου κίνησης πρέπει να είναι όσο το δυνατόν διαφανής διαδικασία και στα δύο αυτά επίπεδα.

Η επιβάρυνση του δικτύου από μηνύματα που ανταλλάσσονται για τον έλεγχο εισόδου κίνησης πρέπει να είναι όσο το δυνατόν μικρότερη και σε λογικά επίπεδα. Στην περίπτωση που υιοθετείται καταναμημένο μοντέλο η διατήρηση συνεπούς και ακριβής εικόνας μεταξύ των διαφορετικών σημείων μπορεί να αποδειχθεί ιδιαίτερα πολύπλοκη, συνεπώς η απαραίτητη επικοινωνία μεταξύ τους θα πρέπει να σχεδιαστεί προσεκτικά. Επιπλέον οι μετρήσεις που προϋποθέτει ο κάθε μηχανισμός για την εκτίμηση της πραγματικής κατάστασης στο δίκτυο πρέπει να είναι περιορισμένες καθώς επιβαρύνουν σε επεξεργασία τους κόμβους του δικτύου και σε εύρος ζώνης για τη συγκέντρωσή τους στα σημεία επεξεργασίας.

Αντίστοιχα, οι χρόνοι απόκρισης στις αιτήσεις υπηρεσίας προς τους χρήστες πρέπει να είναι μικροί έτσι ώστε είτε η επιτυχής ενεργοποίηση είτε η απόρριψη μίας αίτησης να μην καθυστερεί σημαντικά.

Η αποτελεσματικότητα του ίδιου του αλγορίθμου κρίνεται από το μέγεθος της πληροφορίας και την ισχύ επεξεργασίας που απαιτεί.

A.2.1.3.2 Επίλυση Συμφόρησης

Λόγω της ύπαρξης υπηρεσιών με εγγυήσεις στατιστικής φύσεως αναμένεται πως το δίκτυο περιστασιακά θα παρουσιάζει συμφόρηση. Η αναγκαιότητα επίλυσης της συμφόρησης προκύπτει από πολλούς παράγοντες [MR99]. Ο πιο σημαντικός είναι η φύση του πρωτοκόλλου TCP το οποίο, όταν διαγνώσει απώλεια πακέτων προβαίνει σε αναμεταδόσεις επαυξάνοντας έτσι την συμφόρηση στο δίκτυο. Ένας άλλος παράγοντας είναι η συμπεριφορά των χρηστών. Συχνά, όταν η ποιότητα υπηρεσίας είναι πολύ κακή ο χρήστης εγκαταλείπει την εφαρμογή με αποτέλεσμα την σπατάλη χωρίς κανένα όφελος των πόρων που ήδη χρησιμοποιήθηκαν ακριβώς τη στιγμή που είναι πιο απαραίτητοι και που επιπλέον θα χρησιμοποιηθούν ξανά στην επόμενη απόπειρα χρήσης της υπηρεσίας.

Ο έλεγχος εισόδου κίνησης επιφορτίζεται με την ευθύνη της επίλυσης αυτής της συμφόρησης ανακόπτοντας την ενεργοποίηση νέων αιτήσεων υπηρεσιών και μειώνοντας την κίνηση που διοχετεύεται στο δίκτυο από τις ενεργές ροές είτε διακόπτοντας κάποιες (pre-emption) είτε μειώνοντας την έντασή τους. Η επίλυση της συμφόρησης πρέπει να γίνεται άμεσα και αποτελεσματικά έτσι ώστε αφενός το δίκτυο να βγαίνει γρήγορα από την κατάσταση συμφόρησης αφετέρου να επιβάλλονται οι λιγότερες δυνατές κυρώσεις. Η πολυπλοκότητα αυτής της διαδικασίας όταν η επίλυση της συμφόρησης γίνεται με καταναμημένο τρόπο είναι προφανής.

A.2.1.3.3 Αμεροληψία

Μία άλλη σημαντική ιδιότητα του ελέγχου εισόδου κίνησης αφορά την αμεροληψία του. Δεδομένων των πολιτικών του παροχέα που ενδέχεται να ευνοούν συγκεκριμένους πελάτες ή τύπους υπηρεσιών, η αποδοχή ή απόρριψη αιτήσεων υπηρεσιών, η ποιότητα υπηρεσίας που λαμβάνουν τελικά και οι κυρώσεις σε περίπτωση συμφόρησης θα πρέπει να επιβάλλονται με δίκαιο τρόπο μεταξύ των υπηρεσιών της ίδιας κατηγορίας σε όλη την έκταση του δικτύου. Παράγοντες όπως η διάρκεια των υπηρεσιών, το μήκος του μονοπατιού ή οι απαιτούμενοι πόροι δεν θα πρέπει να είναι έμφυτοι του αλγορίθμου και να επηρεάζουν τη λειτουργία του παρά μόνο εάν έτσι καθορίζεται από τις πολιτικές του παροχέα.

A.2.1.3.4 Ευστάθεια και Ευρωστία

Επιπλέον, όπως κάθε μηχανισμός διαχείρισης, ο έλεγχος εισόδου κίνησης θα πρέπει να είναι ευσταθής (stable) και εύρωστος (robust). Καθώς η πραγματική κατάσταση του δικτύου επηρεάζει την λειτουργία του και η λειτουργία του την κατάσταση του δικτύου θα πρέπει να εξασφαλίζεται πως αυτή η αλληλεπίδραση δεν θα οδηγεί σε ταλαντώσεις τον αλγόριθμο και κατά συνέπεια το σύστημα. Ακόμα, θα πρέπει να σχεδιαστεί ώστε να αντιμετωπίζει όλες τις ενδεχόμενες περιπτώσεις χωρίς να αποτυγχάνει ακόμα και όταν αντιμετωπίζει μη πιστοποιημένους ή/και κακόβουλους χρήστες.

A.2.1.3.5 Πρακτική Εφαρμογή

Τέλος, για να είναι πρακτικά χρήσιμος ένας αλγόριθμος ελέγχου εισόδου κίνησης θα πρέπει να μπορεί να εφαρμοστεί χωρίς να απαιτεί νέες ειδικές λειτουργίες από το υλικό και το λογισμικό των κόμβων αλλά βασιζόμενος στις ήδη προδιαγεγραμμένες και υλοποιημένες.

A.2.2 Βασικές Αρχές Γνωστών Προσεγγίσεων

Οι γνωστές προσεγγίσεις διακρίνονται με βάση τα εξής κριτήρια:

- Κριτήριο μεθόδου ελέγχου εισόδου κίνησης.
- Κριτήριο τοπολογικής ή/και λειτουργικής κατανομής.
- Κριτήριο μεθόδου κατανομής των πόρων.
- Κριτήριο στατικής ή δυναμικής λειτουργίας.

Στις επόμενες ενότητες παρουσιάζονται αναλυτικά οι βασικές αρχές σχετικά με καθένα από τα παραπάνω κριτήρια.

A.2.2.1 Μέθοδος ελέγχου εισόδου κίνησης

Ο έλεγχος εισόδου κίνησης είναι μια λειτουργία που μελετήθηκε διεξοδικά αρχικά στα δίκτυα τεχνολογίας ATM και στη συνέχεια στα δίκτυα IntServ. Και οι δύο αυτές τεχνολογίες χρησιμοποιούν ανά κόμβο σηματοδότηση (hop-by-hop signalling) για την εγκατάσταση μίας κλήσης ή αλλιώς την ενεργοποίηση μιας υπηρεσίας. Έτσι, οι πρώτες και θεμελιώδεις μελέτες που έχουν διεξαχθεί σχετικά είναι μονού κόμβου (single-node).

Οι πόροι που χρησιμοποιεί ο έλεγχος εισόδου κίνησης σε ένα κόμβο είναι γνωστοί και αμετάβλητοι, ίσοι με τη χωρητικότητα των ζεύξεων του κόμβου. Έτσι, ο έλεγχος εισόδου περιορίζεται στον καθορισμό των απαραίτητων πόρων για τη συσσωματωμένη κίνηση και σύγκριση των απαραίτητων πόρων με τη γνωστή

χωρητικότητα της ζεύξης. Ανάλογα με τη μέθοδο που χρησιμοποιείται για τον καθορισμό των απαραίτητων πόρων οι μελέτες single-node διακρίνονται σε *βασιζόμενες σε μοντέλα* (model-based) ή *αλλιώς βασιζόμενες σε παραμέτρους* (parameter-based) και *βασιζόμενες σε μετρήσεις* (measurement-based).

Στις model-based ή parameter-based μεθόδους η κίνηση που παράγει η κάθε ροή περιγράφεται από μαθηματικά μοντέλα και με βάση κάποιες a priori καθορισμένες παραμέτρους. Με βάση αυτή την πληροφορία που καταχωρείται ανά ροή με τη χρήση των κατάλληλων τεχνικών συσσωμάτωσης (aggregation) καθορίζονται κάθε φορά οι απαραίτητοι πόροι για την εξυπηρέτηση του συνόλου των ενεργών και της νέας υπηρεσίας. Στις measurement-based μεθόδους οι απαραίτητοι πόροι για τη συσσωματωμένη ροή συνάγονται απευθείας από μετρήσεις της πραγματικής κίνησης χωρίς τη χρήση a priori παραδοχών για τα χαρακτηριστικά της κίνησης των ροών. Είναι φανερό πως στις measurement-based μεθόδους δεν υπάρχει μαθηματική ακρίβεια τέτοια ώστε να επιτρέπει τη διασφάλιση της ποιότητας υπηρεσίας. Συνεπώς οι parameter-based μέθοδοι χρησιμοποιούνται για εγγυημένες υπηρεσίες και υπηρεσίες στατιστικών εγγυήσεων ενώ οι measurement-based μέθοδοι για προβλέψιμες υπηρεσίες.

Όταν αντίστοιχες μέθοδοι χρησιμοποιούνται σε multi-node περιβάλλοντα και η λειτουργία του ελέγχου εισόδου κίνησης κατανέμεται στα άκρα του δικτύου και δεν γίνεται με σηματοδοσία ανά κόμβο (hop-by-hop signalling) τότε απαιτείται να είναι γνωστή η χωρητικότητα από άκρο σε άκρο (end-to-end). Η χωρητικότητα από άκρο σε άκρο όμως εξαρτάται από τη δρομολόγηση και την ένταση της κίνησης και από άλλους ακραίους κόμβους. Η εικόνα χωρητικότητας που βλέπει κάθε δείγμα (instance) του κατανεμημένου ελέγχου εισόδου καθορίζεται από τη μέθοδο κατανομής πόρων (βλέπε ενότητα A.2.2.3).

Μια μεταγενέστερη κατηγορία μεθόδων ελέγχου εισόδου κίνησης καταργεί το παραδοσιακό μοντέλο της έμμεσης εκτίμησης των διαθέσιμων πόρων από το υπόλοιπο της γνωστής χωρητικότητας μείον τους πόρους που καταναλώνουν οι ήδη ενεργές ροές και υιοθετεί ένα νέο μοντέλο άμεσης εκτίμησης των διαθέσιμων πόρων με τη χρήση ενεργών μετρήσεων (active measurements). Δοκιμαστικά πακέτα (probes) αποστέλλονται προς τον προορισμό και στη συνέχεια συγκεντρώνονται μετρήσεις της ποιότητας υπηρεσίας που έλαβαν η οποία συγκρίνεται με την απαιτούμενη ποιότητα υπηρεσίας και το αποτέλεσμα καθορίζει την αποδοχή ή απόρριψη της αίτησης υπηρεσίας. Οι μέθοδοι αυτοί ονομάζονται *βασιζόμενες σε δοκιμή* (probe-based) και όπως και οι measurement-based μέθοδοι χρησιμοποιούνται για προβλέψιμες υπηρεσίες.

Οι βασικές αρχές των τριών μεθόδων περιγράφονται εκτενέστερα στις ακόλουθες ενότητες.

A.2.2.1.1 Parameter-Based

Παραδοσιακά από τα τηλεφωνικά δίκτυα και τα δίκτυα ATM ο έλεγχος εισόδου κίνησης βασίζεται σε parameter-based μεθόδους.

Τα χαρακτηριστικά της κίνησης της κάθε ροής και άρα της κάθε εφαρμογής που την παράγει περιγράφονται a priori από ντετερμινιστικά ή στοχαστικά μαθηματικά μοντέλα και με βάση κάποιες καθορισμένες παραμέτρους. Κατάλληλες τεχνικές συσσωμάτωσης με βάση τα a priori καθορισμένα χαρακτηριστικά της κάθε ροής υπολογίζουν την ποιότητα υπηρεσίας που θα λάβει η κίνηση που θα προκύψει συνολικά με την προσθήκη της υπό εξέταση αίτησης υπηρεσίας δεδομένης της φυσικής χωρητικότητας της ζεύξης του κόμβου. Η ποιότητα υπηρεσίας περιγράφεται είτε ντετερμινιστικά είτε στατιστικά σε αντιστοιχία με τις εγγυημένες υπηρεσίες και τις υπηρεσίες στατιστικών εγγυήσεων. Η μελέτη [KS99] παρέχει μία πλήρη σύνοψη και ανάλυση των single-node parameter-based μεθόδων για την παροχή ποιότητας υπηρεσίας με στατιστικές εγγυήσεις.

Η δυνατότητα a priori χαρακτηρισμού με ακρίβεια της κίνησης μιας ροής είναι προϋπόθεση για τη λειτουργία των parameter-based σχημάτων η οποία συχνά δυστυχώς δεν ισχύει. Σε αρκετές περιπτώσεις οι εφαρμογές δεν είναι σε θέση να περιγράψουν a priori την κίνηση που θα παραχθεί, όταν για παράδειγμα το περιεχόμενο παράγεται δυναμικά ή και λόγω μη προσδιορίσιμης παραμόρφωσης του παραγόμενου traffic profile από τους κόμβους που ενδεχομένως παρεμβάλλονται μεταξύ της πηγής και του ακραίου κόμβου του δικτύου. Επιπλέον, σε κάποιες περιπτώσεις η συμπεριφορά της παραγόμενης κίνησης είναι ιδιαίτερα πολύπλοκη με διακυμάνσεις πολλαπλής περιοδικότητας και άρα αδύνατο να χαρακτηριστεί με τα συνήθη απλά μοντέλα όπως το token bucket. Η χρήση περισσότερο πολύπλοκων μοντέλων για το χαρακτηρισμό τέτοιας κίνησης είναι πρακτικά απαγορευτική εξαιτίας αφενός της σχετικής επιβάρυνσης της διαδικασίας ελέγχου κίνησης και αφετέρου της δυσκολίας καθορισμού του policing μηχανισμού που χρησιμοποιείται κατά την εξυπηρέτηση της ροής για τη διασφάλιση του δικτύου. Σε ένα περιβάλλον όπως το σημερινό διαδίκτυο όπου η συμπεριφορά των διάφορων εφαρμογών είναι ιδιαίτερα ετερογενής και νέες εφαρμογές αναπτύσσονται με μεγάλο ρυθμό, ο περιορισμός του a priori χαρακτηρισμού της κίνησης είναι ακόμα περισσότερο σημαντικός.

Δεδομένης της αδυναμίας ακριβή a priori χαρακτηρισμού της κίνησης η τακτική που ακολουθείται συνήθως για τη διασφάλιση ποιότητας υπηρεσίας είναι ο χαρακτηρισμός της κίνησης βάσει της συμπεριφοράς της στη χειρότερη περίπτωση (worst-case). Αυτό έχει ως αποτέλεσμα τη δέσμευση πόρων που τον περισσότερο χρόνο δεν χρησιμοποιούνται και άρα χαμηλή αξιοποίηση των πόρων, ιδιαίτερα όσο περισσότερο καταγιστική (bursty) είναι η κίνηση.

Ένα άλλο βασικό μειονέκτημα των parameter-based σχημάτων είναι η επιβάρυνση σε σηματοδότηση και επεξεργαστική ισχύ που προκύπτει από την καταχώρηση και διαχείριση πληροφορίας για κάθε ροή στους κόμβους του δικτύου. Ο έλεγχος εισόδου κίνησης για κάθε αίτηση υπηρεσίας που εξυπηρετεί εκτελεί συχνά πολύπλοκους υπολογισμούς επί του συνόλου των χαρακτηριστικών των ροών για την εξαγωγή των χαρακτηριστικών της συσσωματωμένης κίνησης που προκύπτει με την νέα αίτηση υπηρεσίας, επιβαρύνοντας σημαντικά την απόδοση της διαδικασίας με αποτέλεσμα να μην είναι εφαρμόσιμη σε μεγάλη κλίμακα.

Ωστόσο, η παροχή εγγυημένων υπηρεσιών και υπηρεσιών στατιστικών εγγυήσεων γίνεται παραδοσιακά με parameter-based σχήματα. Όταν η φύση της υπηρεσίας επιτρέπει κάποια απόκλιση στην παρεχόμενη ποιότητα υπηρεσίας όπως συμβαίνει για τις προβλέψιμες υπηρεσίες τότε μπορούν να εφαρμοστούν measurement-based σχήματα.

A.2.2.1.2 Measurement-Based

Στα measurement-based σχήματα ο έλεγχος εισόδου κίνησης βασίζεται στις πραγματικές ιδιότητες της κίνησης όπως προκύπτουν από τις μετρήσεις και όχι σε χαρακτηριστικά που καθορίζονται a priori από τους χρήστες και σε ανακριβή μοντέλα που αναπαριστούν με συντηρητικό τρόπο την συμπεριφορά των πηγών. Η λειτουργία του ελέγχου εισόδου κίνησης απλοποιείται σημαντικά. Αφενός δεν απαιτείται πληροφορία ανά ροή αφού οι μετρήσεις γίνονται για το σύνολο των ενεργών ροών και αφετέρου η πολύπλοκη ανάλυση και επεξεργασία αυτής της πληροφορίας για κάθε αίτηση υπηρεσίας.

Συγκριτικά με τα parameter-based, τα measurement-based σχήματα επιτυγχάνουν κατά πολύ πιο υψηλή αξιοποίηση των πόρων χάρη στην άρση των συντηρητικών a priori εκτιμήσεων της συμπεριφοράς των πηγών. Ωστόσο, καθώς το στιγμιότυπο της κίνησης που προκύπτει από κάθε μέτρηση μπορεί να διαφέρει σημαντικά από τις πραγματικές ιδιότητές της δεν μπορεί να θεωρηθεί και αρκετά αξιόπιστο για την

διασφάλιση καθορισμένων εγγυήσεων. Έτσι, αντίθετα με τα parameter-based, τα measurement-based σχήματα είναι κατάλληλα μόνο για προβλέψιμες υπηρεσίες.

Η μελέτη [BJS00] παρέχει μια σύνοψη και αξιολόγηση των single-node measurement-based μεθόδων. Οι σχετικές μελέτες διαφέρουν ως προς τις μεθόδους μέτρησης και τους αλγόριθμους απόρριψης ή αποδοχής των αιτήσεων υπηρεσιών. Η μέθοδος μέτρησης που χρησιμοποιείται για την εκτίμηση του φόρτου του δικτύου μπορεί να είναι απλές περιοδικές μετρήσεις του ρυθμού άφιξης (simple point sampling), να βασίζεται σε εκθετικό με βάρη υπολογισμό του μέσου όρου προηγούμενων δειγμάτων (exponential weighted moving average), ή και μέθοδοι που λαμβάνουν υπόψη και τη διακύμανση του ρυθμού άφιξης. Κάποιοι από τους αλγόριθμους απόρριψης ή αποδοχής των αιτήσεων υπηρεσιών μπορεί να έχουν μαθηματικό υπόβαθρο στοχεύοντας σε μια πιο δομημένη εκτίμηση των πραγματικών ιδιοτήτων της κίνησης και άρα της δυνητικής εξέλιξης της διαδικασίας άφιξης.

Στους measurement-based αλγορίθμους, περισσότερο συντηρητική εκτίμηση των πραγματικών ιδιοτήτων της κίνησης με βάση τις τιμές των μετρήσεων έχει ως αποτέλεσμα μικρότερη αξιοποίηση των πόρων ενώ αντίθετα περισσότερο αισιόδοξη εκτίμηση αποφέρει μεγαλύτερες αποκλίσεις από την προσδοκώμενη ποιότητα υπηρεσίας. Η πλειονότητα των αλγορίθμων ρυθμίζει τη λειτουργία του από λιγότερο σε περισσότερο συντηρητική με τη χρήση κάποιας παραμέτρου. Σε κάποιους αλγορίθμους η ρύθμιση της τιμής της παραμέτρου επαφίεται στον παροχέα ενώ άλλοι έχουν την απαραίτητη λογική για να την προσαρμόζουν κατάλληλα με το περιβάλλον και την παρατηρούμενη απόδοση του αλγορίθμου.

Σύμφωνα με τους [BJS00] η απόδοση των αλγορίθμων που μελετήθηκαν ως προς τη σχέση της αξιοποίησης των πόρων και της υποβάθμισης της ποιότητας είναι σχεδόν ταυτόσημη, ανεξάρτητα με το μαθηματικό υπόβαθρο και την πολυπλοκότητα του καθενός. Επιπλέον, σε όσους αλγορίθμους επιχειρείται ακριβής αντιστοίχιση καθορισμένης απόκλισης από την ποιότητα της υπηρεσίας με συγκεκριμένη τιμή της παραμέτρου ρύθμισης για την περισσότερο ή λιγότερο συντηρητική εκτίμηση των ιδιοτήτων της κίνησης αποτυγχάνει, καθιστώντας έτσι αναγκαία κάποια περίοδο παρακολούθησης από τον παροχέα της πραγματικής συμπεριφοράς του αλγορίθμου στο συγκεκριμένο περιβάλλον ώστε να είναι σε θέση εμπειρικά να εφαρμόσει τις κατάλληλες ρυθμίσεις.

A.2.2.1.3 Probe-based

Στις probe-based μεθόδους η αποδοχή ή απόρριψη μιας αίτησης υπηρεσίας αποφασίζεται μετά από μια περίοδο δοκιμής (probing period). Κατά την περίοδο δοκιμής δοκιμαστικά πακέτα (probes) αποστέλλονται προς τον προορισμό προσομοιάζοντας κατά το δυνατό τα χαρακτηριστικά κίνησης της ροής που θα χρησιμοποιήσει την υπηρεσία. Στο τέλος της δοκιμαστικής περιόδου συντάσσεται στον προορισμό μια αναφορά με τις σχετικές μετρήσεις και αποστέλλεται πίσω στον αποστολέα όπου και εκτιμάται αν η ποιότητα που παρέχει το δίκτυο εκείνη τη στιγμή είναι αποδεκτή ή όχι με κριτήριο την απαιτούμενη ποιότητα της συγκεκριμένης υπηρεσίας, αποφασίζοντας κατά συνέπεια και για την αποδοχή ή απόρριψη της αίτησης υπηρεσίας.

Καθώς με τις μεθόδους probe-based η ποιότητα υπηρεσίας που παρέχει το δίκτυο μπορεί να υπολογιστεί χωρίς να είναι απαραίτητη καμία άλλη πληροφορία για τη δομή και τη χωρητικότητα του δικτύου ο έλεγχος εισόδου κίνησης μπορεί να τοποθετηθεί και έξω από το δίκτυο ή αλλιώς στους τερματικούς κόμβους του χρήστη (βλέπε ενότητα A.2.2.2.2). Σε αυτή την περίπτωση ο ίδιος ο τερματικός κόμβος ή αλλιώς η ίδια η εφαρμογή αποφασίζει εάν η ποιότητα είναι κατάλληλη και με το πέρας της περιόδου δοκιμής προχωρά στη φάση μετάδοσης. Η αλληλεπίδραση με τις λειτουργίες διαχείρισης του δικτύου είναι μηδαμινή.

Κάθε μελέτη χρησιμοποιεί διαφορετικά κριτήρια και διαφορετικές παραμέτρους της ποιότητας υπηρεσίας. Άλλες μελέτες βασίζονται στο ποσοστό απώλειας των δοκιμαστικών πακέτων, άλλες στην καθυστέρηση, τη διακύμανση της καθυστέρησης κ.λπ. Κάποιες μελέτες χρησιμοποιούν ενδείξεις για την κατάσταση συμφόρησης του δικτύου αντί για μετρήσεις των καθορισμένων παραμέτρων της ποιότητας υπηρεσίας. Τέτοιες ενδείξεις είναι το ποσοστό των πακέτων που αντιμετώπισαν συμφόρηση σε τουλάχιστον ένα από τους ενδιάμεσους κόμβους, ή αλλιώς το ποσοστό των μαρκαρισμένων πακέτων θεωρώντας πως οι κόμβοι κορμού έχουν τη δυνατότητα ανίχνευσης της συμφόρησης τοπικά και σημείωσης (marking) των πακέτων που διέρχονται από τον κόμβο όταν βρίσκεται σε κατάσταση συμφόρησης. Το πιο τυπικό τέτοιο σχήμα είναι η *Ρητή Ειδοποίηση Συμφόρησης* (Explicit Congestion Notification – ECN) [F94].

Το δίκτυο πρέπει να διαθέτει τους μηχανισμούς να προστατέψει την κίνηση των ενεργών ροών από την κίνηση που προκύπτει από τα δοκιμαστικά πακέτα κατά την περίοδο δοκιμής. Το φαινόμενο της στέρισης πόρων από τις ενεργές ροές ονομάζεται *κλοπή εύρους ζώνης* (stealing bandwidth). Διακρίνουμε σε *δοκιμή εντός ζώνης* (in-band probing) και *δοκιμή εκτός ζώνης* (out-of-band probing). Στο in-band probing τα πακέτα δεδομένων των ενεργών ροών και τα δοκιμαστικά πακέτων των νέων υπηρεσιών ανήκουν στην ίδια κλάση και χρησιμοποιούν τους πόρους της κλάσης από κοινού. Στο out-of-band probing τα δοκιμαστικά πακέτα χρησιμοποιούν ξεχωριστή κλάση χαμηλότερης προτεραιότητας, για την οποία δεν έχουν δεσμευτεί πόροι αλλά της αποδίδονται μόνο όταν περισσεύουν από την αντίστοιχη κλάση των πακέτων δεδομένων των ενεργών ροών. Με το out-of-band probing οι ενεργές ροές δεν επηρεάζονται από τις νέες αιτήσεις και οι νέες αιτήσεις έχουν μια αντιπροσωπευτική εικόνα των πόρων που είναι διαθέσιμοι για νέες υπηρεσίες επιπλέον των ήδη ενεργών.

Το μεγαλύτερο μειονέκτημα των probe-based μεθόδων είναι οι μεγάλοι χρόνοι απόκρισης του συστήματος εξαιτίας της περιόδου δοκιμής. Όσο μεγαλύτερη η διάρκεια της περιόδου δοκιμής τόσο πιο αξιόπιστες οι μετρήσεις της ποιότητας υπηρεσίας και συνεπώς αναπόφευκτη η αντίστοιχη επιβάρυνση στον χρόνο απόκρισης. Για κάποιες εφαρμογές μεγάλοι χρόνοι απόκρισης είναι απαγορευτικοί. Τέτοια εφαρμογή είναι η VoIP όπου οι χρήστες είναι εξοικειωμένοι με τους χρόνους απόκρισης των αντίστοιχων δικτύων τηλεφωνίας.

Ένα άλλο πρόβλημα αυτών των μεθόδων είναι η άρνηση παραχώρησης της υπηρεσίας (denial of service) λόγω του φαινομένου *κατατρόπωσης* (thrashing). Όταν η ζήτηση για πόρους ξεπερνά τους διαθέσιμους, ή αλλιώς όταν πολλές νέες ροές αποστέλλουν δοκιμαστικά πακέτα ταυτόχρονα και οι διαθέσιμοι πόροι επαρκούν μόνο για μερικές από αυτές, τότε το δίκτυο επιδεικνύει πλασματική συμφόρηση από τα δοκιμαστικά πακέτα και όλες οι ροές απορρίπτονται χωρίς να αξιοποιούνται οι διαθέσιμοι πόροι για πακέτα δεδομένων.

Όπως και οι measurement-based μέθοδοι έτσι και εδώ, η εκτίμηση της διαθεσιμότητας των πόρων βάσει μετρήσεων μπορεί να χρησιμοποιηθεί ως ένδειξη της πραγματικής κατάστασης και δεν μπορεί να θεωρηθεί αξιόπιστο κριτήριο για την παροχή εγγυήσεων. Συνεπώς τα probe-based σχήματα υποστηρίζουν μόνο προβλέψιμες υπηρεσίες.

A.2.2.2 Τοπολογική και λειτουργική κατανομή

Διακρίνουμε μελέτες συγκεντρωτικού και κατανεμημένου ελέγχου εισόδου κίνησης. Οι μελέτες συγκεντρωτικού ελέγχου εισόδου κίνησης υιοθετούν την αρχιτεκτονική *Διαμεσολαβητή Εύρους Ζώνης* (Bandwidth Broker – BB), η διαχείριση των πόρων του δικτύου και ο έλεγχος εισόδου κίνησης γίνονται από ένα κεντρικό σημείο ανά επικράτεια, το οποίο επιφορτίζεται με όλη την σχετική πολυπλοκότητα, καθιστώντας αυτή την προσέγγιση δύσκολα εφαρμόσιμη σε μεγάλα δίκτυα. Αντίθετα στις μελέτες

κατανεμημένου ελέγχου εισόδου κίνησης η πολυπλοκότητα καταμερίζεται τόσο τοπολογικά όσο και λειτουργικά. Τοπολογικά η διαχείριση των αιτήσεων υπηρεσίας μπορεί να γίνεται από κατανεμημένες αυτόνομες οντότητες ελέγχου εισόδου κίνησης ανά ακραίο κόμβο εισόδου ή ανά τερματικό κόμβο χρήστη. Λειτουργικά μπορεί για παράδειγμα ο έλεγχος διαθεσιμότητας των πόρων να πραγματοποιείται από τους κόμβους κορμού και η συγκέντρωση και επεξεργασία των αποτελεσμάτων να γίνεται στους ακραίους κόμβους.

A.2.2.2.1 Συγκεντρωτικές Μελέτες

Στα πλαίσια της αρχιτεκτονικής DiffServ αναπτύχθηκε ένα μοντέλο που συγκεντρώνει τη διαχείριση της πληροφορίας και τις διαδικασίες που απαιτούνται για την παροχή ποιότητας υπηρεσίας σε ένα κεντρικό διαμεσολαβητή ανά επικράτεια δικτύου που ονομάζεται Bandwidth Broker (BB) [DS-TB]. Σύμφωνα με αυτό το μοντέλο, ο έλεγχος εισόδου κίνησης, η κατανομή των πόρων και άλλες λειτουργίες που υλοποιούν πολιτικές του παροχέα εκτελούνται από το κεντρικό BB. Πέραν της διαχείρισης της επικρατείας του το BB επιφορτίζεται ακόμα με την επικοινωνία με ομότιμους BBs των παρακείμενων επικρατειών με σκοπό τη διαπραγμάτευση και σύναψη SLAs μεταξύ των επικρατειών (inter-domain) για την εξυπηρέτηση της κίνησης που ξεπερνά τα όρια της συγκεκριμένης επικρατείας.

Η πολυπλοκότητα των κόμβων κορμού μειώνεται καθώς αυτοί απαλλάσσονται από τις πολύπλοκες λειτουργίες, τη διατήρηση και επεξεργασία πληροφορίας που απαιτούνται για την παροχή ποιότητας υπηρεσίας. Επιπλέον, καθώς η απαραίτητη πληροφορία παραμένει συγκεντρωμένη και δεν αναπαράγεται σε διαφορετικά σημεία για την εκτέλεση διαφορετικών ή κατανεμημένων λειτουργιών δεν υπάρχει ανάγκη και για περίπλοκους μηχανισμούς συγχρονισμού μεταξύ διαδικασιών. Τέλος, η παροχή ποιότητας υπηρεσίας αποσυνδέεται σε μεγάλο βαθμό από το επίπεδο δεδομένων (data plane) και συγκεντρώνεται σε ένα μόνο σημείο καθιστώντας έτσι απλή την αντικατάσταση ή αναβάθμιση των σχετικών μηχανισμών.

Ενώ με αυτή την προσέγγιση στην εξυπηρέτηση μιας αίτησης υπηρεσίας δεν εμπλέκονται όλοι οι κόμβοι στο μονοπάτι από άκρο σε άκρο όπως συμβαίνει στο RSVP, αλλά μόνο μία κεντρική διαδικασία, ωστόσο ακριβώς η συγκέντρωση όλης της πολυπλοκότητας σε αυτή τη διαδικασία δημιουργεί προβλήματα σε μεγάλη κλίμακα. Αφενός η διαχείριση του συνόλου των πόρων της επικρατείας αντί του διαμερισμού τους σε διαχειρίσιμα υποσύνολα καθιστά τη διαδικασία ιδιαίτερα πολύπλοκη και χρονοβόρα και αφετέρου η σειριακή εξυπηρέτηση όλων των αιτήσεων υπηρεσιών από τους πελάτες αυτής της επικρατείας αντί του παραλληλισμού την καθιστά σημείο συνωστισμού (bottleneck). Σε κάποιες περιπτώσεις είναι δυνατόν η καθυστέρηση στην εξυπηρέτηση των αιτήσεων να είναι μεγαλύτερη από τη διάρκεια των ροών, έχοντας ως αποτέλεσμα την αχρησία των πόρων του δικτύου και επομένως την μειωμένη αξιοποίησή τους. Επιπλέον, σχετικά με την ευρωστία του συστήματος, το συγκεντρωτικό μοντέλο συνεπάγεται και κατάρρευση του συστήματος σε περίπτωση βλάβης του BB ή ακόμα και βλάβης ζεύξης στο μονοπάτι προς τον BB.

Διάφορες μελέτες ακολούθησαν με σκοπό την αποσαφήνιση των σχετικών θεμάτων που έμειναν ανοικτά από το αρχικό μοντέλο του [DS-TB], όπως οι μηχανισμοί κατανομής πόρων και ελέγχου εισόδου κίνησης από τους [ZDGH00] και την επαύξησή του με σκοπό την αντιμετώπιση των προβλημάτων εφαρμογής σε μεγάλη κλίμακα και ευρωστίας [ZDH01, CSKJ00].

A.2.2.2.2 Κατανεμημένες Μελέτες

Η λειτουργία ελέγχου εισόδου κίνησης κατανέμεται τοπολογικά σε πολλαπλά αυτόνομες οντότητες (instances) καθένα από τα οποία επιτελεί την ίδια λειτουργία για ένα υποσύνολο των αιτήσεων υπηρεσίας. Το φάσμα των σχετικών μελετών περιλαμβάνει κατανομή ανά κόμβο, όπως και στα παραδοσιακά σχήματα

σηματοδοσίας ανά κόμβο της τεχνολογίας ATM και IntServ, κατανομή ανά ακραίο κόμβο και κατανομή ανά τερματικό κόμβο χρήστη.

Σύμφωνα με τα παραδοσιακά single-node σχήματα, η αίτηση υπηρεσίας συντάσσεται στην πηγή και προωθείται με τη σειρά σε κάθε κόμβο του μονοπατιού προς τον προορισμό. Ο κάθε κόμβος επεξεργάζεται την αίτηση υπηρεσίας και αποφαινεται εάν είναι διαθέσιμοι τοπικά οι αντίστοιχοι απαραίτητοι πόροι. Η διαφορά των μελετών κατανεμημένου ελέγχου εισόδου κίνησης για δίκτυα DiffServ με τα παραδοσιακά σχήματα ATM και IntServ έγκειται στο γεγονός πως στα τελευταία οι κόμβοι κορμού λειτουργούν διατηρώντας και χρησιμοποιώντας πληροφορία ανά ροή ενώ στα πρώτα ανά κλάση.

Στις μελέτες κατανομής στους ακραίους κόμβους, οι αιτήσεις υπηρεσίας εξυπηρετούνται συνήθως στον ακραίο κόμβο εισόδου (ingress edge node). Η αντίστοιχη οντότητα ελέγχου εισόδου κίνησης μπορεί να βασίζεται αποκλειστικά σε πληροφορία που είναι διαθέσιμη τοπικά ή να ανταλλάσσει πληροφορία με τις ομότιμες οντότητες των άλλων ακραίων κόμβων, ή ακόμα και να επικοινωνεί με τους κόμβους κορμού ώστε να ανακτά πληροφορία για την κατάσταση συμφόρησης του δικτύου. Η κατάσταση συμφόρησης μπορεί να ανακτάται είτε άμεσα με probe-based σχήματα είτε έμμεσα με παθητικές μετρήσεις (passive measurements) τοπικά στους κόμβους κορμού και συλλογή των αποτελεσμάτων στους ακραίους κόμβους. Σε αυτή την περίπτωση κατανεμημένη είναι και η λειτουργία μέτρησης και εκτίμησης της κατάστασης συμφόρησης του δικτύου.

Η κατανομή στους τερματικούς κόμβους (endpoint admission control) είναι εφικτή χάρη στα probe-based σχήματα.

A.2.2.3 Μέθοδος κατανομής των πόρων

Όταν συμβατικές single-node parameter-based και measurement-based μέθοδοι χρησιμοποιούνται στους ακραίους κόμβους τότε απαιτείται να είναι γνωστή η χωρητικότητα από άκρο σε άκρο (end-to-end). Η χωρητικότητα από άκρο σε άκρο όμως εξαρτάται από τη δρομολόγηση και την ένταση της κίνησης και από άλλους ακραίους κόμβους.

Με τη μέθοδο partitioning (βλέπε ενότητα A.2.1.2) σε κάθε ακραίο κόμβο αντιστοιχίζεται προκαταβολικά ένα μερίδιο των πόρων το οποίο καθορίζει και τη γνωστή χωρητικότητα. Ο a priori διαμερισμός των πόρων γίνεται στο παρασκήνιο (offline) κατά τη φάση αρχικοποίησης του συστήματος έτσι ώστε τα μερίδια που προκύπτουν να αντανακλούν κατά το δυνατό την προσδοκώμενη ζήτηση για κάθε ακραίο κόμβο. Τα αποτελέσματα χρησιμοποιούνται από τις κατανεμημένες οντότητες ελέγχου εισόδου κατά τη λειτουργία του συστήματος (runtime). Όταν διαπιστώνεται πως η απόκλιση της πραγματικής κίνησης από την προσδοκώμενη είναι σημαντική και κατ' επέκταση πως η κατανομή των πόρων σε μερίδια δεν ανταποκρίνεται στις ανάγκες του συστήματος τότε προκαλείται νέος διαμερισμός των πόρων ώστε να γίνουν οι κατάλληλες προσαρμογές.

Όταν χρησιμοποιείται η μέθοδος sharing η χωρητικότητα δεν είναι προκαθορισμένη. Για κάθε αίτηση υπηρεσίας η οντότητα ελέγχου εισόδου ανταγωνίζεται με τις ομότιμες οντότητες για την κατοχύρωση των αντίστοιχων απαραίτητων πόρων μέσα από το σύνολο των κοινόχρηστων πόρων του συστήματος. Με αυτό τον τρόπο οι διαθέσιμοι πόροι αξιοποιούνται σε κάθε περίπτωση αντίθετα με τη μέθοδο partitioning όπου οι πόροι κατακερματίζονται σε απομονωμένα μερίδια. Ωστόσο, η πολυπλοκότητα και κατά συνέπεια ο χρόνος απόκρισης του συστήματος αυξάνονται σημαντικά εξαιτίας του αναγκαίου συγχρονισμού μεταξύ των κατανεμημένων ομότιμων οντοτήτων για την κατανομή των κοινόχρηστων πόρων.

Υβριδικές μελέτες χρησιμοποιούν τη μέθοδο partitioning κατά την αρχικοποίηση του συστήματος και τη μέθοδο sharing κατά τη λειτουργία του. Με αυτό τον τρόπο, όσο η πραγματική κίνηση είναι στα πλαίσια της προσδοκώμενης κάθε οντότητα λειτουργεί αυτόνομα χωρίς την ανάγκη συγχρονισμού με τις ομότιμές της. Όταν η πραγματική κίνηση υπερβαίνει την προσδοκώμενη τότε οι επιπλέον απαραίτητοι πόροι αναζητούνται στους διαθέσιμους πόρους των μεριδίων των ομότιμων οντοτήτων. Όταν η απόκλιση είναι σημαντική και το σύστημα βρίσκεται περισσότερο σε κατάσταση συγχρονισμού τότε προκαλείται ανακατανομή των πόρων όπως και στις αμιγώς partitioning μεθόδους.

A.2.3 Σχετικές Μελέτες

Στις παρακάτω ενότητες εξετάζονται με λεπτομέρεια μερικές από τις πιο χαρακτηριστικές μελέτες για κάθε μία από τις παραπάνω περιπτώσεις. Ο Πίνακας A-2 παρουσιάζει μια ταξινόμηση των σχετικών μελετών ως προς τα κριτήρια που παρουσιάστηκαν στην ενότητα A.2.2.

Κατανομή	Μέθοδος Ελέγχου Εισόδου	Μελέτη	Μέθοδος Ελέγχου Εισόδου			Διαμερισμός Πόρων	Προσαρμοστικότητα			
			Parameter	Measurement	Probe			Ακράτος Κόμβος	Κόμβος Κορμού	Τερματικός Κόμβος
Συγκεντρωτικές		[ZDGH00]	✓							
		[ZDH01]	✓							
		[CSKJ00]	✓							
		[MHOC01]	✓	✓						
Κατανεμημένες		[SZ99]	✓			▽		■	■	
		[XLBCZ00]	✓			▽		■	■	
		[CB01]	✓			▽		■	■	
		[BN03]	✓			▽		■	■	
		[MPCC00]		✓				■	■	
		[BBFP01]		✓				■	■	
		[PSZG02]		✓				■	■	
		[CKK01]		✓				■	■	
		[MK01]			✓			■	■	■
		[SVK01]			✓			■	■	■
		[BK01]			✓		□	■	■	■

Πίνακας A-2: μελέτες κατανεμημένου ελέγχου εισόδου κίνησης σε δίκτυα DiffServ

A.2.3.1 Συγκεντρωτικές Μελέτες

A.2.3.1.1 Μελέτη [ZDGH00]

A.2.3.1.1.1 Περιγραφή

Η μελέτη των [ZDGH00] περιγράφει τον αλγόριθμο ελέγχου εισόδου κίνησης για την εξυπηρέτηση *guaranteed services* με βάση το Σύστημα Αναφοράς Εικονικού Χρόνου (Virtual Time Reference System – VTRS) [ZDH00].

Το VTRS είναι ένα σύστημα αναπαράστασης των ιδιοτήτων των διαφορετικών μηχανισμών χρονοπρογραμματισμού κάθε κόμβου με ένα κοινό τρόπο έτσι ώστε καθίσταται δυνατός ο λογισμός επί των συνολικών ιδιοτήτων μιας ακολουθίας κόμβων σε ένα μονοπάτι από άκρο σε άκρο. Συγκεκριμένα, το PHB που προσφέρει ένας κόμβος αναπαρίσταται χρησιμοποιώντας την *παράμετρο λάθους* (error term), με βάση την οποία είναι δυνατό να υπολογιστεί η μέγιστη καθυστέρηση για ροές που διασχίζουν οποιαδήποτε ακολουθία κόμβων σε μια επικράτεια δικτύου.

Ο BB είναι υπεύθυνος για τον έλεγχο εισόδου κίνησης, τη δρομολόγηση και την εφαρμογή των πολιτικών του παροχέα. Η λειτουργία δρομολόγησης συγκεντρώνει τη φυσική τοπολογία από τους κόμβους του δικτύου και είναι υπεύθυνη για την επιλογή και τη διευθέτηση των μονοπατιών. Η λειτουργία ελέγχου εισόδου κίνησης είναι υπεύθυνη και για τη δέσμευση των πόρων και διαχειρίζεται πληροφορία για τις ροές, την κατάσταση ποιότητας υπηρεσίας των κόμβων και των μονοπατιών όπως αυτά προκύπτουν από τη λειτουργία της δρομολόγησης.

Κατά την άφιξη μιας αίτησης υπηρεσίας ο ακραίος κόμβος εισόδου τη μεταβιβάζει στον κεντρικό BB. Εκεί, ο έλεγχος εισόδου κίνησης πραγματοποιεί την *εξέταση δυνατότητας αποδοχής* (admissibility test) όπου εξετάζεται από τα χαρακτηριστικά κίνησης και απαιτήσεις ποιότητας σε καθυστέρηση της υπό εξέταση ροής συνάγονται οι πόροι που χρειάζεται να δεσμευτούν για την εξυπηρέτησή της και εν συνεχεία γίνεται έλεγχος στην κατάσταση ποιότητας υπηρεσίας του μονοπατιού που θα διατρέξει ώστε να διαπιστωθεί αν οι πόροι αυτοί είναι διαθέσιμοι.

Συγκεκριμένα, ο αλγόριθμος ελέγχου εισόδου κίνησης με βάση το VTRS διακρίνει τους κόμβους σε αυτούς που χρησιμοποιούν μηχανισμό χρονοπρογραμματισμού *βασισμένο σε ρυθμό* (rate-based) ή *βασισμένο σε καθυστέρηση* (delay-based). Για ένα μονοπάτι που αποτελείται αποκλειστικά από rate-based κόμβους η διαθεσιμότητα των πόρων προκύπτει απευθείας από το *εναπομένον εύρος ζώνης* (residual bandwidth) που είναι γνωστό από την πληροφορία κατάστασης ποιότητας υπηρεσίας του μονοπατιού. Στην περίπτωση που το μονοπάτι περιλαμβάνει και delay-based κόμβους η διαθεσιμότητα των απαραίτητων πόρων υπολογίζεται εκ νέου από την πληροφορία κατάστασης των κόμβων και τις ήδη ενεργές ροές που διατρέχουν καθένα από αυτούς, συμπεριλαμβάνοντας στο σύνολο των ροών και τη νέα ροή. Και στις δύο περιπτώσεις ο αλγόριθμος ελέγχου εισόδου κίνησης διασφαλίζει πως η μέγιστη καθυστέρηση δεν υπερβαίνει τα όρια που απαιτούνται από την κάθε ροή. Στην περίπτωση που χρησιμοποιούνται κλάσεις ποιότητας υπηρεσίας, τα όρια της μέγιστης καθυστέρησης είναι κοινά ανά κλάση και η πολυπλοκότητα του αλγορίθμου μειώνεται σημαντικά.

Εάν υπάρχουν οι διαθέσιμοι πόροι τότε ενημερώνεται η πληροφορία κατάστασης του μονοπατιού που έχει επιλεγεί, των ζεύξεων και των κόμβων που διατρέχει, ώστε να αντανakλά τη δέσμευση πόρων για τη ροή που έγινε αποδεκτή. Το αποτέλεσμα του admissibility test διαβιβάζεται πίσω στον ακραίο κόμβο εισόδου, που σε περίπτωση αποδοχής εγκαθίσταται και οι αντίστοιχες traffic conditioning ρυθμίσεις και εν συνεχεία πίσω στον χρήστη.

A.2.3.1.1.2 Αξιολόγηση

Αυτή η μελέτη αποσαφηνίζει τους μηχανισμούς ελέγχου εισόδου κίνησης με την αρχιτεκτονική BB και διατηρεί τα βασικά πλεονεκτήματά της. Οι κόμβοι κορμού απαλλάσσονται εντελώς από την πολυπλοκότητα που επισύρει η παροχή ποιότητας υπηρεσίας, η εφαρμογή νέων μηχανισμών γίνεται με απλή αντικατάσταση της λογικής της μονάδας του BB χωρίς άλλες επιπλοκές στο δίκτυο. Ο έλεγχος εισόδου κίνησης γίνεται κεντρικά και με βέλτιστο τρόπο για όλο το δίκτυο, αποτελεσματικά ανά μονοπάτι και όχι ανά κόμβο (hop-by-hop) όπως στην αρχιτεκτονική IntServ, και επιπλέον η αποτελεσματικότητα του αλγορίθμου ενισχύεται ακόμα περισσότερο από τη χρήση κλάσεων υπηρεσίας.

Ωστόσο, τα θέματα εφαρμογής σε μεγάλη κλίμακα και ευρωστίας παραμένουν άλυτα. Ο BB παραμένει η μία οντότητα που επιτελεί τις πολύπλοκες λειτουργίες για ολόκληρο το δίκτυο, τη δρομολόγηση, τον έλεγχο εισόδου κίνησης, την δέσμευση πόρων, την επικοινωνία με τους χρήστες και την επικοινωνία με ομότιμους BBs παρακαίμενων επικρατειών και διαχειρίζεται πληροφορία για το σύνολο των στοιχείων, των ροών και των μονοπατιών. Η πολυπλοκότητα αυτή είναι πολύ μεγάλη για μία οντότητα και έτσι για μεγάλα δίκτυα η χρήση κατανεμημένων ή ιεραρχικών μοντέλων θεωρείται αναγκαία. Επιπλέον, η συγκεκριμένη μελέτη δεν καλύπτει την περίπτωση στατιστικών εγγυήσεων ποιότητας υπηρεσίας και δεν ασχολείται με τους μηχανισμούς επικοινωνίας μεταξύ ομότιμων BBs.

A.2.3.1.2 Μελέτη [ZDH01]

A.2.3.1.2.1 Περιγραφή

Η μελέτη των [ZDH01] βασίζεται στην [ZDGH00] και προτείνει μηχανισμούς κατανομής της πολυπλοκότητας έτσι ώστε να αντιμετωπιστούν τα θέματα εφαρμογής σε μεγάλη κλίμακα.

Ως κριτήριο για την αποτελεσματικότητα του BB χρησιμοποιείται η δυνατότητά του να αντεπεξέλθει όταν τροφοδοτείται με αιτήσεις υπηρεσίας με υψηλό ρυθμό. Οι δύο παράγοντες που επηρεάζουν την αποτελεσματικότητα του BB κατά την εξυπηρέτηση μιας αίτησης υπηρεσίας είναι αφενός η καθυστέρηση πρόσβασης στη μνήμη που οφείλεται στην ανάγνωση και καταγραφή της απαραίτητης πληροφορίας και αφετέρου η καθυστέρηση στην επικοινωνία μεταξύ ακραίων κόμβων και BB από το κανάλι επικοινωνίας και από την αναμονή στην ουρά σειριακής εξυπηρέτησης του BB. Ο πρώτος παράγοντας αντιμετωπίζεται με την ανάπτυξη του μηχανισμού δυναμικής κατανομής εύρους ζώνης *ανά μονοπάτι βασιζόμενο σε τεμάχια* (Path-Oriented, Quota-Based – PoQ) και ο δεύτερος παράγοντας με την προδιαγραφή μιας ιεραρχικά κατανεμημένης αρχιτεκτονικής *πολλαπλών BBs* (Multiple BB – MBB).

A.2.3.1.2.1.1 Μηχανισμός PoQ

Ο μηχανισμός ελέγχου εισόδου κίνησης που περιγράφεται από την μελέτη [ZDGH00] κατανέμει τη συνολική χωρητικότητα κάθε ζεύξης σε επίπεδο αιτήσεων υπηρεσίας με τη σειρά, για κάθε υπηρεσία που ενεργοποιείται δεσμεύονται οι απαραίτητοι πόροι από τις ζεύξεις του αντίστοιχου μονοπατιού και αποδεσμεύονται όταν λήγει ώστε να διατεθούν στις επόμενες αιτήσεις υπηρεσίας που θα τους χρειαστούν. Συνεπώς κατά την ενεργοποίηση και λήξη μιας υπηρεσίας ο έλεγχος εισόδου κίνησης εξετάζει και ενημερώνει την πληροφορία ποιότητας υπηρεσίας του μονοπατιού και της κάθε ζεύξης του μονοπατιού. Το βασικό αυτό σχήμα ονομάζεται *ενημέρωσης ζεύξης* (link-update).

Ο μηχανισμός PoQ βασίζεται στην έννοια των quotas. Ένα quota είναι ένα κομμάτι χωρητικότητας σε εύρος ζώνης, επιλεγμένο έτσι ώστε να είναι αρκετά μεγάλο πολλαπλάσιο του μέσου εύρους ζώνης που απαιτείται

από μια υπηρεσία. Η κατανομή της χωρητικότητας των ζεύξεων γίνεται ανά μονοπάτι σε μονάδες quotas, ενώ η πραγματική δέσμευση πόρων γίνεται για κάθε ροή από τα τεμάχια που έχουν κατανεμηθεί στο μονοπάτι που διατρέχει.

Όταν εξυπηρετείται μία αίτηση υπηρεσίας εξετάζονται οι διαθέσιμοι πόροι του μονοπατιού, όπως έχουν προκύψει από τη διαφορά των quotas που έχουν παραχωρηθεί στο μονοπάτι μείον τους πόρους που έχουν δεσμευτεί για τις ήδη ενεργές ροές του συγκεκριμένου μονοπατιού. Στην περίπτωση που οι διαθέσιμοι πόροι του μονοπατιού υπερκαλύπτουν τους πόρους που είναι απαραίτητοι για την νέα αίτηση τότε η αίτηση γίνεται αποδεκτή και ενημερώνεται η καταχώρηση των διαθέσιμων πόρων του μονοπατιού χωρίς να εμπλέκονται οι αντίστοιχες καταχωρήσεις των ζεύξεων. Πρόσβαση στη μνήμη για ανάκτηση και ενημέρωση πραγματοποιείται μία φορά για τις ιδιότητες του μονοπατιού.

Στην περίπτωση που οι διαθέσιμοι πόροι δεν επαρκούν τότε ένα ακόμα quota κατανέμεται στο μονοπάτι και αφαιρείται από τους διαθέσιμους πόρους της κάθε ζεύξης. Αντίστοιχα όταν, λόγω της λήξης κάποιων υπηρεσιών, ελευθερωθεί ένα quota που έχει παραχωρηθεί σε κάποιο μονοπάτι, τότε αυτό επιστρέφεται και προστίθεται ξανά στους διαθέσιμους πόρους των αντίστοιχων ζεύξεων. Αντίστοιχα, πρόσβαση στη μνήμη πραγματοποιείται για τις ιδιότητες του μονοπατιού και για τις ιδιότητες των ζεύξεων που το αποτελούν.

Όταν τα διαθέσιμα quotas μίας ζεύξης τείνουν να εξαντληθούν, ή αλλιώς όταν είναι λιγότερα από κάποιο κατώφλι (threshold), τότε η ζεύξη περνά σε κατάσταση *κρίσιμου φόρτου* (critically loaded). Όταν μία ζεύξη είναι critically loaded, τότε και κάθε μονοπάτι που τη χρησιμοποιεί περνά σε κατάσταση critically loaded. Τα μονοπάτια σε αυτή την κατάσταση λειτουργούν με το σχήμα link-update, όπου δηλαδή παύει η περαιτέρω κατανομή των πόρων σε μονοπάτια ανά quota και για κάθε νέα αίτηση υπηρεσίας δεσμεύονται μόνο οι απαραίτητοι πόροι ανά ζεύξη. Η ζεύξη εξέρχεται από την κατάσταση critically loaded όταν οι διαθέσιμοι πόροι της υπερβαίνουν το κατώφλι. Προφανώς, σε αυτή την κατάσταση πρόσβαση στην μνήμη πραγματοποιείται για το μονοπάτι και τις ζεύξεις και η σχετική καθυστέρηση δεν διαφέρει από το σχήμα link-update.

A.2.3.1.2.1.2 Αρχιτεκτονική MBB

Η αρχιτεκτονική MBB αποτελείται από ένα κεντρικό *BB* (central BB – cBB) και πολλούς ακραίους *BBs* (edge BBs – eBBs). Ο cBB διατηρεί την πληροφορία των ζεύξεων και διαχειρίζεται την κατανομή των quotas μεταξύ των eBBs. Ο κάθε eBB επιτελεί τον έλεγχο εισόδου κίνησης για ένα υποσύνολο των μονοπατιών, τέτοιο ώστε η τομή του με τα υποσύνολα των άλλων eBBs να είναι το κενό σύνολο. Το πλήθος των eBBs είναι αυθαίρετα μεγάλος αριθμός: ένας eBB μπορεί να αντιστοιχεί σε ένα ή και περισσότερους ακραίους κόμβους, ανάλογα με το μέγεθος του δικτύου.

Όταν μία αίτηση υπηρεσίας φτάνει στον ακραίο κόμβο τότε προωθείται στον αντίστοιχο eBB, ο οποίος θα εξετάσει την κατάσταση του αντίστοιχου μονοπατιού. Εάν δεν υπάρχουν διαθέσιμοι πόροι τότε ο eBB ζητά την παραχώρηση ενός quota για το συγκεκριμένο μονοπάτι από τον cBB. Εάν το μονοπάτι είναι σε critically loaded κατάσταση τότε για κάθε αίτηση υπηρεσίας ο eBB επικοινωνεί με τον cBB. Ο eBB λέγεται πως λειτουργεί κατά το μοντέλο *μονοπατιού χωρίς απώλειες* (non-lossy-path).

Σε αντιπαράβολη προτείνεται ένα διαφορετικό μοντέλο, το μοντέλο *μονοπατιού με απώλειες* (lossy-path). Κατά το lossy-path μοντέλο, όταν δεν υπάρχουν διαθέσιμα quotas για μια συγκεκριμένη ζεύξη και η αίτηση παραχώρησης ενός επιπλέον quota για ένα μονοπάτι αποτυγχάνει τότε, αντί η ζεύξη και το μονοπάτι να χαρακτηριστούν critically loaded και το σύστημα να περάσει σε link-update μηχανισμό, η αίτηση υπηρεσίας απλά απορρίπτεται. Με αυτό τον τρόπο, ο cBB απαλλάσσεται από κάθε λειτουργία ελέγχου εισόδου κίνησης και επιτελεί μόνο τη διαχείριση κατανομής quotas μετά των eBBs. Το κόστος για αυτή την απλοποίηση της

διαδικασίας προκύπτει από τη σπατάλη των quotas που έχουν παραχωρηθεί σε άλλα μονοπάτια και δεν χρησιμοποιούνται ολόκληρα.

A.2.3.1.2.1.3 Περαιτέρω βελτιώσεις

Το μέγεθος του quota επηρεάζει σημαντικά την απόδοση της προτεινόμενης λύσης. Αφενός, quota μικρού μεγέθους συνεπάγεται λιγότερες απώλειες από το lossy-path μοντέλο και μικρότερη παραμονή στην κατάσταση critically loaded όταν το δίκτυο είναι φορτωμένο στο non-lossy-path μοντέλο, συνεπώς πιο σύντομη link-update λειτουργία και άρα καλύτερη απόδοση. Αφετέρου, quota μεγαλύτερου μεγέθους συνεπάγεται λιγότερο συχνές αιτήσεις παραχώρησης quota και άρα καλύτερη απόδοση, ιδιαίτερα σε δίκτυα με χαμηλό φόρτο. Επιπλέον, όταν σε ένα εκτεταμένο δίκτυο ο αριθμός των μονοπατιών που χρησιμοποιούν μία ζεύξη είναι μεγάλος, το μέγεθος του quota πρέπει να είναι αρκετά μικρό ώστε η συνολική χωρητικότητα να κατανέμεται ομαλά μεταξύ τους.

Δύο μηχανισμοί που στηρίζονται στις παραπάνω παρατηρήσεις και βελτιώνουν την απόδοση του συστήματος είναι ο μηχανισμός *PoQ με υστέρηση* (PoQ with hysteresis) και *PoQ με μεταβλητό μέγεθος quota* (PoQ with Variable Quota Size – VQS).

Ο μηχανισμός PoQ με υστέρηση στηρίζεται στην παρατήρηση πως η αίτηση για επιπλέον quota ή η επιστροφή ενός ελεύθερου quota σε κάποιες περιπτώσεις αντικατοπτρίζουν σύντομες διακυμάνσεις στη ζήτηση για πόρους από τις υπηρεσίες, που γρήγορα αντιστραφούν για να επιστρέψουν στην πρότερη πιο αντιπροσωπευτική κατάσταση, με ένα λιγότερο ή με ένα περισσότερο quota αντίστοιχα. Στον μηχανισμό PoQ με υστέρηση η επιστροφή του διαθέσιμου quota δεν γίνεται αμέσως όταν ελευθερώνεται το quota αλλά εάν και όταν οι πόροι που είναι διαθέσιμοι στο μονοπάτι χωρίς το quota υπερβαίνουν ένα προκαθορισμένο threshold. Αντίστοιχα, η αίτηση για νέο quota δεν γίνεται μόλις διαπιστώνεται η ανεπάρκεια σε διαθέσιμους πόρους από μια νέα αίτηση υπηρεσίας, αλλά όταν οι διαθέσιμοι πόροι γίνουν λιγότεροι από ένα άλλο προκαθορισμένο threshold. Ενώ μεγάλες τιμές των thresholds ενισχύουν την αποτελεσματικότητα του σχήματος, ωστόσο μπορεί να συντελέσουν σε αδικαιολόγητα παρατεταμένη παρακράτηση των quotas και αποκλεισμό μονοπατιών που τα έχουν περισσότερο ανάγκη.

Κατά το μηχανισμό PoQ με VQS το μέγεθος του quota είναι μεταβλητό και εξαρτάται από τους διαθέσιμους πόρους των ζεύξεων. Η χωρητικότητα κάθε ζεύξης χωρίζεται σε διαστήματα, το καθένα από τα οποία συσχετίζεται με ένα μέγεθος quota. Όταν οι διαθέσιμοι πόροι μιας ζεύξης, όπως προκύπτουν από τη διαφορά της συνολικής χωρητικότητας μείον τα quotas που έχουν ήδη κατανεμηθεί, είναι αρκετοί σε σχέση με τη συνολική χωρητικότητα τότε το μέγεθος του quota είναι μεγάλο και μικραίνει όσο οι διαθέσιμοι πόροι τείνουν να εξαντληθούν.

A.2.3.1.2.2 Αξιολόγηση

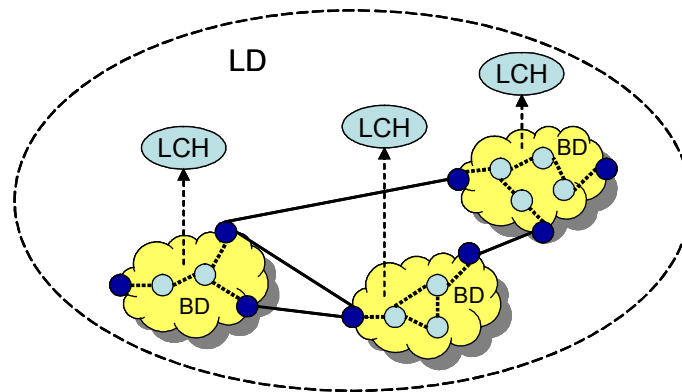
Η προτεινόμενη λύση διατηρεί τα πλεονεκτήματα και επιπλέον αντιμετωπίζει τα προβλήματα εφαρμογής σε μεγάλη κλίμακα της μελέτης [ZDGH00]. Τα θέματα επικοινωνίας μεταξύ ομότιμων BBs και στατιστικών εγγυήσεων ποιότητας υπηρεσίας ωστόσο παραμένουν ανοικτά και η ευρωστία του συστήματος εξακολουθεί να είναι ευπαθής λόγω της εξάρτησης των ακραίων BBs από τον κεντρικό BB.

A.2.3.1.3 Μελέτη [CSKJ00]

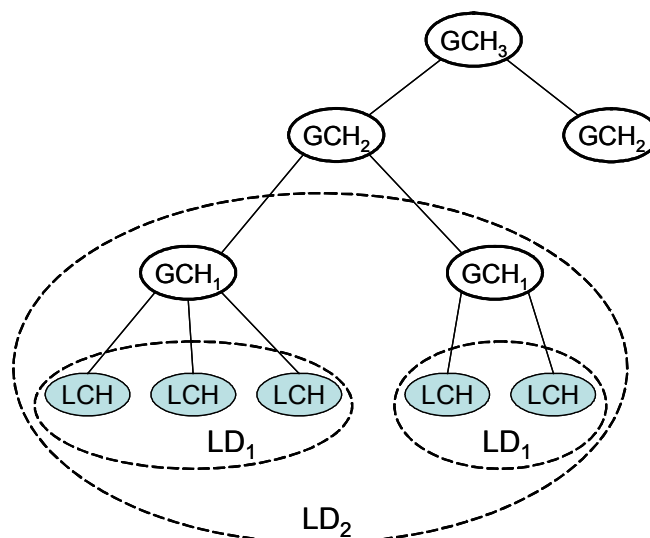
A.2.3.1.3.1 Περιγραφή

Η μελέτη [CSKJ00] προτείνει μία ιεραρχικά δομημένη αρχιτεκτονική BBs μεταξύ επικρατειών (inter-domain) για την παροχή στατιστικών εγγυήσεων ποιότητας υπηρεσίας με βάση το σχήμα *Εκκαθαριστικού Οίκου* (Clearing House – CH) που χρησιμοποιήθηκε αρχικά σε τραπεζικά συστήματα.

Διαφορετικές παρακείμενες στοιχειώδεις επικράτειες (basic domains – BDs) ενώνονται και διαμορφώνουν λογικές επικράτειες (logical domains – LDs) (βλέπε Σχήμα A-5). Σε ένα δεύτερο επίπεδο οι λογικές επικράτειες ενώνονται και διαμορφώνουν μεγαλύτερες λογικές επικράτειες κ.ο.κ. Έτσι, δημιουργείται μία ιεραρχική δομή από λογικές επικράτειες, σε κάθε μία από τις οποίες αντιστοιχεί ένας κόμβος CH που για τις στοιχειώδεις επικράτειες ονομάζεται *τοπικός κόμβος CH* (Local CH – LCH) ενώ για τις λογικές επικράτειες ανώτερων επιπέδων ονομάζεται *συνολικός κόμβος CH* (Global CH – GCH) (βλέπε Σχήμα A-6).



Σχήμα A-5: διαμόρφωση λογικής επιφάνειας



Σχήμα A-6: ιεραρχική δομή κόμβων CH

Κάθε κόμβος CH λειτουργεί ως BB για την λογική επικράτεια που του αντιστοιχεί. Οι αιτήσεις υπηρεσιών που εκτείνονται τοπολογικά πέρα από τα όρια της λογικής επικράτειας ομαδοποιούνται και προωθούνται για εξυπηρέτηση στον κόμβο CH του ανώτερου επιπέδου.

Αντί για τα μαθηματικά μοντέλα που χρησιμοποιούνται στις προηγούμενες μελέτες για την παροχή εγγυημένων υπηρεσιών, αυτή τη μελέτη παρέχει στατιστικές εγγυήσεις ποιότητας υπηρεσίας χρησιμοποιώντας μετρήσεις με βάση τις οποίες εκτιμάται η μελλοντική ζήτηση για πόρους και πραγματοποιούνται *προκαταβολικές κρατήσεις* (advance reservations) στις ζεύξεις έτσι ώστε ο έλεγχος εισόδου κίνησης είναι λιγότερο χρονοβόρος καθώς βρίσκει το σύστημα προετοιμασμένο να εξυπηρετήσει τις νέες αιτήσεις υπηρεσίας.

Κάθε ακραίος κόμβος λαμβάνει μετρήσεις για την συνολική κίνηση εντός και εκτός της επικράτειας. Στο LCH συγκεντρώνονται οι μετρήσεις για την κίνηση εντός της στοιχειώδους επικρατείας, εκτιμάται η μελλοντική ζήτηση για πόρους και γίνονται οι απαραίτητες προκαταβολικές κρατήσεις στις φυσικές ζεύξεις της επικράτειας. Οι μετρήσεις για την κίνηση από και προς τις παρακείμενες επικράτειες προωθούνται στον GCH όπου και, με βάση τις αντίστοιχες μετρήσεις των άλλων υποκείμενων LCHs, εκτιμάται η μελλοντική ζήτηση μεταξύ των αντίστοιχων στοιχειωδών επικρατειών και γίνονται οι κατάλληλες προκαταβολικές κρατήσεις στις ζεύξεις που ενώνουν τους ακραίους κόμβους των στοιχειωδών επικρατειών. Οι αιτήσεις υπηρεσιών που εκτείνονται πέρα από τα όρια της στοιχειώδους επικράτειας συγκεντρώνονται στον LCH και προωθούνται για εξυπηρέτηση στον GCH που με τη σειρά του συγκεντρώνει τις αιτήσεις υπηρεσιών που ξεπερνούν τα όρια της λογικής επικράτειας και της προωθεί στον GCH του ανώτερου επιπέδου κ.ο.κ.

Οι ακραίοι κόμβοι πραγματοποιούν μετρήσεις για το μέσο ρυθμό m και τη διακύμανση σ^2 ανά χρονικές περιόδους διάρκειας T_{mes} . Με την παραδοχή πως η κατανομή του ρυθμού της κίνησης όλων των ροών προς κάποιο προορισμό είναι Gaussian κατανομή, το εύρος ζώνης \hat{B} που θα είναι μελλοντικά αναγκαίο είναι δυνατό να εκτιμηθεί από τον τύπο $\hat{B} = m + \alpha\sigma$, όπου α είναι ο παράγοντας ποιότητας υπηρεσίας που καθορίζει το βαθμό πρόβλεψης και εγγυήσεων για την ποιότητα υπηρεσίας των διακυμάνσεων της κίνησης. Το διάστημα T_{mes} επηρεάζει σημαντικά την ακρίβεια αυτής της εκτίμησης καθώς η πραγματική κατανομή της κίνησης είναι δυνατό να αλλάξει πιο γρήγορα ή πιο αργά χαρακτηριστικά από όσο το διάστημα T_{mes} μπορεί να παρακολουθήσει. Όταν το \hat{B} αποδεικνύεται μικρότερο από την πραγματική ζήτηση η υποβάθμιση της ποιότητας υπηρεσίας εντείνεται ενώ όταν αποδεικνύεται μεγαλύτερο μικραίνει η αξιοποίηση των πόρων.

Ο έλεγχος εισόδου κίνησης εξετάζει αν οι πόροι που έχουν δεσμευτεί από τις προκαταβολικές κρατήσεις επαρκούν για την εξυπηρέτηση των νέων αιτήσεων υπηρεσιών. Όταν δεν επαρκούν οι αιτήσεις υπηρεσίας απορρίπτονται. Εναλλακτικά είναι δυνατό να πραγματοποιούνται *άμεσες κρατήσεις* (immediate reservations), δηλαδή ανεύρεση και δέσμευση διαθέσιμων πόρων επιπλέον αυτών που έχουν διατεθεί κατά τις προκαταβολικές κρατήσεις και με γνώμονα την πραγματική ζήτηση της κάθε αίτησης υπηρεσίας.

A.2.3.1.3.2 Αξιολόγηση

Ο καταμερισμός της πολυπλοκότητας της κατανομής πόρων και του ελέγχου εισόδου κίνησης μεταξύ διαφορετικών επικρατειών που επιτυγχάνεται χάρη στο ιεραρχικό και ευέλικτο μοντέλο οργάνωσης των κόμβων CH καθιστά την προτεινόμενη λύση εφαρμόσιμη σε μεγάλη κλίμακα. Κάθε επίπεδο της ιεραρχίας αναλαμβάνει τη διαχείριση υποσυνόλου της κίνησης και των φυσικών ζεύξεων, το οποίο μπορεί να είναι οσοδήποτε μικρό ή μεγάλο ανάλογα με τον αριθμό των επιπέδων. Επιπλέον, η διαχείριση συσσωματώσεων ροών επιτρέπει καλύτερη αξιοποίηση των πόρων χάρη στο συνυπολογισμό της επικάλυψης των αναγκών σε πόρους που προκύπτει από την πολυπλεξία των ροών (multiplexing gain). Χάρη στο σύστημα προκαταβολικών κρατήσεων, ο χρόνος απόκρισης σε μια αίτηση υπηρεσίας είναι κατά πολύ συντομότερος σε σχέση με άλλα μοντέλα που χρησιμοποιούν μόνο άμεσες κρατήσεις.

Αυτή η μελέτη περιορίζεται στην παροχή υπηρεσιών με στατιστικές εγγυήσεις. Η δέσμευση πόρων και η εξυπηρέτηση των αιτήσεων υπηρεσιών με βάση αποκλειστικά και μόνο τις ενδεικτικές μετρήσεις έχει σημαντικά μειονεκτήματα. Αφενός, η συμπεριφορά των χρηστών μπορεί να αλλάζει πιο γρήγορα από όσο παρακολουθείται και προβλέπεται από το σύστημα έχοντας ως αποτέλεσμα της υποβάθμιση της ποιότητας υπηρεσίας των ενεργών ροών. Η πιο εντατική παρακολούθηση και συχνότερη προσαρμογή στις διακυμάνσεις της κίνησης συνεισφέρει στην ακρίβεια των προβλέψεων και στην εξασφάλιση της ποιότητας υπηρεσίας συνεπάγεται όμως πιο συχνούς ελέγχους και ανακατανομή των κρατήσεων και κατ' επέκταση σημαντική επιβάρυνση σε συστήματα μεγάλης κλίμακας και είναι δυνατό να αποσταθεροποιήσει το σύστημα. Αφετέρου, η αποδοχή νέων αιτήσεων υπηρεσιών με βάση προηγούμενα στιγμιότυπα της κατάστασης του δικτύου μπορεί να οδηγήσει σε εξάντληση των πόρων ιδιαίτερα όταν οι ροές δεν είναι μικρής διάρκειας.

A.2.3.1.4 Μελέτη [MHOC01]

A.2.3.1.4.1 Περιγραφή

Με βάση το BB όπως περιγράφεται στο [DS-TB] ορίζεται το *σημείο διαχείρισης εύρους ζώνης* (Bandwidth Management Point – BMP). Ο βασικός ρόλος ενός BMP είναι η διαχείριση των πόρων εντός της επικράτειας και η διαπραγμάτευση των SLAs με τις παρακείμενες επικράτειες για τη χρησιμοποίηση πόρων εκτός της επικράτειας.

Τα SLAs μεταξύ επικρατειών γίνονται ανά περιοχή προορισμού χρησιμοποιώντας για παράδειγμα το IP prefix της επικράτειας προορισμού. Κάθε BMP συλλέγει τις αιτήσεις για πόρους προς μια επικράτεια προορισμού από τους πελάτες της επικράτειάς του και από τις παρακείμενες επικράτειες, τις ενοποιεί και συνάπτει με την παρακείμενη επικράτεια προς την οποία δρομολογείται η κίνηση με τον συγκεκριμένο προορισμό ένα SLA για τη συσσωματωμένη κίνηση με συνδέσεις τύπου σωλήνα (pipes) ανά DSCP. Έτσι, καθώς το μονοπάτι από διαφορετικές πηγές προς ένα συγκεκριμένο προορισμό συγχωνεύεται (merge) συσσωματώνονται και τα αντίστοιχα pipes.

Τα SLAs τροποποιούνται δυναμικά ανάλογα με τις αυξομειώσεις της ζήτησης. Ο αλγόριθμος προσαρμογής των SLAs ορίζει την *περιοχή λειτουργίας* (Operation Region – OR) έντασης της κίνησης για κάθε pipe. Όταν η ένταση της κίνησης ξεπερνά το άνω όριο της OR τότε ο αλγόριθμος ενεργοποιείται και διαπραγματεύεται την ενίσχυση του SLA με περισσότερο εύρος ζώνης, ενώ όταν αντίστοιχα η ένταση της κίνησης πέφτει κάτω από το κάτω όριο της OR το SLA αναπροσαρμόζεται παραχωρώντας ένα μέρος του εύρους ζώνης. Το άνω όριο της OR είναι μικρότερο από το δεδομένο εύρος ζώνης του SLA έτσι ώστε η διαπραγμάτευση για νέους πόρους να προλαμβάνει την εξάντληση των υπαρχόντων. Το μέγεθος της OR επηρεάζει την αποδοτικότητα του αλγορίθμου. Μεγάλες περιοχές συνεπάγονται λιγότερο συχνές προσαρμογές και κατ' επέκταση μικρότερη επιβάρυνση του συστήματος και μεγαλύτερη ευστάθεια με κόστος σε αξιοποίηση των πόρων που δεν χρησιμοποιούνται και δεν παραχωρούνται από κάποια pipes σε άλλα που είναι ενδεχομένως απαραίτητοι. Έτσι, το μέγεθος της OR διαμορφώνεται και αυτό δυναμικά ανάλογα με τη συχνότητα των προσαρμογών του SLA. Όταν η ένταση της κίνησης γρήγορα ξεπερνά την περιοχή OR τότε στην επόμενη προσαρμογή η περιοχή OR γίνεται μεγαλύτερη επιτυγχάνοντας μεγαλύτερη σταθερότητα και αντίθετα όταν η κίνηση είναι πιο σταθερή η περιοχή OR μικραίνει επιτυγχάνοντας μεγαλύτερη αξιοποίηση των πόρων χωρίς να διακυβεύεται η σταθερότητα του συστήματος.

Η ένταση της κίνησης είναι δυνατό να υπολογίζεται είτε με parameter-based είτε με measurement-based μοντέλα. Στην περίπτωση του της parameter-based μεθόδου η ένταση της κίνησης υπολογίζεται με βάση τον

αριθμό και το traffic profile των ενεργών ροών ενώ στην περίπτωση της measurement-based μεθόδου εξάγεται από τις κατάλληλες μετρήσεις. Και στις δύο περιπτώσεις η χωρητικότητα των pipes θεωρείται σταθερή όπως προκύπτει από τη διαστασιοποίηση του δικτύου με βάση τα υπάρχοντα SLAs. Ο έλεγχος εισόδου κίνησης εκτιμά τους διαθέσιμους πόρους ως τη διαφορά της σταθερής χωρητικότητας του pipe και της έντασης της κίνησης και εκτιμά αν επαρκούν για την αποδοχή της νέας αίτησης υπηρεσίας.

Τα measurement-based σχήματα επιτυγχάνουν υψηλότερη αξιοποίηση των πόρων αλλά παρέχουν στατιστικές μόνο εγγυήσεις με κάποια μεγάλη πιθανότητα και θεωρούνται περισσότερο κατάλληλα για τις υπηρεσίες με στατιστικές εγγυήσεις. Αντίθετα, τα parameter-based σχήματα παρέχουν ντετερμινιστικές εγγυήσεις ποιότητας με υψηλό κόστος σε αξιοποίηση των πόρων και εφαρμόζονται περισσότερο για απόλυτα εγγυημένες υπηρεσίες. Μια λεπτομερέστερη ανάλυση των χαρακτηριστικών των δύο σχημάτων γίνεται στις ενότητες A.2.3.2 και A.2.3.2.4.2.

A.2.3.1.4.2 Αξιολόγηση

Σε αυτή την μελέτη αντιμετωπίζονται τα θέματα της επικοινωνίας μεταξύ BBs διαφορετικών επικρατειών με τρόπο εφαρμόσιμο σε μεγάλη κλίμακα και ορίζονται δύο εναλλακτικές προσεγγίσεις για την παροχή υπηρεσιών στατιστικών και απόλυτων εγγυήσεων. Δεν προεκτείνεται ωστόσο στον καθορισμό του απαραίτητου κοινού πλαισίου που να περιβάλλει και τις δύο δυνατότητες και να προσφέρει τη δυνατότητα ρύθμισης ανάλογα με τις πολιτικές του παροχέα. Επιπλέον, η κατανομή των πόρων του δικτύου σε pipes και ο λογισμός του ελέγχου εισόδου κίνησης με partitioning λογική αναστέλλει το πλεονέκτημα των συγκεντρωτικών BB να λαμβάνουν αποφάσεις με βέλτιστο τρόπο για όλη την έκταση της επικράτειας του δικτύου.

A.2.3.2 Κατανεμημένες Parameter-Based Μελέτες

A.2.3.2.1 Μελέτη [SZ99]

A.2.3.2.1.1 Περιγραφή

Ο στόχος της μελέτης [SZ99] είναι η παροχή εγγυημένων υπηρεσιών με την ακρίβεια του μοντέλου υπηρεσιών IntServ πάνω από ένα δίκτυο που δεν καταχωρεί πληροφορία ανά ροή και άρα με τρόπο εφαρμόσιμο και σε μεγάλη κλίμακα. Αυτό επιτυγχάνεται με την τεχνική της *δυναμικής κατάστασης πακέτου* (Dynamic Packet State – DPS) σύμφωνα με την οποία η απαραίτητη πληροφορία ανά ροή δεν καταχωρείται στους κόμβους κορμού αλλά μεταφέρεται στην επικεφαλίδα των ίδιων των πακέτων δεδομένων της ροής.

Οι ακραίοι κόμβοι διατηρούν πληροφορία ανά ροή σύμφωνα με την οποία αρχικοποιούν τα κατάλληλα πεδία στην επικεφαλίδα των πακέτων κατά την είσοδό τους στο δίκτυο. Στους κόμβους κορμού η ταξινόμηση των πακέτων σε ροές ή κλάσεις δεν είναι πλέον αναγκαία. Η μεταχείριση του κάθε πακέτου καθορίζεται μόνο από την πληροφορία που μεταφέρει στην επικεφαλίδα του. Η πληροφορία αυτή αντανακλά τη δυναμική κατάσταση του πακέτου και προσαρμόζεται κατάλληλα σε κάθε κόμβο πριν προωθηθεί στον επόμενο.

Ο έλεγχος εισόδου κίνησης γίνεται με σηματοδосία ανά κόμβο (hop-by-hop signalling). Σε κάθε κόμβο περιοδικά υπολογίζεται μια εκτίμηση του *συνολικού δεσμευμένου εύρους ζώνης* (aggregate reserved rate) R_{bound} χρησιμοποιώντας την πληροφορία των πακέτων που έχουν εξυπηρετηθεί σε κάθε περίοδο. Η εκτίμηση είναι τέτοια ώστε το πραγματικό απαραίτητο εύρος ζώνης που αντιστοιχεί σε όλες τις ενεργές ροές που διέρχονται από τον συγκεκριμένο κόμβο να μην υπερβαίνει το R_{bound} . Κάθε αίτηση υπηρεσίας που φτάνει σε

ένα κόμβο κορμού προσδιορίζει και το εύρος ζώνης r_{new} που είναι απαραίτητο για την εξυπηρέτησή της. Αν το άθροισμα R_{bound} συν r_{new} είναι μικρότερο από τη χωρητικότητα της αντίστοιχης ζεύξης τότε η αίτηση υπηρεσίας γίνεται αποδεκτή από τον συγκεκριμένο κόμβο και προωθείται στον επόμενο.

A.2.3.2.1.2 Αξιολόγηση

Η μελέτη [SZ99] υποστηρίζει αποκλειστικά εγγυημένες υπηρεσίες με το κόστος που αυτό συνεπάγεται στην αξιοποίηση των πόρων. Επιπλέον, υπερεκτίμηση του R_{bound} σε σχέση με το πραγματικού απαραίτητο εύρος ζώνης συνεπάγεται ακόμα χαμηλότερη αξιοποίηση των πόρων.

Ενώ οι κόμβοι κορμού απαλλάσσονται από την πληροφορία ανά ροή ωστόσο επιβαρύνονται με επιπλέον υπολογισμούς για την εξυπηρέτηση και την ενημέρωση της επικεφαλίδας του κάθε πακέτου αλλά και για τον υπολογισμό του R_{bound} . Ακόμα, η πρακτική εφαρμογή της μεθόδου είναι αμφίβολη καθώς προϋποθέτει αφενός ελαφρώς παραλλαγμένη χρήση της επικεφαλίδας του IP πρωτοκόλλου και αφετέρου συγκεκριμένους συμβατούς μηχανισμούς χρονοπρογραμματισμού στους κόμβους κορμού που δεν διατίθενται στις γνωστές υλοποιήσεις σήμερα. Τέλος, για την ορθή λειτουργία του μηχανισμού τίθενται κάποιες προϋποθέσεις σχετικά με το μέγιστο χρόνο μεταξύ της αποστολής δύο διαδοχικών πακέτων μιας ροής, παρακολουθώντας και εισάγοντας από τον κόμβο εισόδου κενά (dummy) πακέτα κάθε φορά που η ίδια η ροή δεν αντεπεξέρχεται σε αυτές τις προϋποθέσεις.

A.2.3.2.2 Μελέτη [DSGLBZ98]

A.2.3.2.2.1 Περιγραφή

Η μελέτη [DSGLBZ98] σχεδιάστηκε στα πλαίσια της τεχνολογίας ATM με σκοπό τη δημιουργία μιας προηγμένης μεθόδου ελέγχου αποδοχής σύνδεσης (Connection Admission Control – CAC) που να προσαρμόζει δυναμικά την παρεχόμενη ποιότητα υπηρεσίας ανάλογα με τις αυξομειώσεις των διαθέσιμων πόρων.

Με τους παραδοσιακούς CAC αλγορίθμους, άπαξ και η σύνδεση γίνει αποδεκτή λαμβάνει σταθερή ποιότητα υπηρεσίας καθ' όλη τη διάρκεια ζωής της. Για πολλές ακόμα και real-time εφαρμογές που λειτουργούν ικανοποιητικά για ένα φάσμα διαφορετικών τιμών η καθορισμένη και αμετάβλητη ποιότητα υπηρεσίας είναι αδικαιολόγητα περιοριστική. Έτσι, αντί της μοναδικής τιμής, οι παράμετροι ποιότητας υπηρεσίας αποτιμώνται από ένα σύνολο επιτρεπτών τιμών που προσδιορίζεται στη συμφωνία από την ελάχιστη αποδεκτή και τη μέγιστη επιθυμητή τιμή.

Κατά την αποδοχή μιας νέας σύνδεσης καλείται η διαδικασία *συρρίκνωσης ποιότητας υπηρεσίας* (QoS Shrinkage). Στην αίτηση αποδοχής σύνδεσης μαζί με την επιθυμητή ποιότητα δίνονται και κατευθύνσεις για τη διαδικασία συρρίκνωσης με βάση τις οποίες επιλέγονται οι συνδέσεις στις οποίες θα αναζητηθούν οι απαραίτητοι πόροι για τη νέα σύνδεση. Με το υποσύνολο αυτό των ενεργών συνδέσεων και τη νέα σύνδεση καλείται η διαδικασία υπολογισμού της μέσης καθυστέρησης (delay computation) η οποία αποφαίνεται εάν με την αποδοχή της νέας σύνδεσης η καθυστέρηση κάθε σύνδεσης παραβιάζει την ελάχιστη αποδεκτή τιμή οπότε και η νέα σύνδεση απορρίπτεται. Εάν όχι τότε γίνεται εγκατάσταση της νέας σύνδεσης για την οποία δεσμεύονται οι απαραίτητοι πόροι που έχουν πριν ελευθερωθεί από το αντίστοιχο υποσύνολο των ενεργών συνδέσεων. Η παρεχόμενη ποιότητα υπηρεσίας αυτών των συνδέσεων μειώνεται αλλά παραμένει αποδεκτή.

Κατά τον τερματισμό μιας ενεργής σύνδεσης καλείται η διαδικασία *διεύρυνσης ποιότητας υπηρεσίας* (QoS Expansion). Επιλέγεται και πάλι ένα υποσύνολο των ενεργών συνδέσεων οι οποίες και απορροφούν τους πόρους που ελευθερώνονται από τον τερματισμό της σύνδεσης. Η παρεχόμενη ποιότητα υπηρεσίας αυτών

των συνδέσεων αυξάνεται χωρίς να φράσσεται από κάποιο άνω όριο, χρησιμοποιώντας κατ' αυτό τον τρόπο όλους τους διαθέσιμους πόρους.

A.2.3.2.2.2 Αξιολόγηση

Οι παραδοσιακοί CAC αλγόριθμοι αφήνοντας ανεκμετάλλευτους τους πόρους που ελευθερώνονται κατά τον τερματισμό συνδέσεων δεν αξιοποιούν τη δυναμική του συστήματος. Επιπλέον, απορρίπτοντας νέες συνδέσεις όταν η αντίστοιχη υποβάθμιση της ποιότητας είναι αποδεκτή παραγνωρίζουν την ευελιξία των εφαρμογών εις βάρος της αξιοποίησης των πόρων.

Αντίθετα, η μελέτη [DSGLBZ98] κάνοντας αποδεκτές περισσότερες συνδέσεις και εκμεταλλεζόμενη ανά πάσα στιγμή τους διαθέσιμους πόρους επιτυγχάνει αφενός σημαντικά μεγαλύτερη αξιοποίηση των πόρων και αφετέρου καλύτερη ποιότητα υπηρεσίας, για κάποιες συνδέσεις υψηλότερη και από την καθορισμένη μέγιστη επιθυμητή τιμή.

Το τίμημα για αυτή τη βελτίωση είναι η αυξημένη πολυπλοκότητα και η αντίστοιχη καθυστέρηση στους χρόνους απόκρισης του συστήματος σε μια αίτηση σύνδεσης. Ο μηχανισμός συρρίκνωσης απαιτεί υπολογισμούς για την καθυστέρηση περισσότερο πολύπλοκους από τους απλούς μηχανισμούς ελέγχου αποδοχής σύνδεσης και ο τερματισμός μιας σύνδεσης δεν γίνεται αυτόματα αλλά απαιτεί εξίσου πολύπλοκους υπολογισμούς με τη διαδικασία ενεργοποίησής της.

Στα δίκτυα DiffServ όπως και με την τεχνολογία ATM κάθε ροή που λαμβάνει ποιότητα καλύτερη από best effort ρυθμίζεται κατά την είσοδό της στο δίκτυο από το traffic conditioning (βλέπε ενότητα A.1.4.2.2.1) έτσι ώστε να συμμορφώνεται σύμφωνα με κάποιο συμφωνημένο traffic profile, να διοχετεύει δηλαδή κίνηση στο δίκτυο με κάποιο συμφωνημένο μέγιστο ρυθμό. Η διαδικασία συρρίκνωσης και διεύρυνσης της ποιότητας αντιστοιχίζεται έτσι σε αύξηση και μείωση του μέγιστου ρυθμού που επιβάλλεται σε κάθε ροή από το traffic conditioning.

Η καινοτομία της μελέτης έγκειται στην ιδέα πως οι εφαρμογές είναι δυνατό να λειτουργούν ικανοποιητικά για ένα φάσμα τιμών ποιότητας και συνεπώς η περιγραφή της προσφερόμενης ποιότητας να μην περιορίζεται στη μοναδική τιμή και το καθορισμένο ποσοστό επιτρεπτής παραβίασής της αλλά να περιλαμβάνει το σύνολο των επιτρεπτών τιμών.

A.2.3.2.3 Μελέτη [CB01]

A.2.3.2.3.1 Περιγραφή

Η μελέτη [CB01] είναι χαρακτηριστική μέθοδος ελέγχου εισόδου κίνησης κατανεμημένου στους ακραίους κόμβους με partitioning κατά την αρχικοποίηση και sharing κατά τη λειτουργία του συστήματος. Η έμφαση δίνεται στους μηχανισμούς διαμερισμού με partitioning κατά την αρχικοποίηση.

Κατά τη φάση αρχικοποίησης του συστήματος οι πόροι του δικτύου διανέμονται ανά *καταληκτικά δέντρα* (sink trees) ανάλογα με την προσφορά της κίνησης. Τα sink trees χρησιμοποιούνται ως δομές για το λογισμό επί συσσωματώσεων ξεχωριστών ροών από οπουδήποτε στο δίκτυο προς ένα συγκεκριμένο ακραίο κόμβο εξόδου. Η ρίζα ενός sink tree είναι ο ακραίος κόμβος εξόδου (egress node) ενώ τα φύλλα είναι οι ακραίοι κόμβοι εισόδου (ingress nodes). Σε κάθε κλάδο από ένα φύλλο προς τη ρίζα του δέντρου αποδίδεται η χωρητικότητα της διαδρομής από τον αντίστοιχο ακραίο κόμβο εισόδου προς τον κόμβο εξόδου. Η προσδοκώμενη κίνηση ανά κόμβο εισόδου τροφοδοτείται στον αλγόριθμο διαμερισμού πόρων, ο οποίος υπολογίζει τα sink trees στο δίκτυο έτσι ώστε η μέγιστη καθυστέρηση στη χειρότερη περίπτωση (maximum

worst-case delay) σε κάθε μονοπάτι να μην υπερβαίνει ένα άνω όριο που περιγράφει την ποιότητα της προσφερόμενης υπηρεσίας και καθορίζεται από τον παροχέα. Τα sink trees μπορεί να σχεδιαστούν με τη μέθοδο partitioning ή με τη μέθοδο sharing σε επίπεδο μονοπατιού, σε επίπεδο δέντρου ή σε επίπεδο ζεύξης (σχήματα no-sharing, path-sharing, tree-sharing και link-sharing). Τα διαφορετικά επίπεδα sharing καθορίζουν και τη δυνατότητα χρησιμοποίησης πόρων κατά τη λειτουργία του συστήματος που έχουν αρχικά ανατεθεί σε άλλα μερίδια μετά από την απαραίτητη διαδικασία κατοχύρωσης και συγχρονισμού μεταξύ των ακραίων κόμβων.

A.2.3.2.3.2 Αξιολόγηση

Η συγκεκριμένη μελέτη στοχεύει στην παροχή εγγυημένων υπηρεσιών. Όλη η απαραίτητη πληροφορία βρίσκεται στον κόμβο εισόδου του συστήματος εκτός από τις περιπτώσεις sharing και για εκείνη την κίνηση για την οποία δεν επαρκούν οι πόροι του αντίστοιχου μεριδίου. Έτσι, η επιβάρυνση σε σηματοδοσία λόγω αναγκαίου συγχρονισμού μεταξύ των ακραίων κόμβων και οι χρόνοι απόκρισης του συστήματος στις αιτήσεις υπηρεσιών είναι ιδιαίτερα χαμηλοί και το σύστημα εφαρμόσιμο σε μεγάλη κλίμακα.

Οι μηχανισμοί συγχρονισμού μεταξύ των ακραίων κόμβων και τα ζητήματα επιλογής του κατάλληλου επιπέδου sharing παραμένουν ανοικτά.

A.2.3.2.4 Μελέτη [BN03]

A.2.3.2.4.1 Περιγραφή

Η μελέτη [BN03] περιγράφει τους μηχανισμούς συγχρονισμού μεταξύ των οντοτήτων ελέγχου εισόδου των ακραίων κόμβων σε μια μέθοδο αμιγώς sharing. Οι ακραίοι κόμβοι οργανώνονται σε ένα νοητό δακτύλιο και ο συγχρονισμός των λειτουργιών κατανομής των κοινόχρηστων πόρων γίνεται με την κυκλοφορία κουπονιού (token) στο δακτύλιο των ακραίων κόμβων.

Το κουπόνι περιέχει πληροφορία για τη διαθέσιμη χωρητικότητα ανά ζεύξη και μεταδίδεται με προκαθορισμένη σειρά σε κάθε ακραίο κόμβο του νοητού δακτυλίου. Κάθε αίτηση υπηρεσίας που φτάνει σε ένα ακραίο κόμβο μπαίνει σε κατάσταση αναμονής και εξυπηρετείται όταν φτάσει το κουπόνι με την ενημερωμένη κατάσταση διαθεσιμότητας. Κάθε ακραίος κόμβος γνωρίζει την τοπολογία του δικτύου και τη δρομολόγηση των πακέτων. Έτσι, οι απαραίτητοι πόροι για κάθε αίτηση υπηρεσίας αντιστοιχίζονται στις ζεύξεις του μονοπατιού προς τον προορισμό για τη συγκεκριμένη υπηρεσία. Εάν οι διαθέσιμοι πόροι των ζεύξεων επαρκούν τότε η αίτηση γίνεται αποδεκτή, οι αντίστοιχοι απαραίτητοι πόροι αφαιρούνται από τους διαθέσιμους των ζεύξεων και το κουπόνι ενημερώνεται με τις νέες τιμές διαθεσιμότητας ανά ζεύξη. Αφού εξυπηρετηθούν όλες οι αιτήσεις το ενημερωμένο κουπόνι στέλνεται στον επόμενο ακραίο κόμβο στη σειρά του νοητού δακτυλίου. Μια αίτηση υπηρεσίας είναι δυνατό να εξυπηρετηθεί και πριν την άφιξη του κουπονιού στην περίπτωση που ο τερματισμός άλλων υπηρεσιών του ίδιου ακραίου κόμβου έχει ελευθερώσει τους απαραίτητους πόρους. Ουσιαστικά, σε κάθε ακραίο κόμβο αποδίδονται εκείνοι οι διαθέσιμοι κοινόχρηστοι πόροι που είναι απαραίτητοι για την εξυπηρέτηση των υπηρεσιών τη στιγμή που το κουπόνι βρίσκεται στον ακραίο κόμβο και κατά τη διάρκεια του κύκλου του κουπονιού ο ακραίος κόμβος διαχειρίζεται ελεύθερα αυτούς τους κατοχυρωμένους πόρους. Η μόνη περίπτωση απόρριψης μιας αίτησης υπηρεσίας ενώ υπάρχουν διαθέσιμοι πόροι είναι όταν οι απαραίτητοι πόροι έχουν ελευθερωθεί από τον τερματισμό υπηρεσιών σε κάποιο ακραίο κόμβο μετά την έλευση του κουπονιού. Εάν η διάρκεια του κύκλου του κουπονιού είναι μικρή τότε το κουπόνι ενημερώνεται πιο γρήγορα και η πιθανότητα τέτοιας λανθασμένης απόρριψης μικραίνει.

Με αυτό το βασικό μοντέλο ο χρόνος απόκρισης του συστήματος στη χειρότερη περίπτωση είναι ίσος με τη διάρκεια κύκλου του κουπονιού μεταξύ των ακραίων κόμβων. Για να βελτιωθεί ο χρόνος απόκρισης χρησιμοποιούνται δύο βελτιώσεις στο βασικό μοντέλο, ο μηχανισμός *προκαταβολικής δέσμευσης* (anticipatory reservation) και ο μηχανισμός *σημείωσης ζεύξης* (link marking).

Με τον μηχανισμό anticipatory reservation κάθε ακραίος κόμβος επιπλέον των απαραίτητων πόρων σε κάθε άφιξη κουπονιού δεσμεύει προκαταβολικά και ένα κλάσμα της διαθέσιμης χωρητικότητας έτσι ώστε να είναι σε θέση να εξυπηρετήσει άμεσα μελλοντικές αιτήσεις υπηρεσίας χωρίς να αναμένει την επόμενη άφιξη του κουπονιού. Είναι φανερό πως όταν οι διαθέσιμοι πόροι είναι λιγосτοί, η προκαταβολική δέσμευση πόρων από ένα κόμβο είναι πολύ πιθανό να τους στερήσει από κάποιο άλλο με αποτέλεσμα την απόρριψη αιτήσεων ενώ υπάρχουν διαθέσιμοι πόροι. Όσο μεγαλύτερο το κλάσμα που δεσμεύεται προκαταβολικά τόσο μικρότερος ο μέσο χρόνος απόκρισης του συστήματος και τόσο μεγαλύτερη η πιθανότητα λανθασμένης απόρριψης. Το κατάλληλο κλάσμα καθορίζεται με βάση τον επιθυμητό χρόνο απόκρισης με κόστος σε αξιοποίηση των πόρων.

Όταν ο φόρτος του δικτύου είναι χαμηλός οι διαθέσιμοι πόροι αφθονούν, η πιθανότητα δηλαδή να υπάρχουν οι απαραίτητοι πόροι υψηλή και άρα η αναμονή της άφιξης του κουπονιού για την κατοχύρωσή τους άσκοπη. Με τον μηχανισμό link marking οι ακραίοι κόμβοι χρησιμοποιούν τους πόρους των ζεύξεων με χαμηλό φόρτο χωρίς να περιμένουν την έλευση του κουπονιού. Όταν οι διαθέσιμοι πόροι μιας ζεύξης λιγοστεύουν τότε η ζεύξη μαρκάρεται και η κατανομή των πόρων της γίνεται και πάλι με το μηχανισμό anticipatory reservation. Αντίστοιχα όταν διαπιστώνεται εκ νέου αύξηση των διαθέσιμων πόρων η ζεύξη επιστρέφει στη μη μαρκαρισμένη κατάσταση. Καθώς η δέσμευση πόρων για τις μη μαρκαρισμένες ζεύξεις δε γίνεται πια σε σειρά αλλά παράλληλα από όλους τους κόμβους, για να μην υπάρχουν επικαλύψεις κάθε κόμβος μπορεί να χρησιμοποιήσει ένα μέρος από τους διαθέσιμους πόρους. Αυτό το μερίδιο υπολογίζεται δυναμικά κάθε φορά που λαμβάνεται το κουπόνι, αναλογικά με τις ανάγκες του κάθε κόμβου και έτσι ώστε να διασφαλίζεται πως το σύνολο των μεριδίων δεν υπερβαίνει τη φυσική χωρητικότητα της ζεύξης.

A.2.3.2.4.2 Αξιολόγηση

Χάρη στη μέθοδο sharing η αξιοποίηση των πόρων είναι ιδιαίτερα υψηλή με κόστος την επιμήκυνση του χρόνου απόκρισης. Κατάλληλες βελτιώσεις επιτυγχάνουν τη μείωση του χρόνου απόκρισης με μικρό κόστος σε αξιοποίηση των πόρων.

Έμφαση δίνεται και στην ευρωστία του μοντέλου. Κάθε κόμβος περιοδικά επιβεβαιώνει πως οι παρακείμενοι σε αυτόν στον δακτύλιο παραμένουν συνδεδεμένοι και σε λειτουργία. Σε περίπτωση που διαγνωσθεί βλάβη ο κόμβος που είναι εκτός λειτουργίας παρακάμπτεται και ο δακτύλιος κλείνει ξανά. Σε περίπτωση που η βλάβη συμβαίνει στον κόμβο που είχε το κουπόνι τότε το κουπόνι αναπαράγεται στον προηγούμενο κόμβο και η κυκλοφορία του συνεχίζεται. Ο δακτύλιος με αυτό τον τρόπο χαρακτηρίζεται ως λογικό *εύρωστο υπερκείμενο δίκτυο* (Resilient Overlay Network – RON). Αντίστοιχες συγκεντρωτικές μέθοδοι υψηλής αξιοποίησης των πόρων χάρη στη μέθοδο sharing υστερούν σημαντικά σε ευρωστία και σε δυνατότητα εφαρμογής σε μεγάλη κλίμακα.

Η μελέτη αυτή υποστηρίζει εγγυημένες υπηρεσίες. Η επέκτασή της σε προβλέψιμες υπηρεσίες είναι εφικτή και η μελέτη των απαραίτητων τροποποιήσεων αποτελεί θέμα προς περαιτέρω έρευνα.

A.2.3.3 Κατανεμημένες Measurement-Based Μελέτες

A.2.3.3.1 Μελέτη [MPCC00]

A.2.3.3.1.1 Περιγραφή

Η μελέτη [MPCC00] παρουσιάζει μια απλή μεταφορά των single-node measurement-based σχημάτων σε multi-node περιβάλλοντα.

Ο έλεγχος εισόδου κίνησης γίνεται στο επίπεδο του πρωτοκόλλου TCP. Το έναυσμα για τη συγκεκριμένη μελέτη δεν είναι η παροχή συγκεκριμένης ποιότητας υπηρεσίας αλλά πηγάζει από τη διαπίστωση πως η απόδοση του δικτύου βελτιώνεται με τον έλεγχο της συμφόρησης ακόμα και για ροές δεδομένων με βάση το πρωτόκολλο TCP που διαθέτει εγγενείς μηχανισμούς για την αντιμετώπιση της συμφόρησης [S95]. Έτσι, της μετάδοσης των δεδομένων δεν προηγείται μια αίτηση υπηρεσίας. Οι εφαρμογές ξεκινούν απευθείας τη μετάδοση των δεδομένων και το δίκτυο πρέπει να διαθέτει μηχανισμούς για την ανίχνευση των νέων ροών και για την αναχαίτιση εκείνων των νέων ροών που απορρίπτονται. Οι νέες ροές ανιχνεύονται παρακολουθώντας τα SYN ή SYN/ACK TCP πακέτα. Η αναχαίτιση των ροών γίνεται είτε διαγράφοντας αυτά τα πακέτα είτε με εναλλακτικές μεθόδους όπως αποστολή πακέτων RST, ειδικών ICMP πακέτων κ.ά. Επειδή η λειτουργία του μοντέλου βασίζεται στους ήδη υπάρχοντες μηχανισμούς του TCP και είναι έτσι διάφανη στις εφαρμογές, το σχήμα αναφέρεται και ως *σιωπηρός έλεγχος εισόδου κίνησης* (implicit admission control).

Σε κάθε κόμβο μια λειτουργία *εκτίμησης φόρτου* (load estimator) καθορίζει εάν ο κόμβος βρίσκεται σε κατάσταση αποδοχής ή απόρριψης (accepting/rejecting state) ανά ζεύξη. Εάν κατά την ανίχνευση μιας νέας ροής ο κόμβος βρίσκεται σε κατάσταση απόρριψης τότε ακολουθείται η διαδικασία αναχαίτισης της ροής. Στην κατάσταση αποδοχής δεν πραγματοποιείται καμία επέμβαση. Ο load estimator χρησιμοποιεί κάποια single-node measurement-based τεχνική (βλέπε ενότητα A.2.2.1.2) και εκτιμά την πραγματική συνολική κίνηση με βάση τις μετρήσεις. Η μετάβαση στην κατάσταση απόρριψης καθορίζεται από κάποιο κατώφλι (threshold) η τιμή του οποίου ρυθμίζεται από τον παροχέα.

A.2.3.3.1.2 Αξιολόγηση

Η μελέτη [MPCC00] δεν αποτελεί ολοκληρωμένη μελέτη για την παροχή ποιότητας υπηρεσίας καθώς δεν προτείνει μηχανισμούς για τη διασφάλιση συγκεκριμένης ποιότητας υπηρεσίας. Επιπλέον περιορίζεται σε ροές που χρησιμοποιούν το πρωτόκολλο TCP και δεν είναι άμεσα εφαρμόσιμη καθώς προϋποθέτει ειδικές δυνατότητες στο λογισμικό ή το υλικό των κόμβων ώστε να είναι δυνατή η αποκωδικοποίηση και επεξεργασία των IP πακέτων με βάση τα περιεχόμενα της επικεφαλίδας ανώτερων πρωτοκόλλων.

Η αξιολόγηση συνεισφορά αυτής της μελέτης έγκειται στη βασική ιδέα της παρακολούθησης και χαρακτηρισμού των επιπέδων συμφόρησης ανά κόμβο κορμού χρησιμοποιώντας τοπικά απλούς measurement-based μηχανισμούς. Ο έλεγχος εισόδου κίνησης τότε μπορεί να γίνεται όπως εδώ ανά κόμβο (hop-by-hop) ή εναλλακτικά στους ακραίους ή τερματικούς κόμβους όπου συγκεντρώνονται και επεξεργάζονται κατάλληλα οι ενδείξεις συμφόρησης όπως υπολογίζονται στους κόμβους κορμού. Αρκετές από τις μελέτες που εξετάζονται παρακάτω χρησιμοποιούν αυτή την τεχνική.

A.2.3.3.2 Μελέτη [BBFP01]

A.2.3.3.2.1 Περιγραφή

Η μελέτη αυτή χρησιμοποιεί τη βασική αρχή της μεθόδου [MPCC00] σε μια πιο ολοκληρωμένη προσέγγιση για την παροχή ποιότητας σε δίκτυα DiffServ.

Ο βασικός μηχανισμός είναι απλός. Κάθε κόμβος κορμού ελέγχει τα επίπεδα συμφόρησης τοπικά χρησιμοποιώντας ένα measurement-based μηχανισμό και ακολουθώντας ένα κριτήριο απόφασης μεταβαίνει σε κατάσταση αποδοχής ή απόρριψης (βλέπε ενότητα A.2.3.3.1.1). Ο τερματικός κόμβος του αποστολέα πριν από τη χρήση της υπηρεσίας στέλνει ένα δοκιμαστικό πακέτο (probe) προς τον προορισμό και ξεκινά ένα χρονομετρητή (timer). Οι ενδιαμέσοι κόμβοι του μονοπατιού προς τον προορισμό προωθούν ή διαγράφουν το probe πακέτο ανάλογα με το αν βρίσκονται σε κατάσταση αποδοχής ή απόρριψης αντίστοιχα. Αν το πακέτο φτάσει στον προορισμό τότε εκεί δημιουργείται ένα πακέτο επιβεβαίωσης και στέλνεται προς τον τερματικό κόμβο του αποστολέα. Η λήψη του πακέτου επιβεβαίωσης στον αποστολέα σημαίνει αποδοχή της αίτησης υπηρεσίας και η μετάδοση των δεδομένων μπορεί να ξεκινήσει. Σε αντίθετη περίπτωση ο timer λήγει, η αίτηση θεωρείται πως έχει απορριφθεί και η χρήση της υπηρεσίας ματαιώνεται. Αυτό το βασικό σχήμα φέρει το όνομα GRIP (Gauge & Gate Reservation with Independent Probing) [GRIP-TR].

Η μελέτη [BBFP01] ασχολείται με την παροχή εγγυημένων υπηρεσιών και υπηρεσιών στατιστικών εγγυήσεων. Για να καταστεί δυνατή η ακριβής εκτίμηση της πραγματικής κίνησης μέσω των μετρήσεων που λαμβάνονται απαιτείται να είναι γνωστή και καθορισμένη η συμπεριφορά των ροών.

Έτσι, κάθε ροή κατά την είσοδό της στο δίκτυο ρυθμίζεται από ένα μηχανισμό *διπλού διαρρέοντος δοχείου* (Dual Leaky Bucket – DLB) και σύμφωνα με τις παραμέτρους που αντιστοιχούν στα χαρακτηριστικά της ροής. Τα χαρακτηριστικά κίνησης δεν καθορίζονται ανά ροή κατά την αίτηση για την αντίστοιχη υπηρεσία αλλά συνάγονται εμμέσως από το DSCP των πακέτων της ροής. Στα DSCPs αντιστοιχούν κλάσεις με διαφορετικά προκαθορισμένα χαρακτηριστικά κίνησης και κατ' επέκταση διαφορετικές τιμές των παραμέτρων των DLBs για τις ροές της κάθε κλάσης.

Ο βασικός μηχανισμός DLB επιβάλλει ένα άνω όριο στη ροή της κίνησης που ρυθμίζει. Αυτό όμως δεν είναι αρκετό για την εκτίμηση της πραγματικής κίνησης μέσω μετρήσεων. Όταν η ροή της κίνησης δεν είναι και κάτω φραγμένη τότε το δείγμα μπορεί να είναι ανεξέλεγκτα χαμηλό και άρα να μην αντιπροσωπεύει την πραγματική εν δυνάμει κίνησης. Για να καταστεί δυνατή η εκτίμηση της πραγματικής κίνησης ο μηχανισμός DLB επανξάνεται έτσι ώστε όταν η ροή των δεδομένων πέφτει κάτω από τον προβλεπόμενο ρυθμό τότε εισάγονται και αποστέλλονται κενά (dummy) πακέτα για να συμπληρώσουν το κενό.

Δεδομένης της δυνατότητας εκτίμησης της πραγματικής κίνησης από τις μετρήσεις με ακρίβεια η παρεχόμενη ποιότητα υπηρεσίας εξαρτάται από το κριτήριο απόφασης. Οι κανόνες, οι παράμετροι και οι τιμές των παραμέτρων στο κριτήριο απόφασης καθορίζονται από τον παροχέα. Το κριτήριο απόφασης που εξετάζεται από τους [BBFP01] για την παροχή υπηρεσιών στατιστικών εγγυήσεων εκτιμά τον αριθμό των ενεργών ροών από τις μετρήσεις και με βάση τα χαρακτηριστικά κίνησης κάθε κλάσης. Το επιθυμητό πλήθος των ενεργών ροών ορίζεται από τον παροχέα. Ένας κόμβος είναι σε κατάσταση αποδοχής όσο το δείγμα της συνολικής κίνησης εκτιμάται πως προέρχεται από πλήθος μικρότερο από το επιθυμητό. Είναι φανερό πως όσο μεγαλύτερο το πλήθος των ταυτόχρονων ενεργών ροών τόσο μικρότερη η παρεχόμενη ποιότητα υπηρεσίας.

Όταν ένα στιγμιότυπο της κίνησης έχει θέσει τον κόμβο σε κατάσταση αποδοχής είναι πιθανό τα δοκιμαστικά πακέτα που θα περάσουν πριν αλλάξει το στιγμιότυπο να αντιστοιχούν σε περισσότερες ροές

από όσες μπορεί ο κόμβος να υποστηρίξει επιπλέον των ήδη ενεργών. Το αντίκτυπο των νέων ροών που γίνονται ταυτόχρονα αποδεκτές δεν μπορεί να ανιχνευθεί από τις μετρήσεις παρά μόνο αφού τελειώσει ο κύκλος ενεργοποίησης και τα πρώτα πακέτα δεδομένων των νέων ροών φτάσουν στον κόμβο. Έτσι τα δοκιμαστικά πακέτα που περνούν όταν ο κόμβος βρίσκεται σε κατάσταση αποδοχής καταγράφονται και συνυπολογίζονται στον αριθμό των ενεργών ροών για κάποιο διάστημα οπότε και θεωρείται πως η κίνηση των αντίστοιχων ροών καταγράφεται πλέον από τις μετρήσεις.

A.2.3.3.2 Αξιολόγηση

Εδώ επιχειρείται η παροχή εγγυημένων υπηρεσιών χωρίς την καταχώρηση και διαχείριση πληροφορίας ανά ροή όπως στις παραδοσιακές μεθόδους σηματοδότησης ανά κόμβο ATM και IntServ (stateful) αλλά αποκλειστικά βάσει τοπικής και άμεσα ανακτήσιμης πληροφορίας (stateless).

Ο σχεδιασμός μιας stateless προσέγγισης απαγορεύει τη χρήση ενός parameter-based σχήματος το οποίο προϋποθέτει την καταχώρηση και διαχείριση πληροφορίας ανά ροή. Η παροχή εγγυημένων υπηρεσιών με measurement-based σχήματα προϋποθέτει δυνατότητα εκτίμησης της πραγματικής κίνησης με ακρίβεια. Οι σχετικές παραδοχές ωστόσο επιφέρουν υψηλό κόστος στην αξιοποίηση των πόρων.

Η αποστολή dummy πακέτων για τη διασφάλιση πως μια ροή μεταδίδει πάντα με τον προσδοκώμενο ρυθμό στερεί το σύστημα από το κέρδος πολυπλεξίας. Επιπλέον, όπως και στις parameter-based μεθόδους (βλέπε ενότητα A.2.2.1.1), ο a priori worst-case χαρακτηρισμός της κίνησης έχει ως αποτέλεσμα περαιτέρω κόστος σε αξιοποίηση των πόρων. Περαιτέρω κόστος σημειώνεται και εξαιτίας της μεθόδου συνυπολογισμού των δοκιμαστικών πακέτων στις ενεργές ροές. Δοκιμαστικά πακέτα που περνούν από ένα κόμβο είναι πιθανό να διαγραφούν από άλλους μεταγενέστερους στο μονοπάτι προς τον προορισμό κόμβους και έτσι η ροή εσφαλμένα θα θεωρείται για κάποιο διάστημα ενεργή αποκλείοντας άλλες ροές από τη χρήση των αντίστοιχων πόρων.

Η προϋπόθεση μερικών τύπων εφαρμογών και αντίστοιχων χαρακτηριστικών κίνησης που να αντιστοιχούν στις πεπερασμένες κλάσεις που προκύπτουν από τις διαφορετικές τιμές του DSCP μειώνει την ευελιξία του μοντέλου. Επιπλέον σύμφωνα με τον ορισμό του προτύπου [DS-ARCH] το DSCP χρησιμοποιείται για να καταδεικνύει το PHB, δηλαδή την ποιότητα υπηρεσίας που απαιτεί η ροή από το δίκτυο και όχι τα χαρακτηριστικά κίνησης της ροής.

A.2.3.3.3 Μελέτη [KHAWG98]

A.2.3.3.3.1 Περιγραφή

Αυτή η μελέτη ασχολείται αποκλειστικά με τον έλεγχο ροής των ενεργών υπηρεσιών και βασίζεται στο σχήμα *Ρητής Ειδοποίηση Συμφόρησης* (Explicit Congestion Notification – ECN) [F94]. Πρόκειται για ένα σχήμα επαύξησης της αρχιτεκτονικής DiffServ με ανάδραση (feedback enhanced DiffServ architecture). Χρησιμοποιώντας όπως και με τη μέθοδο ECN ένα bit της επικεφαλίδας των πακέτων οι κόμβοι κορμού σηματοδοτούν την κατάσταση συμφόρησης του δικτύου και οι ακραίοι κόμβοι προσαρμόζουν ανάλογα τους traffic conditioners των ενεργών ροών ελέγχοντας έτσι την ένταση της κίνησης.

Συγκεκριμένα, οι κόμβοι κορμού κατά την ανίχνευση συμφόρησης τοπικά, μαρκάρουν το ECN bit. Σε περίπτωση που στους τερματικούς κόμβους χρησιμοποιείται συμβατό πρωτόκολλο τότε η πηγή στον τερματικό κόμβο του αποστολέα θα μειώσει κατάλληλα τον ρυθμό μετάδοσης πακέτων (end-to-end ECN). Σε αντίθετη περίπτωση την ευθύνη αναλαμβάνει ο ακραίος κόμβος εισόδου (edge-to-edge ECN). Ο ακραίος κόμβος εξόδου εντοπίζει τα μαρκαρισμένα πακέτα και συντάσσει ένα ειδικό μήνυμα ειδοποίησης προς τον

τερματικό κόμβο του αποστολέα. Λαμβάνοντας αυτό το ειδικό μήνυμα ο ακραίος κόμβος εισόδου συμπεραίνει από την IP διεύθυνση του μηνύματος τη ροή στην οποία ανήκουν τα μαρκαρισμένα πακέτα και που κατ' επέκταση συμβάλλει στη συμμόρφωση και προσαρμόζει κατάλληλα τις ρυθμίσεις στον αντίστοιχο traffic conditioner μειώνοντας τελικά το ρυθμό της κίνησης που διοχετεύεται στο δίκτυο.

A.2.3.3.3.2 Αξιολόγηση

Αυτή η μελέτη περιγράφει ένα στοιχειώδη μηχανισμό διαχείρισης της συμμόρφωσης μέσω του ελέγχου εισόδου κίνησης με την κατάλληλη προσαρμογή των traffic conditioners. Δεν αναφέρεται με ποια κριτήρια το σύστημα επιστρέφει επανέρχεται αποκαθιστώντας τις αρχικές ρυθμίσεις των επηρεασμένων ροών, δεν αναλύεται το αποτέλεσμα του μηχανισμού στα επίπεδα συμμόρφωσης και δεν εξετάζεται η αλληλεπίδραση με τη διαδικασία αποδοχής ή απόρριψης αιτήσεων υπηρεσιών.

Παρά τις σημαντικές ελλείψεις της, η μελέτη αυτή συνεισφέρει την ιδέα της επιφόρτισης των λειτουργιών ελέγχου εισόδου κίνησης όχι μόνο με την πρόβλεψη και αποφυγή της συμμόρφωσης με τον αποκλεισμό των ροών που πιθανώς οδηγήσουν το δίκτυο σε συμμόρφωση αλλά και για την επίλυση της συμμόρφωσης σε περίπτωση κάποιας λανθασμένης ή σκοπίμως αισιόδοξης εκτίμησης για το αντίκτυπο των ροών που έγιναν αποδεκτές. Επισημαίνουμε ακόμα ότι μηχανισμοί διαχείρισης της συμμόρφωσης είναι εγγενείς σε κάποιες εφαρμογές, αποδεικνύονται ωστόσο ανεπαρκείς σε κάποιες περιπτώσεις δημιουργώντας την ανάγκη για λήψη επιπρόσθετων μέτρων σε ανώτερο επίπεδο από τον έλεγχο εισόδου κίνησης.

A.2.3.3.4 Μελέτη [PSZG02]

A.2.3.3.4.1 Περιγραφή

Οι [PSZG02] προτείνουν ένα σχήμα σύμφωνα με το οποίο οι κόμβοι κορμού πραγματοποιούν τις μετρήσεις και την ανάλυσή τους τοπικά και η λειτουργία ελέγχου εισόδου κίνησης στους ακραίους κόμβους συλλέγει τα αποτελέσματα βάσει των οποίων αποφαινεται την αποδοχή ή την απόρριψη της αίτησης υπηρεσίας ή, στην περίπτωση που υπάρχουν διαθέσιμοι πόροι αλλά δεν επαρκούν για να καλύψουν τις απαιτήσεις της αίτησης, την αποδοχή μιας κατάλληλα τροποποιημένης εναλλακτικής υπηρεσίας.

Σε κάθε κόμβο κορμού εκτιμάται ανά κλάση το εύρος ζώνης που είναι διαθέσιμο για την εξυπηρέτηση νέων υπηρεσιών έτσι ώστε, δεδομένης της έντασης της διερχόμενης κίνησης από τις ενεργές ροές όπως προκύπτει από τις μετρήσεις, να διασφαλίζεται η ποιότητα υπηρεσίας που αντιστοιχεί στην κλάση. Εδώ, η ποιότητα υπηρεσίας που αντιστοιχεί σε κάθε κλάση χαρακτηρίζεται από το άνω όριο της καθυστέρησης και την πιθανότητα παραβίασης του άνω ορίου. Λαμβάνοντας μια αίτηση υπηρεσίας ο ακραίος κόμβος εισόδου στέλνει πακέτα προς τον κόμβο εξόδου σύμφωνα με το πρωτόκολλο συλλογής αποτελεσμάτων. Στα πακέτα αυτά κάθε κόμβος κορμού αναγράφει το εκτιμώμενο διαθέσιμο εύρος ζώνης. Ο ακραίος κόμβος εξόδου, αφού αναγράφει το τοπικό εκτιμώμενο διαθέσιμο εύρος ζώνης, στέλνει το πακέτο πίσω στον κόμβο εισόδου όπου και υπολογίζεται το διαθέσιμο εύρος ζώνης από άκρο σε άκρο και συγκρίνεται με το ρυθμό της αίτησης υπηρεσίας. Σε περίπτωση που το διαθέσιμο εύρος ζώνης δεν επαρκεί αλλά δεν είναι μηδενικό τότε είναι δυνατό να γίνει αποδεκτή η αίτηση υπηρεσίας για μειωμένο ρυθμό και ίσο με το διαθέσιμο εύρος ζώνης.

A.2.3.3.4.2 Αξιολόγηση

Η μέθοδος αυτή χρησιμοποιεί κατ' ουσία την τεχνική των single-node measurement-based σχημάτων (βλέπε ενότητα A.2.2.1.2) και ανά κόμβο σηματοδοσίας με λίγο διαφορετική προσέγγιση, η απόφαση απόρριψης ή

αποδοχής, ή αλλιώς η σύγκριση των διαθέσιμων με τους απαραίτητους για τη νέα υπηρεσία πόρους, δεν λαμβάνεται ανά κόμβο αλλά συνολικά από τον ακραίο κόμβο εισόδου αφού έχει συλλέξει τα αποτελέσματα της διαδικασίας εκτίμησης των διαθέσιμων πόρων σε κάθε κόμβο του μονοπατιού. Η επιφόρτιση των κόμβων κορμού με την αντίστοιχη πολυπλοκότητα και ακόμα περισσότερο η δυσκολία στην πρακτική εφαρμογή της λόγω των νέων λειτουργιών και πρωτοκόλλων που δεν υποστηρίζονται από τις υπάρχουσες υλοποιήσεις στους δρομολογητές αποτελούν τα βασικά μειονεκτήματα της μεθόδου.

A.2.3.3.5 Μελέτη [CKK01]

A.2.3.3.5.1 Περιγραφή

Η μελέτη [CKK01] προτείνει ένα μοντέλο ελέγχου εισόδου κίνησης αμιγώς κατανεμημένου σε ακραίους κόμβους με measurement-based μέθοδο. Η απόφαση αποδοχής ή απόρριψης λαμβάνεται στον ακραίο κόμβο εξόδου (egress admission control) βάσει μετρήσεων της διαθέσιμης ποιότητας από κάθε ακραίο κόμβο εισόδου προς τον συγκεκριμένο κόμβο εξόδου.

Στον κόμβο εξόδου λαμβάνονται μετρήσεις με την τεχνική των *φακέλων κίνησης* (traffic envelopes) για τον υπολογισμό του *φακέλου αφίξεων* (arrival envelope) βάσει του μέγιστου ρυθμού αφίξεων και του *φακέλου υπηρεσίας* (service envelope) βάσει του ελάχιστου ρυθμού εξυπηρέτησης που υπολογίζεται μετρώντας τη συνολική από άκρο σε άκρο καθυστέρηση των πακέτων προς το μέγεθος των πακέτων.

Ο αλγόριθμος ελέγχου εισόδου κίνησης ελέγχει εάν η κατ' ελάχιστο διαθέσιμη ποιότητα υπηρεσίας όπως αποτυπώνεται στον φάκελο υπηρεσίας επαρκεί για την ικανοποιητική εξυπηρέτηση της νέας και των ήδη ενεργών υπηρεσιών όπως αποτυπώνονται στον φάκελο αφίξεων. Οι σχετικές μαθηματικές εξισώσεις διαμορφώνουν αυτό που ονομάζεται *συνθήκη προγραμματισμού* (schedulability condition) η οποία χαρακτηρίζεται από ένα *επίπεδο εμπιστοσύνης* (confidence level) ως προς την ακρίβεια των υπολογισμών, δεδομένης της αβεβαιότητας που υπεισέρχεται από τη χρήση μετρήσεων όπως σε κάθε measurement-based μέθοδο.

Η διακύμανση στους καταμετρημένους φακέλους κίνησης χρησιμοποιείται για την μακροπρόθεσμη εκτίμηση του στατιστικού λάθους στις μετρήσεις έτσι ώστε να προσαρμόζεται κατάλληλα το confidence level με στόχο να αντανakλά καλύτερα την πραγματική εν δυνάμει συμπεριφορά και διακύμανση της κίνησης και κατ' επέκταση να ρυθμίζεται κατάλληλα η αυστηρότητα του schedulability condition.

Η πληροφορία για τη συνολική από άκρο σε άκρο καθυστέρηση των πακέτων για τον υπολογισμό του φακέλου υπηρεσίας δεν είναι άμεσα διαθέσιμη στον κόμβο εξόδου που βλέπει μόνο τις αναχωρήσεις των πακέτων από το άκρο εξόδου του δικτύου. Για την απόκτηση αυτής της πληροφορίας προτείνονται δύο εναλλακτικοί τρόποι. Εάν οι ακραίοι κόμβοι του δικτύου έχουν συγχρονισμένα ρολόγια (synchronized clocks) τότε με την χρονο-εντύπωση (time-stamping) κάθε πακέτου στον ακραίο κόμβο εισόδου ο ακραίος κόμβος εξόδου υπολογίζει την καθυστέρηση με μια απλή αφαίρεση. Εάν ο συγχρονισμός των ρολογιών δεν είναι εφικτός ή δεν είναι ακριβής σε κλίμακα χρόνου συγκρίσιμη με την καθυστέρηση των πακέτων τότε εναλλακτικά προτείνεται η αναγραφή σε ένα πεδίο της επικεφαλίδας του πακέτου του συνολικού χρόνου αναμονής σε όλους τους κόμβους. Ο ακραίος κόμβος εισόδου και κάθε ενδιάμεσος κόμβος κορμού υπολογίζει τον χρόνο αναμονής που υπέστη το πακέτο τοπικά, τον προσθέτει στο συνολικό χρόνο αναμονής σε όλους τους προηγούμενους κόμβους και ανανεώνει την τιμή του πεδίου. Ο ακραίος κόμβος εξόδου στην τιμή του πεδίου προσθέτει τη σταθερή καθυστέρηση από το φυσικό επίπεδο των ζεύξεων στο συγκεκριμένο μονοπάτι και καταλήγει με αυτό τον τρόπο στη συνολική από άκρο σε άκρο καθυστέρηση για το κάθε πακέτο.

A.2.3.3.5.2 Αξιολόγηση

Το δίκτυο αντιμετωπίζεται σαν μαύρο κουτί (black box), και η προσφερόμενη ποιότητα συνάγεται άμεσα, όπως και με τις probe-based μεθόδους, μετρώντας τη συνολική από άκρο σε άκρο καθυστέρηση των πακέτων. Ωστόσο, συγκριτικά με τις probe-based μεθόδους η συλλογή παθητικών μετρήσεων στους ακραίους κόμβους υπερτερεί σημαντικά σε ακρίβεια χάρη στη συνεχή και διαρκή συλλογή και καταγραφή των χαρακτηριστικών της κίνησης από το σύνολο των διερχόμενων πακέτων δεδομένων.

Η βασική αυτή διαφορά σε σύγκριση με τις παραδοσιακές measurement-based μεθόδους που βασίζονται στη μέτρηση της έντασης της κίνησης που διοχετεύεται πάνω από γνωστή και σταθερή χωρητικότητα αποτελεί θεμελιώδη καινοτομία και το βασικότερο πλεονέκτημα της μελέτης. Αφενός άρονται οι ενίοτε αυθαίρετες ή προσεγγιστικές και συχνά δεσμευτικές παραδοχές των μοντέλων των μηχανισμών χρονοπρογραμματισμού και διαχείρισης καταχωρητή των ενδιάμεσων κόμβων κορμού.

Αφετέρου, εξαλείφοντας από τους υπολογισμούς την παράμετρο της χωρητικότητας άρεται και η ανάγκη για προσδιορισμό των διαθέσιμων πόρων, χρησιμοποιώντας είτε partitioning με εκ των προτέρων κατακερματισμό των πόρων σε μη επικαλυπτόμενα μερίδια που αποτελούν και την από άκρο σε άκρο χωρητικότητα, είτε sharing χρησιμοποιώντας ως χωρητικότητα το σύνολο των πόρων και συνυπολογίζοντας την ένταση της διασταυρωνόμενης κίνησης (cross-traffic), δηλαδή της κίνησης από άλλους ακραίους κόμβους που διασταυρώνεται με την κίνηση στο υπό εξέταση από άκρο σε άκρο μονοπάτι. Εδώ δεν κατακερματίζονται οι πόροι επιτυγχάνοντας τη μέγιστη αξιοποίηση των πόρων ενώ το αντίκτυπο των αυξομειώσεων της έντασης της διασταυρωνόμενης κίνησης αποτυπώνεται εμμέσως στις μετρήσεις του φακέλου υπηρεσίας αποφεύγοντας την πολυπλοκότητα της μεθόδου sharing που απαιτεί πληροφορία για τη δρομολόγηση και την ένταση των ροών κάθε κόμβου και κατ' επέκταση και την επικοινωνία μεταξύ των κόμβων για την εξαγωγή αυτής της πληροφορίας.

Είναι προφανές πως, πέραν της δυνατότητας προσαρμογής της αυστηρότητας του αλγορίθμου στο περιβάλλον μέσω της ρύθμισης του confidence level έτσι ώστε να εξασφαλίζεται κατά το δυνατό η ποιότητα υπηρεσίας, ωστόσο κάθε αβεβαιότητα δεν είναι δυνατό να εξαλειφθεί, καθιστώντας την προτεινόμενη λύση ακατάλληλη για εγγυημένες υπηρεσίες και υπηρεσίες στατιστικών εγγυήσεων. Όπως και στις probe-based μεθόδους έτσι και εδώ εμφανίζεται το πρόβλημα του stealing bandwidth (βλέπε ενότητα A.2.2.1.3) μεταξύ διαφορετικών κλάσεων. Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί εφαρμόζοντας επιπλέον ειδικές συνθήκες για την αποδοχή μιας νέας αίτησης πέραν του τυπικού schedulability condition, έχοντας όμως ως αποτέλεσμα τον κατακερματισμό των πόρων μεταξύ των διαφορετικών κλάσεων.

Ωστόσο, το βασικότερο μειονέκτημα της μελέτης έγκειται στη δυσκολία της πρακτικής εφαρμογής της εφόσον προϋποθέτει είτε το συγχρονισμό μεταξύ των ακραίων κόμβων είτε ειδικούς μηχανισμούς υπολογισμού και αναγραφής του χρόνου αναμονής σε κάθε κόμβο του δικτύου. Σημαντικό μειονέκτημα είναι και η αντίστοιχη επιβάρυνση σε επεξεργασία ανά πακέτο στους ακραίους κόμβους αλλά και στους κόμβους κορμού με τη δεύτερη μέθοδο.

A.2.4 Ανακεφαλαίωση

Το επίπεδο των εγγυήσεων εξασφάλισης της ποιότητας υπηρεσίας που απαιτούνται από την κάθε εφαρμογή διαφέρει ανάλογα με την ανεκτικότητα της σε περιστασιακή υποβάθμιση. Η διαφοροποίηση αυτή αντανακλάται στις υπηρεσίες με το διαχωρισμό τους σε εγγυημένες υπηρεσίες, υπηρεσίες στατιστικών εγγυήσεων και προβλέψιμες υπηρεσίες.

Η απαίτηση για παροχή εγγυήσεων ποιότητας, είτε απόλυτων είτε στατιστικών, επιβάλλει τον συνυπολογισμό όλων των παραμέτρων που συμβάλλουν στην παραβίαση της ποιότητας. Στα δίκτυα DiffServ οι κόμβοι κορμού δεν διαχειρίζονται πληροφορία ανά ροή χωρίς την οποία ο υπολογισμός της πιθανότητας παραβίασης της ποιότητας δεν μπορεί να είναι ακριβής. Αναγκαστικά ο έλεγχος εισόδου κίνησης δεν γίνεται πλέον ανά κόμβο (single-node admission control) με το παραδοσιακό σχήμα της ανά κόμβο σηματοδότησης (hop-by-hop signalling). Η οντότητα που αναλαμβάνει τη λειτουργία ελέγχου εισόδου κίνησης καλείται να υπολογίσει την πιθανότητα παραβίασης της ποιότητας για κάθε κόμβο συγκρίνοντας τις απαιτήσεις των ενεργών ροών που διέρχονται από τον κόμβο έναντι της χωρητικότητας του κόμβου. Ο όγκος της πληροφορίας που απαιτείται και η επεξεργαστική ισχύ που αναλώνεται σε αυτό το έργο είναι απαγορευτικά μεγάλα για την εφαρμογή ενός τέτοιου συστήματος σε μεγάλη κλίμακα.

Η διαδικασία απλοποιείται θέτοντας a priori παραδοχές και κατανέμοντας ανάλογα τους πόρους. Ο υπολογισμός της πιθανότητας παραβίασης απλοποιείται καθώς κάποιες από τις παραμέτρους είναι τώρα σταθερές και δεν υπολογίζονται δυναμικά και επιπλέον η λειτουργία ελέγχου κίνησης μπορεί να διασπαστεί σε μικρότερες παράλληλες αυτόνομες διεργασίες. Όσο μεγαλύτερη απόκλιση έχει η πραγματική ζήτηση από την a priori κατανομή των πόρων τόσο μεγαλύτερο είναι το κόστος σε αξιοποίηση των πόρων. Η δυνατότητα αφομοίωσης αυτής της απόκλισης με την κατάλληλη δυναμική ανακατανομή των πόρων μειώνει το σχετικό κόστος, δημιουργεί όμως την ανάγκη συντονισμού μεταξύ των αυτόνομων διεργασιών με αντίστοιχο κόστος σε ανταλλαγή μηνυμάτων.

Εν κατακλείδι, το πρόβλημα του ελέγχου εισόδου κίνησης έγκειται στη διασφάλιση της ποιότητας υπηρεσίας πάνω από ένα DiffServ δίκτυο όπου οι κόμβοι κορμού δεν διαχειρίζονται πληροφορία ανά ροή, με τρόπο βέλτιστο για την αξιοποίηση των πόρων, πλην όμως εφαρμόσιμο σε μεγάλη κλίμακα.

Οι παραδοχές πάνω στις οποίες οικοδομείται το μοντέλο υπηρεσιών που χρησιμοποιεί κάθε προσέγγιση είναι σημαντικό να μην είναι δεσμευτικές και να επιτρέπουν τον καθορισμό του τύπου των υπηρεσιών, της ποιότητας υπηρεσίας και των σχετικών εγγυήσεων από τον παροχέα ανάλογα με τις πολιτικές του. Επιπλέον, πρόσθετες απαιτήσεις τίθενται σχετικά με την αποτελεσματικότητα, τη δυνατότητα επίλυσης της συμφόρησης, την αμεροληψία, την ευστάθεια, την ευρωστία και την πρακτική εφαρμογή κάθε προσέγγισης.

Η αρχική προσέγγιση του προβλήματος έγινε με το συγκεντρωτικό μοντέλο του Bandwidth Broker (BB).

Αντίθετα με την αρχιτεκτονική IntServ οι κόμβοι του δικτύου απαλλάσσονται από κάθε επιφόρτιση εξαιτίας των λειτουργιών δέσμευσης πόρων και ελέγχου εισόδου κίνησης χωρίς ωστόσο αυτό να λύνει το πρόβλημα εφαρμογής σε μεγάλη κλίμακα καθώς η σχετική πολυπλοκότητα για ολόκληρη την επικράτεια συγκεντρώνεται σε μία οντότητα. Ο BB γίνεται σημείο συνωστισμού (bottleneck) στην εξυπηρέτηση των αιτήσεων υπηρεσιών αυξάνοντας το χρόνο απόκρισης σε βαθμό που σε κάποιες περιπτώσεις βλάπτει ακόμα και την αξιοποίηση των πόρων, ενώ σε περίπτωση βλάβης το σύστημα καταρρέει χωρίς καμία διασφάλιση της ευρωστίας του.

Στο άλλο άκρο του συγκεντρωτικού μοντέλου βρίσκουμε το μοντέλο κατανεμημένου ελέγχου εισόδου κίνησης στους τερματικούς κόμβους των χρηστών (endpoint admission control), εφικτό χάρη στις probe-based μεθόδους.

Οι probe-based μέθοδοι εκτιμούν τη δυνατότητα του δικτύου να προσφέρει την απαραίτητη ποιότητα άμεσα με την αποστολή δοκιμαστικών πακέτων χωρίς να απαιτούν καμία πληροφορία σχετικά με τη δομή και τη χωρητικότητα του δικτύου. Οι κριτές της καταλληλότητας της προσφερόμενης σε σχέση με την απαιτούμενη ποιότητα γίνονται οι ίδιες οι εφαρμογές και το δίκτυο απαλλάσσεται εντελώς από την επιβάρυνση της λειτουργίας του ελέγχου εισόδου. Ενώ τα φαινόμενα thrashing και stealing bandwidth επηρεάζουν αρνητικά την αξιοποίηση των πόρων το μεγαλύτερο μειονέκτημα των probe-based μεθόδων θεωρείται η επιμήκυνση

του χρόνου απόκρισης του συστήματος εξαιτίας της περιόδου δοκιμής. Σύντομες περίοδοι δοκιμής είναι ανεπαρκείς για την εκτίμηση της κατάστασης του δικτύου αποδυναμώνοντας ακόμα περισσότερο την ούτως ή άλλως περιορισμένη αξιοπιστία της μεθόδου.

Ανάμεσα στις δύο αυτές προσεγγίσεις βρίσκεται το μοντέλο καταναμημένου ελέγχου εισόδου κίνησης στους ακραίους κόμβους του δικτύου. Κάποιες από τις σχετικές μελέτες προϋποθέτουν και την ανάμιξη των κόμβων κορμού.

Οι λειτουργίες που απαιτούνται τοπικά σε κάθε κόμβο κορμού, αν και σε κάθε περίπτωση χωρίς να χρησιμοποιούν πληροφορία ανά ροή (stateless), αφενός επιβαρύνουν την απόδοση του κόμβου και αφετέρου δεν υποστηρίζονται από τις υπάρχουσες υλοποιήσεις στους δρομολογητές δυσχεραίνοντας την πρακτική εφαρμογή αυτών των μελετών. Ακόμα μεγαλύτερο πρόβλημα πρακτικής εφαρμογής δημιουργείται όταν, σε κάποιες περιπτώσεις, τα πρωτόκολλα συλλογής των αποτελεσμάτων των λειτουργιών από τους κόμβους κορμού στους ακραίους κόμβους απαιτούν τροποποιήσεις της TCP/IP στοίβας πρωτοκόλλων.

Στον καταναμημένο στους ακραίους κόμβους έλεγχο εισόδου κίνησης εκτός από τις probe-based μεθόδους χρησιμοποιούνται και parameter-based ή measurement-based μέθοδοι.

Οι parameter-based μέθοδοι παρέχουν την απαραίτητη μαθηματική ακρίβεια για την παροχή εγγυημένων υπηρεσιών και υπηρεσιών στατιστικών εγγυήσεων με σημαντικό ωστόσο και συχνά αδικαιολόγητα υψηλό κόστος σε αξιοποίηση των πόρων εξαιτίας της αδυναμίας επακριβούς a priori χαρακτηρισμού της κίνησης. Επιπλέον, οι πολύπλοκοι υπολογισμοί που απαιτούνται ανά αίτηση υπηρεσίας μειώνουν σημαντικά την αποτελεσματικότητα της μεθόδου.

Οι measurement-based μέθοδοι εκτιμούν τον εν δυνάμει φόρτο του δικτύου με βάση τις μετρήσεις της πραγματικής συνολικής κίνησης. Έτσι, απλοποιούνται οι υπολογισμοί, βελτιώνεται η αποτελεσματικότητα της μεθόδου και επιπλέον αποφεύγεται το κόστος σε αξιοποίηση των πόρων από την κατάργηση των συντηρητικών a priori παραδοχών. Οι μετρήσεις ωστόσο μπορούν να είναι μόνο ενδεικτικές της κατάστασης του δικτύου. Η αβεβαιότητα που υπεισέρχεται στους υπολογισμούς καθιστά τις measurement-based μεθόδους ακατάλληλες για εγγυημένες υπηρεσίες και υπηρεσίες στατιστικών εγγυήσεων. Η εκτίμηση του εν δυνάμει φόρτου ρυθμίζεται να είναι περισσότερο ή λιγότερο συντηρητική ανάλογα με την πολιτική του παροχέα επί της σχέσης μεταξύ κέρδους σε αξιοποίηση και κόστους σε υποβάθμιση της ποιότητας. Αποδεικνύεται πως οποιαδήποτε επίδοξη αντιστοίχιση της παραμέτρου ρύθμισης με συγκεκριμένη πιθανότητα παραβίασης της ποιότητας υπήρξε αποτυχημένη.

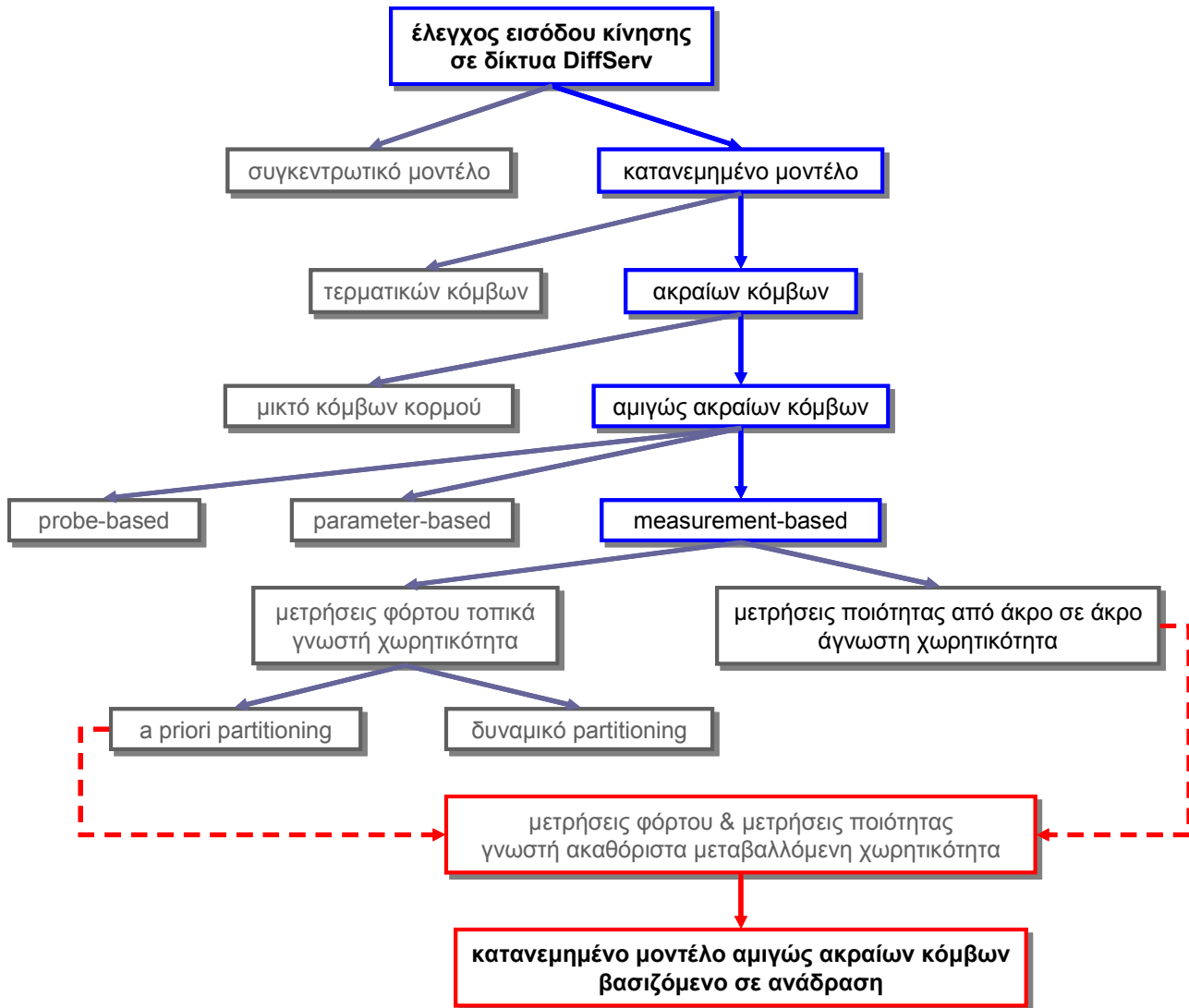
Όταν οι συμβατικές single-node parameter-based και measurement-based μέθοδοι των τεχνολογιών ATM και IntServ χρησιμοποιούνται στους ακραίους κόμβους τότε απαιτείται να είναι γνωστή η χωρητικότητα από άκρο σε άκρο (end-to-end). Η χωρητικότητα από άκρο σε άκρο όμως καθορίζεται από τη δρομολόγηση και την ένταση της κίνησης και από άλλους ακραίους κόμβους. ***Η χωρητικότητα που βλέπει κάθε ακραίος κόμβος διαμορφώνεται είτε a priori με διαμερισμό των πόρων σε μη επικαλυπτόμενα μερίδια ανά ακραίο κόμβο, είτε δυναμικά με κατανομή των κοινόχρηστων πόρων στους ανταγωνιζόμενους ακραίους κόμβους ανάλογα με την πραγματική ζήτηση.*** Με την πρώτη προσέγγιση κάθε ακραίος κόμβος λειτουργεί εντελώς ανεξάρτητα αλλά η αξιοποίηση των πόρων είναι μειωμένη λόγω κατακερματισμού (fragmentation) των πόρων σε απομονωμένα μερίδια, τόσο περισσότερο μάλιστα όσο λιγότερο αντιστοιχεί ο a priori διαμερισμός με την πραγματική ζήτηση. Με τη δεύτερη προσέγγιση η αξιοποίηση των πόρων είναι βέλτιστη καθώς η κατανομή των πόρων αντανακλά σε κάθε στιγμή την πραγματική ζήτηση αλλά η πολυπλοκότητα και ο χρόνος απόκρισης του συστήματος αυξάνονται σημαντικά εξαιτίας του αναγκαίου συγχρονισμού μεταξύ των ανταγωνιζόμενων ακραίων κόμβων. Μια υβριδική λύση χρησιμοποιεί a priori διαμερισμό των πόρων κατά την αρχικοποίηση του συστήματος και δυναμική κατανομή κατά τη λειτουργία του.

Κάποιες μελέτες εξερευνούν τη διάσταση της δυναμικής προσαρμογής της παρεχόμενης ποιότητας υπηρεσίας ανάλογα με τις μεταβολές του φόρτου του δικτύου. Σε κατάσταση συμφόρησης η επέμβαση του ελέγχου εισόδου κίνησης με τη μείωση του επιτρεπτού ρυθμού ροής των ενεργών υπηρεσιών κατά την είσοδό τους στο δίκτυο διευκολύνει την άμεση επίλυση της συμφόρησης. Αντίστοιχα, σε περίπτωση που με τον τερματισμό κάποιων ενεργών ροών έχουν δημιουργηθεί αρκετοί διαθέσιμοι πόροι, η αύξηση του επιτρεπτού ρυθμού ροής των ενεργών υπηρεσιών αυξάνει αντίστοιχα την παρεχόμενη ποιότητα υπηρεσίας.

ΜΕΡΟΣ Β. ΠΕΡΙΓΡΑΦΗ

B.1 Τοποθέτηση επί του Προβλήματος

B.1.1 Το Βασικό Μοντέλο



Σχήμα B-1: το βασικό μοντέλο βήμα προς βήμα

B.1.1.1 Καταναμημένο Μοντέλο Αμιγώς Ακραίων Κόμβων

Το συγκεντρωτικό μοντέλο της αρχιτεκτονικής Bandwidth Broker αποδείχθηκε προβληματικό ως προς την εφαρμογή σε μεγάλη κλίμακα, την αποτελεσματικότητα και την ευρωστία του. Συνεπώς οδηγούμαστε σε καταναμημένο μοντέλο.

Το μοντέλο ελέγχου εισόδου κίνησης καταναμημένου στους τερματικούς κόμβους (endpoint admission control) είναι σύμφυτο με τις probe-based μεθόδους. Τα ζητήματα εμπιστοσύνης και ασφάλειας που προκύπτουν με την ανάθεση λειτουργιών διαχείρισης όπως ο έλεγχος εισόδου κίνησης σε οντότητες εξωτερικές προς το δίκτυο, εδώ στους τερματικούς κόμβους, δημιουργούν περαιτέρω προβλήματα

εφαρμογής. Επιπλέον, τα εγγενή προβλήματα των probe-based μεθόδων και ιδιαίτερα η επιμήκυνση του χρόνου απόκρισης εξαιτίας της δοκιμαστικής περιόδου σε τέτοιο βαθμό που να γίνεται απαγορευτικά μεγάλος για κάποιες εφαρμογές καθιστούν αυτό το μοντέλο ακατάλληλο για ένα μεγάλο μέρος περιπτώσεων.

Η εναλλακτική ενδιάμεση προσέγγιση είναι το μοντέλο ελέγχου εισόδου κίνησης κατανεμημένου στους ακραίους κόμβους. Σε κάποιες από τις σχετικές μελέτες λειτουργίες επιτελούνται και στους κόμβους κορμού και ειδικά πρωτόκολλα αναλαμβάνουν την επικοινωνία μεταξύ των κατανεμημένων λειτουργιών. Οι λειτουργίες αυτές δεν συναντώνται στις γνωστές υλοποιήσεις και τα πρωτόκολλα επικοινωνίας είναι συχνά ασύμβατα με τις υπάρχουσες προδιαγραφές δημιουργούνται προβλήματα πρακτικής εφαρμογής. Η απαλλαγή των κόμβων κορμού από την υλοποίηση νέων πολύπλοκων λειτουργιών και από τη σχετική επιβάρυνση σε απόδοση θεωρείται η καλύτερη προσέγγιση.

Για τους λόγους που αναφέρονται παραπάνω **στην παρούσα διατριβή έχει επιλεγεί και χρησιμοποιείται μοντέλο κατανεμημένου ελέγχου εισόδου κίνησης στους ακραίους κόμβους χωρίς την ανάμιξη των κόμβων κορμού.**

B.1.1.2 Μοντέλο Βασιζόμενο σε Ανάδραση

Στον κατανεμημένο στους ακραίους κόμβους έλεγχο εισόδου κίνησης εκτός από τις probe-based μεθόδους είναι δυνατό να χρησιμοποιηθούν και parameter-based ή measurement-based μέθοδοι.

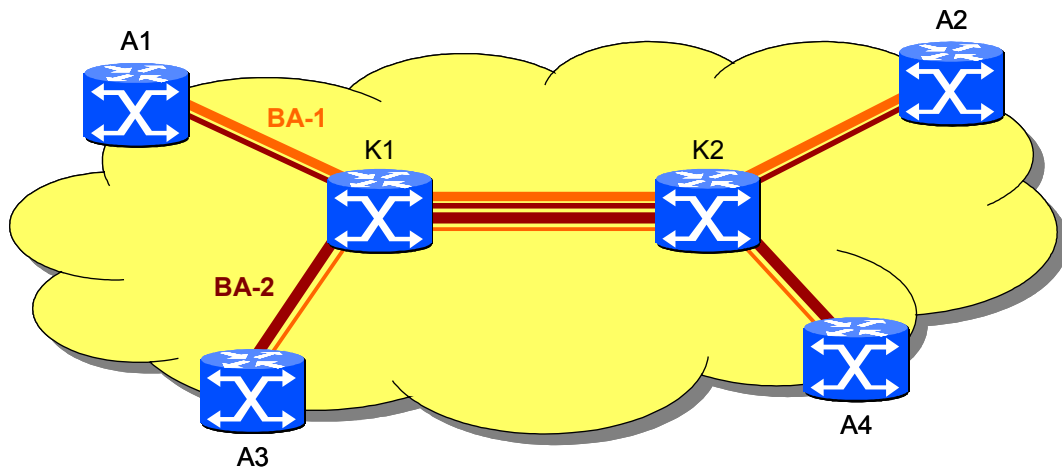
Ο μόνος τρόπος εξασφάλισης εγγυημένης απόλυτης ή στατιστικής ικανοποίησης της ποιότητας υπηρεσίας είναι με τη χρήση parameter-based μεθόδων, που συνεπάγεται την υιοθέτηση worst-case μοντέλων για το χαρακτηρισμό της συμπεριφοράς των πηγών και στην περίπτωση των εγγυημένων υπηρεσιών real-allocation σχημάτων για τη δέσμευση πόρων. Το κόστος σε αξιοποίηση των πόρων είναι απαγορευτικά υψηλό, ιδιαίτερα για κίνηση καταιγιστικής φύσεως, και στις περισσότερες περιπτώσεις αδικαιολόγητο καθώς η ποιότητα υπηρεσίας επιδέχεται περιστασιακή υποβάθμιση. Επισημαίνουμε ακόμα πως, όταν οι parameter-based μέθοδοι εφαρμόζονται από άκρο σε άκρο όπως στο μοντέλο ελέγχου εισόδου κίνησης κατανεμημένου στους ακραίους κόμβους, τότε η εξασφάλιση απόλυτων εγγυήσεων προϋποθέτει τον διαμερισμό των πόρων σε απομονωμένα μερίδια, συντελώντας έτσι σε ακόμα υψηλότερο κόστος από τον κατακερματισμό των πόρων. Σε αυτά τα προβλήματα προστίθεται και οι ιδιαίτερα χρονοβόροι και δαπανηροί υπολογισμοί που απαιτούνται ανά αίτηση υπηρεσίας και έχουν ως αποτέλεσμα τη μείωση της αποτελεσματικότητας του συστήματος.

Για τους παραπάνω λόγους **η παρούσα διατριβή υιοθετεί την measurement-based προσέγγιση.**

Η πλειονότητα των σχετικών μελετών βασίζεται στη μέτρηση της έντασης της κίνησης, ή αλλιώς του φόρτου (load) που διοχετεύεται πάνω από γνωστή χωρητικότητα (capacity). Η ένταση της κίνησης που διοχετεύεται στο δίκτυο μπορεί να μετρηθεί τοπικά στον ακραίο κόμβο και η χωρητικότητα των ζεύξεων κορμού κατανέμεται σε απομονωμένα μη επικαλυπτόμενα μερίδια τα οποία αποτελούν την χωρητικότητα από άκρο σε άκρο που διατίθεται σε κάθε ακραίο κόμβο. Κάθε τέτοιο μερίδιο είναι είτε σταθερό και διαμορφώνεται a priori με τη μέθοδο διαμερισμού των πόρων (a priori partitioning), είτε κατ' αίτηση μεταβαλλόμενο και διαμορφώνεται δυναμικά με την επιβλεπόμενη και συντονισμένη δέσμευση και αποδέσμευση κοινόχρηστων πόρων ανάλογα με την πραγματική ζήτηση μεταξύ των ανταγωνιζόμενων ακραίων κόμβων (δυναμικό partitioning).

Με την πρώτη προσέγγιση κάθε ακραίος κόμβος λειτουργεί εντελώς ανεξάρτητα αλλά η αξιοποίηση των πόρων είναι μειωμένη λόγω κατακερματισμού (fragmentation) των πόρων σε απομονωμένα μερίδια, τόσο

περισσότερο μάλιστα όσο λιγότερο αντιστοιχεί ο a priori διαμερισμός με την πραγματική ζήτηση. Με τη δεύτερη προσέγγιση η αξιοποίηση των πόρων είναι βέλτιστη καθώς η κατανομή των πόρων αντανακλά κάθε στιγμή την πραγματική ζήτηση. Η πολυπλοκότητα και ο χρόνος απόκρισης του συστήματος όμως αυξάνονται σημαντικά εξαιτίας του αναγκαίου συντονισμού μεταξύ των ανταγωνιζόμενων ακραίων κόμβων.



		C	D	AL		
				a-priori-p	class-dynamic-p	dynamic-p
C1	A1-A2	4	5	4	5	5
	A3-A4	4	2	2	2	2
	Σύνολο	8	7	6	7	7
C2	A1-A2	4	3	3	3	3
	A3-A4	4	6	4	5	6
	Σύνολο	8	9	7	8	9

Σχήμα B-2: παράδειγμα κατανομής της ζήτησης

Στο Σχήμα B-2 απεικονίζεται ένα παράδειγμα κατανομής της κίνησης μεταξύ ακραίων κόμβων. Θεωρώντας δύο κλάσεις κίνησης έστω C1 και C2, που αντιστοιχούν σε δύο PHBs και κατ' επέκταση σε δύο διαφορετικά DSCPs διακρίνουμε τις *συσσωματώσεις συμπεριφοράς* (Behaviour Aggregate – BA) BA-1 και BA-2. Το BA ορίζεται ως "η συλλογή των πακέτων με ίδιο DSCP που διασχίζουν μία ζεύξη προς μία κατεύθυνση" [DS-ARCH]. Στον πίνακα του σχήματος με C (capacity) συμβολίζεται η χωρητικότητα του μεριδίου, με D (demand) η πραγματική ζήτηση και με AL (admitted load) η κίνηση που γίνεται αποδεκτή κάτω από τα σχήματα a priori partitioning (a-priori-p), δυναμικό partitioning σε επίπεδο κλάσης (class-dynamic-p) και αμιγώς δυναμικό partitioning (dynamic-p).

Έστω η ζεύξη K1-K2 χωρητικότητας C_{K1-K2} 16 μονάδων κατανέμεται με a priori partitioning σε ίσα μερίδια των 4 μονάδων ανά ζεύγος A1-A2 και A3-A4 και ανά κλάση C1, C2 αντίστοιχα. Σε επίπεδο κλάσης, π.χ. C1, το φαινόμενο κατακερματισμού των πόρων εξαιτίας του διαμερισμού τους σε μερίδια ανά ζεύγος ακραίων κόμβων έχει ως αποτέλεσμα την απόρριψη μίας από τις πέντε μονάδες της D_{A1-A2}^{C1} ενώ υπάρχουν

στην ίδια κλάση διαθέσιμοι πόροι $(C_{A1-A2}^{C1} + C_{A3-A4}^{C1}) - (AL_{A1-A2}^{C1} + AL_{A3-A4}^{C1}) = (4 + 4) - (4 + 2) = 2$. Η συνολική αξιοποίηση των πόρων ισούται με:

$$U_{K1-K2} = \frac{AL_{A1-A2}^{C1} + AL_{A1-A2}^{C2} + AL_{A3-A4}^{C1} + AL_{A3-A4}^{C2}}{C_{K1-K2}} = \frac{4 + 3 + 2 + 4}{16} = \frac{13}{16} = 0,8125$$

Εάν σε επίπεδο κλάσης χρησιμοποιηθεί η μέθοδος δυναμικού partitioning τότε η συνολική αξιοποίηση βελτιώνεται σε $U_{K1-K2} = \frac{5 + 3 + 2 + 5}{16} = \frac{15}{16} = 0,9375$ ενώ και πάλι ο κατακερματισμός των πόρων κοστίζει μία μονάδα από τη D_{A3-A4}^{C2} που έχει παραχωρηθεί αλλά δεν χρησιμοποιείται στην κλάση C1.

Με οποιαδήποτε από τις παραπάνω μεθόδους, η χωρητικότητα από άκρο σε άκρο είναι γνωστή και παραχωρείται αποκλειστικά. Οι αυξομειώσεις της διασταυρωνόμενης κίνησης (cross-traffic), δηλαδή της κίνησης από άλλους ακραίους κόμβους που χρησιμοποιεί τις ίδιες ζεύξεις με μια μακρο-ροή (macro-flow), δεν έχουν κανένα αντίκτυπο στην από άκρο σε άκρο χωρητικότητα για αυτή τη μακρο-ροή. Παράδειγμα μακρο-ροής είναι η κίνηση από τον A1 προς τον A2 για την οποία διασταυρωνόμενη κίνηση θεωρείται η κίνηση από τον A3 στον A4. Με τη μέθοδο δυναμικού partitioning ωστόσο η από άκρο σε άκρο χωρητικότητα προσαρμόζεται ανάλογα με τις αυξομειώσεις της ζήτησης, δηλαδή οι αυξομειώσεις της διασταυρωνόμενης ζήτησης έχουν αντίκτυπο στη δυνατότητα κατοχύρωσης επιπλέον χωρητικότητας για μια μακρο-ροή.

Οι αδυναμίες των δύο μεθόδων που χρησιμοποιούν μετρήσεις φόρτου τοπικά και θεωρούν γνωστή από άκρο σε άκρο χωρητικότητα καθιστούν αναγκαία την εξεύρεση εναλλακτικής λύσης. Έτσι, στη μελέτη [CKK01] το δίκτυο αντιμετωπίζεται σαν μαύρο κουτί (black box) και η χωρητικότητα από άκρο σε άκρο είναι άγνωστη. Η διαθεσιμότητα των πόρων είναι συνάρτηση της τοπικής από τον εκάστοτε ακραίο κόμβο κίνησης και της διασταυρωνόμενης κίνησης και συνάγεται άμεσα χρησιμοποιώντας μετρήσεις της καθυστέρησης από άκρο σε άκρο όπως και με τις probe-based μεθόδους. Η διαφορά είναι πως, αντί για μετρήσεις επί των δοκιμαστικών πακέτων κατά την περιορισμένη περίοδο δοκιμής, εδώ γίνεται συνεχής συλλογή και καταγραφή των χαρακτηριστικών της κίνησης από μετρήσεις επί του συνόλου των διερχόμενων πακέτων δεδομένων, αυξάνοντας σημαντικά την ακρίβεια των εκτιμήσεων.

Η άρση κάθε παραδοχής σχετικά με τη μοντελοποίηση των μηχανισμών του δικτύου που τώρα αντιμετωπίζεται σαν μαύρο κουτί και η εξάλειψη των προβλημάτων κατακερματισμού των πόρων σε επίπεδο κλάσης και πολυπλοκότητας με τις μεθόδους a priori και δυναμικού partitioning αντίστοιχα, αποτελούν σημαντικά πλεονεκτήματα. Στο αντίβαρο βρίσκονται οι προϋποθέσεις που επιβάλλονται για την πρακτική εφαρμογή της και η επιβάρυνση του συστήματος με δαπανηρή ανά πακέτο επεξεργασία.

Στην παρούσα εργασία προτείνουμε ένα διαφορετικό μοντέλο που ονομάζουμε μοντέλο βασισμένο σε ανάδραση (feedback-based).

Σε κάθε ακραίο κόμβο παραχωρείται ένα μερίδιο των πόρων κορμού αποκλειστικά δεσμευμένο (hard-reserved) και ένα συμπληρωματικό κοινόχρηστο μερίδιο (soft-reserved) με a priori partitioning. Εάν υπάρχει αρκετή ζήτηση κάθε κόμβος επιτρέπεται να υπερβεί το hard-reserved αλλά και το soft-reserved μερίδιό του και να χρησιμοποιήσει πόρους από τα soft-reserved μερίδια των άλλων κόμβων χωρίς συντονισμένη δέσμευση και αποδέσμευση μεταξύ των εμπλεκόμενων κόμβων αλλά με ελεγχόμενο ρίσκο. Αντίστοιχα, το soft-reserved μερίδιο κάθε κόμβου δεν είναι εξασφαλισμένα διαθέσιμο αλλά μπορεί να χρησιμοποιείται από άλλους κόμβους. Σε κάθε κόμβο χρησιμοποιώντας μετρήσεις του φόρτου τοπικά υπολογίζεται το μέρος από το hard-reserved ή και το soft-reserved ή ακόμα και από το επιπλέον του soft-reserved μεριδίου που

χρησιμοποιείται από την κίνηση που διοχετεύεται στο δίκτυο από τον συγκεκριμένο κόμβο, το οποίο χρησιμεύει ως ένδειξη για το ρίσκο παραβίασης της ποιότητας λόγω της επικάλυψης των soft-reserved μεριδίων. Το φαινόμενο είναι αντίστοιχο με το stealing bandwidth των probe-based μεθόδων. Μαζί με τις μετρήσεις του φόρτου τοπικά χρησιμοποιούνται και περιοδικές μετρήσεις της ποιότητας από άκρο σε άκρο. Όταν αυτές οι μετρήσεις καταδείξουν παραβίαση της ποιότητας τότε κατάλληλοι μηχανισμοί ανάδρασης ανακόπτουν την ένταση με την οποία διοχετεύεται η κίνηση μέχρι την ομαλή επίλυση της συμφόρησης που προκάλεσε την παραβίαση της ποιότητας.

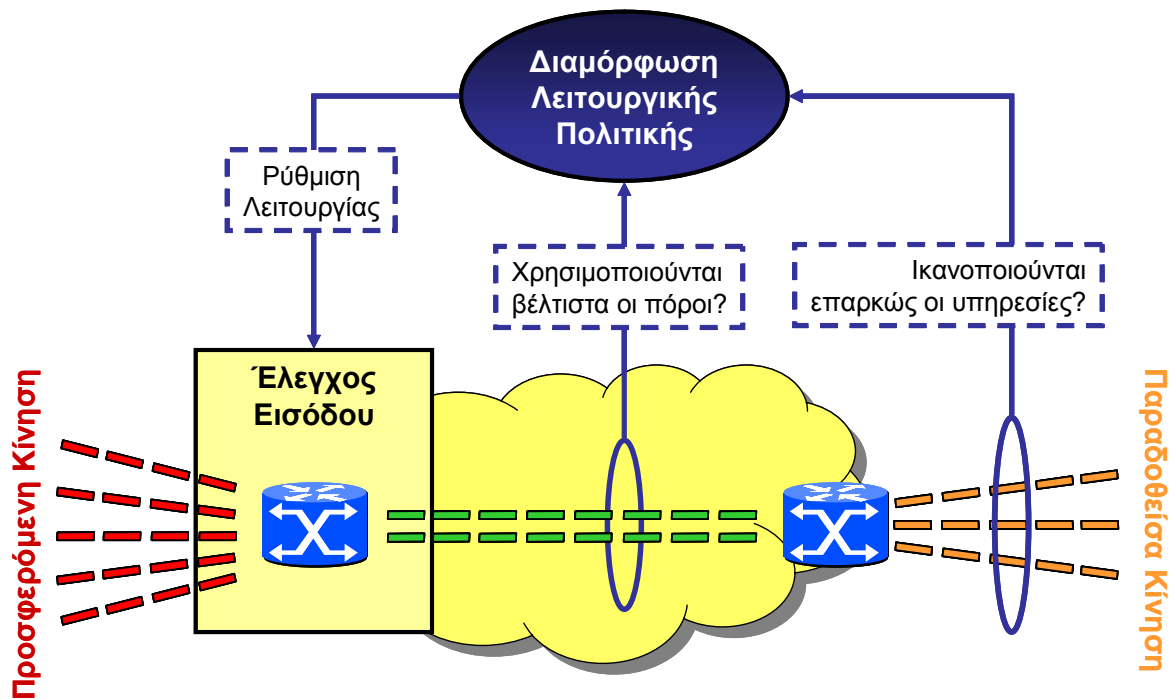
B.1.1.3 Ρύθμιση Από Λειτουργική Πολιτική

Η βασική ιδέα πίσω από το feedback-based μοντέλο είναι η προετοιμασία του συστήματος με a priori κατανομή των πόρων και η ευέλικτη δυναμική προσαρμογή του στις πραγματικές συνθήκες χωρίς συμβιβασμό στην αξιοποίηση των πόρων από εμμονή στις a priori παραδοχές και εκτιμήσεις, χωρίς την επιβάρυνση του συντονισμού για την κατάλληλη ανακατανομή των πόρων μεταξύ των ακραίων κόμβων και τέλος χωρίς υπερβολική επιβάρυνση εξαιτίας αυξημένων ή και ανεδαφικών απαιτήσεων σε ένταση και είδος μετρήσεων.

Το τίμημα για τα προηγούμενα πλεονεκτήματα είναι ο μειωμένος έλεγχος επί της a priori διασφάλισης της ποιότητας καθώς το μοντέλο είναι σχεδιασμένο να λειτουργεί μόνο κατά ένα μέρος επί εξασφαλισμένου μεριδίου και με ελεγχόμενο ρίσκο κατά το υπόλοιπο. Ελεγχόμενο ρίσκο σημαίνει ελεγχόμενα συντηρητική ή επιθετική συμπεριφορά στη χρησιμοποίηση των soft-reserved μεριδίων, η διαθεσιμότητα των οποίων δεν ελέγχεται με κανένα τρόπο a priori καθώς δεν ανταλλάσσονται μηνύματα για τον συντονισμό της επικάλυψης μεταξύ των ανταγωνιζομένων ακραίων κόμβων. Συνεπώς, το ελεγχόμενο ρίσκο δεν αντιστοιχεί με κανένα τρόπο σε μαθηματικά προβλέψιμη πιθανότητα παραβίασης της ποιότητας. Μαθηματικά βέβαιη είναι μόνο η μονοτονική σχέση μεταξύ ρίσκου και πιθανότητας παραβίασης της ποιότητας, όσο μεγαλύτερο το επιτρεπτό ρίσκο τόσο μεγαλύτερη η πιθανότητα παραβίασης της ποιότητας.

Αυτό ωστόσο έχει αποδειχθεί πως ισχύει για κάθε measurement-based αλγόριθμο [BJS00]. Συγκεκριμένα, έχει αποδειχθεί πως, σε όσους measurement-based αλγόριθμους επιχειρείται ακριβής αντιστοίχιση καθορισμένης απόκλισης από την ποιότητα της υπηρεσίας με συγκεκριμένη τιμή της εκάστοτε παραμέτρου ρύθμισης για την περισσότερο ή λιγότερο συντηρητική συμπεριφορά του αλγόριθμου, αποτυγχάνει. Καθίσταται έτσι αναγκαία η παρακολούθηση από το παροχέα της πραγματικής συμπεριφοράς του αλγόριθμου στο συγκεκριμένο περιβάλλον έτσι ώστε εμπειρικά να εφαρμόσει τις κατάλληλες ρυθμίσεις.

Έτσι, υιοθετούμε την πρακτική της εμπειρικής ρύθμισης του αλγόριθμου (βλέπε Σχήμα B-3). Ωστόσο εδώ, οι παράμετροι που ρυθμίζουν την περισσότερο ή λιγότερο συντηρητική συμπεριφορά του αλγόριθμου αντανακλούν τη *λειτουργική πολιτική* (operational policy) του παροχέα και εκφράζονται με σχετικό και περιφραστικό τρόπο χωρίς να επιχειρείται μάταια σύνδεση με συγκεκριμένη μαθηματική a priori διασφάλιση της απόκλισης από τις απόλυτες εγγυήσεις και κατ' επέκταση του ποσοστού παραβίασης της ποιότητας. Η ρύθμιση της λειτουργίας του συστήματος γίνεται ανάλογα με την αντίληψη του παροχέα για το τι σημαίνει *επαρκής ικανοποίηση* των υπηρεσιών και για ποιο συνεπαγόμενο κόστος επί της *βέλτιστης αξιοποίησης* των πόρων, έχοντας αφομοιώσει και εμμέσως συνεκτιμήσει την απρόβλεπτη συμπεριφορά των χρηστών και το στατιστικό λάθος των όποιων σχετικών υποκείμενων μαθηματικών μοντέλων χρησιμοποιούνται.



Σχήμα B-3: εμπειρική ρύθμιση λειτουργίας βασικού μοντέλου

B.1.1.4 Λογισμός επί της Ικανοποίησης

Για να καταστεί δυνατός ο μαθηματικός λογισμός επί της *επαρκούς ικανοποίησης* των υπηρεσιών εισάγουμε την έννοια της *ικανοποίησης της υπηρεσίας* (service satisfaction), η οποία ορίζεται ως το κλάσμα του ρυθμού ροής που παραδίδεται με τη συμφωνημένη ποιότητα προς τον ρυθμό ροής της προσφερόμενης συμμορφούμενης με τους όρους της υπηρεσίας κίνησης.

Είναι γνωστό πως ο βαθμός ανεκτικότητας σε περιστασιακή υποβάθμιση της ενδεδειγμένης ποιότητας ποικίλλει μεταξύ των διαφορετικών εφαρμογών ανάλογα με το πόσο *προσαρμοστικές* (adaptive) είναι. Αυτή η διαπίστωση αντανακλάται και στο διαχωρισμό μεταξύ EF και AF PHBs (βλέπε ενότητα A.1.4.2.2.2). Για την αποτύπωση του διαφορετικού βαθμού ανεκτικότητας χρησιμοποιούνται οι έννοιες της *σχεδόν ικανοποιημένης* (almost satisfied) και της *πλήρως ικανοποιημένης* (fully satisfied) υπηρεσίας. Οι υπηρεσίες κρίνονται ως fully satisfied όταν απολαμβάνουν τη συμφωνημένη ποιότητα στον συμφωνημένο ρυθμό (contractual rate) και almost satisfied όταν η ποιότητα διατηρείται για ρυθμό μικρότερο από τον συμφωνημένο αλλά αποδεκτό για τη συγκεκριμένη εφαρμογή. Αντίστοιχα ορίζουμε τον *σχεδόν ικανοποιητικό ρυθμό εξυπηρέτησης* (almost satisfied service rate – SR_{AS}) και τον *πλήρως ικανοποιητικό ρυθμό εξυπηρέτησης* (fully satisfied service rate – SR_{FS}) ως τον ρυθμό ροής που, εάν εξυπηρετηθεί με τη συμφωνημένη ποιότητα, η υπηρεσία χαρακτηρίζεται ως σχεδόν και πλήρως ικανοποιημένη αντίστοιχα. Η σχέση μεταξύ πλήρως και σχεδόν ικανοποιητικού ρυθμού εξυπηρέτησης αντανακλά και το βαθμό ανεκτικότητας μιας εφαρμογής.

Σύμφωνα με τους παραπάνω ορισμούς, η ικανοποίηση της υπηρεσίας, δεδομένης σταθερής και απαράβατης ποιότητας μεταβάλλεται ανάλογα με τον ρυθμό που εξυπηρετείται με αυτή την ποιότητα. Αυτή η προσέγγιση είναι σύμφωνη με την βασική διαπίστωση πως, σε περιόδους συμφόρησης για την αποφυγή της κατάρρευσης του συστήματος η προτιμότερη πρακτική είναι η μείωση του ρυθμού ροής [S95].

Η αντικειμενική εκτίμηση των σχεδόν και πλήρως ικανοποιητικών ρυθμών ανά τύπο εφαρμογής που χρησιμοποιεί την κάθε υπηρεσία επαφίεται των παραδοχών και των μοντέλων που ορίζονται στον παροχέα.

Περαιτέρω διαφοροποίηση ως προς τον βαθμό ανεκτικότητας εξαρτάται και από την υποκειμενική αντίληψη του παροχέα ως προς το είδος των υπηρεσιών που επιθυμεί να προσφέρει, ευνοώντας ενδεχομένως συγκεκριμένους τύπους υπηρεσιών ή/και πελατών. Αυτή η σύμφωνη με τα κριτήρια του παροχέα διαφοροποίηση αποτυπώνεται στην παράμετρο της *προτεραιότητας υπηρεσίας* (service priority), η αποτίμηση της οποίας ανά τύπο υπηρεσίας είναι ζήτημα εσωτερικής επιχειρησιακής πολιτικής και η ερμηνεία της οποίας μπορεί π.χ. να καθορίζει την προτεραιότητα ικανοποίησης της υπηρεσίας σε περίπτωση συμφόρησης. Θεωρούμε πως ο αλγόριθμος ελέγχου εισόδου κίνησης θα πρέπει να παρέχει τη δυνατότητα διαφοροποίησης της προτεραιότητας μεταξύ των υπηρεσιών ενώ παράλληλα να στοχεύει σε αμερόληπτη αντιμετώπιση των υπηρεσιών του ίδιου επιπέδου προτεραιότητας (βλέπε ενότητα A.2.1.3.3).

B.1.1.5 Επίλυση της Συμφόρησης

Το feedback-based μοντέλο βασίζεται σε μετρήσεις της ποιότητας από άκρο σε άκρο για την ανίχνευση της συμφόρησης. Η συμφόρηση στη γενική περίπτωση δημιουργείται σε μια ζεύξη συνωστισμού (bottleneck link) στο δίκτυο κορμού από κίνηση από περισσότερους από έναν ακραίο κόμβο. Η επίλυση της συμφόρησης επομένως απαιτεί την καλά συντονισμένη δράση όλων των ανά ακραίο κόμβο αντίστοιχων υπεύθυνων οντοτήτων ελέγχου εισόδου κίνησης έτσι ώστε το δίκτυο να φεύγει κατά το δυνατόν άμεσα από την κατάσταση συμφόρησης αλλά ταυτόχρονα και με ομαλό τρόπο χωρίς να απειλείται η σταθερότητα του συστήματος.

Η διαδικασία της a priori κατανομής των πόρων του δικτύου σε απομονωμένα μερίδια εξασφαλίζει πως για την επίλυση της συμφόρησης αρκεί κάθε κόμβος να περιορίσει την κίνηση που διοχετεύει στο δίκτυο στο μερίδιο που του αναλογεί. Το ίδιο αποκλειστικά παραχωρημένο μερίδιο χρησιμεύει σαν πολύτιμος γνώμονας για τη σταθεροποίηση του συστήματος.

Είναι τέλος προφανές πως η διαδικασία επίλυσης της συμφόρησης προϋποθέτει την μείωση της έντασης της κίνησης που διοχετεύεται στο δίκτυο, ή αλλιώς του επιτρεπτού ρυθμού ροής, ή αλλιώς της ικανοποίησης των εμπλεκόμενων ενεργών υπηρεσιών. Με άλλα λόγια, το feedback-based μοντέλο υιοθετεί τη λογική της δυναμικής προσαρμογής της παρεχόμενης ποιότητας υπηρεσίας ανάλογα με τις μεταβολές του φόρτου του δικτύου.

B.1.2 Περιβάλλον Πλαίσιο

B.1.2.1 Μοντέλο Υπηρεσιών

Στην αρχιτεκτονική DiffServ, οι παράμετροι που περιγράφουν τις διαφορετικές πλευρές των δεσμεύσεων του παροχέα και των υποχρεώσεων του πελάτη σε μια υπηρεσία σύνδεσης επιπέδου μεταφοράς σχηματίζουν τη δομή SLA, ενώ το υποσύνολο των παραμέτρων που αφορούν τα τεχνικά χαρακτηριστικά ποιότητας υπηρεσίας σχηματίζει τη δομή SLS (βλέπε ενότητα A.1.4.2.2). Ωστόσο, οι διαδικασίες που σχετίζονται με μια υπηρεσία, το ακριβές περιεχόμενο, η σύνταξη και η σημασιολογία των παραμέτρων των δομών SLA και SLS που την περιγράφουν και η ανάλυση και αντιστοίχισή της στα δομικά στοιχεία του δικτύου παραμένουν ανοικτά θέματα στην προτυποποίηση της αρχιτεκτονικής DiffServ.

Στις επόμενες ενότητες περιγράφεται το διαδικαστικό μοντέλο, προσδιορίζεται το περιεχόμενο της υπηρεσίας και παρουσιάζεται ένα ιεραρχικό δομικό μοντέλο που συνδέει τη δομή SLA με το επίπεδο των PHBs και των Traffic Conditioners που αναπαριστούν τα δομικά στοιχεία του δικτύου στην αρχιτεκτονική DiffServ.

B.1.2.1.1 Διαδικαστικό Μοντέλο

Διακρίνουμε μεταξύ των φάσεων *εγγραφής* (subscription) και *κλήσης* (invocation) της υπηρεσίας [GVT02].

Η φάση εγγραφής περιλαμβάνει τη διαδικασία διαπραγμάτευσης (negotiation) ενός συμβολαίου υπηρεσίας μεταξύ πελάτη και παροχέα, η επιτυχής κατάληξη της οποίας έχει ως αποτέλεσμα τη σύναψη του SLA, που καθιστά τον πελάτη δικαιούχο στη χρήση της υπηρεσίας. Ο πελάτης κάθε φορά που επιθυμεί να χρησιμοποιήσει την υπηρεσία επικαλείται το δικαίωμά του πραγματοποιώντας μια αίτηση *κλήσης* της υπηρεσίας σύμφωνα με τους όρους του συμβολαίου. Εφόσον πληρούνται οι όροι, ο παροχέας είναι υποχρεωμένος να ανταποκριθεί στην αίτηση *κλήσης ενεργοποιώντας* (activating) την υπηρεσία. Η διάκριση αυτή είναι απαραίτητη κυρίως για λόγους εξακρίβωσης γνησιότητας, εξουσιοδότησης και χρέωσης (Authentication-Authorisation-Accounting – AAA) και συμφωνεί με την επικρατούσα πρακτική των επιχειρήσεων.

Επισημαίνουμε πως ο πελάτης, ως νομική οντότητα, μπορεί να είναι ένας ομότιμος παροχέας υπηρεσιών επιπέδου μεταφοράς, αλλιώς και *Παροχέας Υπηρεσιών Διαδικτύου* (Internet Service Provider – ISP), παροχέας υπηρεσιών επιπέδου εφαρμογής (Application Service Provider – ASP), ένας οργανισμός ή ένας μεμονωμένος οικιακός χρήστης. Το συμβόλαιο υπηρεσίας συνάπτεται με τον πελάτη αλλά δύναται να περιλαμβάνει ένα σύνολο εγγεγραμμένων χρηστών που υπόκεινται με κάποιο τρόπο στη νομική οντότητα του πελάτη και έχουν δικαίωμα κλήσης και χρήσης της υπηρεσίας, όπως για παράδειγμα οι υπάλληλοι ενός οργανισμού.

Η φάση της εγγραφής και της κλήσης υπηρεσίας εξετάζεται στις παρακάτω ενότητες.

B.1.2.1.1.1 Εγγραφή Υπηρεσίας

Η φάση εγγραφής αρχικοποιείται από την αποστολή μιας αίτησης εγγραφής από τον πελάτη στον παροχέα όπου περιγράφεται η υπό διαπραγμάτευση υπηρεσία. Η αίτηση εγγραφής σχηματίζεται βάσει της *Δομής Εγγραφής Υπηρεσίας* (Service Subscription Structure – SSS) (βλέπε ενότητα B.1.2.1.2). Η τελική μορφή της SSS που συμφωνείται μεταξύ των διαπραγματευόμενων μερών αποτελεί και το συμβόλαιο υπηρεσίας, ή αλλιώς το SLA.

Το πρωτόκολλο διαπραγμάτευσης και η λογική διαπραγμάτευσης της πλευράς του πελάτη είναι έξω από το πεδίο της παρούσας εργασίας.

Η λογική διαπραγμάτευσης για την αποδοχή μιας αίτησης εγγραφής στον παροχέα περιλαμβάνει τα στάδια εξακρίβωσης γνησιότητας (authentication) και εξουσιοδότησης (authorisation). Η εξουσιοδότηση για την αποδοχή μιας αίτησης εγγραφής υπόκειται σε κριτήρια που υποδεικνύονται από τις σχετικές πολιτικές του παροχέα (policy authorisation) και σε κριτήρια που αφορούν την διαθεσιμότητα των πόρων (resource authorisation).

Οι διαδικασίες εξακρίβωσης γνησιότητας και εξουσιοδότησης βάσει των πολιτικών του παροχέα είναι έξω από το πεδίο της παρούσας εργασίας.

Η εξουσιοδότηση βάσει της διαθεσιμότητας των πόρων προϋποθέτει την μετάφραση και αντιστοίχιση των παραμέτρων που περιγράφουν την υπηρεσία σε πόρους του δικτύου και εν συνεχεία τον έλεγχο εισόδου κίνησης επί των απαιτούμενων προς τους διαθέσιμους πόρους.

Η διαδικασία μετάφρασης και αντιστοίχισης χρησιμοποιεί το ιεραρχικό δομικό μοντέλο υπηρεσίας (βλέπε ενότητα B.1.2.1.3) και εξαρτάται από τις συμβάσεις που χρησιμοποιούνται στη δομή SSS και την ερμηνεία που τους αποδίδει κάθε παροχέας. Για παράδειγμα, η ποιότητα υπηρεσίας μπορεί να χαρακτηρίζεται ως

χρυσή, ασημένια ή χάλκινη (golden, silver, bronze) και η μετάφραση σε συγκεκριμένες τιμές μέγιστης καθυστέρησης και απώλειας πακέτων να είναι θέμα του παροχέα.

Αντί των επισφαλών και αναγκαστικά περιορισμένης χρησιμότητας προβλέψεων, ο παροχέας είναι σε θέση να εκτιμήσει την κατανομή, τις απαιτήσεις ποιότητας και την ένταση της κίνησης που θα πρέπει να εξυπηρετήσει μελλοντικά βάσει της πληροφορίας που βρίσκεται καταγεγραμμένη στα συμβόλαια υπηρεσίας. Η πληροφορία αυτή είναι ζωτική για τον αποτελεσματικό σχεδιασμό και διαχείριση των πόρων του δικτύου συμβάλλοντας έτσι στην αύξηση του κέρδους του παροχέα (βλέπε ενότητα Α.1.4.3). Ο πελάτης από την πλευρά του επωφελείται χάρη στις εγγυήσεις που του παρέχει η εγγραφή του στην υπηρεσία πως θα είναι διαθέσιμη όταν τελικά θα χρειαστεί να τη χρησιμοποιήσει.

B.1.2.1.1.2 Κλήση Υπηρεσίας

Σε αντίθεση με την εγγραφή σε μια υπηρεσία, με την κλήση της υπηρεσίας οι πόροι δεν κατοχυρώνονται για μελλοντική χρήση αλλά παραχωρούνται άμεσα. Με την επιτυχή περάτωση της διαδικασίας ο χρήστης ξεκινά να διοχετεύει κίνηση στο δίκτυο.

Διακρίνουμε τις περιπτώσεις *ρητής κλήσης* (explicit invocation) και *έμμεσης κλήσης* (implicit invocation).

Κατά την έμμεση κλήση, η υπηρεσία ενεργοποιείται από τον παροχέα ως αποτέλεσμα της εγγραφής της υπηρεσίας και σύμφωνα με το χρονοδιάγραμμα που ορίζεται στο συμβόλαιο. Οι χρήστες είναι σε θέση να χρησιμοποιήσουν την υπηρεσία χωρίς άλλη διαδικασία κλήσης.

Αντίθετα, στη ρητή κλήση η ενεργοποίηση της υπηρεσίας γίνεται μόνο κατ' αίτηση του χρήστη. Η αίτηση κλήσης διαβιβάζεται στον παροχέα όπου πραγματοποιείται με τη σειρά εξακρίβωση γνησιότητας και εξουσιοδότηση. Η διαδικασία εξουσιοδότησης αποφαινεται καταρχήν εάν υπάρχει για την υπηρεσία που καλείται υπάρχει αντίστοιχη εγγραφή. Αφού βρεθεί η αντίστοιχη εγγραφή, περαιτέρω έλεγχοι, π.χ. έλεγχος του χρονοδιαγράμματος, γίνονται ανάλογα με τα περιεχόμενα του συμβολαίου. Εάν η αίτηση κλήσης πληροί όλους τους όρους του συμβολαίου τότε γίνεται εξουσιοδότηση με κριτήριο τη διαθεσιμότητα των πόρων και η διαδικασία ελέγχου εισόδου κίνησης βάσει των μεταφρασμένων κατά την εγγραφή της υπηρεσίας απαιτήσεων σε πόρους αποφασίζει εάν η αίτηση γίνεται αποδεκτή ή απορρίπτεται. Το αποτέλεσμα διαβιβάζεται πίσω στον χρήστη. Εδώ, ο τερματισμός χρήσης της υπηρεσίας δεν ανιχνεύεται αυτόματα από το σύστημα αλλά γνωστοποιείται με ειδικό μήνυμα *ρητού τερματισμού* (explicit termination). Σε περίπτωση που για μια ενεργή υπηρεσία ρητής κλήσης δεν σταλεί τέτοιο μήνυμα η υπηρεσία τερματίζεται από τον παροχέα με το πέρας του συμβεβλημένου χρονοδιαγράμματος και στέλνεται ειδοποίηση στο χρήστη. Η επικοινωνία αυτή γίνεται με ανταλλαγή μηνυμάτων ενός ειδικά σχεδιασμένου πρωτοκόλλου σηματοδότησης.

Τα δικαιώματα χρήσης και κατ' επέκταση κλήσης και ενεργοποίησης μιας υπηρεσίας διαφέρουν και σχηματίζουν τρεις διαφορετικούς τύπους υπηρεσιών, τις *υπηρεσίες μόνιμης ενεργοποίησης* (permanent services), τις *υπηρεσίες ενεργοποίησης ανά αίτηση* (on demand services) και τις *υπηρεσίες ευέλικτης ενεργοποίησης* (flexible services).

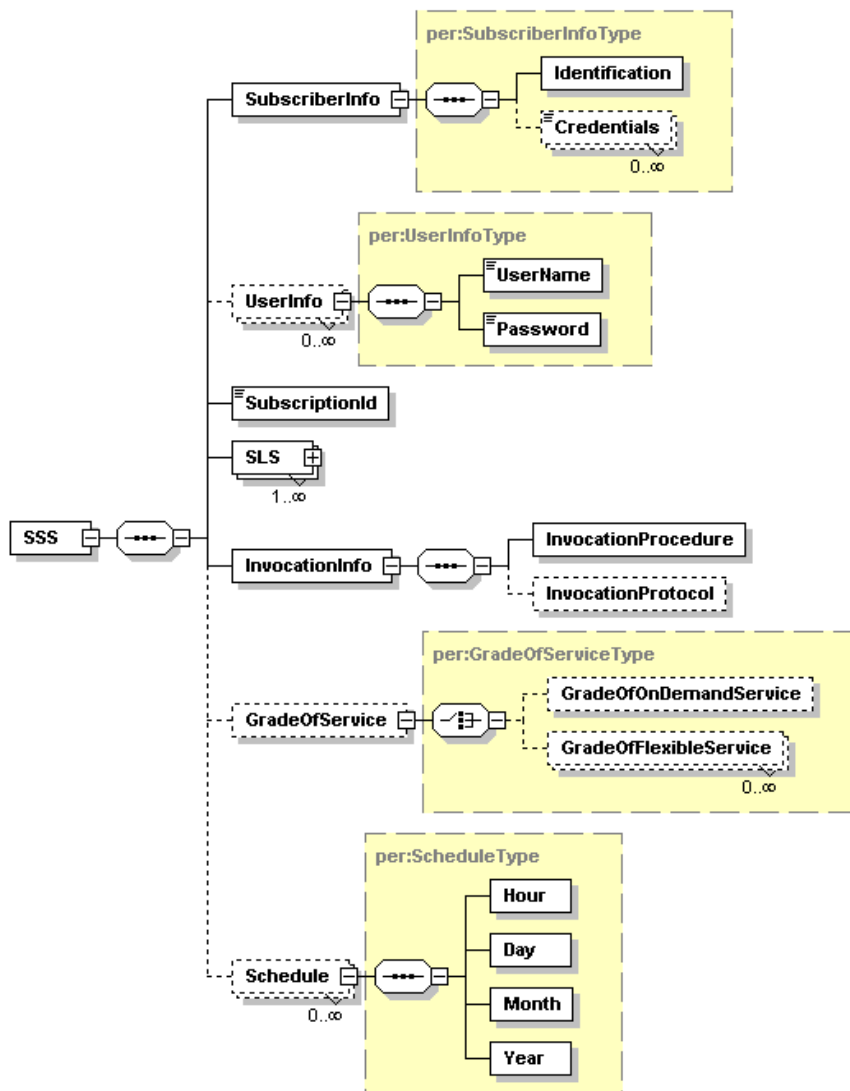
Οι permanent υπηρεσίες είναι μονίμως ενεργές εντός των πλαισίων του χρονοδιαγράμματός τους και είθισται να χρησιμοποιούν έμμεση κλήση. Τυπικό παράδειγμα τέτοιων υπηρεσιών είναι οι υπηρεσίες *Πλασματικού Ιδιωτικού Δικτύου* (Virtual Private Network – VPN). Στις υπηρεσίες on demand επιτρέπονται πολλαπλές ταυτόχρονες κλήσεις της υπηρεσίας και κάθε επιτυχής κλήση δημιουργεί μία νέα ενεργή υπηρεσία. Τυπικό παράδειγμα τέτοιων υπηρεσιών είναι οι υπηρεσίες για εφαρμογές *Τηλεφωνίας μέσω Διαδικτύου* (Voice over IP – VoIP). Η υπηρεσία αντιστοιχεί σε μια τηλεφωνική κλήση με συγκεκριμένα χαρακτηριστικά ποιότητας και πολλαπλές ταυτόχρονες κλήσεις μεταξύ διαφορετικών χρηστών της

υπηρεσίας μπορεί να είναι ενεργές. Στις flexible υπηρεσίες κάθε κλήση ενεργοποιεί ένα μέρος, πιο συγκεκριμένα ένα ποσοστό επιτρεπτού ρυθμού πρόσβασης, της υπηρεσίας έως ότου ενεργοποιηθεί ολόκληρη η υπηρεσία, όταν δηλαδή ο επιτρεπτός ρυθμός πρόσβασης φτάσει στο 100% του ρυθμού που έχει συμφωνηθεί κατά την εγγραφή της υπηρεσίας. Οι flexible και οι on demand υπηρεσίες χρησιμοποιούν ρητή κλήση. Είναι φανερό πως, τόσο οι παράμετροι που περιλαμβάνονται στην αίτηση κλήσης όσο και η ερμηνεία τους διαφέρουν ανάλογα με τον τύπο ενεργοποίησης της υπηρεσίας.

Η ενεργοποίηση της υπηρεσίας γίνεται με τη δημιουργία και εφαρμογή των κατάλληλων κανόνων ταξινόμησης και ρύθμισης της κίνησης (βλέπε ενότητα A.1.4.2.2.1) στον ακραίο κόμβο εισόδου έτσι ώστε η ροή του χρήστη να αναγνωρίζεται, να ρυθμίζεται ώστε να εξασφαλίζεται πως δεν γίνεται παράβαση των υποχρεώσεων του χρήστη και τέλος να σημειώνεται (marked) με το DSCP που μέσα στο δίκτυο θα της εξασφαλίσει τη συμφωνημένη ποιότητα.

B.1.2.1.2 Μοντέλο Περιεχομένου

Η διαπραγμάτευση ενός συμβολαίου επιπέδου μεταφοράς κατά τη φάση της εγγραφής γίνεται χρησιμοποιώντας τη Δομή Εγγραφής Υπηρεσίας (Service Subscription Structure – SSS). Η τελική έκδοση του εγγράφου αποτελεί και το συμβόλαιο υπηρεσίας, ή αλλιώς το SLA σύμφωνα με τη DiffServ ορολογία.



Σχήμα B-4: Service Subscription Structure

Μια υπηρεσία μεταφοράς προσδιορίζεται πλήρως από ένα έγγραφο SSS (βλέπε Σχήμα B-4) το οποίο, πέρα από τα τεχνικά χαρακτηριστικά της υπηρεσίας, περιλαμβάνει πληροφορία για:

- την πιστοποίηση του πελάτη (*subscriber info*): πληροφορία σχετικά με τη νομική οντότητα του πελάτη, χρησιμοποιείται κατά τη φάση εγγραφής από τη διαδικασία εξακρίβωσης γνησιότητας.
- τους δικαιούχους χρήστες (*user info*): για κάθε χρήστη που δικαιούται να χρησιμοποιήσει την υπηρεσία στοιχεία πιστοποίησης που χρησιμοποιούνται από τη διαδικασία εξακρίβωσης γνησιότητας κατά τη φάση κλήσης.
- τη μέθοδο κλήσης (*invocation info*): πρόκειται για τον τύπο ενεργοποίησης βάσει του οποίου καθορίζεται και εάν χρησιμοποιείται ρητή ή έμμεση μέθοδος κλήσης (βλέπε ενότητα B.1.2.1.1.2) και παίρνει τις τιμές permanent, on-demand ή flexible. Στην περίπτωση ρητής κλήσης προσδιορίζεται και το πρωτόκολλο σηματοδότησης που χρησιμοποιείται (π.χ. RSVP).
- το βαθμό διαθεσιμότητας (*grade of service*): δεν εφαρμόζεται σε permanent υπηρεσίες ενώ για τις μεν flexible υπηρεσίες αντιστοιχεί σε ζεύγη πιθανότητας επιτυχούς ενεργοποίησης ανά ποσοστό ενεργοποίησης της υπηρεσίας και για υπηρεσίες on demand σε ελάχιστο αριθμό ταυτόχρονων ενεργοποιήσεων, μέγιστο αριθμό ταυτόχρονων ενεργοποιήσεων και πιθανότητα επιτυχούς ενεργοποίησης μεταξύ ελάχιστου και μέγιστου αριθμού.
- το χρονοδιάγραμμα χρήσης (*schedule*): προσδιορίζει το έτος, τους μήνες, τις ημέρες της εβδομάδας και τις ώρες της ημέρας κατά τις οποίες η υπηρεσία πρέπει να είναι διαθέσιμη.

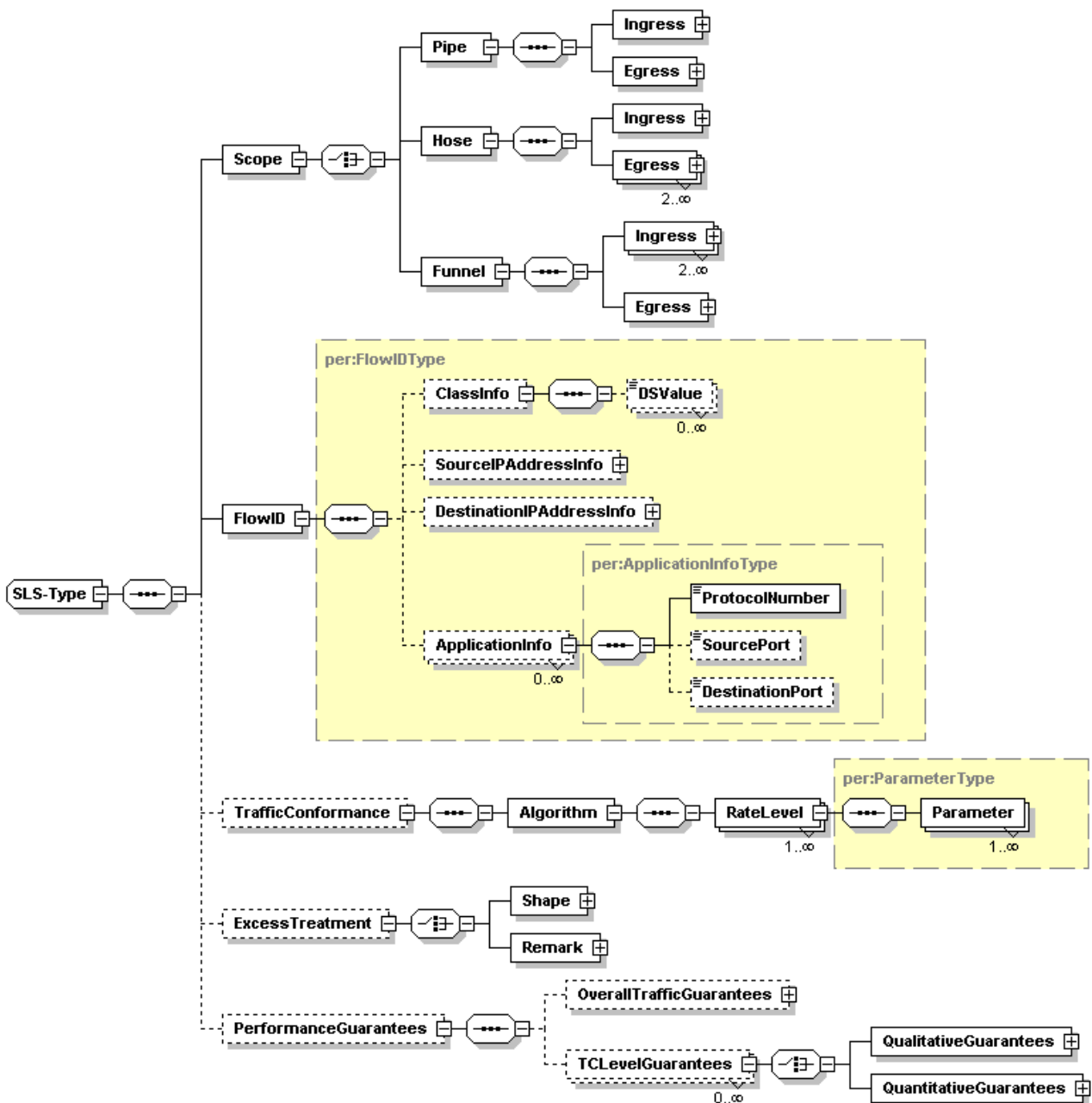
Στοιχεία σχετικά με την τιμολόγηση και τη χρέωση της υπηρεσίας δεν συμπεριλαμβάνονται καθώς δεν εξετάζονται στην παρούσα εργασία.

Οι τεχνικές παράμετροι της υπηρεσίας, σύμφωνα με τη DiffServ ορολογία, περιλαμβάνονται στο SLS. Σε μια υπηρεσία σύνδεσης (*connectivity service*) μπορεί να εμπλέκονται περισσότεροι από δύο τόποι (*sites*). Κάθε μοναδικά προσδιορίσιμη συσχέτιση τόπων σε μια υπηρεσία σύνδεσης ονομάζεται μέλος σύνδεσης (*connectivity leg*). Θεωρούμε πως κάθε μέλος σύνδεσης μονής φοράς (*unidirectional connectivity leg*) που συνθέτει μια υπηρεσία αντιστοιχεί και περιγράφεται από ένα SLS. Μία υπηρεσία VoIP για παράδειγμα περιγράφεται από δύο συμμετρικά SLSs, ένα για κάθε φορά της σύνδεσης μεταξύ των δύο ομιλητών.

Το SSS, αποτελούμενο από τα ως άνω μη τεχνικά χαρακτηριστικά και το σύνολο των ανά μέλος SLSs κάθε υπηρεσίας, δίνει μια πλήρη περιγραφή της υπηρεσίας. Το περιεχόμενο του SLS, η σύνταξη και η σημασιολογία των παραμέτρων που το αποτελούν περιγράφονται στην *SLS φόρμα* (SLS Template – SLS-T) των [SLS-T] (βλέπε Σχήμα B-5).

Οι παράμετροι που αποτελούν το SLS οργανώνονται στις ομάδες πεδίου εφαρμογής (*scope*), ταυτοποίησης ροής (*flow identification*), χαρακτηριστικών κίνησης (*traffic conformance*), μεταχείριση πλεονάζουσας κίνησης (*excess treatment*) και εγγυήσεων ποιότητας (*performance guarantees*).

Το *scope* καθορίζει τα σημεία εισόδου και εξόδου των πακέτων του πελάτη στην επικράτεια του παροχέα, τα σημεία δηλαδή και την τοπολογία εφαρμογής της υπηρεσίας. Καθορίζει δηλαδή, μέσω των κόμβων εισόδου (*ingress*) και εξόδου (*egress*), το συγκεκριμένο *unidirectional connectivity leg*. Κάθε σημείο προσδιορίζεται μοναδικά, είτε άμεσα δίνοντας τη διεύθυνση του ακραίου κόμβου, είτε έμμεσα δίνοντας τη διεύθυνση του εκτός δικτύου site που τελικά συμμετέχει στην υπηρεσία σύνδεσης και βρίσκοντας μέσω της διαδικασίας μετάφρασης και αντιστοίχισης την διεύθυνση του ακραίου κόμβου στον οποίο είναι προσαρτημένο. Ανάλογα με τη σχέση *ingress* προς *egress* η σύνδεση μπορεί να είναι τύπου *σαλήνα* (*pipe*) για 1:1, τύπου *μάνικας* (*hose*) για 1:N και τύπου *χοάνης* (*funnel*) για N:1.



Σχήμα B-5: Service Level Specification

Το flow identification χρησιμοποιείται για την αναγνώριση των πακέτων για τα οποία ισχύει η υπηρεσία. Ορίζει τις τιμές των πεδίων της IP επικεφαλίδας (DSCP, διεύθυνση πηγής, κ.λπ.) που προσδιορίζουν μοναδικά τη ροή στον ταξινομητή (classifier).

Το traffic conformance καθορίζει τα χαρακτηριστικά που πρέπει να πληροί η ροή του χρήστη (traffic profile) ώστε να λάβει τις εγγυήσεις ποιότητας που περιγράφονται από την ομάδα παραμέτρων performance guarantees. Βάσει της μεθόδου και των παραμέτρων που καθορίζονται εδώ, η λειτουργία του μετρητή (meter) στον κόμβο εισόδου κατατάσσει τα πακέτα της ροής σε συμμορφούμενα (in-profile) και μη συμμορφούμενα (out-of-profile), ή εναλλακτικά σε πολλαπλά διαφορετικά επίπεδα συμμόρφωσης. Τα out-of-profile πακέτα στη μία περίπτωση ή τα πακέτα του τελευταίου επιπέδου στην περίπτωση πολλαπλών επιπέδων αποτελούν την *πλεονάζουσα κίνηση* (excess traffic).

Το excess treatment καθορίζει την ενέργεια ρύθμισης επί της κίνησης που έχει χαρακτηριστεί πλεονάζουσα από την εφαρμογή των κανόνων του traffic conformance. Η πλεονάζουσα κίνηση μπορεί να υποστεί

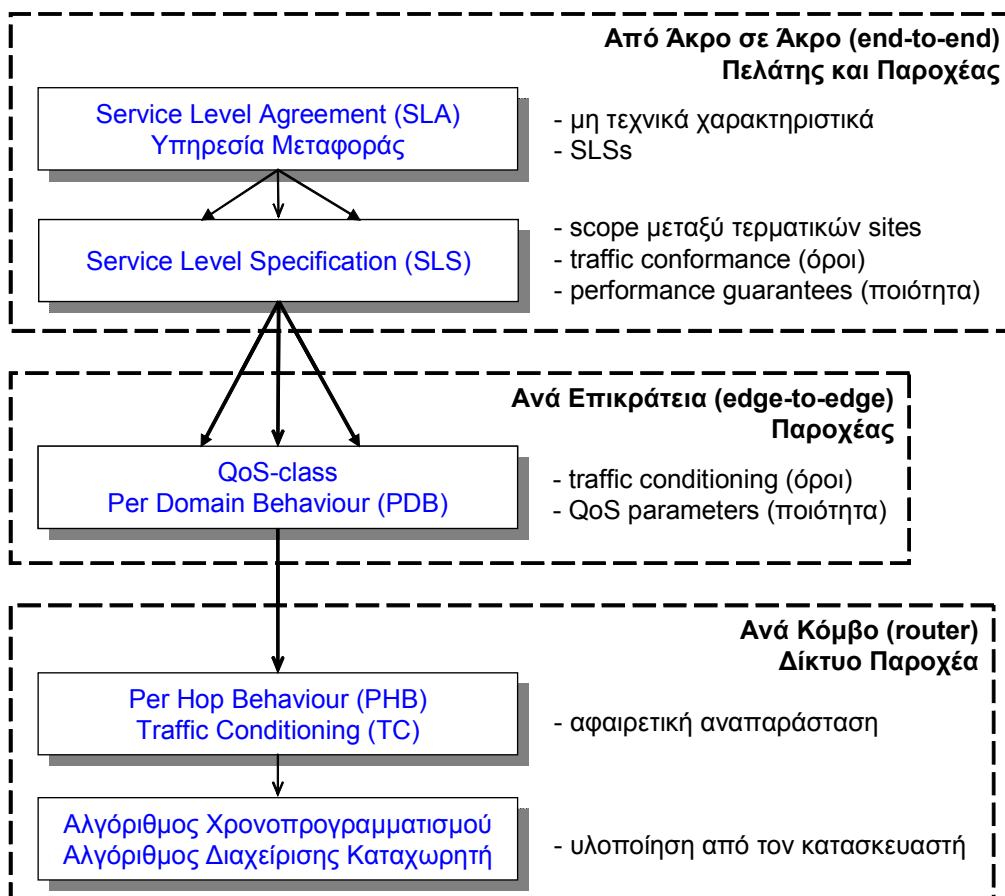
διαγραφή (dropping), μορφοποίηση (shaping) ή επανασημείωση (re-marking) με DSCP που αντιστοιχεί σε κατώτερη κλάση υπηρεσίας.

Το performance guarantees καθορίζει την ποιότητα υπηρεσίας που λαμβάνουν τα πακέτα της ροής του χρήστη ανάλογα με το επίπεδο συμμόρφωσής τους. Ορίζεται συνολικά ο ελάχιστος ρυθμός ροής (throughput) και για κάθε επίπεδο συμμόρφωσης η μέγιστη καθυστέρηση (delay), η διακύμανση καθυστέρησης (jitter) και το ποσοστό απώλειας (loss). Κάποιες από αυτές τις παραμέτρους είναι δυνατό να εκφράζονται με ποσόστωση (quantile) αντί για απόλυτη τιμή. Τέλος, οι τιμές των παραμέτρων μπορεί να είναι ποσοτικές (quantitative) ή ποιοτικές (qualitative), ανάλογα με τον αν εκφράζονται αριθμητικά ή περιγραφικά χρησιμοποιώντας προδιαγεγραμμένες βαθμωτές τιμές, π.χ. χρυσή/ασημένια/χάλκινη ή υψηλή/μέτρια/χαμηλή, η μετάφραση των οποίων εξαρτάται από τον παροχέα και γίνεται συνήθως σχετικά και όχι απόλυτα, π.χ. υψηλή είναι δύο φορές καλύτερη από τη μέτρια.

Όλες οι λεπτομέρειες και η ακριβής σύνταξη του SLS-T και του SSS δίνονται στο [TEQ-SRV].

B.1.2.1.3 Ιεραρχικό Δομικό Μοντέλο

Στο ιεραρχικό δομικό μοντέλο που παρουσιάζεται παρακάτω (βλέπε Σχήμα B-6), η υπηρεσία με εγγυήσεις ποιότητας αναλύεται και αντιστοιχίζεται στα δομικά στοιχεία ποιότητας υπηρεσίας του δικτύου [GVT02]. Κάθε κατώτερο επίπεδο παρέχει κάποιες δυνατότητες ποιότητας υπηρεσίας στα ανώτερα επίπεδα και αντιστρόφως, κάθε ανώτερο επίπεδο χρησιμοποιεί τις πιο στοιχειώδεις δυνατότητες των υποκειμένων επιπέδων για να σχηματίσει περισσότερο πολύπλοκες συσχετίσεις και δυνατότητες ποιότητας υπηρεσίας.



Σχήμα B-6: ιεραρχικό δομικό μοντέλο υπηρεσιών

Το στοιχειώδες δομικό στοιχείο παροχής ποιότητας υπηρεσίας στην αρχιτεκτονική DiffServ είναι το PHB που περιγράφει τα χαρακτηριστικά της ποιότητας που παρέχει ο δρομολογητής με τρόπο ανεξάρτητο της υποκείμενης υλοποίησης. Αντίστοιχα με την παρεχόμενη υπηρεσία στο δίκτυο κορμού που αναπαρίσταται από το PHB, οι κανόνες και οι απαραίτητες ρυθμίσεις της κίνησης του χρήστη που επιβάλλονται στους κόμβους εισόδου για να διασφαλιστούν οι υποχρεώσεις του χρήστη, αναπαρίσταται από τις δομές του *ρυθμιστή κίνησης* (Traffic Conditioner – TC). Διαφορετικοί αλγόριθμοι χρονοπρογραμματισμού (π.χ. WFQ, CBQ) και διαχείρισης καταχωρητή επιλέγονται για την υλοποίηση ενός PHB ανάλογα με τις ειδικές δυνατότητες του υλικού (hardware) του κάθε δρομολογητή. Τα δύο αυτά επίπεδα αναπαράστασης της ποιότητας και υλοποίησης στα στοιχεία του δικτύου, αποτελούν και τη βάση του ιεραρχικού δομικού μοντέλου.

Τα ανώτερα επίπεδα του δομικού μοντέλου υπηρεσιών αναπαριστούν την υπηρεσία στην επικοινωνία μεταξύ πελάτη και παροχέα. Στην αρχιτεκτονική DiffServ, οι παράμετροι που περιγράφουν τις διαφορετικές πλευρές των δεσμεύσεων του παροχέα και των υποχρεώσεων του πελάτη μιας υπηρεσίας σύνδεσης επιπέδου μεταφοράς σχηματίζουν τη δομή SLA, ενώ το υποσύνολο των παραμέτρων που αφορούν τα τεχνικά χαρακτηριστικά ποιότητας υπηρεσίας σχηματίζει τη δομή SLS. Το ακριβές περιεχόμενο των SLA και SLS και η σύνδεση μεταξύ τους και με το επίπεδο των PHBs παραμένουν ανοικτά θέματα στην προτυποποίηση της αρχιτεκτονικής DiffServ.

Στην προηγούμενη ενότητα περιγράψαμε το περιεχόμενο και τη σύνδεση μεταξύ SLA και SLS. Η αντιστοίχιση μεταξύ της παρεχόμενης ποιότητας όπως περιγράφεται στο SLS-T και της ποιότητας που παρέχεται με ένα PHB σε επίπεδο δρομολογητή γίνεται μέσω της *κλάσης ποιότητας υπηρεσίας* (QoS-class).

Τα PHBs αναπαριστούν τη σχετική διαφοροποίηση της ποιότητας υπηρεσίας σε επίπεδο δρομολογητή. Η τελική τιμή της ποιότητας εξαρτάται από την ένταση της κίνησης για αυτό το PHB που με τη σειρά της ρυθμίζεται από τα TCs των ακραίων κόμβων. Η QoS-κλάση καλύπτει αυτές τις συσχετίσεις και αναπαριστά την αριθμητικώς καθορισμένη ποιότητα υπηρεσίας που είναι δυνατό να παρέχεται από άκρο σε άκρο σε επίπεδο επικράτειας για μια κλάση κίνησης. Η ποιότητα εκφράζεται από την μέγιστη από άκρο σε άκρο καθυστέρηση (delay), το ποσοστό απώλειας (loss) και προαιρετικά τη διακύμανση της καθυστέρησης (jitter) και είναι δυνατό να προσδιορίζεται με απόλυτη τιμή ή με ποσόστωση (quantile). Η κλάση κίνησης εσωτερικά σε μια επικράτεια DiffServ αναγνωρίζεται από το ή τα DSCPs. Έτσι, η QoS-κλάση ορίζεται από τις παραμέτρους:

Ordered Aggregate (OA), delay, loss, [jitter]

όπου το OA αναπαριστά την ομάδα των PHBs (συγκεκριμένα των αντίστοιχων BAs) που υπόκεινται σε ένα περιορισμό διάταξης [DS-TERM], καθορίζει κατ' επέκταση την κλάση της κίνησης που θα εξυπηρετηθεί με την συγκεκριμένη ποιότητα και μπορεί να πάρει τις τιμές EF, AF1, AF2, AF3, AF4 ή BE.

Οι παράμετροι ποιότητας μπορούν θεωρητικά να πάρουν οποιαδήποτε τιμή. Το πεπερασμένο σύνολο των QoS-κλάσεων που υποστηρίζει ένας παροχέας δημιουργείται καθορίζοντας συγκεκριμένες τιμές για το delay και το loss. Η επιλογή αυτών των τιμών είναι καθοριστικής σημασίας εφόσον οι υπηρεσίες που είναι σε θέση να υποστηρίξει ο παροχέας οικοδομούνται βάσει των QoS-κλάσεων που υποστηρίζει στο δίκτυό του. Οι τιμές των παραμέτρων ποιότητας που είναι εφικτές καθορίζονται από τις δυνατότητες του εξοπλισμού και της τοπολογίας του δικτύου. Δεδομένου αυτού του περιορισμού, η εξασφάλιση της ποιότητας των QoS-κλάσεων όπως ορίζονται από τον παροχέα προϋποθέτει την εφαρμογή των κατάλληλων PHBs και μηχανισμών διαχείρισης της κίνησης (traffic engineering) εντός της επικράτειας και ελέγχου (admission control) και ρύθμισης της κίνησης (traffic conditioning) που αποκτά πρόσβαση στο δίκτυο στους ακραίους κόμβους της επικράτειας.

Επισημαίνουμε πως η έννοια της QoS-κλάσης είναι συγγενής με την DiffServ έννοια της *συμπεριφοράς ανά επικράτεια* (Per Domain Behaviour – PDB) [DS-PDB]. Το PDB εκφράζει την ποιότητα υπηρεσίας που αναμένεται να λάβει μια ξεχωριστή προσδιορίσιμη ομάδα πακέτων από άκρο σε άκρο μιας DiffServ επικράτειας. Σε κάθε PDB αντιστοιχίζονται ένα ή περισσότερα PHBs και Traffic Conditioning κανόνες. Η ομάδα πακέτων προσδιορίζεται από τα DSCP(s) και οι μετρίσιμες παράμετροι που χαρακτηρίζουν την ποιότητα υπηρεσίας πρέπει να μεταφράζονται και να αντιστοιχούν με κάποιο τρόπο στις σχετικές παραμέτρους που χρησιμοποιούνται στα SLSs. Η προδιαγραφή [DS-PDB] παγιώθηκε μετά το πέρας του σχεδιασμού του μοντέλου υπηρεσιών που χρησιμοποιείται στην παρούσα εργασία και γι' αυτό το λόγο εδώ χρησιμοποιείται η συγγενής έννοια της QoS-κλάσης.

Οι QoS-κλάσεις είναι ο συνδετικός κρίκος μεταξύ υπηρεσιών και δικτύου. Η αντιστοίχιση της QoS-κλάσης σε PHBs εξαρτάται από τις επιλογές του παροχέα και τις δυνατότητες του δικτύου. Η αντιστοίχιση των υπηρεσιών σε QoS-κλάσεις γίνεται μέσω των παραμέτρων της ομάδας performance guarantees του SLS-T. Κάθε επίπεδο συμμόρφωσης της κίνησης του χρήστη όπως προκύπτει από το traffic conformance του SLS-T σημειώνεται με διαφορετικό DSCP και λαμβάνει αντίστοιχα διαφορετική ποιότητα η οποία καθορίζεται από τις αντίστοιχες παραμέτρους της ομάδας performance guarantees. Κάθε τέτοιο επίπεδο ποιότητας του SLS αντιστοιχίζεται σε μία από τις προκαθορισμένες QoS-κλάσεις που υποστηρίζει ο παροχέας της υπηρεσίας. Η αντιστοίχιση γίνεται μέσω των παραμέτρων delay και loss.

Ενώ το SLA χρησιμοποιείται για την περιγραφή της υπηρεσίας μεταξύ πελατών και εσωτερικά από τον παροχέα η QoS-κλάση χρησιμοποιείται από τον παροχέα για τη μετάφραση και αντιστοίχιση των υπηρεσιών στις δυνατότητες του δικτύου του. Επιπλέον, ενώ η ποιότητα στο συμβόλαιο υπηρεσίας SLA εκφράζεται με τρόπο ανεξάρτητο προς τις υποκείμενες τεχνολογίες, η QoS-κλάση είναι συνυφασμένη με την τεχνολογία DiffServ καθώς εκφράζεται από τα αντίστοιχα PHBs που την υλοποιούν. Τέλος, η QoS-κλάση περιορίζεται στα όρια της επικράτειας του παροχέα ενώ η ποιότητα που περιγράφεται σε ένα συμβόλαιο υπηρεσίας ισχύει μεταξύ των τερματικών κόμβων της υπηρεσίας που, στη γενική περίπτωση, συνδέονται μέσω περισσότερων της μίας επικράτειας.

Η παροχή ποιότητας υπηρεσίας επομένως συνοψίζεται (α) στον καθορισμό των QoS-κλάσεων, (β) στη διαστασιοποίηση του δικτύου μέσω των λειτουργιών διαχείρισης της κίνησης (traffic engineering) έτσι ώστε να εξασφαλίζεται η ποιότητα των QoS-κλάσεων, δεδομένων πάντα των σχετικών περιορισμών για την ένταση της κίνησης, και τέλος (γ) στην αντιστοίχιση των υπηρεσιών στις κατάλληλες QoS-κλάσεις εξασφαλίζοντας παράλληλα μέσω των λειτουργιών διαχείρισης υπηρεσιών (service management) πως δεν παραβιάζονται οι περιορισμοί στην ένταση της κίνησης και κάνοντας ταυτόχρονα βέλτιστη αξιοποίηση των πόρων.

Περισσότερα για τις λειτουργίες παροχής ποιότητας υπηρεσίας και τη συσχέτιση μεταξύ τους αναφέρονται στην επόμενη ενότητα.

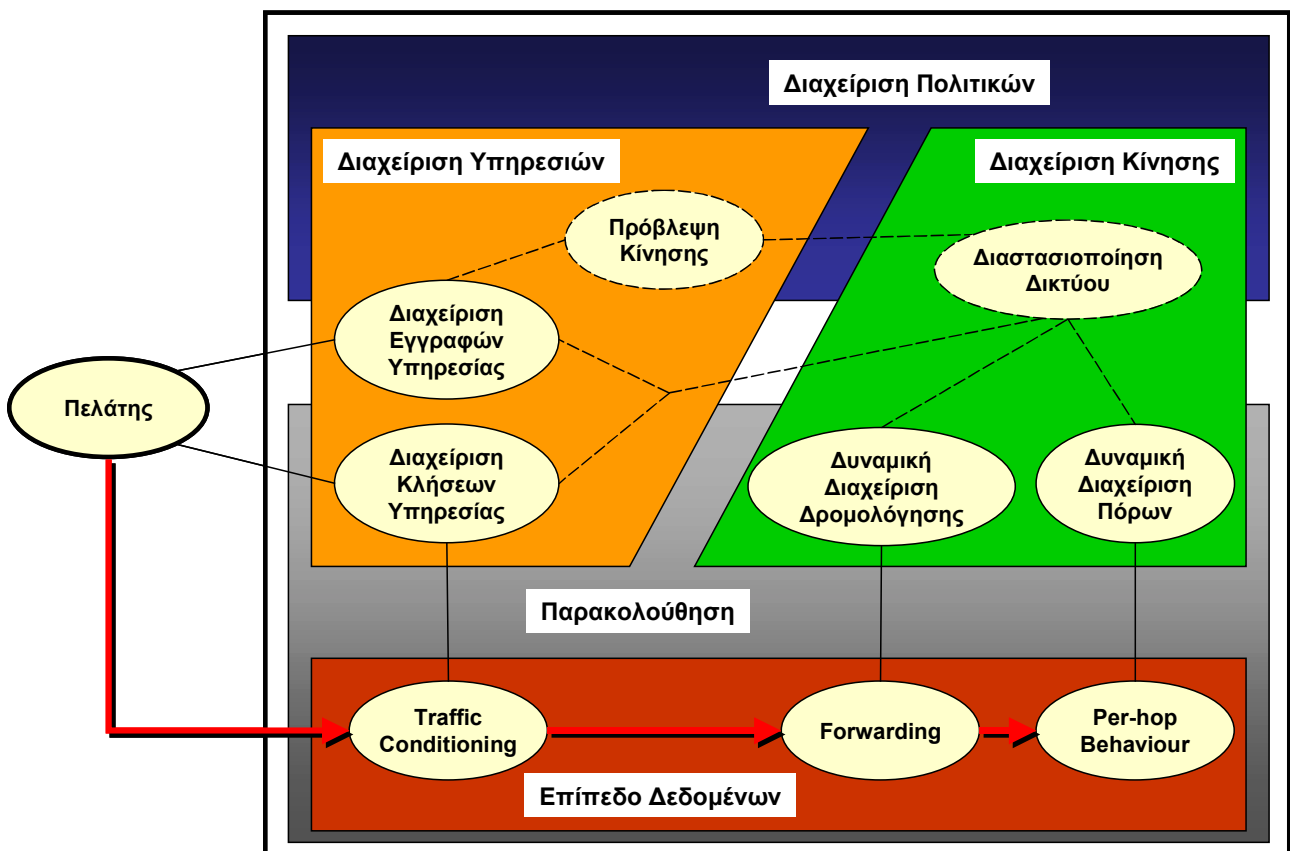
B.1.2.2 Λειτουργικό Μοντέλο

Το βασικό μοντέλο ανάδρασης (βλέπε ενότητα B.1.1.2) ορίζει πως οι πόροι του δικτύου κατανέμονται με a priori partitioning. Αναφέρεται ακόμα πως μετρήσεις του φόρτου τοπικά και περιοδικές μετρήσεις της ποιότητας από άκρο σε άκρο χρησιμοποιούνται για να κατευθύνουν τους μηχανισμούς ανάδρασης του ελέγχου εισόδου κίνησης. Από το διαδικαστικό μοντέλο υπηρεσιών (βλέπε ενότητα B.1.2.1.1) προκύπτει πως η διαδικασία ενεργοποίησης μιας υπηρεσίας εμπεριέχει δύο φάσεις, τη φάση εγγραφής και τη φάση κλήσης, σε καθεμία από τις οποίες ασκείται έλεγχος εισόδου κίνησης. Πριν αναπτύξουμε τον προτεινόμενο αλγόριθμο είναι απαραίτητο να απομονώσουμε τη λειτουργία που επιτελεί στο βασικό μοντέλο, να

καθορίσουμε τις παραδοχές σχετικά με τις συγγενείς και βοηθητικές προς τον αλγόριθμο λειτουργίες και τις αλληλεπιδράσεις μεταξύ των λειτουργιών.

Το λειτουργικό μοντέλο που υιοθετούμε εξυπηρετεί τις ανάγκες και είναι συμβατό με τις επιταγές του βασικού μοντέλου ανάδρασης, είναι επίτευγμα δε του ερευνητικού έργου TEQUILA² (Traffic Engineering for Quality of Service in the Internet, at Large Scale). Στα πλαίσια του έργου TEQUILA το λειτουργικό μοντέλο υλοποιήθηκε, επαληθεύτηκε και δοκιμάστηκε με αξιόλογα αποτελέσματα σε εμπορικούς και πειραματικούς δρομολογητές αλλά και χρησιμοποιώντας προσομοίωση [TEQ-RES]. Πλήρης και αναλυτική περιγραφή όσων λειτουργιών σκιαγραφούνται ή αναφέρονται μόνο εν συντομία παρακάτω μπορεί να βρεθεί στο [TEQ-ARCH].

B.1.2.2.1 Συνοπτική Περιγραφή



Σχήμα B-7: λειτουργικό μοντέλο ολοκληρωμένου συστήματος παροχής ποιότητας

Το λειτουργικό μοντέλο (βλέπε Σχήμα B-7) καλύπτει όλο το φάσμα των λειτουργιών για την παροχή ποιότητας υπηρεσίας σε μια επικράτεια. Η βασική ιδέα είναι πως η επικοινωνία με τον πελάτη και η κατάλληλη ρύθμιση των λειτουργιών του επιπέδου δεδομένων (data plane) γίνεται από τις δύο βασικές ομάδες λειτουργιών, τις λειτουργίες διαχείρισης υπηρεσιών (service management) και τις λειτουργίες διαχείρισης κίνησης (traffic engineering), οι οποίες υποστηρίζονται από τις λειτουργίες παρακολούθησης (monitoring) και καθοδηγούνται μέσω των λειτουργιών διαχείρισης πολιτικών (policy management).

Οι λειτουργίες του επιπέδου δεδομένων προκύπτουν άμεσα από τις προδιαγραφές της αρχιτεκτονικής DiffServ. Η λειτουργία Traffic Conditioning στους ακραίους κόμβους υλοποιεί τις λειτουργίες ταξινόμησης

² Βλέπε www.ist-tequila.org

(classification), *μέτρησης* (metering), *σημείωσης* (marking), *μορφοποίησης* (shaping) και *διαγραφής* (dropping) που καθορίζουν ποια πακέτα και μετά από ποιες ρυθμίσεις και για ποια κλάση ποιότητας, δηλαδή με ποιο DSCP, διοχετεύονται στο δίκτυο (βλέπε ενότητα A.1.4.2.2.1). Η λειτουργία Per Hop Behaviour σε κάθε κόμβο του δικτύου υλοποιεί τις ιδιότητες κάθε κλάσης ποιότητας χρησιμοποιώντας τους εγγενείς αλγόριθμους χρονοπρογραμματισμού και διαχείρισης καταχωρητή και σύμφωνα με την κατανομή των πόρων κάθε φυσικής ζεύξης ανά κλάση όπως καθορίζεται από υπερκείμενες λειτουργίες. Η λειτουργία προώθησης (forwarding) δρομολογεί τα πακέτα στον κατάλληλο προσκείμενο δρομολογητή προς τον προορισμό τους. Η διαμόρφωση των μονοπατιών και των κανόνων δρομολόγησης γίνεται από την υπερκείμενη λειτουργία της δρομολόγησης (routing), μέρος των λειτουργιών διαχείρισης της κίνησης.

Οι λειτουργίες παρακολούθησης υλοποιούν τους μηχανισμούς παθητικής μέτρησης (passive measurement) τοπικά στους κόμβους και ενεργητικής μέτρησης (active measurement) με την αποστολή δοκιμαστικών πακέτων από άκρο σε άκρο. Οι στοιχειώδεις μηχανισμοί μέτρησης πραγματοποιούν τη συλλογή των πρωτογενών αποτελεσμάτων (raw data) και περιβάλλονται από προηγμένους μηχανισμούς επεξεργασίας των αποτελεσμάτων και αποστολής αναφορών (reporting). Η επεξεργασία των αποτελεσμάτων περιλαμβάνει εξαγωγή στατιστικών μεγεθών όπως μέσος όρος, διακύμανση, EWMA μέσος όρος κ.λπ. ανά καθοριζόμενα χρονικά διαστήματα. Η αποστολή αναφορών μπορεί να γίνεται περιοδικά ή με ειδοποιήσεις (notifications) επί προκαθορισμένων γεγονότων (events) όπως η υπέρβαση ενός κατωφλιού (threshold crossing). Τα μεγέθη (metrics) που μετρώνται είναι η καθυστέρηση, η απώλεια, ο ρυθμός εξυπηρέτησης κ.λπ. ενώ η μέτρηση μπορεί να γίνεται ανά ροή, ανά κλάση, ανά μονοπάτι, κ.λπ. Οι λειτουργίες μέτρησης είναι υποστηρικτικές των λειτουργιών διαχείρισης κίνησης και διαχείρισης υπηρεσιών. Όλες οι λεπτομέρειες για τις δυνατότητες και τις προδιαγραφές των λειτουργιών μέτρησης δίνονται στα [AVJ01,TEQ-MON].

Οι λειτουργίες διαχείρισης πολιτικών συμβάλλουν στην αυτόματη μετάφραση, διάδοση και επιβολή των πολιτικών του παροχέα που επηρεάζουν και καθοδηγούν τη λειτουργία του όλου συστήματος. Οι παράμετροι πολιτικής (policy parameters) που κάνουν νόημα και υποστηρίζονται από κάθε λειτουργία διαχείρισης κίνησης και διαχείρισης υπηρεσιών είναι συνυφασμένες με τις προδιαγραφές και τη σχεδίαση των λειτουργιών. Οι πολιτικές του παροχέα ωστόσο είναι ανεξάρτητες από τη διάρθρωση του υποκείμενου συστήματος που τις υλοποιεί και κατ' επέκταση των λειτουργιών που το αποτελούν και ως εκ τούτου διατυπώνονται σε ένα ανώτερο επίπεδο. Η αυτόματη μετάφραση και διάδοση των πολιτικών του παροχέα που εκφράζουν την επιθυμητή συμπεριφορά του συστήματος σε τιμές των εκάστοτε παραμέτρων πολιτικής που ρυθμίζουν την συμπεριφορά των λειτουργιών διαχείρισης επιτρέπει την ευέλικτη προσαρμογή του συστήματος στο περιβάλλον και στις επιχειρηματικές επιδιώξεις του παροχέα.

Οι λειτουργίες traffic engineering και service management είναι και οι κατ' ουσία υπεύθυνες για την παροχή ποιότητας υπηρεσίας. Ο έλεγχος εισόδου κίνησης είναι μέρος του service management (βλέπε ενότητα A.1.4.3.2.2) και πραγματοποιείται σε δύο επίπεδα σύμφωνα με το διαδικαστικό μοντέλο υπηρεσιών (βλέπε ενότητα B.1.2.1.1), από τις αντίστοιχες λειτουργίες *διαχείρισης αιτήσεων εγγραφής* (service subscription management) και *διαχείρισης αιτήσεων κλήσης* υπηρεσίας (service invocation management).

Οι λειτουργίες traffic engineering είναι υπεύθυνες για την κατανομή των πόρων (βλέπε ενότητα A.1.4.3.2.1). Το βασικό feedback-based μοντέλο ελέγχου εισόδου κίνησης (βλέπε ενότητα B.1.1.2) υπαγορεύει την α priori κατανομή των πόρων και την ευέλικτη δυναμική προσαρμογή του συστήματος στις πραγματικές συνθήκες. Έτσι, κατά την αρχικοποίηση του συστήματος στο παρασκήνιο (offline), η λειτουργία *πρόβλεψης κίνησης* (traffic forecast) βάσει της πληροφορίας που βρίσκεται στα συμβόλαια εγγραφής υπηρεσίας εκτιμά την αναμενόμενη κίνηση με δικαίωμα πρόσβασης στο δίκτυο. Οι εκτιμήσεις αυτές τροφοδοτούνται στη λειτουργία *διαστασιοποίησης του δικτύου* (network dimensioning) η οποία κατανέμει κατάλληλα τους

πόρους και το σύστημα είναι έτοιμο να υποδεχθεί την κίνηση. Η δυναμική προσαρμογή στις πραγματικές συνθήκες κατά τη λειτουργία του συστήματος (runtime) που χαρακτηρίζει το feedback-based μοντέλο ελέγχου εισόδου, επιτάσσει και τη δυναμική προσαρμογή της κατανομής των πόρων έτσι ώστε να αφομοιώνονται οι αποκλίσεις που παρατηρούνται. Οι σχετικές λειτουργίες *δυναμικής διαχείρισης δρομολόγησης* (dynamic route management) και *δυναμικής διαχείρισης πόρων* (dynamic resource management) αποτελούν το δυναμικό μέρος του traffic engineering ενώ η λειτουργία του network dimensioning υλοποιεί το στατικό traffic engineering (βλέπε ενότητα A.1.4.3.2.1).

Είναι προφανές πως έντονες ή/και επίμονες αποκλίσεις πλήττουν την αποτελεσματικότητα και την απόδοση του συστήματος, καθιστώντας αναγκαία τη διεξαγωγή νέων εκτιμήσεων της κίνησης και νέα κατανομή των πόρων τέτοια που να ανταποκρίνεται καλύτερα στις πραγματικές συνθήκες. Η ενορχήστρωση των αλληλεπιδράσεων μεταξύ των λειτουργιών service management και traffic engineering γίνεται στα πλαίσια του *κύκλου πρόβλεψης πόρων* (Resource Provisioning Cycle – RPC). Το RPC και οι εμπλεκόμενες λειτουργίες είναι μέρος της προτεινόμενης λύσης και παρουσιάζονται διεξοδικά στις επόμενες ενότητες.

B.1.2.2.2 Resource Provisioning Cycle

Η ανά πάσα στιγμή ευθυγράμμιση των λειτουργιών κατανομής των πόρων και εξυπηρέτησης της κίνησης που προσφέρεται από τις αιτήσεις υπηρεσιών (offered load), ή αλλιώς η άρρηκτη σύνδεση μεταξύ των λειτουργιών traffic engineering και service management έχει αποδειχθεί πως επιφέρει σημαντικές επιβαρύνσεις καταλήγοντας να είναι απαγορευτικά δαπανηρή σε συστήματα μεγάλης κλίμακας (βλέπε ενότητες A.2.1.2 και A.2.2.2.1). Έτσι προκύπτει ο διαχωρισμός και η αυτονόμηση των λειτουργιών traffic engineering και service management [TPF03].

Οι λειτουργίες traffic engineering και service management δρουν μεν αυτόνομα αλλά δεν είναι δυνατό να δρουν απομονωμένα. Το service management καθορίζει την κίνηση που αποκτά πρόσβαση στους πόρους του δικτύου διαμορφώνοντας τις απαιτήσεις σε πόρους (traffic demand) που το traffic engineering, δεδομένων των φυσικών πόρων, καλείται να καλύψει με βέλτιστο τρόπο. Το traffic engineering από την πλευρά του, δημιουργεί τις βάσεις πάνω στις οποίες στηρίζεται το service management για να εξυπηρετήσει τις αιτήσεις υπηρεσίας με βέλτιστο τρόπο.

Έτσι, αντί του αδιάκοπου άρρηκτου συντονισμού των λειτουργιών traffic engineering και service management, ο συνεχής χρόνος χωρίζεται σε διακριτές χρονικές περιόδους στις οποίες παρατηρείται ομοιογενής συμπεριφορά των χρηστών. Κάθε τέτοια χρονική περίοδος ονομάζεται *χρονική περίοδος πρόβλεψης πόρων* (Resource Provisioning Time Period – RPTP). Για παράδειγμα, ο συνεχής χρόνος μπορεί να διασπάται στο RPTP του Σαββατοκύριακου, στο RPTP των ωρών αιχμής, π.χ. 9:00 με 17:00, και στο RPTP των υπόλοιπων ωρών των εργάσιμων ημερών της εβδομάδας.

Χάρη στην ομοιογένεια της συμπεριφοράς των χρηστών είναι δυνατή η ανά RPTP στατιστική εκτίμηση και περιγραφή της προσφερόμενης κίνησης. Έχοντας θεωρήσει τους φυσικούς πόρους του δικτύου αμετάβλητους, η συμπεριφορά των χρηστών που καθορίζει και την προσφερόμενη κίνηση (offered load) είναι η μόνη ελεώθηρη (uncontrollable) μεταβλητή του περιβάλλοντος στο οποίο λειτουργεί ένα σύστημα παροχής ποιότητας υπηρεσίας.

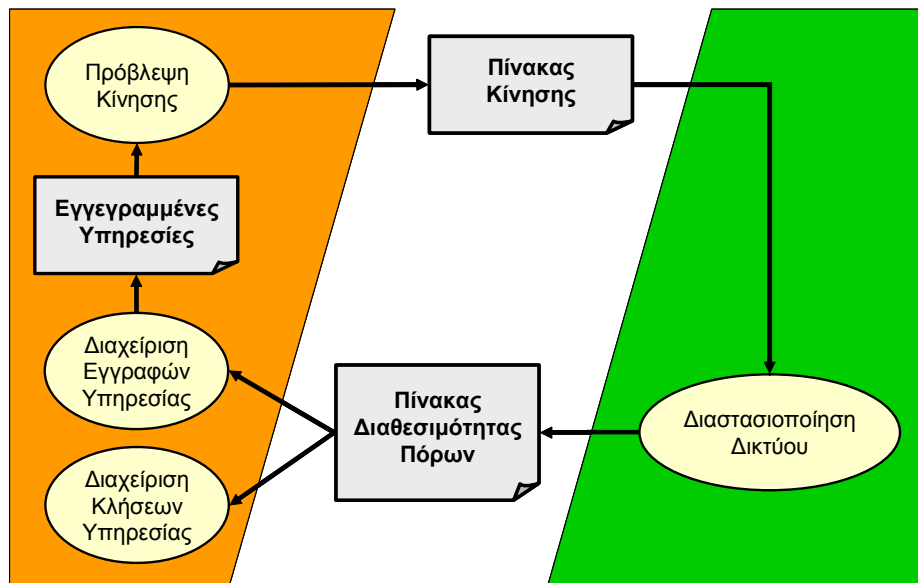
Πριν από την έναρξη ενός RPTP οι λειτουργίες traffic engineering και service management συντονίζουν τις ρυθμίσεις τους έτσι ώστε να εξυπηρετήσουν με βέλτιστο τρόπο την εκτιμώμενη προσφερόμενη κίνηση. Κατά τη διάρκεια του RPTP κάθε λειτουργία αυτόνομα και χωρίς άλλο ρητό συντονισμό προσαρμόζεται, μέσα στα προκαθορισμένα κατά την αρχικοποίηση του RPTP περιθώρια, έτσι ώστε να αφομοιώνει δυναμικά τις στατιστικές διακυμάνσεις της προσφερόμενης κίνησης. Είναι φανερό πως, όσο περισσότερο ομοιογενή

τα χαρακτηριστικά της προσφερόμενης κίνησης, τόσο λιγότερες ή καλύτερα υπολογισμένες οι διακυμάνσεις της από τις a priori υπολογισμένες τιμές της, συνεπώς τόσο καλύτερα εναρμονισμένες οι ενέργειες του service management και του traffic engineering και τόσο πιο ευσταθής η λειτουργία του συστήματος (βλέπε ενότητα A.1.4.3.2.1). Και άρα τόσο πιο σπάνια το service management παίρνει λανθασμένες αποφάσεις, είτε λανθασμένα αποδέχεται αιτήσεις και δημιουργεί συμφόρηση και άρα υποβάθμιση της ποιότητας υπηρεσίας, είτε λανθασμένα απορρίπτει αιτήσεις ενώ υπάρχουν διαθέσιμοι πόροι μειώνοντας την αξιοποίησή τους. Είναι ακόμα φανερό πως όσο μικρότερες οι περιόδους RPTP τόσο μεγαλύτερη η ομοιογένεια κατά τη διάρκειά τους. Ωστόσο, η επιλογή της διάρκειας των RPTPs πρέπει να γίνεται προσεκτικά και λαμβάνοντας υπόψη πως κάθε νέο RPTP επιφέρει νέα διαστασιοποίηση του δικτύου και ανακατανομή των πόρων στους δρομολογητές ολόκληρης της επικράτειας, έργο επίπονο και πολύπλοκο που στις περισσότερες περιπτώσεις σημαίνει πως κάποιες περιοχές χρειάζεται να τεθούν προσωρινά εκτός λειτουργίας.

Πέραν των διακυμάνσεων της συμπεριφοράς των χρηστών που μπορούν να αφομοιωθούν από τη δυναμική προσαρμογή του συστήματος κατά τη διάρκεια του κάθε RPTP, η προσφερόμενη κίνηση είναι δυνατό να παρουσιάζει αποκλίσεις από τις εκτιμώμενες τιμές λόγω σημαντικών αλλαγών είτε στο μέγεθος είτε στη σύνθεση του πληθυσμού των χρηστών. Όταν διαπιστώνονται περισσότερο συχνές ή και έντονες αποκλίσεις της πραγματικής από την προβλεπόμενη προσφερόμενη κίνηση και ιδιαίτερα όταν τείνουν περισσότερο προς μια συγκεκριμένη κατεύθυνση τότε είναι χαρακτηριστική ένδειξη πως οφείλονται σε αλλαγή στον πληθυσμό των χρηστών και πως πλέον η αρχικά προβλεπόμενη προσφερόμενη κίνηση δεν αντιπροσωπεύει την πραγματική συμπεριφορά των χρηστών και πρέπει να επανεκτιμηθεί. Η χρονική περίοδος στην οποία ο πληθυσμός των χρηστών παρουσιάζει ομοιογένεια και για την οποία θεωρούνται έγκυρες οι ανά RPTP εκτιμήσεις της προσφερόμενης κίνησης ονομάζεται *κύκλος πρόβλεψης πόρων* (Resource Provisioning Cycle – RPC). Όταν οι αποκλίσεις που παρατηρούνται υποδεικνύουν σημαντικές αλλαγές στον πληθυσμό τότε οι ανά RPTP εκτιμήσεις της προσφερόμενης κίνησης υπολογίζονται εκ νέου, μαζί με τις αντίστοιχες ρυθμίσεις των λειτουργιών service management και traffic engineering, έτσι ώστε να αντανakλούν τον πληθυσμό όπως έχει διαμορφωθεί, και το σύστημα λέγεται πως μπαίνει σε νέο RPC. Νέο RPC ξεκινά και στην περίπτωση που οι αρχικές εκτιμήσεις αποδεικνύονται άτοπες και δεν επιβεβαιώνονται κατά τη διάρκεια του RPC με αποτέλεσμα να εκδηλώνονται απρόσμενα έντονες αποκλίσεις από τις εκτιμώμενες τιμές. Σε αυτή την περίπτωση τα μαθηματικά μοντέλα που χρησιμοποιούνται στις εκτιμήσεις προσαρμόζονται κατάλληλα βάσει των οποίων εκτιμάται εκ νέου η προσφερόμενη κίνηση για το νέο RPC. Τέλος, ο παροχέας μπορεί περιοδικά να πυροδοτεί την έναρξη νέου RPC έτσι ώστε, έστω και αν το παλιό RPC εξακολουθεί να λειτουργεί ικανοποιητικά, αφομοιώνοντας τις όποιες αλλαγές το νέο RPC να λειτουργήσει ακόμα πιο ικανοποιητικά.

Στο εξής, χωρίς να πλήττεται η γενικότητα της μελέτης, δεχόμαστε πως έχουμε ένα RPTP και αρχικοποίηση του συστήματος με διαφορετικές ρυθμίσεις συμβαίνει μόνο λόγω αλλαγής RPC, πάντοτε δηλαδή μετά από επανεκτίμηση της προσφερόμενης κίνησης.

Οι δομές και η διαδικασία συντονισμού μεταξύ των λειτουργιών service management και traffic engineering πριν από την έναρξη του νέου RPC φαίνονται στο Σχήμα B-8. Η διαδικασία αυτή εκτελείται στο παρασκήνιο (offline) για την αρχικοποίηση του συστήματος και τα αποτελέσματά της χρησιμοποιούνται κατά τη λειτουργία του συστήματος (runtime). Βάσει των καταχωρημένων εγγεγραμμένων υπηρεσιών η λειτουργία πρόβλεψης κίνησης (traffic forecast) εκτιμά την προσφερόμενη κίνηση (traffic demand) από άκρο σε άκρο και ανά QoS-κλάση και σχηματίζει τον *πίνακα κίνησης* (Traffic Matrix – TM). Η εκτιμώμενη για ένα RPC, δηλαδή για μια αρκούντως μεγάλη χρονική περίοδο, προσφερόμενη κίνηση είναι φυσικά στατιστικό μέγεθος και περιγράφεται από μια ελάχιστη και μια μέγιστη τιμή.



Σχήμα B-8: αρχικοποίηση Resource Provisioning Cycle

Στη συνέχεια, η λειτουργία διαστασιοποίησης του δικτύου (network dimensioning) κατανέμει τους φυσικούς πόρους του δικτύου έτσι ώστε να καλύψει με βέλτιστο τρόπο τις απαιτήσεις της προσφερόμενης κίνησης όπως περιγράφεται στο TM. Μαζί με αυτή την αρχική κατανομή των πόρων, κατά την αρχικοποίηση του συστήματος το network dimensioning υπολογίζει και τις τιμές των παραμέτρων που θα χρησιμοποιηθούν κατά τη διάρκεια του RPC από τις δυναμικές λειτουργίες του traffic engineering για την απορρόφηση των διακυμάνσεων της προσφερόμενης κίνησης μεταξύ της ελάχιστης και της μέγιστης τιμής. Βάσει αυτών των στοιχείων το network dimensioning εκτιμά τη διαθεσιμότητα των πόρων (resource availability) από άκρο σε άκρο και ανά QoS-κλάση στο διαστασιοποιημένο δίκτυο και σχηματίζει τον πίνακα διαθεσιμότητας πόρων (Resource Availability Matrix – RAM).

Κατά τη διάρκεια του RPC η κατανομή των πόρων δεν είναι σταθερή αλλά μεταβάλλεται ανάλογα με τις ενέργειες των δυναμικών λειτουργιών του traffic engineering και άρα εξαρτάται από τις τιμές των σχετικών παραμέτρων. Συνεπώς, η διαθεσιμότητα των πόρων στο RAM δεν εκφράζεται σαν ντετερμινιστικό μέγεθος αλλά χρησιμοποιώντας τη λογική των hard-reserved και soft-reserved μεριδίων του βασικού feedback-based μοντέλου (βλέπε ενότητα B.1.1.2). Το RAM αποτελεί με άλλα λόγια τις βάσεις πάνω στις οποίες στηρίζεται το service management για να εξυπηρετήσει τις αιτήσεις υπηρεσίας κατά τη διάρκεια του RPC με βέλτιστο τρόπο.

Όπως αναφέρθηκε προηγουμένως, τα TM και RAM περιέχουν αντίστοιχα το traffic demand και το resource availability από άκρο σε άκρο και ανά QoS-κλάση. Η δομή που χρησιμοποιείται για τη σύνταξή τους είναι ο κορμός κίνησης (Traffic Trunk – TT) και αντιπροσωπεύει μια QoS-κλάση τοποθετημένη σε καθορισμένη από άκρο σε άκρο σύνδεση. Η σύνδεση είναι μονής φοράς, από έναν ακραίο κόμβο εισόδου (ingress) προς ένα ή περισσότερους κόμβους εξόδου (egresses), τύπου σωλήνα (pipe) ή τύπου μάνικας (hose) αντίστοιχα. Το TT περιγράφεται από τις παραμέτρους:

κόμβος εισόδου, κόμβος εξόδου{*}, QoS-κλάση

Με άλλα λόγια, το μέγεθος που αντιστοιχίζεται κάθε φορά σε ένα TT αφορά το σύνολο της κίνησης με κοινές απαιτήσεις ποιότητας που εισέρχονται στο δίκτυο από ένα συγκεκριμένο ακραίο κόμβο και εξέρχονται από έναν άλλο συγκεκριμένο ακραίο κόμβο. Η περίπτωση των περισσοτέρων του ενός κόμβου

εξόδου, δηλαδή της σύνδεσης τύπου hose μεταξύ των ακραίων κόμβων χρησιμεύει στον λογισμό επί των εφαρμογών πολλαπλής αποστολής (multicast) και αναφέρεται για λόγους πληρότητας αλλά δεν εξετάζεται παρακάτω. Η επέκταση των αποτελεσμάτων της παρούσας εργασίας σε συνδέσεις τύπου hose, αν και όχι τετριμμένη, δεν προβλέπεται πολύπλοκη και μπορεί να αποτελέσει θέμα προς περαιτέρω έρευνα. Επισημαίνουμε πως το TT είναι ανεξάρτητο από την τοπολογία του δικτύου κορμού που συνδέει τους ακραίους κόμβους, συγχωνεύοντας έτσι κάθε δυνατό μονοπάτι μεταξύ τους.

Ο διαχωρισμός μεταξύ λειτουργιών traffic engineering και service management είναι εφικτός χάρη στη δομή του TT, ή αλλιώς χάρη στην έννοια της QoS-κλάσης. Σε αντιστοιχία με το δομικό μοντέλο υπηρεσίας (βλέπε ενότητα B.1.2.1.3), όλη η πληροφορία των SLS σε επίπεδο πελάτη και παροχέα περιορίζεται στις λειτουργίες του service management ενώ όλη η πληροφορία της φυσικής τοπολογίας των κόμβων κορμού, των μονοπατιών, των πόρων κάθε κόμβου και της κατανομής τους περιορίζεται στις λειτουργίες του traffic engineering. Το service management βλέπει μόνο πόρους από άκρο σε άκρο ανά TT και το traffic engineering αγνοώντας τις μικρο-ροές δρα επί της ανά TT συσσωματωμένης κίνησης.

Η αποσύνδεση υπηρεσιών και πόρων και η επακόλουθη απλοποίηση των λειτουργιών διαχείρισης είναι, όπως και για την αρχιτεκτονική DiffServ έναντι της αρχιτεκτονικής IntServ, ο μόνος τρόπος για την εφαρμογή ενός συστήματος σε μεγάλη κλίμακα. Η πολυπλοκότητα του συστήματος περιορίζεται τώρα στον αριθμό των TTs, ή αλλιώς κατά μέγιστο στον αριθμό των πιθανών ζευγών ακραίων κόμβων επί τον αριθμό των QoS-κλάσεων, που είναι διαχειρίσιμο μέγεθος ακόμα και σε συστήματα μεγάλης κλίμακας.

Η ακριβής σύνταξη και σημασία του TM και του RAM και η συσχέτιση του τελευταίου με τις παραμέτρους και τις λειτουργίες του δυναμικού traffic engineering αναπτύσσονται με περισσότερη λεπτομέρεια στις ακόλουθες ενότητες.

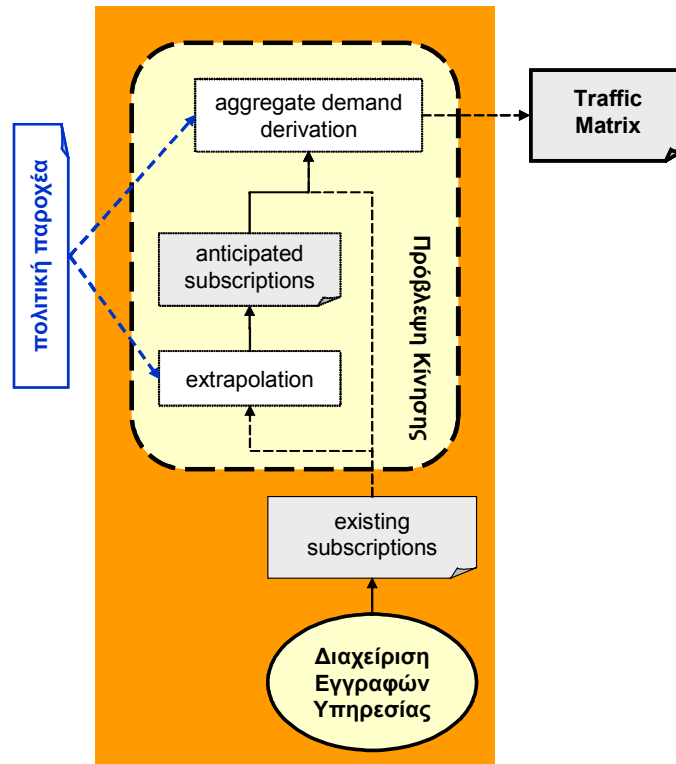
B.1.2.2.3 Αρχικοποίηση – Traffic Forecast και Traffic Matrix

Η λειτουργία πρόβλεψης κίνησης (traffic forecast) καλείται κατά την αρχικοποίηση ενός νέου RPC για την εκτίμηση της προσφερόμενης κίνησης (traffic demand) και το σχηματισμό του Traffic Matrix (βλέπε Σχήμα B-9).

Χάρη στην προϋπόθεση της φάσης εγγραφής πριν τη χρήση της υπηρεσίας γνωρίζουμε πως η κίνηση που θα κληθεί να εξυπηρετήσει το δίκτυο θα είναι μόνο η εξουσιοδοτημένη από τη σύναψη συμβολαίων εγγραφής. Ή αλλιώς, η κίνηση από τις προ της έναρξης του RPC ήδη υπάρχουσες εγγραφές υπηρεσιών (existing subscriptions) και από τις νέες εγγραφές υπηρεσιών που προβλέπεται να γίνουν κατά τη διάρκεια του RPC (anticipated subscriptions).

Οι υπάρχουσες εγγραφές είναι οι αιτήσεις εγγραφής που έχουν γίνει αποδεκτές από τη λειτουργία διαχείρισης εγγραφών υπηρεσίας κατά τους προηγούμενους RPCs. Η πρόβλεψη των νέων εγγραφών γίνεται προεκτείνοντας στο μέλλον (extrapolate) τις τάσεις της αγοράς όπως αποτυπώνονται από την καταγεγραμμένη ιστορία των αιτήσεων εγγραφής. Η διαδικασία πρόβλεψης επηρεάζεται επιπλέον από τις προσδοκίες του παροχέα που εξαρτώνται από την επιχειρησιακή στρατηγική που ακολουθεί όπως για παράδειγμα οι τρέχουσες επενδύσεις σε διαφημιστικές καμπάνιες.

Για το σύνολο των προβλεπόμενων και των υπαρχόντων εγγραφών εξάγεται η προσφερόμενη κίνηση ανά TT (aggregate demand derivation). Εδώ γίνεται η μετάφραση και αντιστοίχιση των υπηρεσιών σε πόρους, δηλαδή η σύνδεση των δομών και της γλώσσας μεταξύ service management και traffic engineering. Τα δεδομένα που τροφοδοτούν τη διαδικασία (input) είναι σε επίπεδο SLS, δηλαδή σε γλώσσα service management, ενώ τα δεδομένα που παράγονται (output) και σχηματίζουν το TM είναι σε επίπεδο TT.



Σχήμα B-9: λειτουργία πρόβλεψης κίνησης

Καθώς η προσφερόμενη κίνηση είναι στατιστικό μέγεθος εκφράζεται από την ελάχιστη και τη μέγιστη τιμή της, TD_{min} και TD_{max} αντίστοιχα. Το TM περιέχει αυτές τις δύο τιμές ανά TT. Το περιεχόμενο μιας γραμμής του TM σε επίπεδο TT και αναλυμένο στις στοιχειώδεις παραμέτρους που το αποτελούν φαίνεται στο Σχήμα B-10.

TT				TD_{min}	TD_{max}
ή					
QoS-κλάση		ingress	egress	TD_{min}	TD_{max}
ή					
OA	loss	delay	ingress	egress	TD_{min} TD_{max}

Σχήμα B-10: περιεχόμενο γραμμής Traffic Matrix

Η μέγιστη προσφερόμενη κίνηση TD_{max} αντιπροσωπεύει το χειρότερο ενδεχόμενο (worst-case) όπου όλες οι υπηρεσίες είναι ενεργές και διοχετεύουν κίνηση με τον μέγιστο επιτρεπτό ρυθμό, συγκεκριμένα με ρυθμό SR_{FS} που αντιστοιχεί στην πλήρη ικανοποίηση των υπηρεσιών (βλέπε ενότητα B.1.1.4). Εάν υπό αυτές τις συνθήκες το δίκτυο δεν παρουσιάζει συμφόρηση, ή αλλιώς εάν η χωρητικότητα του δικτύου επαρκεί για να εξυπηρετήσει την TD_{max} τότε η πιθανότητα παραβίασης της ποιότητας είναι μηδενική για κάθε υπηρεσία, δηλαδή όλες οι υπηρεσίες θα είναι σε κάθε περίπτωση πλήρως ικανοποιημένες.

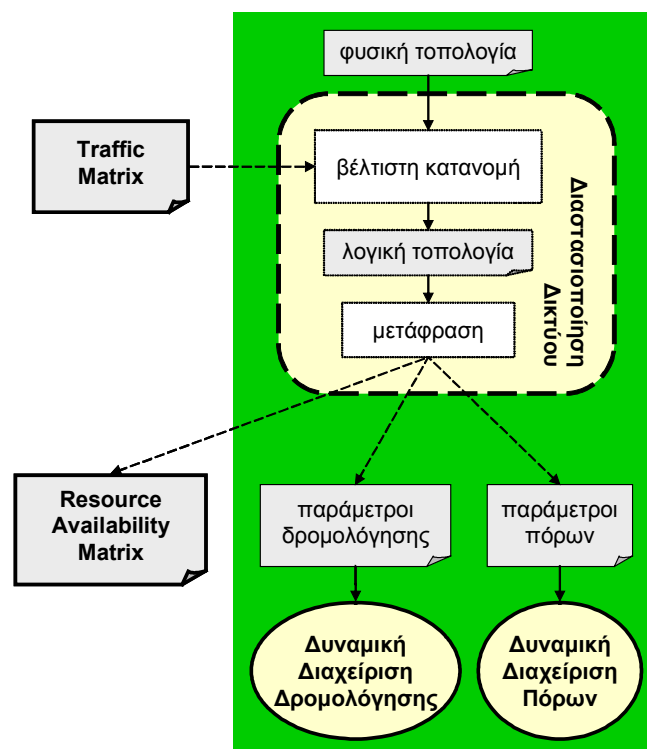
Αντίστοιχα, η ελάχιστη προσφερόμενη κίνηση TD_{min} αντιπροσωπεύει τις κατ' ελάχιστο δεσμεύσεις του παροχέα, δηλαδή το όριο κάτω από το οποίο ο παροχέας δεν μπορεί να εγγυηθεί ούτε καν τον SR_{AS} ρυθμό, δηλαδή υπάρχει μη μηδενική πιθανότητα οι υπηρεσίες να είναι λιγότερο και από σχεδόν ικανοποιημένες.

Οι κατάλληλες πολιτικές του παροχέα καθορίζουν αφενός τον τρόπο υπολογισμού των ρυθμών SR_{FS} και SR_{AS} ανά τύπο υπηρεσίας και αφετέρου τη μέθοδο συνέλιξης (convolution) και συσσωμάτωσής τους (aggregation) για τον υπολογισμό των TD_{max} και TD_{min} αντίστοιχα, ρυθμίζοντας κατάλληλα σε κάθε περίπτωση το κέρδος πολυπλεξίας έναντι των εγγυήσεων υπηρεσίας (βλέπε ενότητα A.2.1.1).

Οι α priori αυτοί υπολογισμοί της προσφερόμενης κίνησης χρησιμοποιούν parameter-based μοντέλα, τα μόνα διαθέσιμα για την περίπτωση. Κατάλληλοι μηχανισμοί επαλήθευσης χρησιμοποιούνται κατά τη διάρκεια του RPC και προσαρμόζουν τις παραμέτρους των μοντέλων όταν διαπιστώνονται σοβαρές αναντιστοιχίες με την πραγματικότητα. Η ακρίβεια των υποκειμένων μοντέλων δεν είναι κρίσιμος παράγοντας για την απόδοση του συστήματος καθώς, αφενός είναι εσωτερικά της λειτουργίας πρόβλεψης κίνησης και δεν επηρεάζουν τις άλλες λειτουργίες και αφετέρου η καλή λειτουργία του συστήματος διασφαλίζεται μέσω της λειτουργικής πολιτικής που έχει σχεδιαστεί ειδικά να αφομοιώνει παρόμοια λάθη (βλέπε ενότητα B.1.1.3).

B.1.2.2.4 Αρχικοποίηση – Network Dimensioning και Resource Availability Matrix

Η λειτουργία διαστασιοποίησης του δικτύου (network dimensioning) καλείται κατά την αρχικοποίηση ενός νέου RPC για την κατανομή των φυσικών πόρων έτσι ώστε να επιτυγχάνεται βέλτιστη εξυπηρέτηση της κίνησης που περιγράφεται στον Traffic Matrix σε όλη την επικράτεια του δικτύου. Το network dimensioning που, όπως και το traffic forecast, μένει ανενεργό κατά τη διάρκεια του RPC, εκτός από αυτή την αρχική βέλτιστη κατανομή των πόρων καθορίζει και τις τιμές των παραμέτρων που χρησιμοποιούν οι δυναμικές λειτουργίες του traffic engineering και οι λειτουργίες του service management κατά τη διάρκεια του RPC για να αφομοιώσουν τις διακυμάνσεις της κίνησης γύρω από τις αρχικά εκτιμώμενες τιμές, η καθεμία με τον καλύτερο και όλες μαζί με συντονισμένο τρόπο (βλέπε Σχήμα B-11).



Σχήμα B-11: λειτουργία διαστασιοποίησης δικτύου

Συγκεκριμένα, με βάση τη φυσική τοπολογία το network dimensioning υπολογίζει σε λογικό επίπεδο τα ανά QoS-κλάση μονοπάτια από άκρο σε άκρο και τους πόρους κάθε ζεύξης κορμού που παραχωρούνται στο κάθε μονοπάτι έτσι ώστε να επιτυγχάνεται βέλτιστη κατανομή της κίνησης του Traffic Matrix στους φυσικούς πόρους του δικτύου και συνεπώς βέλτιστη εξυπηρέτηση.

B.1.2.2.4.1 Βέλτιστη Κατανομή

Από τον ορισμό του Traffic Matrix προκύπτει πως, ενώ είναι βέβαιο πως η πραγματική κίνηση θα ξεπεράσει το TD_{min} , είναι ωστόσο αβέβαιο πόσο παραπάνω από το TD_{min} θα είναι. Επιπλέον, το TD_{max} που αντιπροσωπεύει το worst-case ενδεχόμενο έχει εξαιρετικά μικρή πιθανότητα να συμβεί και άρα δεν μπορεί να χρησιμοποιηθεί σαν απόλυτη τιμή αλλά χρησιμεύει κυρίως σαν βάρος (weight) για την επίλυση των επικαλύψεων μεταξύ των TTs. Συνεπώς το traffic engineering καλείται να καλύψει κατ' ελάχιστο το TD_{min} και το μεγαλύτερο δυνατό κομμάτι του TD_{max} . Εάν το TD_{min} είναι μεγαλύτερο από τη χωρητικότητα του δικτύου τότε είναι βέβαιο πως η ποιότητα υπηρεσίας θα παραβιάζεται συνεχώς και άρα το σύστημα απαιτεί περισσότερους φυσικούς πόρους για να μπορεί να λειτουργήσει. Σε αυτή την περίπτωση ο αλγόριθμος βέλτιστης κατανομής επιστρέφει σφάλμα και ο παροχέας ειδοποιείται σχετικά για να προβεί στις απαραίτητες ενέργειες. Το ποσοστό του TD_{max} επιπλέον του TD_{min} που καλύπτεται από τη δεδομένη χωρητικότητα, ή αλλιώς οι εγγυήσεις για το βαθμό ικανοποίησης από σχεδόν σε εντελώς ικανοποιημένες υπηρεσίες, είναι ανάλογο με το μέγεθος του TD_{max} ή αλλιώς εξαρτάται από το πλήθος των εγγεγραμμένων υπηρεσιών, δηλαδή από τις προηγούμενες ενέργειες του service management. Εφόσον οι προηγούμενες ενέργειες του service management είναι εναρμονισμένες με τους διαθέσιμους φυσικούς πόρους μέσω του RAM του προηγούμενου RPC, το ποσοστό του TD_{max} και οι εγγυήσεις εξαρτώνται από την πολιτική του παροχέα που έχει κατάλληλα ρυθμίσει το service management και άρα δεν απασχολεί το network dimensioning.

Η βέλτιστη κατανομή ακολουθεί την παραπάνω λογική. Στην πρώτη φάση η κίνηση που αντιστοιχεί στα TD_{min} των TTs κατανέμεται στους φυσικούς πόρους έτσι ώστε να εξισορροπηθεί ο φόρτος σε όλη την επικράτεια του δικτύου, δηλαδή το ποσοστό χρήσης (utilisation) να είναι κατά το δυνατό ίσο σε κάθε ζεύξη του δικτύου κορμού. Αφού κατοχυρωθεί αυτή η κατανομή, η δρομολόγηση και η δυνατότητα εξυπηρέτησης των TD_{min} από τη δεδομένη χωρητικότητα, η δεύτερη φάση ασχολείται με την κατανομή του υπολοίπου, TD_{max} μείον TD_{min} . Η χωρητικότητα κάθε ζεύξης που περισσεύει από την πρώτη φάση (residual link capacity) διαμερίζεται στα TTs που έχουν δρομολογηθεί μέσω της ζεύξης χρησιμοποιώντας το TD_{max} σαν βάρος.

Στο Σχήμα B-12 παρουσιάζεται ένα παράδειγμα επίλυσης των επικαλύψεων και τα αντίστοιχα αποτελέσματα της πρώτης και της δεύτερης φάσης της κατανομής για τρία TTs που ανταγωνίζονται σε τρεις διαφορετικές φυσικές τοπολογίες.

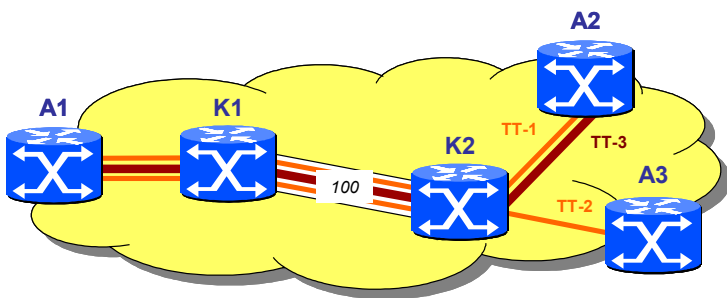
Η πρώτη περίπτωση (II.α) αντιπροσωπεύει ένα απλουστευμένο παράδειγμα όπου ο ανταγωνισμός περιορίζεται σε μία μόνο φυσική ζεύξη K1-K2 και οι υπόλοιπες θεωρείται πως έχουν άπειρη χωρητικότητα. Σε μια περισσότερο ρεαλιστική περίπτωση, ένα TT ανταγωνίζεται για πόρους σε περισσότερες από μία ζεύξεις και οι προαναφερθέντες υπολογισμοί γίνονται επαναληπτικά έτσι ώστε να συνυπολογιστούν οι από άκρο σε άκρο επικαλύψεις για κάθε TT.

Στην περίπτωση (II.β) η ζεύξη K2-A2 έχει χωρητικότητα $C_{K2-A2} = 50$ και η δεύτερη φάση της κατανομής δίνει στα TT-1 και TT-3 5 και 15 μονάδες αντίστοιχα, σύνολο από τις δύο φάσεις 5+5 και 25+15, δηλαδή 50 και επομένως στο TT-2 παραχωρούνται τα υπόλοιπα 20 της ζεύξης K1-K2 που αφαιρέθηκαν από τα άλλα δύο, σύνολο από τις δύο φάσεις 10+40, μαζί με τα 5+5 και 25+15, ίσον με 100 τη χωρητικότητα της ζεύξης

K1-K2. Στην πραγματικότητα δύο ακραίοι κόμβοι είναι δυνατό να συνδέονται μέσω περισσότερων του ενός μονοπατιών και η συνολική κίνηση ενός TT να ισοκατανέμεται σε καθένα από αυτά τα μονοπάτια.

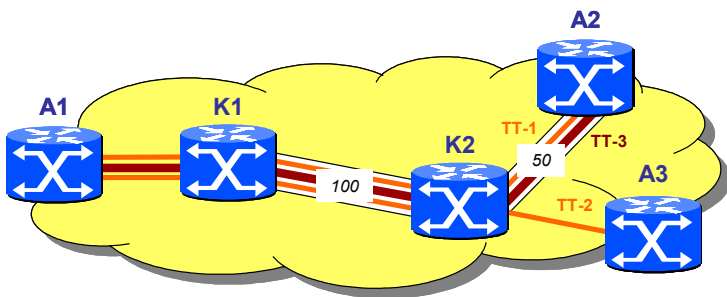
Traffic Matrix					
TT	QoS-κλάση	ingress	egress	TD _{min}	TD _{max}
TT-1	C1	A1	A2	5	35
TT-2	C1	A1	A3	10	70
TT-3	C2	A1	A2	25	115

(I) Traffic Matrix



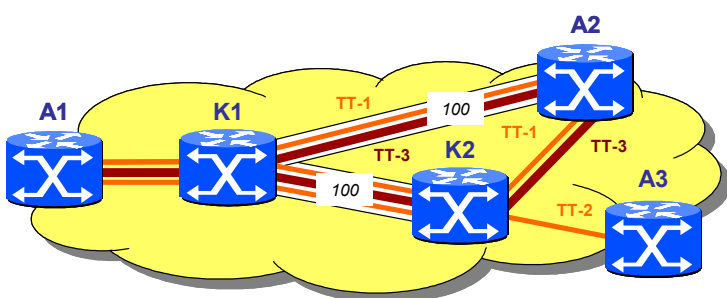
(II.a) ανταγωνισμός σε μία ζεύξη

Λογική Τοπολογία (ND)			
TT	1 ^η κατ/μη	2 ^η κατ/μη	Σύνολο
TT-1	5	10	15
TT-2	10	20	30
TT-3	25	30	55



(II.β) ανταγωνισμός σε πολλαπλές ζεύξεις

Λογική Τοπολογία (ND)			
TT	1 ^η κατ/μη	2 ^η κατ/μη	Σύνολο
TT-1	5	10	10
TT-2	10	20	50
TT-3	25	30	40



(II.γ) ανταγωνισμός σε πολλαπλές ζεύξεις σε πολλαπλά μονοπάτια

Λογική Τοπολογία (ND)			
TT	1 ^η κατ/μη	2 ^η κατ/μη	Σύνολο
TT-1	5	28	33
TT-2	10	48	58
TT-3	25	84	109

(II) φυσική και λογική τοπολογία από το network dimensioning

Σχήμα B-12: παράδειγμα βέλτιστης κατανομής

Στην περίπτωση (II.γ) η νέα ζεύξη K1-A2 εξυπηρετεί μέρος της κίνησης από τον A1 στον A2, δηλαδή ένα μέρος της κίνησης των TT-1 και TT-3. Στην πρώτη κατανομή ισοκατανέμεται ο φόρτος καταλαμβάνοντας 20 από τις 100 μονάδες κάθε ζεύξης. Στη δεύτερη κατανομή το υπόλοιπο των 80 μονάδων της ζεύξης K1-A2 κατανέμεται σε 20 και 60 στα TT-1 και TT-3 αντίστοιχα και το υπόλοιπο των 80 μονάδων της ζεύξης K1-K2

σε 8, 48, 24 στα TT-1, TT-2 και TT-3. Επισημαίνουμε πως στο συγκεκριμένο παράδειγμα χρησιμοποιήσαμε για λόγους περισσότερο κατανοητής απεικόνισης απλουστεύσεις που με τους πραγματικούς αλγορίθμους θα έδιναν ελαφρώς διαφορετικά αποτελέσματα. Κατά τον υπολογισμό των μονοπατιών και σε κάθε φάση της κατανομής συνυπολογίζονται και παράγοντες όπως π.χ. το delay και loss που πρέπει να παραμένει από άκρο σε άκρο φραγμένο στις προκαθορισμένες τιμές της κάθε QoS-κλάσης. Για μια ακριβή περιγραφή σχετικά ο αναγνώστης παραπέμπεται στο [TEQ-ARCH].

Οι πόροι της πρώτης φάσης κατανομής συμβολίζονται στο εξής R_{\min} (minimum resources). Το R_{\min} ισοδυναμεί με το TD_{\min} το οποίο με τη σειρά του αντιστοιχεί στο SR_{AS} , αντιπροσωπεύει δηλαδή τις κατ'ελάχιστο απαιτήσεις για σχεδόν ικανοποιημένες (almost-satisfied) υπηρεσίες ή αλλιώς αντιστοιχεί στους ελάχιστους πόρους που πρέπει να είναι πάντα διαθέσιμοι, αποτελεί με άλλα λόγια το αποκλειστικά δεσμευμένο (hard-reserved) μερίδιο του feedback-based βασικού μοντέλου ελέγχου εισόδου κίνησης. Οι πόροι που περισσεύουν από την εξασφάλιση των hard-reserved μεριδίων, ή αλλιώς οι residual πόροι που κατανέμονται στη δεύτερη φάση, πρέπει να κατανεμηθούν δίκαια με κριτήριο την επίτευξη ισοδύναμου βαθμού ικανοποίησης μεταξύ των υπηρεσιών. Οι απαιτήσεις των υπηρεσιών πέραν του SR_{AS} και έως και τις απόλυτα ικανοποιημένες (fully-satisfied) υπηρεσίες αποτυπώνονται στο SR_{FS} βάσει του οποίου υπολογίζεται το TD_{\max} και χρησιμοποιείται σαν βάρος για τη δίκαιη κατανομή στη δεύτερη φάση. Οι πόροι που προκύπτουν από αυτή τη δεύτερη φάση αποτελούν το συμπληρωματικό κοινόχρηστο (soft-reserved) μερίδιο. Το σύνολο των πόρων από τα δύο μερίδια συμβολίζεται R_{nom} (nominal resources) και είναι το αποτέλεσμα του a priori partitioning σε δίκαια ανά TT μη επικαλυπτόμενα μερίδια.

Η αποτύπωση σε παραμέτρους αρχικοποίησης, η ερμηνεία και η χρήση τους κατά τη λειτουργία του συστήματος για το δυναμικό traffic engineering και το service management έτσι ώστε να επιτυγχάνεται βέλτιστη και συντονισμένη δράση παρουσιάζεται στις επόμενες ενότητες.

B.1.2.2.5 Λειτουργία – Δυναμικό Traffic Engineering

Δεδομένου πως η προσφερόμενη κίνηση είναι στατιστικό μέγεθος και δεδομένου πως το service management, ακολουθώντας τις διακυμάνσεις της, αποδέχεται κίνηση ακόμα και όταν υπερβαίνει τις αρχικές εκτιμήσεις διαθεσιμότητας πόρων R_{nom} , οι λειτουργίες του δυναμικού traffic engineering αποσκοπούν στην απορρόφηση αυτών των διακυμάνσεων της κίνησης με τη βέλτιστη δυναμική ανακατανομή των πόρων γύρω από τις αρχικές τιμές R_{nom} .

Το δυναμικό traffic engineering δεν ακολουθεί τη λογική της αδιάκοπης βελτιστοποίησης (continuous optimisation) αλλά παρεμβαίνει όσο συχνά κρίνεται απαραίτητο για να παραμένει αποτελεσματικό χωρίς να επιφορτίζει σημαντικά την απόδοση του συστήματος και χωρίς να το αποσταθεροποιεί. Αυτό σημαίνει πως οι παρεμβάσεις στην κατανομή των πόρων θα πρέπει έχουν ένα χρονικό ορίζοντα μεγαλύτερο από το τυχαίο στιγμιότυπο. Οι παράμετροι αρχικοποίησης όπως έχουν υπολογιστεί από το network dimensioning χρησιμοποιούνται για αυτό το σκοπό σαν κατευθυντήριες γραμμές (guidelines), απαλλάσσοντας το δυναμικό traffic engineering από τους σχετικούς υπολογισμούς και εφοδιάζοντάς το με μια άποψη του συστήματος περισσότερο προηγμένη από την τοπικά και χρονικά περιορισμένη των κατανεμημένων δυναμικών λειτουργιών του. Επιπλέον, η ανακατανομή των πόρων υιοθετεί την αρχή των hard-reserved και soft-reserved μεριδίων για να είναι συντονισμένη με τις παράλληλες ενέργειες του service management. Με άλλα λόγια, η ανακατανομή γίνεται μόνο στους soft-reserved πόρους ενώ τα hard-reserved μερίδια είναι κατοχυρωμένα πάντα σύμφωνα με την αρχική κατανομή.

Οι διακυμάνσεις της τελικά διοχετευόμενης κίνησης απορροφούνται σε επίπεδο ζεύξης από τη λειτουργία δυναμικής διαχείρισης πόρων (dynamic resource management) που βρίσκεται σε κάθε κόμβο του δικτύου. Η

φυσική χωρητικότητα κάθε ζεύξης κατανέμεται μεταξύ των PHBs σύμφωνα με τις ρυθμίσεις του μηχανισμού χρονοπρογραμματισμού (βλέπε ενότητα A.1.4.2.2.1). Η λειτουργία του dynamic resource management ανά ζεύξη επενεργεί σε αυτές τις ρυθμίσεις, και κατ' επέκταση καθορίζει τους πόρους που αποδίδονται σε κάθε PHB. Παρακολουθώντας με τη λήψη τοπικών παθητικών μετρήσεων την ανά PHB διερχόμενη κίνηση προσαρμόζει την κατανομή των πόρων έτσι ώστε να επιτυγχάνεται βέλτιστη εξυπηρέτηση.

Συγκεκριμένα, το hard-reserved μερίδιο των διερχόμενων μονοπατιών των TTs, ή αλλιώς το R_{min} συναθροίζεται ανά QoS-κλάση και σχηματίζει το hard-reserved μερίδιο του αντίστοιχου PHB. Όποια κι αν είναι η ένταση της πραγματικής κίνησης οι πόροι που αποδίδονται κάθε φορά σε ένα PHB δεν θα είναι ποτέ λιγότεροι από το R_{min} που του αντιστοιχεί και ποτέ περισσότεροι από τη χωρητικότητα της ζεύξης μείον το R_{min} των άλλων PHBs. Αυτή η τελευταία τιμή ονομάζεται R_{max} (maximum resources). Επιπλέον, σε περίπτωση συμφόρησης, σε περίπτωση δηλαδή που η συνολική διοχετευόμενη κίνηση είναι μεγαλύτερη από τη χωρητικότητα της ζεύξης, τότε η κατανομή των πόρων "κλειδώνει" στο αντίστοιχο ανά PHB R_{nom} , που αντιπροσωπεύει το δίκαιο μερίδιο των πόρων. Οι τιμές των R_{min} , R_{nom} και R_{max} για τα PHBs που υλοποιούν τις κλάσεις C1 και C2 στη ζεύξη K1-K2 στην περίπτωση (II.α) του παραδείγματος του Σχήμα B-12 φαίνονται στον Πίνακα B-1.

Στην πραγματικότητα, σε περίπτωση που σε κάποιο κόμβο χρησιμοποιείται not work-conserving μηχανισμός χρονοπρογραμματισμού (βλέπε ενότητα A.1.4.1.1), ένα PHB μπορεί να υπερβεί το R_{max} που του αντιστοιχεί κάνοντας χρήση μέρος ή και του συνόλου των hard-reserved μεριδίων των άλλων PHBs όσο αυτά μένουν ανενεργά. Αυτό που διασφαλίζεται από το dynamic resource management είναι πως, όταν υπάρχει αρκετή κίνηση για ένα PHB τότε αυτή θα λάβει κατ' ελάχιστο σε κάθε περίπτωση το hard-reserved R_{min} μερίδιο του.

παράμετροι αρχικοποίησης (dynamic resource management)			
PHB	R_{min}	R_{nom}	R_{max}
C1	15	45	75
C2	25	55	85

Πίνακας B-1: παράδειγμα παραμέτρων dynamic resource management

Οι διακυμάνσεις της κίνησης απορροφώνται σε επίπεδο μονοπατιού από τη λειτουργία δυναμικής διαχείρισης δρομολόγησης (dynamic route management) που αποσκοπεί στην εξισορρόπηση του συνολικού από άκρο σε άκρο φόρτου (load balancing) μεταξύ των διαφορετικών δυνατών μονοπατιών. Οι παράμετροι αρχικοποίησης και οι σχετικοί μηχανισμοί διαφέρουν ανάλογα με το υποκείμενο σύστημα δρομολόγησης. Διακρίνουμε μεταξύ των συστημάτων *δυναμικής ανακάλυψης δρόμου* (dynamic route computation) βασισμένα στο παραδοσιακό IP (IP-based) και *ρητής δρομολόγησης* (explicit routing) ή *καθήλωσης δρόμου* (route pinning) όπως π.χ. τα συστήματα με βάση με την τεχνολογία Multiprotocol Label Switching (MPLS-based) [MPLS].

Στα MPLS-based συστήματα τα μονοπάτια είναι προκαθορισμένα από άκρο σε άκρο και η διαδικασία δρομολόγησης περιορίζεται στην επιλογή του κατάλληλου μονοπατιού για κάθε ροή έτσι ώστε ο φόρτος των μονοπατιών να παραμένει κατά το δυνατό στα ίδια επίπεδα. Ο καθορισμός των μονοπατιών γίνεται από την κεντρική λειτουργία του network dimensioning ενώ το dynamic route management βρίσκεται σε κάθε

ακραίο κόμβο και συμπεραίνει το φόρτο κάθε μονοπατιού χρησιμοποιώντας τοπικές μετρήσεις για την ένταση της κίνησης που διοχετεύεται σε κάθε μονοπάτι προς τους πόρους που έχουν παραχωρηθεί στο μονοπάτι σύμφωνα με το soft-reserved R_{nom} μερίδιο.

Στα IP-based συστήματα αντίθετα, τα μονοπάτια υπολογίζονται δυναμικά και με κατανεμημένο τρόπο σε κάθε κόμβο του δικτύου με τη λογική του *συντομότερου μονοπατιού* (shortest path) όπου η απόσταση υπολογίζεται βάσει του *κόστους ζεύξης* (link cost). Τα κόστη των ζεύξεων προφανώς επηρεάζουν τα αποτελέσματα των κατανεμημένων αλγορίθμων δρομολόγησης και ρυθμίζονται από την κεντρική λειτουργία του dynamic route management έτσι ώστε να επιτυγχάνεται η επιθυμητή εξισορρόπηση του φόρτου. Για να καταστεί δυνατή διαφοροποίηση των μονοπατιών χρησιμοποιείται διαφορετικό κόστος ανά PHB για κάθε ζεύξη. Η λογική τοπολογία όπως προκύπτει από το network dimensioning χρησιμοποιείται ως σημείο αναφοράς στο οποίο αποσκοπεί να συγκλίνει το dynamic route management αποδίδοντας τις κατάλληλες τιμές ανά κόστος ζεύξης.

B.1.2.2.6 Λειτουργία – Service Management

Το service management συνδιαλέγεται με τους πελάτες σύμφωνα με το διαδικαστικό μοντέλο (βλέπε ενότητα B.1.2.1.1) και ασκεί τον έλεγχο αποδοχής στις προσφερόμενες αιτήσεις εγγραφής και κλήσης με τις λειτουργίες διαχείρισης εγγραφών υπηρεσίας (subscription management) και διαχείρισης κλήσεων υπηρεσίας (invocation management) [MCG03].

Η κίνηση που διοχετεύεται στο δίκτυο προκύπτει από τις αποδεκτές κλήσεις, ο έλεγχος εισόδου κίνησης ως διαδικασία που ρυθμίζει αυτή την κίνηση ανήκει συνεπώς στο invocation management. Ωστόσο, η διαδικασία κλήσης προϋποθέτει έγκυρο συμβόλαιο εγγραφής, με άλλα λόγια οι δικαιούχες αιτήσεις κλήσης ρυθμίζονται από το subscription management. Η αποτελεσματικότητα του ελέγχου εισόδου κίνησης εξαρτάται σε μεγάλο βαθμό από το φόρτο κίνησης που καλείται να διαχειριστεί, δηλαδή από το πλήθος των εγκεκριμένων εγγραφών υπηρεσίας. Έτσι, αν υπάρχει έλεγχος εισόδου και στη διαδικασία εγγραφής, είναι δυνατό να ελεγχθεί το πλήθος και ο τύπος των εγκεκριμένων εγγραφών, γεγονός που επαυξάνει το επίπεδο ελέγχου και προσδίδει ευελιξία.

Θεωρούμε έτσι πως το subscription management επιφορτίζεται με τον έλεγχο αποδοχής των αιτήσεων εγγραφής και το invocation management με τον έλεγχο αποδοχής των αιτήσεων κλήσης, δηλαδή τον έλεγχο εισόδου κίνησης. Θεωρούμε ακόμα πως η λειτουργία τους είναι ανεξάρτητη και αυτόνομη σε βαθμό που ακόμα και η καταστολή του ελέγχου εισόδου σε οποιοδήποτε από τα δύο επίπεδα αν αυτό κριθεί απαραίτητο δεν εμποδίζει τη λειτουργία του άλλου. Ο συντονισμός των ενεργειών τους επιτυγχάνεται, όπως και με το traffic engineering, μέσω των παραμέτρων αρχικοποίησης. Προφανώς, η κάθε λειτουργία επιφορτίζεται και με τις αντίστοιχες διαδικασίες εξακρίβωσης γνησιότητας, εξουσιοδότησης βάσει των πολιτικών του παροχέα και ενεργοποίησης που προβλέπονται από το διαδικαστικό μοντέλο (βλέπε ενότητα B.1.2.1.1).

Το subscription management αποφαίνεται επί της μακροπρόθεσμης δυνατότητας του δικτύου όπως διαμορφώνεται από το traffic engineering, να εξυπηρετήσει την κίνηση που θα προκύψει εν καιρώ από το υπό διαπραγμάτευση επιπλέον συμβόλαιο. Αρκείται συνεπώς να γνωρίζει την κατά μέσο όρο διαθεσιμότητα των πόρων για ένα μεγάλο χρονικό διάστημα και δεν ασχολείται με το στιγμιότυπο της κατάστασης του δικτύου τη στιγμή της διαπραγμάτευσης. Καθώς η συχνότητα των αιτήσεων εγγραφής είναι μικρή και οι υπολογισμοί δεν εμπλέκουν κάποιο εξωτερικό δυναμικά εξελισσόμενο παράγοντα αλλά περιορίζονται στη χρήση των παραμέτρων αρχικοποίησης και των εσωτερικών μεταβλητών, το subscription management θεωρείται απλή και λιγότερο φορτωμένη λειτουργία χωρίς εξαιρετικά υψηλές απαιτήσεις σε χρόνο απόκρισης, γι' αυτό και γίνεται κεντρικά.

Αντίθετα, το invocation management είναι σύμφωνα με το βασικό μοντέλο ελέγχου εισόδου κίνησης (βλέπε ενότητα B.1.1), καταναμημένο στους ακραίους κόμβους του δικτύου και αποφαινεται ως προς την τρέχουσα διαθεσιμότητα των πόρων για την άμεση εξυπηρέτηση της πραγματικής κίνησης. Η τρέχουσα διαθεσιμότητα εκτιμάται από τοπικές και από άκρο σε άκρο μετρήσεις επί της αρχικής κατανομής των πόρων στα hard-reserved και soft-reserved μερίδια.

Η διαθεσιμότητα των πόρων είναι αποτέλεσμα του traffic engineering από το οποίο εκτιμάται και αποτιμάται ανά TT στον πίνακα διαθεσιμότητας πόρων (Resource Availability Matrix – RAM) που προκύπτει από τη λογική τοπολογία του network dimensioning. Η διαθεσιμότητα εκφράζεται για κάθε TT με τρεις διαφορετικές τιμές, το R_{min} που ισοδυναμεί με το hard-reserved μερίδιο, το R_{nom} που ισοδυναμεί με το hard-reserved και το soft-reserved μερίδιο και το R_{max} που ισοδυναμεί με το hard-reserved και τα soft-reserved μερίδια όλων των ανταγωνιζόμενων TTs (βλέπε Σχήμα B-13). Η τριάδα των τιμών για την εκτίμηση της διαθεσιμότητας διαμορφώνει για κάθε TT το διαθέσιμο απόθεμα πόρων (Resource Availability Buffer – RAB). Στο Σχήμα B-14 παρουσιάζεται ο RAM για τις περιπτώσεις ανταγωνισμού σε μία ζεύξη (II.α) και ανταγωνισμού σε πολλαπλές ζεύξεις σε πολλαπλά μονοπάτια του Σχήμα B-12.



Σχήμα B-13: περιεχόμενο γραμμής Resource Availability Matrix ή Resource Availability Buffer

Resource Availability Matrix			
TT	R_{min}	R_{nom}	R_{max}
TT-1	5	15	65
TT-2	10	30	70
TT-3	25	55	85

περίπτωση (II.α)

Resource Availability Matrix			
TT	R_{min}	R_{nom}	R_{max}
TT-1	5	33	165
TT-2	10	58	90
TT-3	25	109	185

περίπτωση (II.γ)

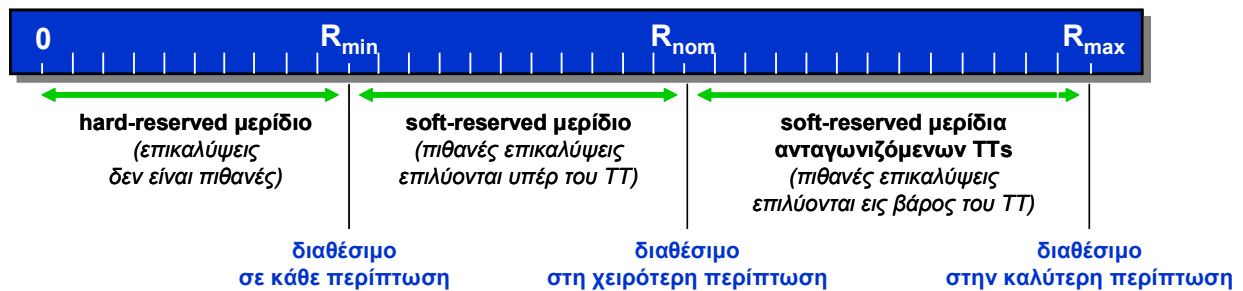
Σχήμα B-14: παράδειγμα Resource Availability Matrix

Το R_{min} είναι το hard-reserved μερίδιο, δηλαδή οι πόροι που είναι σε κάθε περίπτωση διαθέσιμοι για το συγκεκριμένο TT. Οι πόροι αυτοί αντιστοιχούν στις κατ' ελάχιστο απαιτήσεις για σχεδόν ικανοποιημένες (almost-satisfied) υπηρεσίες.

Το R_{nom} είναι το δίκαιο μερίδιο των πόρων που διατίθεται στο συγκεκριμένο TT στη χειρότερη περίπτωση, δηλαδή σε περίπτωση συμφόρησης όπου όλα τα TTs καλύπτουν ή υπερβαίνουν το αντίστοιχο δικό τους R_{nom} . Οι πόροι αυτοί προκύπτουν από τη δίκαιη ανάλογα με τις απαιτήσεις για πλήρως ικανοποιημένες (fully-satisfied) υπηρεσίες κατανομή των υπόλοιπων (residual) πόρων σε μη επικαλυπτόμενα μερίδια.

Το R_{max} είναι οι περισσότεροι πόροι που μπορεί να διατεθούν στο συγκεκριμένο TT στην καλύτερη περίπτωση, δηλαδή στην περίπτωση που η κίνηση από τα άλλα TTs δεν ξεπερνά τα hard-reserved μερίδιά

τους αφήνοντας ελεύθερους όλους τους υπόλοιπους (residual) πόρους, ή αλλιώς είναι το R_{nom} μερίδιο συν όλα τα soft-reserved μερίδια των άλλων TTs.



Σχήμα B-15: ερμηνεία Resource Availability Buffer

Επισημαίνουμε πως οι παραπάνω ορισμοί αφορούν την ερμηνεία που πρέπει να αποδοθεί σε κάθε τιμή από τις λειτουργίες του service management, ερμηνεία που δεν αντιστοιχεί πάντα στην πραγματικότητα. Έτσι, το πραγματικό μέγιστο διαθέσιμο για ένα TT συμπεριλαμβάνει και τα hard-reserved μερίδια άλλων TTs όταν αυτά δεν χρησιμοποιούνται και ειδικότερα εκείνων των TTs της ίδιας QoS-κλάσης, αλλά και των TTs διαφορετικών QoS-κλάσεων στην περίπτωση χρονοπρογραμματισμού μη σταθερού έργου (βλέπε ενότητα A.1.4.1.1). Επιπλέον, επικαλύψεις είναι δυνατό να συμβούν και στο hard-reserved μερίδιο R_{min} . Αν και τα hard-reserved μερίδια δεν δίνονται για κοινή χρήση εφόσον δεν υπολογίζονται στο R_{max} , ωστόσο η δυνατότητα πολλαπλής χρησιμοποίησης από διαφορετικά TTs του ίδιου soft-reserved μεριδίου ενός τρίτου TT ισοδυναμεί με δανεισμό από το hard-reserved μερίδιο αυτού του TT. Στην περίπτωση (II.α) του Σχήμα B-14 εάν το TT-1 χρησιμοποιεί 40 και το TT-2 50 μονάδες εύρους ζώνης στα όρια των R_{max} τους, ουσιαστικά χρησιμοποιούν 25 και 20 μονάδες αντίστοιχα από το soft-reserved μερίδιο του TT-3 το οποίο όμως συνολικά ανέρχεται στις 30 μονάδες, με αποτέλεσμα την επικάλυψη μεταξύ των TTs εάν το TT-3 ξεπεράσει τις 10 και όχι τις 25 μονάδες του hard-reserved R_{min} μεριδίου του.

Η μεταξύ των QoS-κλάσεων, ή αλλιώς των αντίστοιχων PHBs, διαφύλαξη των hard-reserved μεριδίων και η σύγκλιση στα R_{nom} σε περίπτωση συμφόρησης επιτυγχάνεται χάρη στις ενέργειες του δυναμικού traffic engineering (βλέπε ενότητα B.1.2.2.5). Αντίστοιχα, η επιβολή της διαφύλαξης των hard-reserved μεριδίων και της σύγκλισης στα R_{nom} σε περίπτωση συμφόρησης μεταξύ των διαφορετικών TTs σε κάθε QoS-κλάση επιτυγχάνεται από τη συνεργία της ανά ακραίο κόμβο κατανεμημένης λειτουργίας του invocation management (βλέπε ενότητα B.2.2). Σηριζόμενο στη διαφύλαξη αυτών των αρχών από το invocation management, το subscription management είναι σε θέση να εφαρμόσει λογισμό επί της μακροπρόθεσμης διαθεσιμότητας των πόρων για την αποδοχή ενός συμβολαίου (βλέπε ενότητα B.2.1).

Λόγω των ενεργειών του δυναμικού traffic engineering και της επιτρεπτής επικάλυψης σε επίπεδο TT και QoS-κλάσης οι τιμές του RAB δεν αντιστοιχούν σε απόλυτα εγγυημένα διαθέσιμα μερίδια αλλά χρησιμεύουν ως κατευθυντήριες γραμμές (guidelines) έτσι ώστε οι ενέργειες του service management να είναι εναρμονισμένες μεταξύ τους και με τις ενέργειες του δυναμικού traffic engineering.

Συγκεκριμένα, το RAM παρέχει στη φάση αρχικοποίησης του συστήματος το κοινό σημείο αναφοράς μεταξύ των κατανεμημένων δειγμάτων (instances) του invocation management που επιτρέπει τον εναρμονισμό των ενεργειών τους χωρίς περαιτέρω συντονισμό κατά τη λειτουργία του συστήματος. Κατ' αναλογία, η αντιστοιχία του RAM με τις παραμέτρους αρχικοποίησης των κατανεμημένων λειτουργιών του δυναμικού traffic engineering, χάρη στην κοινή τους προέλευση από τη λογική τοπολογία του network dimensioning, διασφαλίζει τον εναρμονισμό των ενεργειών μεταξύ service management και traffic

engineering. Ο τρόπος μεταβολής της κατανομής και συνεπώς της διαθεσιμότητας των πόρων από το δυναμικό traffic engineering κατά τη λειτουργία του συστήματος είναι ο αναμενόμενος από το invocation management και επιπλέον όταν και οι δύο ειδοποιούνται μέσω των κατάλληλων μετρήσεων για την εκδήλωση συμφόρησης οι ενέργειές τους από κοινού συγκλίνουν στην επίλυσή της.

Το subscription management και το invocation management χρησιμοποιούν παράλληλα τους πόρους του RAB, το καθένα υπό το δικό του πρίσμα, το μεν subscription management για την εξυπηρέτηση της κίνησης που θα προκύψει εν καιρώ και κατά μέσο όρο από το υπό διαπραγμάτευση συμβόλαιο, το δε invocation management για την εξυπηρέτηση της κίνησης που θα προκύψει άμεσα από την πραγματική ροή δεδομένων του χρήστη που καλεί και ενεργοποιεί την υπηρεσία.

Η εξασφάλιση της ποιότητας των υπηρεσιών και κατ' επέκταση η ικανοποίησή τους αντιστοιχεί στην εξασφάλιση της διαθεσιμότητας των πόρων η οποία εξ ορισμού διαφέρει για τα hard-reserved και soft-reserved μερίδια. Η ρύθμιση της συμπεριφοράς των αντίστοιχων αλγορίθμων από τη λειτουργική πολιτική του παροχέα ώστε να επιτυγχάνεται η ενδεδειγμένη ικανοποίηση των υπηρεσιών (βλέπε ενότητα B.1.1.3) μεταφράζεται συνεπώς στον καθορισμό εκείνης της περιοχής του RAB που θεωρείται αβλαβές να καταλαμβάνεται από τα κυμαινόμενα επίπεδα της κίνησης.

Οι παράμετροι πολιτικής, η ερμηνεία τους και οι αλγόριθμοι παρουσιάζονται στις επόμενες ενότητες. Η παρουσίαση του subscription management διευκολύνει την κατανόηση εκείνων των πτυχών που επηρεάζουν τον αλγόριθμο του invocation management που αποτελεί και το θέμα της παρούσας διατριβής.

B.1.3 Ανακεφαλαίωση

Μετά από την ανάλυση των πλεονεκτημάτων και των μειονεκτημάτων των διαφορετικών προσεγγίσεων ***ιοθετούμε το κατανεμημένο μοντέλο αμιγώς ακραίων κόμβων***. Αποκλείοντας τις parameter-based μεθόδους, αναγνωρίζουμε στις γνωστές measurement-based μεθόδους δύο διαφορετικές προσεγγίσεις.

Στην πρώτη προσέγγιση, η από άκρο σε άκρο χωρητικότητα θεωρείται γνωστή και διαμορφώνεται από τον a priori ή το δυναμικό και συντονισμένο διαμερισμό των πόρων του δικτύου σε απομονωμένα μη επικαλυπτόμενα μερίδια. Μετρήσεις του φόρτου που διοχετεύεται τοπικά προς το δίκτυο επί της γνωστής χωρητικότητας χρησιμοποιούνται για την εκτίμηση της διαθεσιμότητας των πόρων. Στη δεύτερη προσέγγιση η από άκρο σε άκρο χωρητικότητα είναι άγνωστη, οι πόροι δεν διαμερίζονται, το δίκτυο αναπαριστάται ως μαύρο κουτί και η διαθεσιμότητα των πόρων εκτιμάται άμεσα χρησιμοποιώντας απευθείας μετρήσεις της ποιότητας από άκρο σε άκρο.

Διακρίνοντας τις αδυναμίες των δύο διαφορετικών προσεγγίσεων ***προτείνουμε μια νέα υβριδική measurement-based μέθοδο που συνδυάζει χαρακτηριστικά των γνωστών προσεγγίσεων και αποκαλούμε μοντέλο βασιζόμενο σε ανάδραση (feedback-based)***.

Σε κάθε ακραίο κόμβο παραχωρείται ένα μερίδιο των πόρων κορμού αποκλειστικά δεσμευμένο (hard-reserved) και ένα συμπληρωματικό κοινόχρηστο μερίδιο (soft-reserved) με a priori partitioning. Εάν υπάρχει αρκετή ζήτηση κάθε κόμβος επιτρέπεται να υπερβεί το hard-reserved αλλά και το soft-reserved μερίδιό του και να χρησιμοποιήσει πόρους από τα soft-reserved μερίδια των άλλων κόμβων χωρίς συντονισμένη δέσμευση και αποδέσμευση μεταξύ των εμπλεκόμενων κόμβων αλλά με ελεγχόμενο ρίσκο. Σε κάθε κόμβο χρησιμοποιώντας μετρήσεις του φόρτου τοπικά υπολογίζεται το μέρος από το hard-reserved ή και το soft-reserved ή ακόμα και από το επιπλέον του soft-reserved μεριδίου που χρησιμοποιείται από την κίνηση που διοχετεύεται στο δίκτυο από τον συγκεκριμένο κόμβο, το οποίο χρησιμεύει ως ένδειξη για το ρίσκο παραβίασης της ποιότητας λόγω της επικάλυψης των soft-reserved μεριδίων. Μαζί με τις μετρήσεις του

φόρτου τοπικά χρησιμοποιούνται και περιοδικές μετρήσεις της ποιότητας από άκρο σε άκρο. Όταν αυτές οι μετρήσεις καταδείξουν παραβίαση της ποιότητας τότε κατάλληλοι μηχανισμοί ανάδρασης ανακόπτουν την ένταση με την οποία διοχετεύεται η κίνηση μέχρι την ομαλή επίλυση της συμφόρησης που προκάλεσε την παραβίαση της ποιότητας.

Όπως και για κάθε measurement-based σχήμα έτσι και στο feedback-based μοντέλο το ελεγχόμενο ρίσκο, δηλαδή η ελεγχόμενα συντηρητική ή επιθετική συμπεριφορά στη χρησιμοποίηση των soft-reserved μεριδίων, δεν είναι δυνατό να αντιστοιχηθεί σε μαθηματικά προβλέψιμη πιθανότητα παραβίαση της ποιότητας. Έτσι, **η ρύθμιση της λειτουργίας του συστήματος γίνεται ανάλογα με την αντίληψη του παροχέα για το τι σημαίνει επαρκής ικανοποίηση των υπηρεσιών και για ποιο συνεπαγόμενο κόστος επί της βέλτιστης αξιοποίησης των πόρων**, έχοντας αφομοιώσει και εμμέσως συνεκτιμήσει τις ιδιομορφίες του εκάστοτε περιβάλλοντος στο οποίο εφαρμόζεται ο αλγόριθμος.

Για να καταστεί δυνατός ο μαθηματικός λογισμός επί της *επαρκούς ικανοποίησης* των υπηρεσιών εισάγουμε τους απαραίτητους ορισμούς, στους οποίους λαμβάνονται υπόψη πλευρές της ικανοποίησης κατά τη χρήση αλλά και ικανοποίησης ως προς τη διαθεσιμότητα της υπηρεσίας, υπό το πρίσμα του διαφορετικού βαθμού ανεκτικότητας σε υποβάθμιση κάθε υπηρεσίας και κάθε εφαρμογής για την οποία προορίζεται.

Διακρίνουμε μεταξύ των φάσεων εγγραφής, κλήσης και ενεργοποίησης της υπηρεσίας. Θεωρούμε τις περιπτώσεις *ρητής κλήσης* και *έμμεσης κλήσης*. Κατά την έμμεση κλήση, η υπηρεσία ενεργοποιείται αυτόματα από τον παροχέα σύμφωνα με το χρονοδιάγραμμα που ορίζεται στο συμβόλαιο. Αντίθετα, στη ρητή κλήση η ενεργοποίηση της υπηρεσίας γίνεται μόνο κατ' αίτηση του χρήστη. Θεωρούμε ακόμα τρεις διαφορετικούς τύπους υπηρεσιών βάσει της μεθόδου ενεργοποίησης, τις *υπηρεσίες μόνιμης ενεργοποίησης* (permanent services), τις *υπηρεσίες ενεργοποίησης ανά αίτηση* (on demand services) και τις *υπηρεσίες ευέλικτης ενεργοποίησης* (flexible services).

Η ανάλυση των παραμέτρων που περιγράφουν την υπηρεσία στο συμβόλαιο και η αντιστοίχισή τους στα δομικά στοιχεία ποιότητας υπηρεσίας του δικτύου και κατ' επέκταση στους απαραίτητους πόρους για την ενεργοποίησή της επιτυγχάνεται με την παρεμβολή της δομής της QoS-κλάσης μεταξύ του SLA και των PHBs. Κάθε παροχέας ορίζει τις QoS-κλάσεις που υποστηρίζει το δίκτυό του και επί των οποίων οικοδομούνται οι υπηρεσίες που παρέχει στους πελάτες του.

Διαχωρίζουμε τη μελέτη μας από τις λειτουργίες του επιπέδου δεδομένων που υλοποιούν τις ρυθμίσεις ενεργοποίησης των υπηρεσιών, τις λειτουργίες παρακολούθησης που υλοποιούν τις υποστηρικτικές λειτουργίες της συγκέντρωσης, επεξεργασίας και σύνταξης αναφορών επί των αποτελεσμάτων των μετρήσεων. **Οριοθετούμε τον έλεγχο εισόδου κίνησης ως μέρος της λειτουργίας διαχείρισης κλήσης** που, μαζί με τη λειτουργία διαχείρισης εγγραφής συνιστούν τις λειτουργίες διαχείρισης υπηρεσιών. Θεωρούμε έλεγχο αποδοχής και κατά τη φάση εγγραφής οπότε εξετάζεται η μακροπρόθεσμη διαθεσιμότητα των πόρων του δικτύου για την εξυπηρέτηση της υπηρεσίας. Θεωρούμε τέλος πως *a priori* υπολογισμός των μεριδίων των πόρων και η δυναμική κατανομή τους για τη βέλτιστη εξυπηρέτηση της κίνησης που διοχετεύεται στο δίκτυο από τον έλεγχο εισόδου κίνησης είναι μέρος των λειτουργιών διαχείρισης της κίνησης.

Ο συνεχής χρόνος χωρίζεται σε διακριτές χρονικές περιόδους στις οποίες παρατηρείται ομοιογενής συμπεριφορά των χρηστών και για τις οποίες είναι δυνατό να γίνει συνεπής *a priori* υπολογισμός των μεριδίων, σύμφωνα με προβλέψεις της προσφερόμενης κίνησης βάσει των συμβολαίων εγγραφής των υπηρεσιών. Τα αποτελέσματα συνθέτουν τις παραμέτρους αρχικοποίησης τόσο των λειτουργιών διαχείρισης υπηρεσιών όσο και των δυναμικών λειτουργιών διαχείρισης κίνησης, διασφαλίζοντας χωρίς άλλο ρητό συντονισμό τον εναρμονισμό των ενεργειών τους κατά τη λειτουργία του συστήματος. Ο υπολογισμός των μεριδίων των πόρων διαμορφώνει τον *πίνακα διαθεσιμότητας πόρων* RAM και γίνεται σε επίπεδο *κορμού*

κίνησης TT, δηλαδή σε επίπεδο QoS-κλάσης τοποθετημένης στο χώρο, σε καθορισμένη από άκρο σε άκρο σύνδεση. Το κοινό σημείο των λειτουργιών διαχείρισης υπηρεσιών και πόρων είναι η δομή TT, περιορίζοντας έτσι την πολυπλοκότητα του συστήματος στον αριθμό των πιθανών ζευγών ακραίων κόμβων επί τον αριθμό των QoS-κλάσεων, καθιστώντας έτσι εφικτή την εφαρμογή του συστήματος σε μεγάλη κλίμακα.

Η διαθεσιμότητα των πόρων εκφράζεται από τις τιμές R_{min} , R_{nom} και R_{max} . Το R_{min} είναι σε κάθε περίπτωση διαθέσιμο για το συγκεκριμένο TT και αντιστοιχεί στις απαιτήσεις για ικανοποίηση του μικρότερου επιτρεπτού βαθμού. Το R_{nom} είναι το δίκαιο μερίδιο των πόρων που διατίθεται στο συγκεκριμένο TT στη χειρότερη περίπτωση, δηλαδή σε περίπτωση συμφόρησης όπου όλα τα TTs καλύπτουν ή υπερβαίνουν το αντίστοιχο δικό τους R_{nom} . Οι πόροι αυτοί προκύπτουν από τη δίκαιη ανάλογα με τις απαιτήσεις για μέγιστη ικανοποίηση κατανομή των υπόλοιπων πόρων σε μη επικαλυπτόμενα μερίδια. Το R_{max} είναι οι περισσότεροι πόροι που μπορεί να διατεθούν στο συγκεκριμένο TT στην καλύτερη περίπτωση, δηλαδή στην περίπτωση που η κίνηση από τα άλλα TTs δεν ξεπερνά τα R_{min} μερίδιά τους.

Η μεταξύ των QoS-κλάσεων, ή αλλιώς των αντίστοιχων PHBs, διαφύλαξη των μεριδίων και η σύγκλιση στα R_{nom} σε περίπτωση συμφόρησης επιτυγχάνεται χάρη στις ενέργειες του δυναμικού traffic engineering. Αντίστοιχα, η επιβολή της διαφύλαξης των μεριδίων και της σύγκλισης στα R_{nom} μεταξύ των διαφορετικών TTs επιτυγχάνεται από τη συνεργία της ανά ακραίο κόμβο κατανεμημένης λειτουργίας διαχείρισης κλήσης. Στηριζόμενο στη διαφύλαξη αυτών των αρχών από τη διαχείριση κλήσης, η διαχείριση εγγραφής είναι σε θέση να εφαρμόσει λογισμό επί της μακροπρόθεσμης διαθεσιμότητας των πόρων για την αποδοχή ενός συμβολαίου.

B.2 Έλεγχος Εισόδου Κίνησης

B.2.1 Διαχείριση Εγγραφής

B.2.1.1 Πρόβλημα

Η λειτουργία διαχείρισης εγγραφής (subscription management) ελέγχει τις διαδικασίες αποδοχής και διαπραγμάτευσης των συμβολαίων εγγραφής σε υπηρεσίες, και ειδικότερα την εξουσιοδότηση βάσει της διαθεσιμότητας των πόρων (βλέπε ενότητα B.1.2.1.1.1), επιδιώκοντας (α) να προστατέψει το δίκτυο από την εν καιρώ υπερφόρτωσή του κατά την μελλοντική ενεργοποίηση των εγγεγραμμένων υπηρεσιών και κατ' επέκταση την υποβάθμιση της ικανοποίησης των τελευταίων ενώ ταυτόχρονα (β) μεγιστοποιεί τα συμβόλαια εγγραφής σε υπηρεσίες με τρόπο που να αντιστοιχεί σε μεγιστοποίηση του κέρδους του παροχέα.

B.2.1.2 Παραδοχές και Περιβάλλον

Όπως έχουμε αναφέρει, η παρούσα διατριβή επικεντρώνεται σε υπηρεσίες επιπέδου μεταφοράς, δεν εξετάζονται επομένως η διαθεσιμότητα των πόρων που αφορούν σε ανώτερα επίπεδα, όπως π.χ. η χωρητικότητα και η κατάσταση των εξυπηρετητών εφαρμογής (application servers).

Θεωρούμε πως ο έλεγχος περιορίζεται στους πόρους των ακραίων κόμβων και των κόμβων κορμού της δεδομένης επικράτειας και δεν περιλαμβάνει το δίκτυο πρόσβασης (access network), όπως π.χ. οι περιορισμοί που προκύπτουν από τον πεπερασμένο αριθμό των προσκείμενων modems.

Θεωρούμε ακόμα πως το πρωτόκολλο διαπραγμάτευσης μιας αίτησης εγγραφής επιτρέπει την αποστολή και διαπραγμάτευση επί εναλλακτικών αντιπροτάσεων υπηρεσίας (alternative services) όταν η υπηρεσία όπως αρχικά περιγράφεται από τον πελάτη δεν είναι δυνατό να παρασχεθεί.

Θεωρούμε τέλος πως το δίκτυο είναι "προγραμματισμένο" από τους μηχανισμούς του traffic engineering να διαθέτει τους πόρους σύμφωνα με τον πίνακα διαθεσιμότητας πόρων RAM (βλέπε ενότητα B.1.2.2.6).

B.2.1.3 Τοποθέτηση

Η διαχείριση εγγραφής είναι κεντρική λειτουργία (βλέπε ενότητα B.1.2.2.6).

Κάθε αίτηση εγγραφής μεταφράζεται και αντιστοιχίζεται βάσει του δομικού μοντέλου (βλέπε ενότητα B.1.2.1.3) στα TTs που εμπλέκονται στην εξυπηρέτηση της κίνησης που θα προκύψει από το αντίστοιχο συμβόλαιο και ο αλγόριθμος ελέγχου εισόδου εφαρμόζεται ανά TT.

Ο αλγόριθμος ελέγχου εισόδου ανά TT αποφαινεται επί της μακροπρόθεσμης δυνατότητας του δικτύου βάσει των διαθέσιμων πόρων του RAB να εξυπηρετήσει τη συνολική κίνηση που θα προκύπτει ανά πάσα στιγμή από τις ήδη εγγεγραμμένες και τη νέα υπηρεσία. Η ανά πάσα στιγμή συνολική ανά TT προσφερόμενη κίνηση είναι στατιστικό μέγεθος και υπολογίζεται βάσει των parameter-based μοντέλων που χρησιμοποιούνται από τη λειτουργία πρόβλεψης της κίνησης κατά τη διαδικασία εξαγωγής της συνολικής προσφερόμενης κίνησης (aggregate demand derivation) (βλέπε ενότητα B.1.2.2.3).

Η φυσική τοπολογία των μονοπατιών από τα οποία δρομολογείται κάθε TT και κατ' επέκταση η συσχέτιση των TTs που ανταγωνίζονται για τους πόρους μιας ζεύξης κορμού δεν είναι γνωστή με αποτέλεσμα το αντίκτυπο της αποδοχής μιας νέας υπηρεσίας να μπορεί να υπολογιστεί μόνο για τα άμεσα εμπλεκόμενα TTs αλλά όχι και για τα εξίσου επιβαρημένα λόγω των έμμεσων επικαλύψεων TTs. Αυτό επιβάλλεται για την

επίτευξη της αποσύνδεσης υπηρεσιών και πόρων με σκοπό τον περιορισμό της πολυπλοκότητας του συστήματος και τελικώς τη δυνατότητα εφαρμογής της μεθόδου σε μεγάλη κλίμακα (βλέπε ενότητα B.1.2.2.2).

Το δίκτυο είναι "προγραμματισμένο" να διαθέτει τους πόρους σε εξασφαλισμένα hard-reserved μερίδια και σε κοινόχρηστα soft-reserved μερίδια για τα οποία όμως οι επικαλύψεις δεν μπορούν να διασταυρωθούν. Η μεγιστοποίηση των συμβολαίων εγγραφής που επιβάλλεται από το σκέλος (β) του προβλήματος, επιτάσσει τη χρήση των soft-reserved μεριδίων όταν υπάρχει ζήτηση ενώ η εξασφάλιση της ικανοποίησης των υπηρεσιών που επιβάλλεται από το σκέλος (α) επιτάσσει την αποφυγή των επικαλύψεων στα soft-reserved μερίδια. Ο βαθμός στον οποίο ο αλγόριθμος προάγει την αξιοποίηση των πόρων εις βάρος της ικανοποίησης των υπηρεσιών, με άλλα λόγια η συμπεριφορά του αλγορίθμου ως προς τη χρήση των soft-reserved μεριδίων με το συνακόλουθο ρίσκο υποβάθμισης της ικανοποίησης των υπηρεσιών, ρυθμίζεται από τη λειτουργική πολιτική του παροχέα (βλέπε ενότητα B.1.1.3).

Η λειτουργική πολιτική του παροχέα αποτυπώνεται στην παράμετρο *επιπέδου ικανοποίησης* (Satisfaction Level – SL) η οποία αναπαριστά το επίπεδο ικανοποίησης που ο παροχέας θεωρεί *επαρκές* να παρέχει στις εγγεγραμμένες υπηρεσίες και εκφράζεται με σχετικό τρόπο στο πεδίο τιμών [-1,1], όπου -1 αντιστοιχεί σε *καμία εγγύηση* (no-guarantees) ικανοποίησης της υπηρεσίας, 0 σε *σχεδόν ικανοποιημένη* (almost-satisfied) υπηρεσία και 1 σε *εντελώς ικανοποιημένη* (fully-satisfied) υπηρεσία.

Η λειτουργική πολιτική του παροχέα αντανακλά τις επιχειρησιακές επιδιώξεις του παροχέα. Ένα τυπικό παράδειγμα διαφοροποίησης είναι μεταξύ επιχειρήσεων μεταπράτησης και χονδρικής πώλησης. Οι επιχειρήσεις μεταπράτησης έχουν κατά κανόνα περισσότερους και πιο μικρούς πελάτες, από τους οποίους συνήθως μόνο ένα μικρό ποσοστό είναι προσοδοφόρο. Σύμφωνα με την ευρέως αποδεκτή αρχή του οικονομολόγου Pareto το 80% των εσόδων μιας εταιρείας προέρχεται από το 20% των πελατών της. Συνηθίζεται έτσι οι επιχειρήσεις μεταπράτησης να στοχεύουν σε περισσότερα συμβόλαια με μικρότερες εγγυήσεις. Αυτές οι επιχειρήσεις είναι λογικό να επιλέξουν μικρές, ακόμα και αρνητικές τιμές για το επίπεδο ικανοποίησης. Αντίθετα, οι επιχειρήσεις χονδρικής πώλησης έχουν λιγότερους και διακεκριμένους πελάτες η ικανοποίηση των οποίων είναι κρίσιμη για την επιτυχία του οργανισμού καθιστώντας αναγκαία τη ρύθμιση του επιπέδου ικανοποίησης σε θετικές τιμές.

Ο βαθμός ικανοποίησης μιας υπηρεσίας εκφράζεται, σύμφωνα με τους ορισμούς της ενότητας B.1.1.4, από το κλάσμα του ρυθμού ροής που παραδίδεται με τη συμφωνημένη ποιότητα προς τον ρυθμό ροής της προσφερόμενης συμμορφούμενης με τους όρους της υπηρεσίας κίνησης. Έτσι, το επίπεδο ικανοποίησης μεταφράζεται στο κλάσμα του ρυθμού ροής που πρέπει σε κάθε περίπτωση με ισχυρή βεβαιότητα να λαμβάνει τη συμφωνημένη ποιότητα, δηλαδή στους πόρους που πρέπει να είναι διαθέσιμοι με ισχυρή βεβαιότητα, δηλαδή στην περιοχή του RAB που θεωρείται επιτρεπτό να καλυφθεί από τη συνολική προσφερόμενη κίνηση (admissible area) έχοντας συνεκτιμήσει το ανάλογο ρίσκο επικάλυψης και άρα υποβάθμισης της ποιότητας.

Επισημαίνουμε πως τα στατιστικά λάθη που υπεισέρχονται στους υπολογισμούς των απαραίτητων και των διαθέσιμων πόρων δεν επιτρέπουν την a priori εξασφάλιση μιας απόλυτης τιμής βαθμού ικανοποίησης. Έτσι, ανάλογα με την αυστηρότητα των μοντέλων χαρακτηρισμού των πηγών και το κέρδος πολυπλεξίας κατά τη συσσωμάτωση των ροών (βλέπε ενότητα A.2.1.1) ενδέχεται να υποτιμούνται οι απαραίτητοι πόροι. Αντίστοιχα, εάν η πραγματική συμπεριφορά των χρηστών είναι σημαντικά πιο ήπια από τις προβλέψεις που έχουν χρησιμοποιηθεί για τη διαστασιοποίηση του δικτύου τότε αναμένεται να είναι σημαντική και η απόκλιση από τις εκτιμήσεις μακροπρόθεσμης διαθεσιμότητας του RAM οπότε το σύστημα θα είναι σε θέση να προσφέρει πολύ υψηλότερη ικανοποίηση. Επιπλέον, είναι προφανές πως ο βαθμός ικανοποίησης θα

διαφέρει για κάθε υπηρεσία ανάλογα με το φόρτο του δικτύου το διάστημα που τυχαίνει να ενεργοποιείται. Συνεπώς, η ρύθμιση του επιπέδου ικανοποίησης δεν προεξοφλεί το βαθμό ικανοποίησης που θα λάβουν τελικά οι υπηρεσίες, ο οποίος μπορεί να είναι μικρότερος ή μεγαλύτερος από τον ενδεδειγμένο κατά το επίπεδο ικανοποίησης, ανάλογα με τα στατιστικά λάθη των μοντέλων και την προβλεψιμότητα και ομοιογένεια στη συμπεριφορά των χρηστών.

Η χρησιμότητα του επιπέδου ικανοποίησης έγκειται στη δυνατότητα εμπειρικής ρύθμισης της συμπεριφοράς του αλγορίθμου. Η πιθανότητα υποβάθμισης της ποιότητας είναι αντιστρόφως ανάλογη προς την τιμή του επιπέδου ικανοποίησης. Συγκεκριμένα, όσο αυξάνεται το επίπεδο ικανοποίησης ο αλγόριθμος θα αποδέχεται λιγότερες αιτήσεις εγγραφής και άρα η πιθανότητα συμφόρησης στο δίκτυο θα μικραίνει. Ωστόσο, τα επίπεδα συμφόρησης στην πραγματικότητα εξαρτώνται από τη συμπεριφορά των χρηστών και συνεπώς η ορθότητα των ενεργειών του ελέγχου εισόδου από την ακρίβεια στη μοντελοποίησή της. Εφόσον δεν υπάρχουν ακριβή και ταυτόχρονα οικονομικά μοντέλα, η όποια ανακρίβεια στη μοντελοποίηση αντιμετωπίζεται από την εμπειρική ρύθμιση της παραμέτρου.

B.2.1.4 Αλγόριθμος

Η αρχικοποίηση του αλγορίθμου συμπίπτει με την έναρξη του RPC. Σε αυτή τη φάση υπολογίζεται βάσει του επιπέδου ικανοποίησης η επιτρεπτή περιοχή (admissible area) από τους πόρους του RAB που θα παραχωρηθούν προς χρήση στη συνολική προβλεπόμενη κίνηση όπως διαμορφώνεται από τις ήδη εγγεγραμμένες από το προηγούμενο RPC και τις νέες υπηρεσίες που θα προκύψουν στη διάρκεια του RPC που πρόκειται να ξεκινήσει. Ενώ η επιτρεπτή περιοχή είναι σταθερή η συνολική προσφερόμενη κίνηση μεταβάλλεται ανάλογα με τις νέες αιτήσεις που γίνονται αποδεκτές αλλά και τη διακοπή ή λήξη συμβολαίων εγγραφής. Μία νέα αίτηση γίνεται αποδεκτή εάν η εκάστοτε τρέχουσα συνολική προσφερόμενη κίνηση μαζί και με την κίνηση της νέας υπηρεσίας δεν υπερβαίνει την επιτρεπτή περιοχή, σε αντίθετη περίπτωση το σύστημα προχωρά σε διαπραγματεύσεις.

Στην παρακάτω αναλυτική περιγραφή θεωρούμε πως έχουν προηγηθεί τα στάδια της μετάφρασης και αντιστοίχισης της υπηρεσίας σε TTs. Ο αλγόριθμος χρησιμοποιείται σε επίπεδο TT και η αίτηση υπηρεσίας γίνεται αποδεκτή μόνο όταν η κίνηση που αντιστοιχεί σε κάθε εμπλεκόμενο TT γίνει αποδεκτή, η απόφαση σε επίπεδο υπηρεσίας είναι με άλλα λόγια το λογικό AND της απόφασης σε επίπεδο TT.

B.2.1.4.1 Αρχικοποίηση

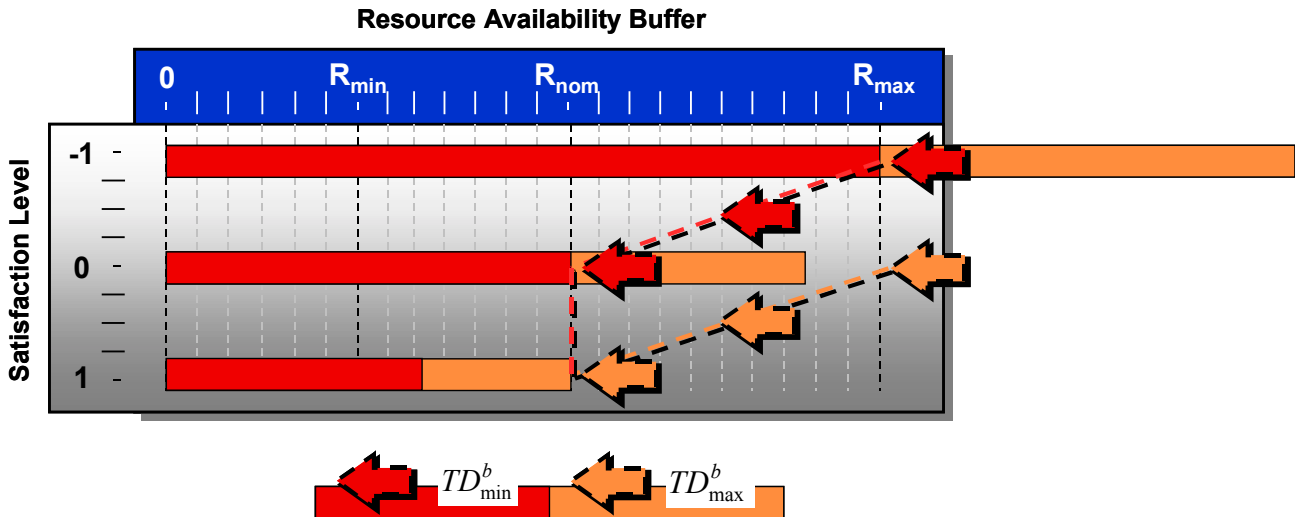
Η τρέχουσα συνολική προσφερόμενη κίνηση εκφράζει την μακροπρόθεσμα προσδοκώμενη κίνηση από τις εκάστοτε εγγεγραμμένες υπηρεσίες και εκφράζεται, όπως και το αντίστοιχο μέγεθος που προκύπτει από τη λειτουργία πρόβλεψης κίνησης από την ελάχιστη και μέγιστη τιμή έστω TD_{\min}^c και TD_{\max}^c που αντιστοιχούν σε ρυθμό SR_{AS} και SR_{FS} δηλαδή σε απαιτήσεις για almost-satisfied και fully-satisfied βαθμό ικανοποίησης.

Η τρέχουσα συνολική προσφερόμενη κίνηση αρχικοποιείται στη συνολική προσφερόμενη κίνηση από τις εγγεγραμμένες υπηρεσίες προ της έναρξης του νέου RPC η οποία προκύπτει από τους υπολογισμούς της λειτουργίας πρόβλεψης κίνησης (βλέπε ενότητα B.1.2.2.3) για τις υπάρχουσες εγγραφές (existing subscriptions) προ της διαδικασίας προέκτασης στο μέλλον (extrapolation).

Η επιτρεπτή περιοχή οριοθετείται μέσω των άνω ορίων των μεταβλητών TD_{\min}^c και TD_{\max}^c , TD_{\min}^b και TD_{\max}^b αντιστοίχως που υπολογίζονται βάσει του RAB και του SL ως εξής:

$$TD_{\min}^b = \begin{cases} R_{nom} + (R_{\max} - R_{nom}) \cdot |SL| & SL \leq 0 \\ R_{nom} & SL > 0 \end{cases} \quad TD_{\max}^b = \begin{cases} +\infty & SL \leq 0 \\ R_{nom} + (R_{\max} - R_{nom}) \cdot (1 - SL) & SL > 0 \end{cases}$$

Η φυσική ερμηνεία των TD_{\min}^b και TD_{\max}^b φαίνεται στο Σχήμα B-16.



Σχήμα B-16: καθορισμός επιτρεπτής περιοχής βάσει του επιπέδου ικανοποίησης

Το SL μεταφράζεται στο κλάσμα της κίνησης που θα πρέπει τη συμφωνημένη ποιότητα με ισχυρή βεβαιότητα. Η βεβαιότητα με την οποία είναι διαθέσιμοι οι πόροι διαφέρει ανά περιοχές του RAB (βλέπε ενότητα B.1.2.2.6 και Σχήμα B-15). Εφόσον το δίκτυο είναι προγραμματισμένο να παραχωρεί τουλάχιστον R_{nom} σε κάθε TT οι πόροι της περιοχής αυτής είναι εξασφαλισμένοι. Επιπλέον πόροι έως και το R_{\max} μπορούν να χρησιμοποιηθούν αλλά χωρίς εγγυήσεις. Το ποσοστό της συνολικής προσφερόμενης κίνησης που εμπίπτει στην περιοχή των εξασφαλισμένων πόρων R_{nom} αντανακλά συνεπώς, δεδομένων των στατιστικών λαθών, και το βαθμό ικανοποίησης.

Όταν το SL ισούται με 0 ο ενδεδειγμένος βαθμός ικανοποίησης είναι almost-satisfied, δηλαδή αρκεί η εξασφάλιση του TD_{\min}^c , ή αλλιώς ο περιορισμός του στην περιοχή R_{nom} , οπότε $TD_{\min}^b = R_{nom}$, ενώ δεν δίνονται περαιτέρω εγγυήσεις για μεγαλύτερο βαθμό ικανοποίησης, θέτοντας έτσι το TD_{\max}^b στο άπειρο. Όταν το SL ισούται με +1 ο ενδεδειγμένος βαθμός ικανοποίησης είναι fully-satisfied, είναι απαραίτητη δηλαδή η εξασφάλιση και του TD_{\max}^c , οπότε $TD_{\max}^b = R_{nom}$. Όταν το SL ισούται με -1 οποιαδήποτε εξασφάλιση θεωρείται περιττή, υιοθετώντας ως μόνο περιορισμό την αποφυγή της ντετερμινιστικά βέβαιης υποβάθμισης της ποιότητας θέτοντας $TD_{\min}^b = R_{\max}$.

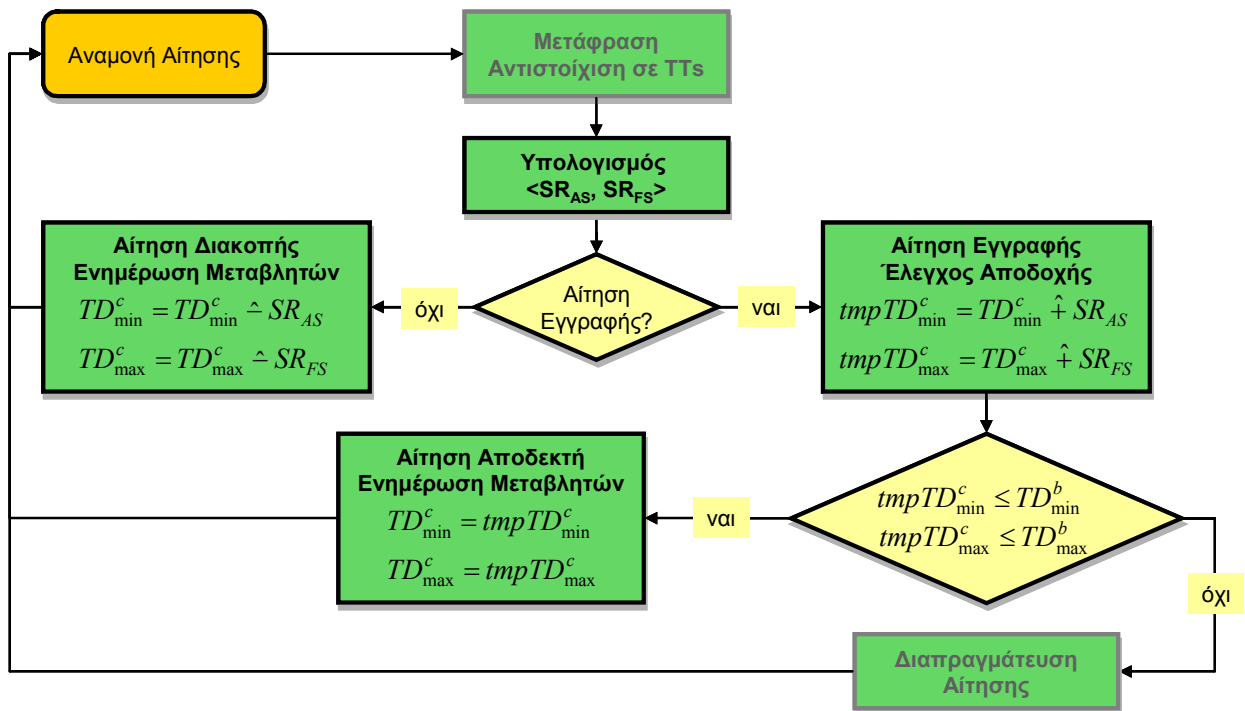
Όσο το SL μετατοπίζεται στη θετική περιοχή προς το +1 τότε, επιπλέον του TD_{\min}^c , και ένα όλο και μεγαλύτερο μέρος του TD_{\max}^c ανάλογο με το SL θα πρέπει να εμπίπτει στην εξασφαλισμένη περιοχή. Αντίστοιχα, όσο το SL μετατοπίζεται στην αρνητική περιοχή προς το -1 τότε όλο και μεγαλύτερο μέρος του TD_{\min}^c επιτρέπεται να ξεπερνά την εξασφαλισμένη περιοχή μέχρι και το απόλυτο όριο του R_{\max} .

Καθίσταται προφανές πως ο αλγόριθμος ελέγχου αποδοχής εγγραφής ρυθμίζεται από το SL να παρέχει από πρακτικά καμία εγγύηση ποιότητας έως απόλυτες εγγυήσεις με τη λογική του χειρότερου σεναρίου (worst-

case) με την οποία υπολογίζονται οι απαραίτητοι κατά το TD_{max}^c πόροι και απομονώνονται στεγανά τα μερίδια.

B.2.1.4.2 Λειτουργία

Ξεκινώντας από τις εγγεγραμμένες υπηρεσίες κατά την έναρξη της λειτουργίας ο αλγόριθμος ενημερώνει τη συνολική προσφερόμενη κίνηση, δηλαδή τις μεταβλητές TD_{min}^c και TD_{max}^c , βάσει των συμβολαίων που τερματίζονται έπειτα από αιτήσεις διακοπής και των νέων συμβολαίων που συνάπτονται με την αποδοχή των αντίστοιχων αιτήσεων εγγραφής (βλέπε Σχήμα B-17). Η προσφερόμενη κίνηση από κάθε υπηρεσία για ένα TT εξάγεται αφού η υπηρεσία αντιστοιχηθεί σε TTs και βάσει των μοντέλων εξαγωγής των SR_{AS} και SR_{FS} , ενώ η συνολική προσφερόμενη κίνηση παράγεται από τη συνέλιξη και συσσωμάτωση των επιμέρους ανά υπηρεσία SR_{AS} και SR_{FS} (βλέπε ενότητα B.1.2.2.3). Σε κάθε προσθήκη ($\hat{+}$) ή αφαίρεση ($\hat{-}$) της κίνησης από μια επιμέρους υπηρεσία η διαδικασία συσσωμάτωσης γίνεται με οικονομία αποφεύγοντας την εκ νέου συσσωμάτωση όλων των παραμενόντων επιμέρους υπηρεσιών. Οι συγκεκριμένες τεχνικές είναι εκτός του πλαισίου της παρούσας εργασίας, περισσότερες λεπτομέρειες δίνονται στο [TEQ-ARCH].



Σχήμα B-17: λειτουργία διαχείρισης εγγραφής

Ο έλεγχος αποδοχής μιας αίτησης εγγραφής συγκρίνει τη συνολική προσφερόμενη κίνηση από τις εγγεγραμμένες υπηρεσίες και την αίτηση εγγραφής ($tmpTD_{min}^c$ και $tmpTD_{max}^c$) με την επιτρεπτή περιοχή όπως έχει οριστεί κατά τη φάση αρχικοποίησης βάσει του RAB και του επιπέδου ικανοποίησης. Σε περίπτωση που η προσφερόμενη κίνηση δεν υπερβαίνει την επιτρεπτή περιοχή τότε η αίτηση εγγραφής γίνεται αποδεκτή και η τρέχουσα συνολική προσφερόμενη κίνηση ενημερώνεται κατάλληλα. Σε αντίθετη περίπτωση το σύστημα προχωρά σε διαπραγματεύσεις με πιθανότητα έστω P_{neg} , η τιμή της οποίας καθορίζεται κατά την αρχικοποίηση. Με πιθανότητα $(1-P_{neg})$ το σύστημα απορρίπτει την αίτηση χωρίς επανεξέταση. Η τιμή της παραμέτρου P_{neg} μπορεί να παραμένει σταθερή καθ' όλη τη διάρκεια της

λειτουργίας του συστήματος ή να μεταβάλλεται αντιστρόφως ανάλογα με το μέγεθος της τρέχουσας συνολικής προσφερόμενης κίνησης βάσει της λογικής του μηχανισμού RED (βλέπε ενότητα A.1.4.1.2).

Η διαδικασία της διαπραγμάτευσης εμπεριέχει τη λογική της διαμόρφωσης και εισήγησης εναλλακτικών προτάσεων (alternatives) που να είναι συμβατές με τις δυνατότητες από τους πόρους της επιτρεπτής περιοχής και κατά το δυνατό πλησιέστερες στις αρχικές προσδοκίες του πελάτη. Τέτοιες εναλλακτικές προτάσεις μπορεί να διαφοροποιούνται:

- (α) κατά την ποιότητα της υπηρεσίας (πεδίο performance guarantees στο SLS-T, βλέπε ενότητα B.1.2.1.2), προτείνοντας μια διαφορετική QoS-κλάση το αντίστοιχο TT της οποίας έχει επάρκεια σε διαθέσιμους πόρους,
- (β) κατά το ρυθμό εξυπηρέτησης (πεδίο traffic conformance στο SLS-T), όπου για την αρχικά αιτούμενη ποιότητα ο παροχέας εγγυάται την εξυπηρέτηση μικρότερου ρυθμού κίνησης,
- (γ) κατά τις εγγυήσεις ικανοποίησης της υπηρεσίας (πεδίο grade of service στο SSS-T), όπου η υπηρεσία παρέχεται αυτούσια αλλά με μικρότερη πιθανότητα επιτυχούς ενεργοποίησης και περάτωσης,
- (δ) κατά το χρονοδιάγραμμα της υπηρεσίας (πεδίο schedule στο SSS-T) όπου η υπηρεσία παρέχεται αυτούσια αλλά με δικαίωμα ενεργοποίησης σε διαφορετικό RPTP από το αιτούμενο (βλέπε ενότητα B.1.2.2.2) και
- (ε) σε συνδυασμό των παραπάνω παραμέτρων.

Η στρατηγική διαμόρφωσης, παραμετροποίησης, ρύθμισης των παραμέτρων ανάλογα με τις συνθήκες φόρτου και διαπραγμάτευσης των εναλλακτικών προτάσεων είναι πέρα από το πλαίσιο της παρούσας εργασίας.

B.2.2 Διαχείριση Κλήσης

Η λειτουργία διαχείρισης κλήσης (invocation management) είναι υπεύθυνη για τον έλεγχο εισόδου κίνησης. Το πρόβλημα του ελέγχου εισόδου κίνησης σε δίκτυα DiffServ περιγράφεται αναλυτικά στην ενότητα A.2.1 ενώ στην ενότητα B.1.1 παρουσιάζεται η τοποθέτηση στο πρόβλημα και οι βασικές αρχές του προτεινόμενου feedback-based μοντέλου. Ο έλεγχος εισόδου αποκόπτεται από τις συναφείς λειτουργίες και τα υποκείμενα μοντέλα τα οποία περιγράφονται στην ενότητα B.1.2 και αποτελούν το περιβάλλον πλαίσιο λειτουργίας. Οι αιτήσεις που παραδίδονται προς έγκριση στον έλεγχο εισόδου κίνησης πρέπει να αντιστοιχούν σε συμβόλαια εγγραφής, αλλιώς απορρίπτονται από το στάδιο της εξουσιοδότησης. Έλεγχος αποδοχής των συμβολαίων εγγραφής εφαρμόζεται και κατά τη φάση εγγραφής από τη λειτουργία διαχείρισης εγγραφής (βλέπε ενότητα B.2.1) που ρυθμίζει με αυτό τον τρόπο τον αριθμό και τον τύπο των αιτήσεων κλήσης που θα κληθεί να εξυπηρετήσει ο έλεγχος εισόδου κίνησης.

Στις ακόλουθες ενότητες εξετάζουμε το πρόβλημα, την τοποθέτηση και την προτεινόμενη λύση στο ειδικότερο πρόβλημα ελέγχου εισόδου όπως διαμορφώνεται δεδομένων των ως άνω επιλογών.

B.2.2.1 Πρόβλημα

Η λειτουργία διαχείρισης κλήσης ρυθμίζει την κίνηση που διοχετεύεται στο δίκτυο ανά πάσα στιγμή ελέγχοντας αφενός τη διαδικασία αποδοχής αιτήσεων κλήσης υπηρεσίας και αφετέρου τις ρυθμίσεις της ποιότητας για τις ενεργές ροές και επιδιώκοντας:

- (α) την αποτροπή της υποβάθμισης της ποιότητας των ενεργών υπηρεσιών λόγω υπερφόρτωσης του δικτύου,
- (β) τη μεγιστοποίηση της ικανοποίησης των υπηρεσιών, δηλαδή του αριθμού των αιτήσεων κλήσης που ενεργοποιούνται επιτυχώς και της ποιότητας που λαμβάνουν καθ' όλη τη διάρκεια χρήσης. Δεδομένης της ζήτησης, η μεγιστοποίηση της ικανοποίησης των υπηρεσιών ισοδυναμεί με μεγιστοποίηση του κέρδους του παροχέα.
- (γ) την άμεση και αποτελεσματική επίλυση ενδεχόμενης συμφόρησης,
- (δ) την αμερόληπτη μεταχείριση των υπηρεσιών σε όλη την επικράτεια του δικτύου.

Η προτεινόμενη λύση είναι απαραίτητο να πληροί τις απαιτήσεις πρακτικής εφαρμογής, ευστάθειας και ευρωστίας (βλέπε ενότητα A.2.1.3) ενώ η αποτελεσματικότητα του συστήματος, ως προς την επιβάρυνση που προκαλεί στους χρήστες και στο δίκτυο, αποτελεί ένα από τα κριτήρια αξιολόγησης της προτεινόμενης λύσης.

B.2.2.2 Παραδοχές και Περιβάλλον

Σύμφωνα και με το διαδικαστικό μοντέλο (βλέπε ενότητα B.1.2.1.1), της κλήσης μιας υπηρεσίας πρέπει απαραίτητα να έχει προηγηθεί εγγραφή στην υπηρεσία. Αποκλείουμε συνεπώς τη δυνατότητα ενεργοποίησης μιας υπηρεσίας χωρίς αντίστοιχο καταχωρημένο συμβόλαιο εγγραφής. Αυτό δεν αποτελεί περιορισμό, αλλά θεωρούμε πως πριν από την κλήση και ενεργοποίηση μιας νέας υπηρεσίας επιβάλλεται να παρεμβληθεί το επιπλέον στάδιο της εγγραφής.

Ισχύουν συνεπώς και οι σχετικές παραδοχές της λειτουργίας διαχείρισης εγγραφής (βλέπε ενότητα B.2.1.2) ως προς την οριοθέτηση των πόρων στους οποίους αφορά η εφαρμογή του ελέγχου εισόδου κίνησης, εξαιρώντας τους πόρους επιπέδων ανώτερων του επιπέδου μεταφοράς και πόρους πέραν του δικτύου κορμού.

Επισημαίνουμε πως η λειτουργία διαχείρισης εγγραφής δεν θεωρείται προϋπόθεση. Ο έλεγχος εισόδου κίνησης είναι σε θέση να λειτουργήσει βάσει των συμβολαίων εγγραφής, ανεξάρτητα με τη διαδικασία που χρησιμοποιείται για τη σύνασή τους και αν αυτή λαμβάνει υπόψη κάποιο κριτήριο για τη διαθεσιμότητα των πόρων. Η επίδραση της λογικής με την οποία συνάπτονται τα συμβόλαια αφορά στον όγκο των εξουσιοδοτημένων αιτήσεων που αξιώνουν ικανοποίησης από τον έλεγχο εισόδου κίνησης και συνεπώς στα επίπεδα ικανοποίησης που πρόκειται να λάβουν εφόσον ανταγωνίζονται για πεπερασμένους πόρους δικτύου.

Θεωρούμε πως η λειτουργία διαχείρισης εγγραφής είναι κατανομημένη στους ακραίους κόμβους και χρησιμοποιεί το feedback-based μοντέλο (βλέπε ενότητα B.1.1).

Θεωρούμε την ύπαρξη ενός κατάλληλου πρωτοκόλλου για την μεταξύ χρηστών και ακραίων κόμβων για την ενεργοποίηση των υπηρεσιών ρητής κλήσης (βλέπε ενότητα B.1.2.1.1.2). Σημειώνουμε πως οι λεπτομέρειες που αφορούν το πρωτόκολλο είναι πέραν του πλαισίου της παρούσας διατριβής.

Σύμφωνα και με το λειτουργικό μοντέλο (βλέπε ενότητα B.1.2.2.1) θεωρούμε τις λειτουργίες Traffic Conditioning και παρακολούθησης.

Η λειτουργία Traffic Conditioning υλοποιεί τους DiffServ μηχανισμούς ταξινόμησης και ρύθμισης της κίνησης στους ακραίους κόμβους (βλέπε ενότητα A.1.4.2.2.1). Δεδομένων των ενεργειών του traffic engineering οι μηχανισμοί αυτοί ρυθμίζουν την ποιότητα στην οποία αποκτούν πρόσβαση οι ροές των πακέτων, συνιστούν με άλλα λόγια τα μέσα υλοποίησης των αποφάσεων του ελέγχου εισόδου κίνησης.

Οι λειτουργίες παρακολούθησης υλοποιούν τους μηχανισμούς μέτρησης και αποστολής αναφορών που χρησιμοποιούνται βάσει του feedback-based μοντέλου για τη λήψη των αποφάσεων ελέγχου εισόδου κίνησης. Ειδικότερα θεωρούμε πως οι λειτουργίες παρακολούθησης είναι σε θέση να παρέχουν ανά TT περιοδικές αναφορές για τα επίπεδα του φόρτου τοπικά (load) και ειδοποιήσεις (notifications) κατά την υπέρβαση ενός κατωφλιού στην καθυστέρηση (delay) και το ποσοστό απώλειας (loss) από άκρο σε άκρο.

Σύμφωνα με το feedback-based μοντέλο θεωρούμε την κατανομή των πόρων σε hard-reserved και soft-reserved μερίδια ανά ακραίο κόμβο με a priori partitioning, ή αλλιώς τη διαστασιοποίηση των πόρων του δικτύου όπως αποτυπώνεται ανά TT στον πίνακα διαθεσιμότητας πόρων RAM (βλέπε ενότητα B.1.2.2.6).

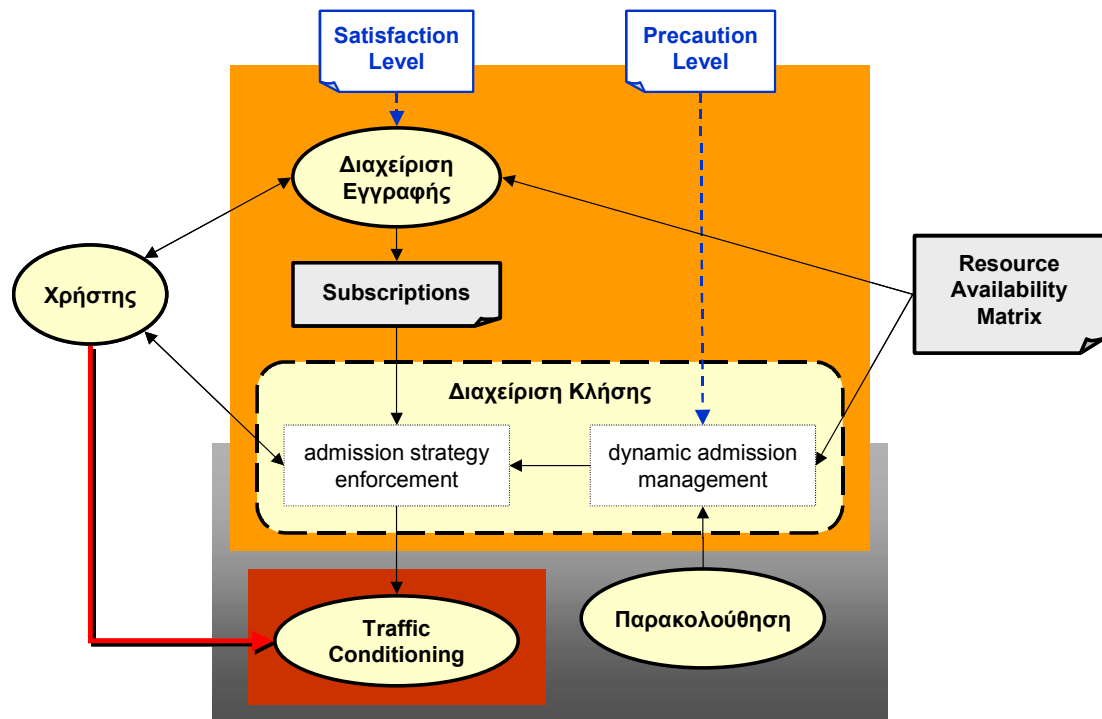
Επισημαίνουμε πως οι λειτουργίες του δυναμικού traffic engineering δεν θεωρούνται προϋπόθεση. Ο βέλτιστος καταμερισμός του φόρτου σε εναλλακτικά μονοπάτια και η αναπροσαρμογή της κατανομής των πόρων ανά PHB στους κόμβους κορμού απλά επαυξάνουν τη δεδομένη δυνατότητα δανεισμού μεταξύ των κοινόχρηστων soft-reserved μεριδίων. Το partitioning των πόρων σε επίπεδο TT γίνεται μόνο σε λογικό επίπεδο καθιστώντας έτσι δυνατή τη χρήση από ένα και μόνο TT των πόρων που έχουν ανατεθεί στο αντίστοιχο PHB σε κάθε ζεύξη ενός μονοπατιού όταν δεν συναντά διασταυρωνόμενη κίνηση (cross-traffic) από άλλα TTs. Το λογικό partitioning των πόρων σε επίπεδο TT είναι απαραίτητο ως κατευθυντήρια γραμμή (guideline) για τη συντονισμένη λειτουργία και επίλυση της συμφόρησης από τα ανά ακραίο κόμβο κατανεμημένα στιγμιότυπα (instances) ελέγχου εισόδου κίνησης.

B.2.2.3 Τοποθέτηση

Ο έλεγχος εισόδου επιδρά αφενός στην αποδοχή κίνησης από τις ενεργές ροές και αφετέρου στην αποδοχή νέων ενεργών ροών. Σύμφωνα με το feedback-based μοντέλο η διαδικασία λήψης αποφάσεων βασίζεται και κατ' επέκταση πυροδοτείται από τη μεταβολή της συνολικής κίνησης και της κατάστασης του δικτύου όπως προκύπτει από τις μετρήσεις. Ωστόσο, η ανεξάρτητη διαδικασία εξυπηρέτησης αιτήσεων κλήσης που υλοποιεί τις αποφάσεις ως προς την αποδοχή νέων ενεργών ροών, δεν συνδέεται χρονικά με τις μετρήσεις της κίνησης αλλά πυροδοτείται από την άφιξη των αιτήσεων κλήσης.

Ονομάζουμε *στρατηγική αποδοχής* (admission strategy) την τοποθέτηση της διαδικασίας λήψης αποφάσεων ως προς κάθε παράγοντα που επηρεάζει την κίνηση που διοχετεύεται στο δίκτυο και διακρίνουμε μεταξύ των διαδικασιών λήψης και υλοποίησης των αποφάσεων οι οποίες ανατίθενται αντίστοιχα στις λειτουργίες *δυναμικής διαχείρισης αποδοχής* (dynamic admission management) και *εφαρμογής στρατηγικής αποδοχής* (admission strategy enforcement) (βλέπε Σχήμα B-18).

Αντιδρώντας στις αναφορές από την παρακολούθηση, η λειτουργία δυναμικής διαχείρισης αποδοχής διαθέτει τη λογική να αποφανθεί την κατάλληλη στρατηγική αποδοχής την οποία μεταθέτει προς υλοποίηση στη λειτουργία εφαρμογής στρατηγικής αποδοχής. Ο αλγόριθμος ελέγχου εισόδου συνοψίζεται στον καθορισμό των συνθηκών μετάβασης σε νέα στρατηγική και της μεθόδου διαμόρφωσης της εκάστοτε ενδεδειγμένης στρατηγικής αποδοχής. Η λειτουργία εφαρμογής στρατηγικής αποδοχής επικοινωνεί με το χρήστη για το χειρισμό των αιτήσεων κλήσης και με τον ακραίο κόμβο για την εφαρμογή των ρυθμίσεων κίνησης (traffic conditioning) που υλοποιούν τις ρυθμίσεις ποιότητας για κάθε ενεργή ροή βάσει του αντίστοιχου συμβολαίου υπηρεσίας και των επιταγών της τρέχουσας στρατηγικής αποδοχής. Η σχέση των λειτουργιών δυναμικής διαχείρισης αποδοχής και εφαρμογής στρατηγικής αποδοχής οριοθετείται από τη στρατηγική αποδοχής, δηλαδή τους παράγοντες οι οποίοι επηρεάζουν την κίνηση που διοχετεύεται στο δίκτυο και είναι υπό τον έλεγχο της λειτουργίας διαχείρισης κλήσης.



Σχήμα B-18: λειτουργία διαχείρισης κλήσης

Ο διαχωρισμός της διαδικασίας εφαρμογής από τη διαδικασία επιλογής της κατάλληλης στρατηγικής επιτρέπει την αποσύνδεση της λογικής ελέγχου εισόδου από τις υποκείμενες τεχνολογίες για τη ρύθμιση της ποιότητας υπηρεσίας, το μοντέλο περιεχομένου της υπηρεσίας, τα πρωτόκολλα επικοινωνίας με το χρήστη, τις διαδικασίες εξουσιοδότησης και ενεργοποίησης, αλλά κυρίως απαλλάσσει τη διαδικασία εξυπηρέτησης αιτήσεων κλήσης από κάθε επιβάρυνση σχετική με τη λήψη αποφάσεων ελαχιστοποιώντας τον χρόνο απόκρισης.

Ο αλγόριθμος ελέγχου εισόδου κίνησης εφαρμόζεται ανά ΤΤ. Η διαθεσιμότητα των πόρων του RAM εκφράζεται ανά ΤΤ και οι μετρήσεις λαμβάνονται τοπικά και από άκρο σε άκρο ανά ΤΤ. Συνεπώς, σε κάθε ΤΤ αντιστοιχεί και εφαρμόζεται διαφορετική στρατηγική αποδοχής ανάλογα με τις επικρατούσες συνθήκες όπως αποτυπώνονται στις μετρήσεις. Το στιγμιότυπο της λειτουργίας διαχείρισης κλήσης σε έναν ακραίο κόμβο είναι σε θέση να αποφανθεί για τα ΤΤs για τα οποία αποτελεί κόμβο εισόδου (ingress). Στη γενική περίπτωση μια υπηρεσία αντιστοιχεί σε περισσότερα από ένα ΤΤs και μάλιστα με διαφορετικούς κόμβους εισόδου. Συνεπώς, η έγκριση για την αποδοχή μιας αίτησης κλήσης υπηρεσίας λαμβάνεται με τη σύγκλιση κάθε στρατηγικής σε κάθε εμπλεκόμενο ακραίο κόμβο. Οι αρχές και οι μηχανισμοί συντονισμού emπίπτουν στην εφαρμογή της στρατηγικής κατά τη διαδικασία εξυπηρέτησης αιτήσεων κλήσης που παρουσιάζεται λεπτομερώς στην ενότητα B.2.2.4.

Η ενδεδειγμένη στρατηγική αποδοχής είναι η περισσότερο παραχωρητική στρατηγική που δεν οδηγεί σε υπερφόρτωση του δικτύου, με άλλα λόγια η στρατηγική που έχει ως αποτέλεσμα τη μέγιστη αξιοποίηση των πόρων χωρίς την υποβάθμιση της ποιότητας των ενεργών υπηρεσιών, ικανοποιώντας αντίστοιχα τις διαστάσεις (β) και (α) του προβλήματος (βλέπε ενότητα B.2.2.1). Δεδομένου του σχεδιασμού του βασικού feedback-based μοντέλου, το ενδεχόμενο επικάλυψης στη χρήση κοινόχρηστων μεριδίων και συνεπώς υπερφόρτωσης και συμφόρησης, δεν είναι δυνατό να προβλεφθεί με ακρίβεια. Υποτίμηση του ρίσκου υπερφόρτωσης και εφαρμογή περισσότερο παραχωρητικών στρατηγικών οδηγεί σε πιο συχνή και μεγαλύτερης διάρκειας συμφόρηση και ισοδυναμεί με σφοδρότερη υποβάθμιση της ποιότητας υπηρεσίας των ενεργών ροών. Αντίθετα, υπερεκτίμηση του ρίσκου υπερφόρτωσης και εφαρμογή περισσότερο

επιφυλακτικών στρατηγικών οδηγεί σε αδικαιολόγητη απόρριψη αιτήσεων που ισοδυναμεί με σφοδρότερη υποβάθμιση της διαθεσιμότητας της υπηρεσίας και λιγότερο εντατική αξιοποίηση των πόρων. Ο βαθμός στον οποίο η προφύλαξη του συστήματος είναι αφενός δικαιολογημένη και αφετέρου επιθυμητή καθορίζεται από τη λειτουργική πολιτική του παροχέα (βλέπε ενότητα B.1.1.3).

Η λειτουργική πολιτική του παροχέα αποτυπώνεται στην παράμετρο *επίπεδο προφύλαξης* (Precaution Level – PL), η οποία περιγράφει το κόστος επί της βέλτιστης αξιοποίησης των πόρων που ο παροχέας θεωρεί αποδεκτό και δικαιολογημένο προκειμένου να διαφυλάσσεται η ποιότητα των ενεργών υπηρεσιών. Το επίπεδο προφύλαξης εκφράζεται με σχετικό τρόπο στο πεδίο τιμών $[0,1]$, όπου 0 αντιστοιχεί σε *επιφυλακτικό* (conservative) και 1 σε *αισιόδοξο* (optimistic).

Το επίπεδο προφύλαξης πρέπει να αντιστοιχεί στην πιθανότητα του ενδεχομένου συμφόρησης όπως εκτιμάται από τον παροχέα βάσει αφενός της συμπεριφοράς των χρηστών και της καταγιστικής ή όχι φύσεως της κίνησης στο συγκεκριμένο περιβάλλον και αφετέρου της επάρκειας των φυσικών πόρων για την κάλυψη των αναγκών της κίνησης από τα συμβόλαια εγγραφής. Στην περίπτωση για παράδειγμα του over-provisioning (βλέπε ενότητα A.1.4.3.3) κάθε προφύλαξη θεωρείται αδικαιολόγητη. Δεδομένου ενός περιορισμένου εξοπλισμού σε φυσικούς πόρους, η επάρκειά τους εξαρτάται από το πλήθος των εγγεγραμμένων υπηρεσιών, το οποίο με τη σειρά του ρυθμίζεται από το επίπεδο ικανοποίησης της λειτουργίας διαχείρισης εγγραφής. Όταν το επίπεδο ικανοποίησης έχει τεθεί σε χαμηλές τιμές, περισσότερες αιτήσεις εγγραφής γίνονται αποδεκτές και στη συνέχεια περισσότερες αιτήσεις κλήσης ανταγωνίζονται για τη χρήση των πεπερασμένων πόρων. Είναι φυσικό σε αυτή την περίπτωση να απαιτείται μεγαλύτερη προφύλαξη κατά τη χρήση των κοινόχρηστων soft-reserved μεριδίων καθώς η πιθανότητα να έχουν ήδη καταληφθεί από άλλες ενεργές ροές είναι μεγαλύτερη. Αντίθετα, υψηλότερο επίπεδο ικανοποίησης θα σημαίνει μικρότερο ανταγωνισμό, και συνεπώς μικρότερη πιθανότητα επικάλυψης στη χρήση των κοινόχρηστων μεριδίων οπότε και περιττή την προληπτική προφύλαξη.

Συνοψίζοντας, ο έλεγχος εισόδου κίνησης καλείται να υλοποιηθεί με βέλτιστο τρόπο τις εγγυήσεις που έχουν παραχωρηθεί από τον παροχέα στους πελάτες μέσω των συμβολαίων εγγραφής. Το επίπεδο ικανοποίησης της λειτουργίας διαχείρισης εγγραφής είναι αυτό που καθορίζει τον εν δυνάμει μέγιστο βαθμό ικανοποίησης των υπηρεσιών. Στα περιθώρια που οριοθετούνται από το επίπεδο ικανοποίησης, το επίπεδο προφύλαξης της λειτουργίας διαχείρισης κλήσης καθορίζει αφενός την απόκλιση από το εν δυνάμει μέγιστο λόγω ασυμφωνίας με την πραγματική πιθανότητα συμφόρησης και αφετέρου τον τύπο της υποβάθμισης της ικανοποίησης, είτε σε ποιότητα υπηρεσίας λόγω συμφόρησης κατά τη χρήση, είτε σε διαθεσιμότητα λόγω αποκλεισμού από τη χρήση με την απόρριψη των αιτήσεων κλήσης.

Η στρατηγική αποδοχής προσαρμόζεται ανάλογα με τα επίπεδα του φόρτου και την αναλόγως κυμαινόμενη πιθανότητα συμφόρησης. Το επίπεδο προφύλαξης καθορίζει αφενός την κρισιμότητα που αποδίδεται στις εκάστοτε συνθήκες του φόρτου και αφετέρου την προφύλαξη που κρίνεται απαραίτητη ανάλογα με την κρισιμότητα της κατάστασης. Το πρώτο επιτυγχάνεται με το "χρωματισμό" του αποθέματος των διαθέσιμων πόρων RAB σε περιοχές διαφορετικής κρισιμότητας ενώ το δεύτερο επηρεάζοντας την αυστηρότητα της στρατηγικής αποδοχής σε κάθε περίπτωση.

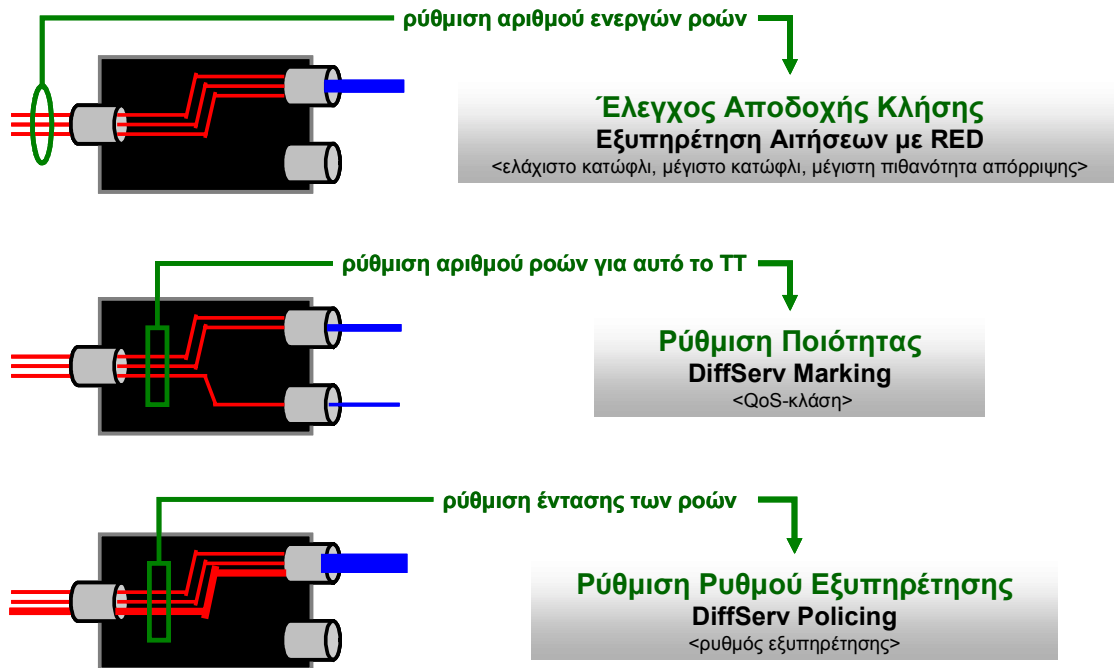
B.2.2.4 Εφαρμογή Στρατηγικής Αποδοχής

Η στρατηγική αποδοχής είναι ο συνδυασμός κρίσης μεταξύ της εξυπηρέτησης των υπηρεσιών και της κατάστασης του δικτύου. Η δυναμική διαχείριση αποδοχής αποφαίνεται πόσο παραχωρητική στρατηγική αποδοχής ενδείκνυται με κριτήριο το φόρτο και τους διαθέσιμους πόρους του δικτύου. Η στρατηγική αποδοχής διαμορφώνεται βάσει των παραγόντων που διέπουν την εξυπηρέτηση των υπηρεσιών. Η επέμβαση

σε αυτούς τους παράγοντες αφενός καθορίζει ποιοτικά και ποσοτικά την ικανοποίηση των υπηρεσιών και αφετέρου συμβάλλει στην μεταβολή του φόρτου του δικτύου στην επιθυμητή κατεύθυνση όπως αποφασίστηκε από τη δυναμική διαχείριση αποδοχής.

B.2.2.4.1 Παράγοντες Στρατηγικής Αποδοχής

Οι παράγοντες που ρυθμίζουν την εξυπηρέτηση των υπηρεσιών είναι η πιθανότητα αποδοχής κλήσης (admission probability) για τις εξουσιοδοτημένες αιτήσεις κλήσης, η ποιότητα υπηρεσίας (quality class) και ο ρυθμός εξυπηρέτησης (service rate) με πρόσβαση στην ποιότητα υπηρεσίας για τις επιτυχώς ενεργοποιημένες αιτήσεις κλήσης (βλέπε Σχήμα B-19).



Σχήμα B-19: επίδραση στρατηγικής αποδοχής στη συνολική διοχετευόμενη κίνηση

B.2.2.4.1.1 Έλεγχος Αποδοχής Κλήσης

Ο έλεγχος αποδοχής κλήσης αφορά τις αιτήσεις κλήσης των υπηρεσιών ρητής κλήσης (explicit invocation) (βλέπε ενότητα B.1.2.1.1.2). Σύμφωνα με το μοντέλο υπηρεσιών που έχουμε υιοθετήσει, υπηρεσίες έμμεσης κλήσης που ενεργοποιούνται αυτόματα από τον παροχέα (implicit invocation) είναι μόνο υπηρεσίες μόνιμης ενεργοποίησης (permanent services), η διαθεσιμότητά των οποίων θεωρείται ντετερμινιστικά και όχι στατιστικά διασφαλισμένη, όπως προκύπτει και από το περιεχόμενο του συμβολαίου που δεν προβλέπει προσδιορισμό του βαθμού διαθεσιμότητας για τις υπηρεσίες μόνιμης ενεργοποίησης (βλέπε ενότητα B.1.2.1.2). Επισημαίνουμε πως αυτή η θεώρηση δεν είναι περιοριστική. Η λειτουργία διαχείρισης κλήσης είναι υπεύθυνη για την ενεργοποίηση αμφοτέρων των υπηρεσιών ρητής και των υπηρεσιών έμμεσης κλήσης. Είναι συνεπώς δυνατό, πέραν της αποδοχής και στη συνέχεια ενεργοποίησης των αιτήσεων κλήσης από το χρήστη, να ελέγχεται και η έμμεση ενεργοποίηση των υπηρεσιών και να υπόκειται σε περιορισμούς κατά τις επιταγές της στρατηγικής αποδοχής.

Στην πλειονότητα των γνωστών measurement-based προσεγγίσεων (βλέπε ενότητα A.2.3.3), ο αλγόριθμος θέτει τη διαδικασία εξυπηρέτησης των αιτήσεων σε κατάσταση αποδοχής ή απόρριψης (accepting/rejecting state) ανάλογα με τις συνθήκες του φόρτου του δικτύου. Γενικεύοντας αυτή τη ντετερμινιστική λογική,

θεωρούμε πως κάθε σύστημα λειτουργεί βάσει μιας πιθανότητας απόρριψης, έστω P_{rej} . Στην προαναφερθείσα κατάσταση αποδοχής ή απόρριψης παίρνει αντίστοιχα τιμές 0 και 1. Είναι λογικό η πιθανότητα αυτή να είναι ανάλογη με το φόρτο του δικτύου, όσο αυτός αυξάνεται τόσο λιγότερες αιτήσεις θα γίνονται αποδεκτές.

Το σχήμα που υιοθετούμε υπολογίζει δυναμικά αυτή την πιθανότητα εφαρμόζοντας τη λογική του μηχανισμού διαχείρισης καταχωρητή RED (βλέπε ενότητα A.1.4.1.2). Η πιθανότητα απόρριψης P_{rej} κυμαίνεται ανάλογα με το φόρτο (load) στο TT, παραμένει μηδενική μέχρι και το *ελάχιστο κατώφλι* του φόρτου, έστω L_{min} , και γίνεται ίση με τη *μέγιστη πιθανότητα απόρριψης*, έστω P_{rej}^{max} , στο *μέγιστο κατώφλι* έστω L_{max} . Συγκεκριμένα,

$$P_{rej} = \begin{cases} 0 & L < L_{min} \\ P_{rej}^{max} \cdot \frac{L - L_{min}}{L_{max} - L_{min}} & L_{min} \leq L \leq L_{max} \\ 1 & L > L_{max} \end{cases}$$

Τα κατώφλια L_{min} και L_{max} τοποθετούνται στα όρια του αποθέματος RAB και μαζί με την P_{rej}^{max} καθορίζουν και την αυστηρότητα του αλγορίθμου. Όσο μικρότερο το L_{min} τόσο νωρίτερα εγκαταλείπεται η ντετερμινιστική αποδοχή. Όσο μικρότερο το L_{max} τόσο νωρίτερα μπαίνει σε ισχύ η ντετερμινιστική απόρριψη. Τέλος όσο μεγαλύτερη η P_{rej}^{max} , τόσο λιγότερο παραχωρητική και ταχέως φθίνουσα η πιθανοτική αποδοχή νέων αιτήσεων. Είναι προφανές ότι ο μηχανισμός RED εκφυλίζεται σε ντετερμινιστική διαδικασία όταν L_{min} τίθεται ίσο με L_{max} .

Ο αριθμός των ενεργών ροών ρυθμίζεται από τη διαδικασία εξυπηρέτησης αιτήσεων κλήσης. Ο αντίστοιχος παράγοντας ελέγχου αποδοχής κλήσης στη στρατηγική αποδοχής γίνεται με τη ρύθμιση των τριών παραμέτρων L_{min} , L_{max} και P_{rej}^{max} .

B.2.2.4.1.2 Ρύθμιση Ποιότητας

Η ποιότητα υπηρεσίας που λαμβάνουν οι ροές είναι αυτή της QoS-κλάσης στην οποία ανατίθενται. Οι λειτουργίες του traffic engineering είναι υπεύθυνες για τη διασφάλιση της ποιότητας υπηρεσίας ανά QoS-κλάση. Διακρίνουμε μεταξύ *ποσοτικών* (quantitative) και *ποιοτικών* (qualitative) υπηρεσιών, ανάλογα με την μέθοδο αποτίμησης των εγγυήσεων ποιότητας (performance guarantees) στο συμβόλαιο υπηρεσίας (βλέπε ενότητα B.1.2.1.2).

Η ανάθεση των ποσοτικών υπηρεσιών σε QoS-κλάσεις γίνεται με μονοσήμαντο και αδιαμφισβήτητο τρόπο αντιστοιχίζοντας την ποιότητα της υπηρεσίας στην ποιότητα της QoS-κλάσης. Θεωρώντας πως το traffic engineering διασφαλίζει την προκαθορισμένη ποιότητα των QoS-κλάσεων, η αντιστοιχία αυτή δεν μεταβάλλεται καθιστώντας απαγορευτική την ανάθεση μιας ποσοτικής υπηρεσίας σε QoS-κλάσεις κατώτερης ποιότητας.

Αντίθετα, η ανάθεση των ποιοτικών υπηρεσιών σε QoS-κλάσεις εξαρτάται από την ερμηνεία που ο παροχέας αποδίδει στις ποιοτικές υπηρεσίες. Ελλείψει ποσοτικών εγγυήσεων, ο παροχέας έχει την ελευθερία, με κριτήριο τους εκάστοτε διαθέσιμους πόρους και τα επίπεδα συμφόρησης σε κάθε QoS-κλάση, να υποβαθμίζει και να αναβαθμίζει την ποιότητα που παρέχεται στις ποιοτικές υπηρεσίες αναθέτοντάς τις σε QoS-κλάσεις κατώτερης και ανώτερης ποιότητας αντίστοιχα με χαμηλότερο φόρτο.

Το επιτρεπτό ανά ποιοτική υπηρεσία φάσμα των QoS-κλάσεων και κατ' επέκταση της ποιότητας ορίζεται εκ των προτέρων. Εάν π.χ. ένας παροχέας διαθέτει τις QoS-κλάσεις *C1*, *C2* και *C3*, οι βαθμωτές ποιοτικές υπηρεσίες *χρυσή*, *ασημένια* και *χάλκινη* θα μπορούσαν να υλοποιούνται από τις [*C1-C2-C3*], [*C2-C3*] και [*C3*] αντίστοιχα. Έτσι, εάν στην QoS-κλάση *C1* συμβεί συμφόρηση τότε οι ενεργές υπηρεσίες με εγγυήσεις *χρυσής* υπηρεσίας θα ανατεθούν στην QoS-κλάση *C2* μέχρι να μειωθεί επαρκώς το επίπεδο του φόρτου στη *C1*. Εάν και η *C2* παρουσιάσει συμφόρηση τότε η *χρυσή* και η *ασημένια* υπηρεσία θα ανατεθεί στην QoS-κλάση *C3*. Με αυτό τον τρόπο ευνοούνται και διαφυλάσσονται καλύτερα οι περισσότερο δεσμευτικές για τον παροχέα ποσοτικές εγγυήσεις.

Η ανά ΤΤ στρατηγική αποδοχής προσδιορίζει εάν μπορούν να ανατεθούν ποιοτικές υπηρεσίες σε αυτό το ΤΤ, δηλαδή στην αντίστοιχη με το ΤΤ QoS-κλάση. Εάν όχι, οι ενεργές ροές ποιοτικών υπηρεσιών θα ανατεθούν στο ΤΤ με QoS-κλάση την επόμενη κατώτερη όπως προκύπτει από το επιτρεπτό φάσμα της κάθε υπηρεσίας. Εάν η QoS-κλάση αυτού του ΤΤ είναι η κατώτερη τότε παραμένουν στο ΤΤ. Στο ως άνω παράδειγμα, έστω και αν η στρατηγική αποδοχής ενός ΤΤ της QoS-κλάσης *C3* αποφανθεί απαλλαγή από τις ποιοτικές υπηρεσίες, καμία αλλαγή δεν πρόκειται να συμβεί. Αντίστροφα, όταν η στρατηγική αποδοχής αποφανθεί πως τα επίπεδα του φόρτου είναι αρκούντως χαμηλά τότε οι ενεργές ποιοτικές υπηρεσίες θα ανακληθούν και θα ανατεθούν εκ νέου στην αντίστοιχη ανώτερη QoS-κλάση. Με αυτό τον τρόπο οι ποιοτικές υπηρεσίες λαμβάνουν ανά πάσα στιγμή την καλύτερη δυνατή ποιότητα χωρίς να ζημιώνουν την ποιότητα των ποσοτικών υπηρεσιών.

Η ανάθεση της ροής μιας υπηρεσίας σε μια QoS-κλάση γίνεται με την *σημείωση* (marking) των πακέτων με το αντίστοιχο DSCP (βλέπε ενότητα A.1.4.2.2.1).

B.2.2.4.1.3 Ρύθμιση Ρυθμού Εξυπηρέτησης

Ο έλεγχος αποδοχής κλήσης και η ρύθμιση ποιότητας ρυθμίζουν τελικά πόσες και ποιες ενεργές ροές θα ανατεθούν σε ένα ΤΤ. Με τη ρύθμιση του ρυθμού εξυπηρέτησης ρυθμίζεται η ένταση της κίνησης όλων των ενεργών ροών για ένα ΤΤ. Στις ενεργές ροές συμπεριλαμβάνονται και αυτές που προκύπτουν από έμμεση κλήση.

Ο ρυθμός εξυπηρέτησης για μια υπηρεσία είναι ο ρυθμός των πακέτων που θα γίνουν αποδεκτά για να λάβουν την προβλεπόμενη ποιότητα υπηρεσίας και καθορίζει τελικά την ικανοποίηση της υπηρεσίας κατά τη χρήση της (βλέπε ενότητα B.1.1.4). Κατ' αντιστοιχία, ο ρυθμός εξυπηρέτησης κυμαίνεται μεταξύ της μηδενικής τιμής, της τιμής που αντιστοιχεί σε almost-satisfied (SR_{AS}) και της τιμής που αντιστοιχεί σε fully-satisfied (SR_{FS}). Οι τιμές αυτές για κάθε υπηρεσία υπολογίζονται από το συμφωνημένο στο συμβόλαιο εγγραφής της υπηρεσίας ρυθμό (contractual rate) και βάσει των κανόνων που έχουν θεσπιστεί από τις πολιτικές του παροχέα ανά τύπο υπηρεσίας.

Η παράμετρος που προσδιορίζει την στρατηγική αποδοχής ως προς τον ρυθμό εξυπηρέτησης, σε αντιστοιχία με το επίπεδο ικανοποίησης (βλέπε ενότητα B.2.1.3), έχει πεδίο τιμών $[-1,1]$, όπου -1 αντιστοιχεί σε μηδενικό, 0 σε SR_{AS} και 1 σε SR_{FS} . Συγκεκριμένα, έστω sp_{SR} η παράμετρος της στρατηγικής αποδοχής για τη ρύθμιση του ρυθμού εξυπηρέτησης, τότε ο τρέχων ρυθμός εξυπηρέτησης μιας υπηρεσίας, έστω SR , υπολογίζεται ως

$$SR = \begin{cases} (1 + sp_{SR}) \cdot SR_{AS} & sp_{SR} \in [-1,0] \\ SR_{AS} + sp_{SR}(SR_{FS} - SR_{AS}) & sp_{SR} \in (0,1] \end{cases}$$

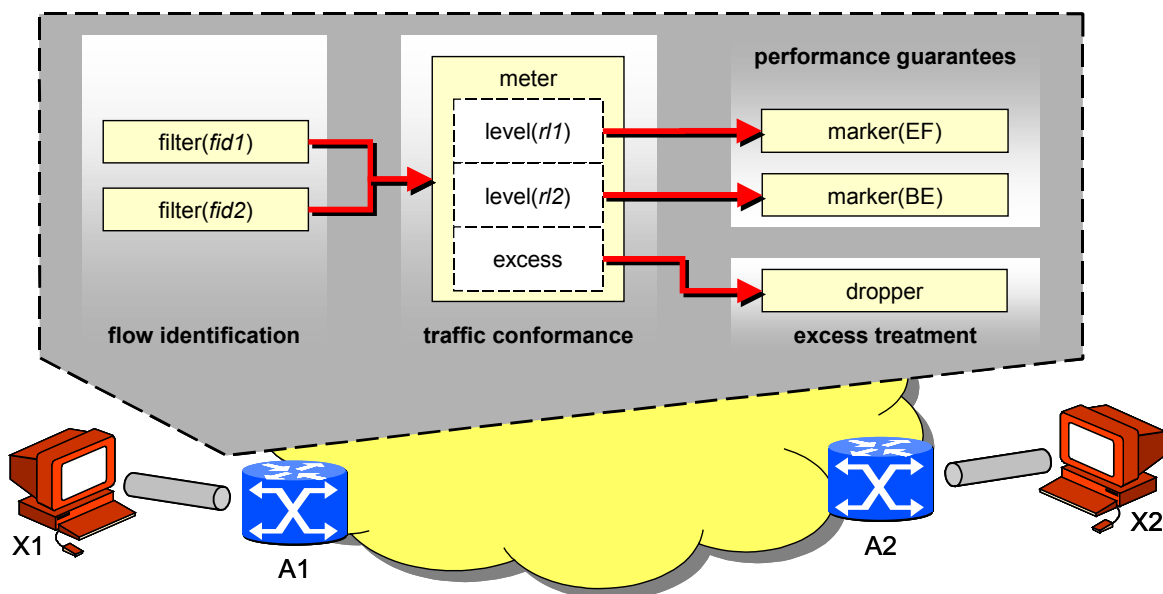
Η μείωση του ρυθμού εξυπηρέτησης δεν ισοδυναμεί αναγκαστικά με διαγραφή των πακέτων που πλεονάζουν μετά τη μείωση. Σε περίπτωση που σε μια υπηρεσία προσδιορίζονται πολλαπλά επίπεδα

συμμόρφωσης, τα πακέτα που βρίσκονται μετά τη μείωση του ρυθμού εξυπηρέτησης να πλεονάζουν, εξωθούνται στο επόμενο επίπεδο συμμόρφωσης στο οποίο, ανάλογα με τους όρους του συμβολαίου, θα λάβουν κατώτερη ποιότητα ή θα διαγραφούν. Ένα τυπικό τέτοιο παράδειγμα είναι η εξυπηρέτηση της πλεονάζουσας κίνησης (excess traffic) με ποιότητα Best Effort αντί της διαγραφής της. Όπως και η αποπομπή ενεργών ροών ποιοτικών υπηρεσιών από ένα TT, έτσι και η μείωση του ρυθμού εξυπηρέτησης προκαλεί μετατόπιση μέρος του φόρτου και αντίστοιχη επιβάρυνση σε TTs κατώτερης ποιότητας.

Η μεταβολή του ρυθμού εξυπηρέτησης γίνεται με την κατάλληλη διεύθυνση του ρυθμού ανά επίπεδο συμμόρφωσης (rate levels) της λειτουργίας μέτρησης (metering) στο Traffic Conditioning (βλέπε ενότητα A.1.4.2.2.1). Η αντιστοίχιση επιπέδου συμμόρφωσης και ποιότητας μέσω της εφαρμογής των κατάλληλων ρυθμίσεων σημείωσης ή διαγραφής παραμένει αμετάβλητη.

B.2.2.4.2 Υλοποίηση στο επίπεδο δεδομένων

Η λειτουργία διαχείρισης κλήσης είναι επιφορτισμένη με την υλοποίηση της υπηρεσίας στον ακραίο κόμβο. Η ροή κίνησης του χρήστη αποκτά πρόσβαση στην υπηρεσία βάσει των κανόνων που περιέχονται στο συμβόλαιο και υλοποιούνται μέσω των μηχανισμών ρύθμισης της κίνησης. Στο επίπεδο δεδομένων θεωρούμε την ύπαρξη των βασικών DiffServ μηχανισμών ταξινόμησης και ρύθμισης της κίνησης (βλέπε ενότητα A.1.4.2.2.1), οι οποίες στο λειτουργικό μοντέλο που υιοθετούμε υλοποιούνται από τη λειτουργία Traffic Conditioning. Η μετάφραση και η αντιστοίχιση των κανόνων για την ενεργοποίηση του συμβολαίου σε ρυθμίσεις των σχετικών μηχανισμών DiffServ περιγράφεται συνοπτικά παρακάτω. Ένα παράδειγμα υλοποίησης συμβολαίου απεικονίζεται στο Σχήμα B-20.



Σχήμα B-20: παράδειγμα υλοποίησης υπηρεσίας στο επίπεδο δεδομένων

Για κάθε ένα SLS που περιέχεται στο συμβόλαιο (βλέπε ενότητα B.1.2.1.2) το πεδίο εφαρμογής (scope) ορίζει τον κόμβο και την πόρτα εισόδου (ingress interface) της ροής που θα λάβει την υπηρεσία. Σε αυτό το interface (π.χ. στον κόμβο A1 με απευθείας γραμμή στον χρήστη X1), εφαρμόζεται ο έλεγχος εισόδου και υλοποιούνται οι ρυθμίσεις πρόσβασης της ροής στους πόρους του δικτύου.

Η αναγνώριση των πακέτων που ανήκουν στη ροή υλοποιείται με τη δημιουργία των κατάλληλων φίλτρων ταξινόμησης (classification filters) βάσει των χαρακτηριστικών που καθορίζονται στο πεδίο ταυτοποίησης

ροής (flow identification) του συμβολαίου. Περισσότερες από μία ταυτότητες καταδεικνύουν περισσότερες από μία ομάδες ροών που μπορούν να χρησιμοποιήσουν την υπηρεσία (π.χ. *fid1* και *fid2*).

Οι εγγυήσεις ποιότητας της υπηρεσίας αλλά και οι δεσμεύσεις του χρήστη ως προς τα χαρακτηριστικά και την ένταση ισχύουν για το σύνολο της κίνησης που προκύπτει από όλες τις αναγνωρισμένες ροές. Το σύνολο της κίνησης περνά από τη λειτουργία του μετρητή (meter) όπου διαπιστώνεται το επίπεδο συμμόρφωσης των πακέτων έναντι του προκαθορισμένου ανά επίπεδο ρυθμού (rate level π.χ. *rl1* και *rl2*) στο πεδίο χαρακτηριστικών κίνησης (traffic conformance). Τα πακέτα που βρίσκονται να είναι μη συμμορφούμενα ακόμα και με το τελευταίο επίπεδο (out-of-profile) λέγεται πως αποτελούν την πλεονάζουσα κίνηση (excess traffic).

Στην κίνηση που προκύπτει από κάθε επίπεδο συμμόρφωσης αντιστοιχεί διαφορετική ποιότητα που καθορίζεται ανά επίπεδο στο πεδίο εγγυήσεων ποιότητας (performance guarantees). Η ποιότητα αυτή έχει μεταφραστεί και αντιστοιχηθεί κατά τη διαδικασία της εγγραφής της υπηρεσίας, στην κατάλληλη QoS-κλάση που παρέχει το δίκτυο. Εντός της επικράτειας του δικτύου η αναγνώριση της QoS-κλάσης ενός πακέτου γίνεται από την τιμή του πεδίου DSCP της επικεφαλίδας. Συνεπώς η κατάταξη ενός πακέτου σε μια QoS-κλάση υλοποιείται με τη σημείωση (marking) του DSCP που της έχει αντιστοιχηθεί (π.χ. τα τυποποιημένα EF και BE DSCPs).

Η μεταχείριση της πλεονάζουσας κίνησης ορίζεται στο αντίστοιχο πεδίο (excess treatment). Εάν δεν οριστεί διαφορετικά, η δεδομένη ενέργεια επί της πλεονάζουσας κίνησης είναι η διαγραφή της (dropping).

B.2.2.4.3 Αντιστοίχιση υπηρεσιών και TTs

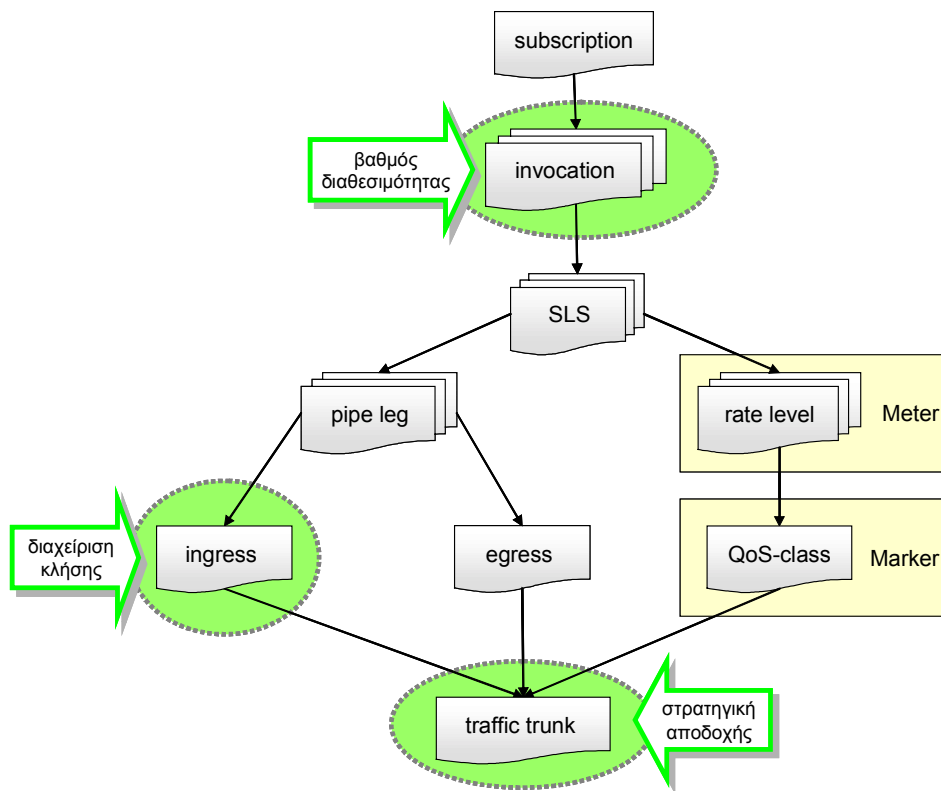
Η υλοποίηση της στρατηγικής ενέχει και την αμφίδρομη αντιστοίχιση μεταξύ υπηρεσιών και ανά TT στρατηγικών αποδοχής. Η εξυπηρέτηση κάθε υπηρεσίας πρέπει να συμβαδίζει με τη στρατηγική αποδοχής του κάθε TT στο οποίο διοχετεύει κίνηση και αντίστροφα, η μεταβολή στη στρατηγική αποδοχής για ένα TT θα πρέπει να εφαρμόζεται σε κάθε υπηρεσία που το χρησιμοποιεί. Οι βασικές δομές για την αντιστοίχιση μεταξύ υπηρεσιών και TTs και οι σχέσεις μεταξύ τους φαίνονται στο Σχήμα B-21.

Στη γενική περίπτωση, περισσότερες από μία κλήσεις (invocations) που έχουν προκύψει από το ίδιο συμβόλαιο εγγραφής (subscription) είναι δυνατό να είναι ταυτόχρονα ενεργές. Ειδικότερα, κατά την αίτηση κλήσης μιας υπηρεσίας, η διαδικασία εξουσιοδότησης αποφαινεται μεταξύ άλλων εάν αυτή η νέα κλήση είναι εξουσιοδοτημένη βάσει του βαθμού διαθεσιμότητας (grade of service) στο συμβόλαιο της υπηρεσίας (βλέπε ενότητα B.1.2.1.2). Όταν για παράδειγμα οι ενεργές κλήσεις μιας υπηρεσίας ευέλικτης ενεργοποίησης (flexible service, βλέπε ενότητα B.1.2.1.1.2) καλύπτουν το 100% του προβλεπόμενου ρυθμού δεν επιτρέπεται περαιτέρω αύξηση με την αποδοχή μιας επιπλέον κλήσης. Αντίστοιχα όταν ο αριθμός των ενεργών κλήσεων μιας υπηρεσίας ενεργοποίησης ανά αίτηση (on demand service) έχει ξεπεράσει τον μέγιστο αριθμό ταυτόχρονων ενεργοποιήσεων (maximum simultaneous uses) τότε η υπό εξέταση αίτηση κλήσης απορρίπτεται.

Κάθε κλήση της υπηρεσίας δημιουργεί ένα νέο στιγμιότυπο για κάθε SLS που περιλαμβάνεται στο συμβόλαιο εγγραφής. Αυτή είναι η πρακτική που ακολουθούμε εδώ χωρίς να αποτελεί κάποιο περιορισμό. Εναλλακτικά, κάθε κλήση θα μπορούσε να ενεργοποιεί και ένα υποσύνολο μόνο από τα εγγεγραμμένα SLS, όταν προστίθεται για παράδειγμα ένα νέο μέλος σύνδεσης (topological leg) λόγω της προσχώρησης ενός χρήστη σε μια τηλεφωνική συνεδρία (conference call).

Κάθε SLS με τη σειρά του αντιστοιχεί σε ένα ή περισσότερα μέλη σύνδεσης τύπου σωλήνα (pipe legs). Έχοντας αποκλείσει από την παρούσα μελέτη τον ορισμό και τη διαχείριση από το traffic engineering TTs

τύπου μάνικας (hose) (βλέπε ενότητα B.1.2.2.2), δεχόμαστε πως οι ενδεχόμενες πολλαπλές σχέσεις N:M μεταξύ κόμβων εισόδου και εξόδου του πεδίου εφαρμογής (scope) ενός SLS, αναλύονται στους αντίστοιχους N×M 1:1 συνδυασμούς διαμορφώνοντας τα N×M μέλη σύνδεσης τύπου σωλήνα που περιγράφουν το SLS.



Σχήμα B-21: πληροφοριακό μοντέλο εφαρμογής στρατηγικής αποδοχής

Σε κάθε SLS περιγράφονται τα χαρακτηριστικά που πρέπει να πληροί η κίνηση των χρηστών που θα φθάσει στους κόμβους εισόδου και η ποιότητα που θα λάβει ανάλογα με το επίπεδο συμμόρφωσης (rate level) ως προς αυτά τα χαρακτηριστικά. Η ποιότητα προσδιορίζεται από την QoS-κλάση στην οποία θα ανατεθεί η κίνηση. Οι λειτουργίες μέτρησης (meter) και σημείωσης (marker) είναι αυτές που θα κληθούν να υλοποιήσουν τις αντίστοιχες ρυθμίσεις της κίνησης στο επίπεδο δεδομένων (βλέπε ενότητα B.2.2.4.2).

Η κίνηση που θα προκύψει για ένα SLS θα ανατεθεί τελικά σε περισσότερες από μία QoS-κλάσεις, όσες και τα επίπεδα συμμόρφωσης όπως ορίζονται στο συμβόλαιο. Για όλο το πεδίο εφαρμογής της υπηρεσίας, ή αλλιώς για κάθε μέλος σύνδεσης μονής φοράς, η κίνηση που θα ανατεθεί σε κάθε QoS-κλάση θα αντιστοιχηθεί σε ένα διαφορετικό TT για τη δεδομένη QoS-κλάση και με άκρα τα άκρα του μέλους μονής σύνδεσης. Στο παράδειγμα του Σχήμα B-20, εάν η υπηρεσία είναι απλής αμφίδρομης σύνδεσης μεταξύ των χρηστών X1 και X2 τότε θα αποτελείται από δύο SLS, ένα για κάθε κατεύθυνση και το καθένα αντιστοιχώντας σε δύο TTs, ένα για την EF QoS-κλάση και ένα για την BE QoS-κλάση.

Η στρατηγική αποδοχής καθορίζεται ανά TT από τη δυναμική διαχείριση στρατηγικής ανάλογα με τις συνθήκες φόρτου στο TT (βλέπε ενότητα B.2.2.5). Η απόφαση για την αποδοχή μιας αίτησης κλήσης υπηρεσίας, αφού περάσει την εξουσιοδότηση βάσει του βαθμού διαθεσιμότητας και των λοιπών όρων του συμβολαίου εγγραφής, περνάει στη φάση της έγκρισης με κριτήριο τη διαθεσιμότητα των πόρων οπότε και απαιτείται η σύγκλιση από τη στρατηγική αποδοχής κάθε TT που παρέχει πόρους για την υπηρεσία. Η απόφαση αποδοχής από τη στρατηγική αποδοχής δίνεται βάσει της πιθανότητας P_{rej} (βλέπε ενότητα B.2.2.4.1.1).

Η λειτουργία διαχείρισης κλήσης είναι κατανεμημένη ανά ακραίο κόμβο. Το στιγμιότυπο διαχείρισης κλήσης σε έναν ακραίο κόμβο ελέγχει εκείνα τα TTs στα οποία διοχετεύει κίνηση, ή αλλιώς για τα οποία αποτελεί κόμβο εισόδου (ingress). Συνεπώς, στη γενική περίπτωση, η έγκριση για την αποδοχή μιας υπηρεσίας δεν μπορεί να παραχωρηθεί αποκλειστικά από τον κόμβο που εξυπηρετεί την αίτηση κλήσης, αλλά μετατίθεται και προωθείται σε κάθε εμπλεκόμενο ακραίο εισόδου. Σε κάθε ακραίο κόμβο συνοψίζεται η σύγκλιση των στρατηγικών αποδοχής για το υποσύνολο των TTs της υπηρεσίας με ingress το συγκεκριμένο κόμβο και αποστέλλεται πίσω στον κόμβο που έλαβε αρχικά την αίτηση. Σε περίπτωση επιτυχούς έγκρισης, σε κάθε κόμβο εισόδου εφαρμόζονται και οι κατάλληλες ρυθμίσεις στο επίπεδο δεδομένων ώστε να είναι έτοιμο να υποδεχθεί την κίνηση που θα προκύψει από την επιτυχή ενεργοποίηση της υπηρεσίας (βλέπε ενότητα B.2.2.4.2). Αφού γίνει η συλλογή όλων των επιμέρους αποτελεσμάτων στον κόμβο που έλαβε την αίτηση κλήσης, λαμβάνεται η απόφαση αποδοχής σε επίπεδο κλήσης και το αποτέλεσμα διαβιβάζεται πίσω στο χρήστη. Σε περίπτωση μερικής απόρριψης, οι κόμβοι που έχουν δώσει έγκριση και έχουν προβεί σε εφαρμογή ρυθμίσεων στο επίπεδο δεδομένων ενημερώνονται ώστε να αποσύρουν τις ρυθμίσεις. Για την απόσυρση των ρυθμίσεων ενημερώνονται όλοι οι εμπλεκόμενοι κόμβοι και κατά τον τερματισμό της υπηρεσίας. Το πρωτόκολλο σηματοδότησης που χρησιμοποιείται μεταξύ των ακραίων κόμβων της επικράτειας (network-to-network-interface – NNI) και μεταξύ κόμβου πρόσβασης και χρήστη (user-to-network-interface – UNI) για αυτό το σκοπό είναι πέρα από το πλαίσιο της παρούσας εργασίας.

Εάν η κίνηση που συγκεντρώνεται στη ζεύξη εξόδου προς το δίκτυο πρόσβασης σε ένα χρήστη (egress access link, interface μεταξύ ακραίου κόμβου A2 και χρήστη X2 στο παράδειγμα του Σχήμα B-20) υπερβαίνει τη φυσική χωρητικότητα της ζεύξης, τότε υποβαθμίζεται η ποιότητα υπηρεσίας που έχει παρασχεθεί εντός της επικράτειας και αναιρούνται οι σχετικές εγγυήσεις. Η κατανεμημένη στους εμπλεκόμενους ακραίους κόμβους απόφαση αποδοχής μιας υπηρεσίας θα μπορούσε μαζί με τη στρατηγική αποδοχής των TTs να εξετάζει και κριτήρια διαθεσιμότητας των πόρων στις ζεύξεις εξόδου. Με αυτό τον τρόπο, εξοικονομούνται πόροι του δικτύου κορμού που αλλιώς σπαταλιούνται σε υπηρεσίες καταδικασμένες να μην ικανοποιηθούν. Ενώ στο σχεδιασμό των αλγορίθμων και των πρωτοκόλλων σηματοδότησης έχουν ληφθεί υπόψη τα σχετικά θέματα, έχοντας θεωρήσει πως δεν εξετάζεται η διαθεσιμότητα των πόρων στο δίκτυο πρόσβασης, αγνοούμε αυτή την παράμετρο εφεξής.

Πέρα από την αποδοχή ή την απόρριψη σε επίπεδο κλήσης, η ανά TT στρατηγική αποδοχής καθορίζει και την ποιότητα και τον ρυθμό εξυπηρέτησης που θα δοθεί στην υπηρεσία κατά την ενεργοποίησή της μέσω των ρυθμίσεων επιπέδου δεδομένων. Ειδικότερα, εάν πρόκειται για ποιοτική (qualitative) υπηρεσία, η έγκρισή της θα παραπεμφθεί στο TT με την ανώτερη στο φάσμα της υπηρεσίας QoS-κλάση και του οποίου η εκάστοτε ισχύουσα στρατηγική αποδέχεται την ανάθεση ποιοτικών υπηρεσιών (βλέπε ενότητα B.2.2.4.1.2).

Ο ρυθμός που θα κληθεί να εξυπηρετήσει κάθε TT από το SLS μιας υπηρεσίας, προκύπτει από το ρυθμό του επιπέδου συμμόρφωσης που αντιστοιχίζεται με την QoS-κλάση του TT. Οι SR_{FS} και SR_{AS} ρυθμοί υπολογίζονται βάσει του συνολικού ρυθμού που θα προκύψει από την υπηρεσία και θα διοχετευθεί στο δίκτυο. Το μέρος που απομένει μετά τους υπολογισμούς σε κάθε επίπεδο συνιστά τους αντίστοιχους ρυθμούς $SR_{FS}(tt)$ και $SR_{AS}(tt)$ του επιπέδου, στους οποίους επενεργεί η στρατηγική του αντίστοιχου TT (βλέπε ενότητα B.2.2.4.1.3).

Στο παράδειγμα του Σχήμα B-20 έστω τα εμπλεκόμενα TTs για μια σύνδεση μονής φοράς από τον χρήστη X1 στο χρήστη X2 είναι τα $TT_1[A1,A2,QC-EF]$ και $TT_2[A1,A2,QC-BE]$ και έστω τα κατά το traffic conformance του συμβολαίου $r11 = 60$, $r12 = 100$. Έστω ακόμα πως, για τον τύπο της συγκεκριμένης υπηρεσίας έχει οριστεί από τον παροχέα $SR_{FS} = 100\%$ (SR_{SS}) και $SR_{AS} = 80\%$ (SR_{SS}), όπου SR_{SS} ο εκάστοτε

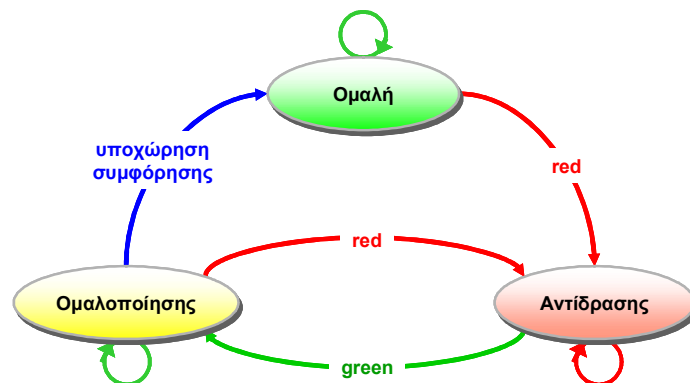
από συμβολαίου συμφωνημένος ρυθμός (contractual rate). Στο παράδειγμα, καθώς η πλεονάζουσα κίνηση διαγράφεται, $SR_{SS} = SR_{FS} = 100$ και $SR_{AS} = 80$. Σε επίπεδο TT και ανάλογα με το ρυθμό των αντίστοιχων επιπέδων συμμόρφωσης ισχύει $SR_{FS}(TT_1) = SR_{AS}(TT_1) = 60$ και $SR_{FS}(TT_2) = 40$, $SR_{AS}(TT_2) = 20$. Εάν οι στρατηγικές αποδοχής των TT_1 και TT_2 επιτάσσουν αντίστοιχα $sp_{SR}(TT_1) = 1$ και $sp_{SR}(TT_2) = 0,5$ τότε $SR(TT_2) = 20 + 0,5 \cdot (40 - 20) = 30$ και οι ρυθμίσεις στο μετρητή γίνονται $rl1 = 60$ και $rl2 = 90$. Εάν $sp_{SR}(TT_1) = -0,4$ και $sp_{SR}(TT_2) = 0,5$ τότε $SR(TT_1) = (1 + (-0,4)) \cdot 60 = 36$ και $SR(TT_2) = 30$ οπότε $rl1 = 36$ και $rl2 = 66$.

B.2.2.5 Δυναμική Διαχείριση Αποδοχής

Ο στόχος της λειτουργίας δυναμικής διαχείρισης αποδοχής είναι η διαμόρφωση της εκάστοτε ενδεδειγμένης στρατηγικής αποδοχής ανά TT. Ενδεδειγμένη στρατηγική θεωρείται η περισσότερο παραχωρητική στρατηγική που δεν οδηγεί σε υπερφόρτωση του δικτύου. Χωρίς τη δυνατότητα πρόβλεψης της υπερφόρτωσης του δικτύου, η μόνη ακριβής και αδιαμφισβήτητη ένδειξη για την ορθότητα των ενεργειών του αλγορίθμου ως προς την προσέγγιση της εκάστοτε ενδεδειγμένης στρατηγικής αποδοχής, είναι η ίδια η απόκριση του συστήματος. Η απόκριση του συστήματος υποδεικνύει και την αναγκαιότητα λήψης διορθωτικών μέτρων και προς ποια κατεύθυνση. Αυτή είναι βασική αρχή κάθε συστήματος που λειτουργεί βασισμένο σε ανάδραση.

Η απόκριση του συστήματος εδώ αφορά στην κατάσταση φόρτου του δικτύου που αποτιμάται με δυαδικό τρόπο ανά TT, καταδεικνύοντας την ύπαρξη επιπλέον διαθέσιμων πόρων ή την εξάντληση των διαθέσιμων πόρων, δηλαδή την υπερφόρτωση του δικτύου με αποτέλεσμα τη συμφόρηση, την παραβίαση των ορίων καθυστέρησης ή/και απώλειας της αντίστοιχης QoS-κλάσης και την υποβάθμιση της ποιότητας. Το TT λέγεται πως βρίσκεται αντίστοιχα σε *πράσινη* (green) ή *κόκκινη* (red) κατάσταση.

Η απόκριση του συστήματος πυροδοτεί και κατευθύνει τον αλγόριθμο. Οι μεταβάσεις της κατάστασης του TT επιτάσσουν κεφαλαιώδεις μεταβολές στρατηγικής. Διακρίνουμε τρεις καταστάσεις λειτουργίας του αλγορίθμου, την *ομαλή* κατάσταση (normal), την κατάσταση *αντίδρασης* (reactive) και την κατάσταση *ομαλοποίησης* (normalisation) (βλέπε Σχήμα B-22).



Σχήμα B-22: διάγραμμα καταστάσεων αλγορίθμου δυναμικής διαχείρισης αποδοχής

Στην ομαλή κατάσταση, το TT παραμένει σε κατάσταση green. Αυτό αποτελεί ένδειξη πως η ισχύουσα στρατηγική αποδοχής ενδεχομένως να είναι λιγότερο παραχωρητική από την ενδεδειγμένη. Σε αυτή την περίπτωση η τροποποίηση της στρατηγικής αποδοχής είναι προαιρετική και εξαρτάται από το εκτιμώμενο ρίσκο υπερφόρτωσης και το επίπεδο προφύλαξης που έχει οριστεί από τον παροχέα. Το ρίσκο υπερφόρτωσης εκτιμάται συναρτήσει του τοπικού φόρτου που διοχετεύεται στο δίκτυο, προβλεβημένου επί του μεριδίου των πόρων όπως αποτιμάται στο απόθεμα πόρων RAB από το traffic engineering.

Ο αλγόριθμος μεταβαίνει στην κατάσταση αντίδρασης όταν το TT μεταβεί σε κατάσταση red, οπότε και η ισχύουσα στρατηγική αποδοχής υπήρξε περισσότερο παραχωρητική από την ενδεδειγμένη. Σε αυτή την περίπτωση είναι επιτακτικό να ληφθούν οι απαραίτητες ενέργειες με την κατάλληλη προσαρμογή της ισχύουσας στρατηγικής αποδοχής ώστε να επιλυθεί η συμφόρηση.

Η κατάσταση ομαλοποίησης μεσολαβεί για την ασφαλή μετάβαση από την κατάσταση αντίδρασης στην ομαλή κατάσταση. Η επίλυση της συμφόρησης και η μετάβαση του TT σε κατάσταση green οφείλεται στις ενέργειες αναχαίτισης από τη στρατηγική αποδοχής στην κατάσταση αντίδρασης και όχι στην υποχώρηση της προσφερόμενης κίνησης από τους χρήστες. Συνεπώς, η άμεση επιστροφή στην ομαλή κατάσταση θα προκαλέσει την ίδια απόκριση και θα οδηγήσει και πάλι σε συμφόρηση. Γνωρίζουμε δηλαδή πως το ρίσκο υπερφόρτωσης είναι αυξημένο λόγω της πρόσφατης συμφόρησης και αντίστοιχα αυξημένη οφείλει να είναι η προφύλαξη με την οποία διαμορφώνονται οι στρατηγικές αποδοχής. Όταν η κατάσταση ομαλοποίησης αποφανθεί την υποχώρηση της προσφερόμενης κίνησης και κατ' επέκταση της εν δυνάμει συμφόρησης, τότε το σύστημα επιστρέφει στην ομαλή κατάσταση όπου μόνη ένδειξη για το ρίσκο υπερφόρτωσης θεωρείται το επίπεδο του τοπικού φόρτου.

B.2.2.5.1 Διαμόρφωση στρατηγικών αποδοχής

Η μεταβολή της στρατηγικής αποδοχής, και ειδικότερα η μεταβολή ως προς τους παράγοντες που αφορούν τις ενεργές ροές, εφαρμόζεται μέσω των κατάλληλων τροποποιήσεων της υλοποίησης των ενεργών υπηρεσιών στο επίπεδο δεδομένων του ακραίου κόμβου. Για κάθε υπηρεσία που χρησιμοποιεί το συγκεκριμένο TT, η ρύθμιση του ρυθμού εξυπηρέτησης σημαίνει αλλαγή στο μηχανισμό μέτρησης (meter) ενώ η ρύθμιση της ποιότητας κάθε ποιοτικής (qualitative) υπηρεσίας σημαίνει αλλαγή στη σημείωση (marking) (βλέπε ενότητα B.2.2.4.2). Κάθε τέτοια αλλαγή επιβαρύνει τον κόμβο, αφενός σε επεξεργαστική ισχύ και αφετέρου σε επικοινωνία όταν η λειτουργία διαχείρισης κλήσης δεν είναι ενσωματωμένη στον κόμβο. Επιπλέον, ανάλογα με τις δυνατότητες του υλικού (hardware) του κόμβου, ενδέχεται η μεταβολή της υλοποίησης μιας υπηρεσίας να προξενεί απώλεια ή/και αναδιάταξη των πακέτων της αντίστοιχης ροής, υποβαθμίζοντας έτσι επιπλέον την ποιότητα υπηρεσίας.

Πέραν των θεμάτων πρακτικής εφαρμογής, η συχνότητα των μεταβολών στρατηγικής είναι κρίσιμη και για την ευστάθεια του δυναμικού συστήματος. Η ευστάθεια απειλείται καταρχήν λόγω του βρόχου ανάδρασης. Η κατάσταση του TT ευθύνεται για τη μεταβολή της στρατηγικής, που με τη σειρά της επηρεάζει την κατάσταση του TT. Αυτή η αλληλεπίδραση είναι δυνατό να οδηγήσει σε ταλαντώσεις τον αλγόριθμο. Επιπλέον, στο κατανεμημένο περιβάλλον που έχουμε θεωρήσει, η μεταβολή στη στρατηγική αποδοχής σε ένα ακραίο κόμβο επηρεάζει την κίνηση που διοχετεύεται στο δίκτυο, η οποία διασταυρώνεται στις ζεύξεις κορμού με την κίνηση από άλλους ακραίους κόμβους και προκαλεί ενδεχομένως αλλαγή της κατάστασης των αντίστοιχων TTs, πυροδοτώντας έτσι μεταβολή της στρατηγικής τους. Αυτές οι αλυσιδωτές μεταβολές είναι δυνατό κλιμακωτά να οδηγήσουν σε πλήρη αποσταθεροποίηση του συστήματος.

Για τους παραπάνω λόγους είναι απαραίτητη η επιβολή φραγμών στην ευαισθησία και στη συχνότητα μεταβολής της στρατηγικής αποδοχής από τη δυναμική διαχείριση αποδοχής. Έτσι, περιορίζουμε τις δυνατές ενδεδειγμένες στρατηγικές σε ένα πεπερασμένο σύνολο προδιαγεγραμμένων στατικών στρατηγικών αποδοχής, απόλυτα διατεταγμένων μεταξύ τους ως προς την παραχωρητικότητά τους. Κάθε παράγοντας που μεταβάλλεται μεταξύ δύο παρακείμενων στρατηγικών, μεταβάλλεται προς την ίδια κατεύθυνση, είτε λιγότερο είτε περισσότερο παραχωρητική. Στη γενική περίπτωση, η ισχύουσα στρατηγική επιτρέπεται να αντικατασταθεί μόνο από παρακείμενή της σε αυτή τη διάταξη, περιορίζοντας έτσι τον αλγόριθμο δυναμικής διαχείρισης στην επιλογή της φοράς, είτε προς περισσότερο είτε προς λιγότερο παραχωρητική στρατηγική.

Το πεδίο τιμών των συνθηκών που καθορίζουν το ρίσκο υπερφόρτωσης καταμερίζεται σε περιοχές ευθύνης και καταλληλότητας για το πεπερασμένο αυτό σύνολο στρατηγικών, καθορίζοντας έτσι και την ευαισθησία του αλγορίθμου στην αλλαγή των συνθηκών.

Επιπλέον, ορίζεται ένα ελάχιστο χρονικό διάστημα (time interval), έστω t_{min} , πριν από την πάροδο του οποίου και ανεξάρτητα με τις μεταβολές της κατάστασης του ΤΤ δεν είναι δυνατό να αλλάξει η ισχύουσα στρατηγική. Αυτό επιβάλλεται για να δημιουργηθεί το περιθώριο ώστε το σύστημα να σταθεροποιηθεί στο νέο σημείο ισορροπίας μετά από την τελευταία μεταβολή. Μόνον έτσι, οι ενδείξεις πάνω στις οποίες θα βασιστεί η επόμενη μεταβολή, αν κριθεί απαραίτητη, δεν θα είναι παροδικές αλλά αντιπροσωπευτικές της νέας κατάστασης. Πέραν αυτού, ο μηχανισμός αυτός επιτρέπει την επιβολή ενός άνω ορίου στη συχνότητα των μεταβολών στρατηγικής. Είναι σαφές ωστόσο πως, όσο μεγαλύτερο το t_{min} , τόσο πλήττεται η δυναμική του αλγορίθμου που, παραμένοντας περισσότερο σε στρατηγική διαφορετική από την ενδεδειγμένη, γίνεται και λιγότερο αποτελεσματικός. Το t_{min} προβλέπεται να παίρνει τιμές σε κλίμακα δευτερολέπτων.

Από τα παραπάνω δεν διασφαλίζεται η σύγκλιση και η ευστάθεια του αλγορίθμου και κατ' επέκταση του συστήματος, διασφαλίζεται μόνο ο περιορισμός των ταλαντώσεων μεταξύ προδιαγεγραμμένων στρατηγικών και η ανίχνευσή τους μέσω της αναγνώρισης μοτίβου στην εναλλαγή των στρατηγικών. Όταν ανιχνευτούν ταλαντώσεις, ειδικοί μηχανισμοί αντιμετώπισης ταλαντώσεων μπαίνουν σε ισχύ, εξομαλύνοντας το αντίκτυπό τους στην σταθερότητα του συστήματος. Όπως θα δείξουμε παρακάτω, η σύγκλιση των ενεργειών του αλγορίθμου για τα αλληλεξαρτώμενα ΤΤs επιτυγχάνεται χάρη στην κοινή αναφορά τους στο RAM.

B.2.2.5.2 Διατεταγμένη ακολουθία προδιαγεγραμμένων στρατηγικών αποδοχής

Η μεταβολή ενός παράγοντα στρατηγικής λέγεται πως συνιστά μια *ενέργεια* (action). Κάθε ενέργεια επιφέρει κάποιες *κυρώσεις* (penalty), δηλαδή κάποιο δυνητικό κόστος στην ικανοποίηση των υπηρεσιών στις οποίες επενεργεί, δεδομένου πως το δίκτυο παρέχει την απαιτούμενη ποιότητα. Χρησιμοποιούμε τον όρο *αυστηρότητα* (severity) για να περιγράψουμε τη σφοδρότητα των κυρώσεων που επιβάλλει μια ενέργεια ή συνολικά μια στρατηγική.

			έλεγχος αποδοχής			ποιότητα	ρυθμός εξυπηρέτησης
			ρ_{rej}	t_{min}	t_{max}		sp_{sr}
βασική		b_0	0		R_{max}	✓	1
	ήπιες	m_1	1-PL	R_{nom}			
		m_2				▽	
		m_3					$\frac{1 + 2PL}{4}$
	m_4		R_{min}				
σφοδρές	κύρια	s_0			0		0
	σύγκλισης(n)	s_n					$\left[\left(1 + sp_{sr}^{n-1} \right) \frac{R_{nom}}{load} \right] - 1$

Πίνακας B-2: προδιαγεγραμμένες στρατηγικές

Οι στρατηγικές προδιαγράφονται σε μια απόλυτα διατεταγμένη ακολουθία με κριτήριο την κλιμάκωση της αυστηρότητας και κατατάσσονται σε ομάδες ανάλογα με τις κυρώσεις, στη βασική (standard), στις ήπιες (mild) και στις σφοδρές (severe) στρατηγικές, όπου η ικανοποίηση των υπηρεσιών κατά τη χρήση μετά τις κυρώσεις είναι αντίστοιχα fully-satisfied, τουλάχιστον almost-satisfied και λιγότερο από almost-satisfied (βλέπε Πίνακας B-2).

Η αυστηρότητα και οι κυρώσεις μιας στρατηγικής κρίνονται δικαιολογημένες ανάλογα αφενός με το επίπεδο συμμόρφωσης στο δίκτυο και αφετέρου με το ρίσκο υπερφόρτωσης, δηλαδή με την κατάσταση του αλγορίθμου. Σε κάθε κατάσταση αρμόζουν και αντιστοιχίζονται διαφορετικές ομάδες στρατηγικών και διαφορετική λογική μετάβασης μεταξύ των επιμέρους στρατηγικών. Ενώ η βασική και οι σφοδρές στρατηγικές υποδεικνύονται από τον ίδιο τον αλγόριθμο, οι ήπιες στρατηγικές εξαρτώνται από την εκτίμηση και την άποψη του παροχέα για τη σφοδρότητα των κυρώσεων. Κατά τη προδιαγραφή των στρατηγικών, ανάλογα με τον σκοπό που καλούνται να εξυπηρετήσουν, λαμβάνεται υπόψη και το αντίκτυπο που έχει κάθε παράγοντας στην κίνηση που διοχετεύεται στο δίκτυο.

B.2.2.5.3 Αντίκτυπο ενεργειών

Γενικότερα, η ρύθμιση του ελέγχου αποδοχής είναι καθαρά προληπτική και θεωρείται ως η ηπιότερη ενέργεια καθώς, αφενός δεν επιβάλλει αλλαγή των ρυθμίσεων στο επίπεδο δεδομένων του κόμβου, και αφετέρου δεν θίγει την ικανοποίηση των ενεργών υπηρεσιών αλλά την ικανοποίηση σε διαθεσιμότητα μελλοντικών κλήσεων. Έτσι, ενώ δεν σημειώνονται άμεσες κυρώσεις, ωστόσο δεν επιτυγχάνεται και καμία διόρθωση στην ένταση της κίνησης που διοχετεύεται στο δίκτυο.

Στο άλλο άκρο, η ρύθμιση του ρυθμού εξυπηρέτησης θεωρείται η πλέον δραστική και σφοδρή ενέργεια, τόσο ως προς την ικανοποίηση των ενεργών υπηρεσιών αλλά και ως προς την ένταση της κίνησης που διοχετεύεται στο δίκτυο. Η μείωση του ρυθμού εξυπηρέτησης μιας ενεργής υπηρεσίας υποβαθμίζει άμεσα ωστόσο όχι αυστηρώς αναλογικά την ποιότητα της υπηρεσίας. Η πραγματική υποβάθμιση της ποιότητας εξαρτάται από τα επίπεδα της προσφερόμενης κίνησης στη διάρκεια της μείωσης του ρυθμού εξυπηρέτησης. Εάν δηλαδή τύχει, η ενεργή ροή να διανύει μια περίοδο σιωπής (silent period), ή η ένταση της κίνησης να είναι μικρότερη από την αποδεκτή μετά τη μείωση επί της προβλεπόμενης στο συμβόλαιο, τότε δεν σημειώνεται καμία επίπτωση στην ικανοποίηση της υπηρεσίας. Κατ' αναλογία, το αντίκτυπο της ρύθμισης του ρυθμού εξυπηρέτησης στη συνολική κίνηση που διοχετεύεται στο δίκτυο, δεν αντιστοιχεί απόλυτα στο ποσοστό της επιβαλλόμενης μείωσης, αλλά εξαρτάται από το ποσοστό των ενεργοποιημένων υπηρεσιών οι πηγές των οποίων είναι πράγματι ενεργές και την ένταση με την οποία προσφέρουν κίνηση στο δίκτυο.

Η ρύθμιση της ποιότητας των ποιοτικών υπηρεσιών θεωρείται μια ήπια ενέργεια καθώς η σχετική υποβάθμιση της ικανοποίησης δεν παραβιάζει τους όρους του συμβολαίου και αφετέρου έχει άμεσο αντίκτυπο στην κίνηση που διοχετεύεται στο δίκτυο. Και εδώ, το αντίκτυπο αυτής της ενέργειας εξαρτάται από τον αριθμό των ποιοτικών υπηρεσιών που είναι ενεργοποιημένες, και από το ποσοστό και την ένταση των αντίστοιχων ενεργών πηγών.

B.2.2.5.4 Ομάδες στρατηγικών αποδοχής

B.2.2.5.4.1 Βασική στρατηγική

Η βασική στρατηγική επιφέρει μηδενικές κυρώσεις στις υπηρεσίες που απολαμβάνουν αδιάβλητη την ποιότητα που έχει συμφωνηθεί στο συμβόλαιο και αντιστοιχεί σε fully-satisfied βαθμό ικανοποίησης. Οι μόνοι περιορισμοί είναι αυτοί που προβλέπονται στο συμβόλαιο. Η υλοποίησή της συνεπάγεται την αποδοχή

όλων των εξουσιοδοτημένων αιτήσεων κλήσης, την αποδοχή ανάθεσης ποιοτικών υπηρεσιών και την αποδοχή ακέραιου του ρυθμού εξυπηρέτησης.

B.2.2.5.4.2 Σφοδρές στρατηγικές

Οι σφοδρές στρατηγικές αποτελούνται από την κύρια στρατηγική (prime) και τις στρατηγικές σύγκλισης (convergence).

Η κύρια στρατηγική επιφέρει κυρώσεις που αντιστοιχούν σε almost-satisfied ικανοποίηση κατά τη χρήση και καθολική απόρριψη νέων αιτήσεων κλήσης. Η κύρια στρατηγική χρησιμοποιείται ως το ύστατο σύνορο με την πέρα του αποδεκτού υποβάθμιση της ικανοποίησης.

Σε περίπτωση συμφόρησης που επιμένει και μετά την εφαρμογή της κύριας σφοδρής στρατηγικής, εφαρμόζονται οι στρατηγικές σύγκλισης που στοχεύουν στην αναχαίτιση της κίνησης τόσο ώστε να επιλυθεί άμεσα η συμφόρηση, ανεξάρτητα με το ύψος των κυρώσεων που επιφέρει μείωση στην ικανοποίηση κάτω από almost-satisfied. Από τη διαμόρφωση του αποθέματος των διαθέσιμων πόρων (βλέπε ενότητα B.1.2.2.6) είναι γνωστό πως, για την επίλυση της συμφόρησης, αρκεί κάθε ΤΤ να περιορίσει την κίνηση που διοχετεύει στο δίκτυο (*load*) στο κατ' ελάχιστο διαθέσιμο σε κατάσταση συμφόρησης μερίδιο πόρων R_{nom} , να περιορίσει δηλαδή την κίνηση κατά ένα παράγοντα $\frac{R_{nom}}{load}$.

Καθώς δεν υπάρχει ακριβής αντιστοιχία μεταξύ του ποσοστού μείωσης του ρυθμού εξυπηρέτησης με το ποσοστό αναχαίτισης της κίνησης (βλέπε ενότητα B.2.2.5.3), ενδέχεται να χρειαστεί πολλαπλή επέμβαση επί του εκάστοτε *load* ώστε να πέσει στο επίπεδο του R_{nom} . Κάθε νέα σφοδρότερη στρατηγική σύγκλισης(n) επιχειρεί την περαιτέρω, σε σχέση με την προηγούμενη ισχύουσα στρατηγική σύγκλισης(n-1), μείωση του ρυθμού εξυπηρέτησης κατά τον απαραίτητο παράγοντα. Επισημαίνουμε πως, καθώς μια στρατηγική σύγκλισης έπεται της κύριας σφοδρής στρατηγικής, ο ρυθμός εξυπηρέτησης υποχωρεί κάτω από το SR_{AS} ,

οπότε το εκάστοτε sp_{sr}^n είναι αρνητικό και υπολογίζεται ως $sp_{sr}^n = \left[\left(1 + sp_{sr}^{n-1} \right) \frac{R_{nom}}{load} \right] - 1$.

B.2.2.5.4.3 Ήπιες στρατηγικές

Οι ήπιες στρατηγικές επιφέρουν κυρώσεις που αντιστοιχούν σε αποδεκτά επίπεδα ικανοποίησης μεταξύ του almost- και του fully-satisfied. Η ανάγκη που εξυπηρετούν οι ήπιες στρατηγικές είναι η μεσολάβηση μεταξύ της βασικής και των σφοδρών στρατηγικών, επιτρέποντας ένα ενδιάμεσο στάδιο ελέγχου της έντασης της κίνησης προσβλέποντας στην πρόληψη και αποφυγή της συμφόρησης. Η αποφυγή της συμφόρησης με την προληπτική επιβολή κάποιων ήπιων κυρώσεων συμβάλλει στη σταθερότητα του συστήματος και στη συγκράτηση της επιβολής σφοδρότερων κυρώσεων λόγω συμφόρησης. Η εφαρμογή και η σφοδρότητα των ήπιων στρατηγικών υπόκειται στην αντίληψη του παροχέα για την πιθανότητα του ενδεχομένου της συμφόρησης η οποία αποτυπώνεται στην παράμετρο επιπέδου προφύλαξης (Precaution Level – PL) (βλέπε ενότητα B.2.2.3).

Δεδομένης της περιοχής ευθύνης και εφαρμογής της ομάδας των ήπιων στρατηγικών σχετικά με τις συνθήκες φόρτου, το πλήθος τους καθορίζει και τη συχνότητα εναλλαγής τους, οπότε οφείλει να είναι περιορισμένο. Εδώ θεωρούμε τις ήπιες στρατηγικές m_1 έως m_4 από την λιγότερο στην περισσότερο αυστηρή, η αυστηρότητα των οποίων έχει παραμετροποιηθεί συναρτήσει του PL.

Η στρατηγική m_1 επαυξάνει τη βασική στρατηγική ως προς τη ρύθμιση του ελέγχου εισόδου με μέγιστη πιθανότητα απόρριψης 1-PL. Είναι προφανές πως, όταν το επίπεδο προφύλαξης έχει καθοριστεί ως το πλέον

αισιόδοξο τότε η μέγιστη πιθανότητα απόρριψης είναι μηδενική, αναιρώντας ουσιαστικά την επιβολή οποιοδήποτε κυρώσεων. Επισημαίνουμε πως η περιοχή εφαρμογής του αλγορίθμου αποδοχής RED ορίζεται μεταξύ R_{nom} και R_{max} . Η στρατηγική m_2 επαυξάνει τη στρατηγική m_1 ως προς τη ρύθμιση της ποιότητας. Οι ποιοτικές υπηρεσίες δεν ανατίθενται πλέον στο συγκεκριμένο ΤΤ. Η στρατηγική m_3 επαυξάνει τη στρατηγική m_2 ως προς τη ρύθμιση του ρυθμού εξυπηρέτησης, ο οποίος ρυθμίζεται σε ποσοστό $\frac{1+2PL}{4}$, δηλαδή σε ποσοστό 25% έως 75% μεταξύ SR_{AS} και SR_{FS} ανάλογα με την τιμή του PL. Τέλος, πριν από τη βασική σφοδρή στρατηγική, η στρατηγική m_4 δυσχεραίνει περαιτέρω την αποδοχή νέων κλήσεων διευρύνοντας την περιοχή εφαρμογής του αλγορίθμου αποδοχής, αυξάνοντας έτσι την πιθανότητα απόρριψης.

Στο σχεδιασμό των ήπιων στρατηγικών θεωρήσαμε πως κάθε στρατηγική συνίσταται και από μία μόνη ενέργεια και πως, εκτός από την ανέξοδη σε κόστος εφαρμογή στο επίπεδο δεδομένων ρύθμιση του ελέγχου αποδοχής, καθένας από τους τρεις τύπους ενεργειών απαντάται μόνο σε μία ήπια στρατηγική. Καθώς κάθε ενέργεια επηρεάζει διαφορετικές πλευρές και διαφορετικό πληθυσμό των υπηρεσιών, με αυτό τον τρόπο επιτυγχάνεται σταδιακή εφαρμογή κυρώσεων και αποφεύγεται ο συσχετισμός του αντίκτυπου από κάθε ενέργεια.

B.2.2.5.5 Ένδειξη κατάστασης ΤΤ

Η ένδειξη της τρέχουσας κατάστασης υπολογίζεται από τη λειτουργία της παρακολούθησης. Για το σκοπό αυτό χρησιμοποιούνται ενεργές μετρήσεις με την αποστολή δοκιμαστικών πακέτων από άκρο σε άκρο σε κάθε εναλλακτικό μονοπάτι από τον ακραίο κόμβο εισόδου προς τον ακραίο κόμβο εξόδου του ΤΤ. Η παραβίαση της ποιότητας σε τουλάχιστον ένα μονοπάτι μεταφράζεται σε παραβίαση της ποιότητας στο ΤΤ. Η συνεισφορά των λειτουργιών του δυναμικού traffic engineering, εάν υπάρχουν και είναι ενεργές, έγκειται ακριβώς στη δυνατότητα απορρόφησης περισσότερης κίνησης με τη βέλτιστη ανακατανομή των πόρων σε επίπεδο ζεύξης και σε επίπεδο μονοπατιού. Έτσι, η συμφόρηση που καλείται να αντιμετωπίσει η δυναμική διαχείριση αποδοχής είναι αυτή που το δυναμικό traffic engineering, έχοντας εξαντλήσει κάθε περιθώριο της φυσικής χωρητικότητας του δικτύου, δεν μπορεί να αποτρέψει. Το κέρδος από την απορρόφηση περισσότερης κίνησης και την αποτροπή της συμφόρησης χάρη στο δυναμικό traffic engineering, αυξάνεται όσο περισσότερο καταγιστική (bursty) είναι η κίνηση και όσο μεγαλύτερη η διακύμανση και η απόκλιση από τις προβλεπόμενες στο Traffic Matrix τιμές (βλέπε ενότητα B.1.2.2.3). Επιπλέον, όσο πιο έντονη η ανεπάρκεια των πόρων για την κάλυψη των αναγκών των εγγεγραμμένων υπηρεσιών, τόσο περισσότερο αναγκαία η δυναμική ανακατανομή τους.

B.2.2.5.6 Ο αλγόριθμος

B.2.2.5.6.1 Σημειογραφία

- minimum time interval (t_{min})* : το ελάχιστο χρονικό διάστημα πριν από την πάροδο του οποίου δεν είναι δυνατό να αλλάξει η ισχύουσα στρατηγική αποδοχής.
- precaution level (PL)* : το επίπεδο προφύλαξης που ο παροχέας θεωρεί αποδεκτό και δικαιολογημένο προκειμένου να διαφυλάσσεται η ποιότητα των ενεργών υπηρεσιών. Παίρνει τιμές στο $[0,1]$, όπου 0 αντιστοιχεί σε *επιφυλακτικό* (conservative) και 1 σε *αισιόδοξο* (optimistic).

- load (LD)* : ο φόρτος, ή αλλιώς η ένταση της κίνησης, που διοχετεύεται στο δίκτυο κορμού από ένα συγκεκριμένο TT και παρέχεται από τη λειτουργία παρακολούθησης μέσω παθητικών μετρήσεων τοπικά στον εποπτευόμενο ακραίο κόμβο.
- load trend (LDT)* : η τάση του φόρτου *LD*, που μπορεί να είναι θετική ή αρνητική, υποδεικνύοντας ανοδική ή πτωτική τάση του επιπέδου του φόρτου.
- resources critical level (RCL)* : το όριο επί της επισφαλούς περιοχής R_{nom} έως R_{max} που, σύμφωνα με το επίπεδο προφύλαξης *PL*, η πιθανότητα συμφόρησης φτάνει σε κρίσιμα επίπεδα.

B.2.2.5.6.2 Ομαλή Κατάσταση

Στην ομαλή κατάσταση, η κατάσταση διαθεσιμότητας του TT είναι green και ο αντικειμενικός σκοπός είναι να διατηρηθεί green με τις μικρότερες δυνατές κυρώσεις.

Ενώ μια περισσότερο παραχωρητική στρατηγική ενδεχομένως οδηγήσει σε συμφόρηση, μια περισσότερο αυστηρή στρατηγική οδηγεί με βεβαιότητα σε αυξημένες κυρώσεις στην ικανοποίηση των υπηρεσιών, σε λιγότερο εντατική αξιοποίηση των πόρων και διαφυγόντα κέρδη για τον παροχέα. Συνεπώς, στην ομαλή κατάσταση θεωρείται αδικαιολόγητη η επιβολή κυρώσεων κάτω του καθορισμένου έως αποδεκτού ορίου almost-satisfied, αποκλείοντας έτσι την υιοθέτηση σφοδρών στρατηγικών σύγκλισης για τη συγκράτηση του φόρτου σε ορισμένα επίπεδα. Με άλλα λόγια, ενώ η διατεταγμένη ακολουθία των στρατηγικών θεωρητικά εκτείνεται στο άπειρο, με οσοδήποτε αυστηρές κυρώσεις στις στρατηγικές σύγκλισης, στην ομαλή κατάσταση η επιτρεπτή ακολουθία είναι φραγμένη στην κύρια σφοδρή στρατηγική s_0 . Αφού έχει εφαρμοστεί και αυτή, ο αλγόριθμος θα παραμείνει αδρανής σε όποια επίπεδα κι αν βρίσκεται ο φόρτος, ακόμα και εάν συνεχίζει να αυξάνει.

Για την προληπτική προστασία του δικτύου προβλέπεται η εφαρμογή στρατηγικών, το κόστος των οποίων θεωρείται δικαιολογημένο προκειμένου να αποφευχθεί μεγαλύτερο κόστος λόγω συμφόρησης στο δίκτυο. Όσο αυξάνονται τα επίπεδα του φόρτου, τόσο αυξάνεται η πιθανότητα επικάλυψης με άλλα TTs, δηλαδή υπερφόρτωσης του δικτύου και συμφόρησης. Όσο αυξάνεται η πιθανότητα συμφόρησης, τόσο η στρατηγική αποδοχής δικαιολογείται να γίνεται περισσότερο επιφυλακτική, εφαρμόζοντας περισσότερο αυστηρές κυρώσεις για τη συγκράτηση του φόρτου που διοχετεύεται στο δίκτυο. Αντίστροφα, όταν μειώνονται τα επίπεδα του φόρτου σε σημείο που η πιθανότητα συμφόρησης να θεωρείται αμελητέα, η στρατηγική αποδοχής οφείλει να άρει όποιες ισχύουσες κυρώσεις και να επιτρέψει σε μεγαλύτερο μέρος της προσφερόμενης κίνησης την πρόσβαση στο εκτιμώμενο πλεόνασμα διαθέσιμων πόρων του δικτύου.

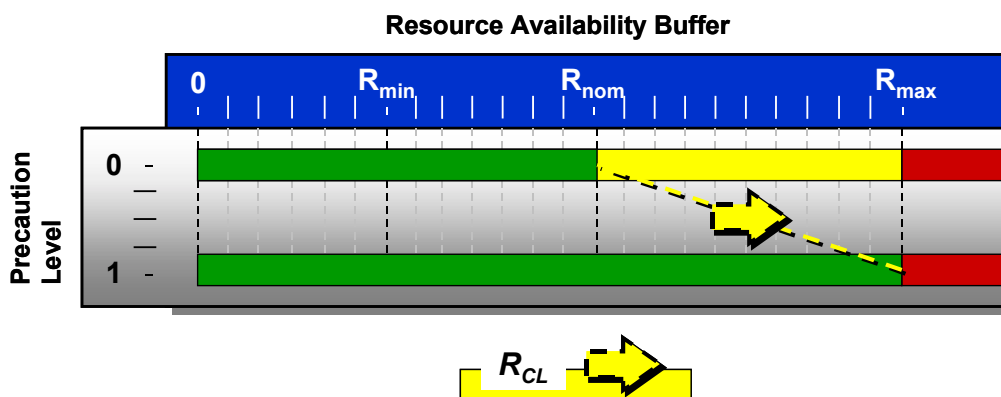
Η πιθανότητα επικάλυψης με άλλα TTs διαφοροποιείται για κάθε περιοχή του RAB (βλέπε ενότητα B.1.2.2.6). Υπενθυμίζουμε πως επικάλυψη και συμφόρηση μπορεί να συμβεί ακόμα και για μηδενικό μεμονωμένο *LD* εφόσον ο φόρτος από τα ανταγωνιζόμενα TTs αρκεί για να υπερκεράσει το δίκτυο, έστω και με τη μηδενική συνδρομή του φόρτου από το υπό εξέταση TT. Είναι προφανές ωστόσο πως, όσο μεγαλύτερο μέρος από τους συνολικούς πόρους καταλαμβάνει ένα TT, τόσο μικραίνει η πιθανότητα αυτοί οι πόροι να μη χρησιμοποιούνται από κανένα άλλο μεταξύ των ανταγωνιζόμενων TTs.

Η διαφοροποίηση ανά περιοχές του RAB αφορά στην επίλυση των επικαλύψεων σε περίπτωση συμφόρησης. Από το σύνολο του φόρτου *LD* η περιοχή έως και το R_{nom} θα παραμείνει στο TT. Αντίθετα, όσο από το υπερβάλλον το R_{nom} χρησιμοποιείται και είναι τώρα απαραίτητο στα TTs από τα οποία ήταν

δανεικό, θα εξαναγκαστεί σε υποχώρηση. Σύμφωνα με αυτή τη θεώρηση, η περιοχή έως και το R_{nom} κρίνεται ασφαλής και επισφαλής η περιοχή από το R_{nom} έως το R_{max} .

Η πιθανότητα της συμφόρησης εξαρτάται στην πραγματικότητα αποκλειστικά από την πιθανότητα της ταυτόχρονης χρησιμοποίησης των κοινόχρηστων μεριδίων R_{nom} , η οποία με τη σειρά της εξαρτάται από τη ένταση της ζήτησης για τους πόρους αυτούς. Η ένταση της ζήτησης και ο συναφής ανταγωνισμός για τα κοινόχρηστα μερίδια, πέρα από τις ιδιομορφίες του περιβάλλοντος ως προς τη συμπεριφορά των χρηστών και των πηγών κίνησης, εξαρτάται, δεδομένων των φυσικών πόρων, από το πλήθος των εγγεγραμμένων υπηρεσιών και άρα από την παράμετρο επιπέδου ικανοποίησης της λειτουργίας διαχείρισης εγγραφής (βλέπε ενότητα B.2.2.3). Συνεκτιμώντας τις ιδιομορφίες του περιβάλλοντος και την επάρκεια των φυσικών πόρων σε σχέση με τις απαιτήσεις των εγγεγραμμένων υπηρεσιών, η πιθανότητα της συμφόρησης και η βαρύτητα που δεδομένου του κόστους προφύλαξης είναι επιθυμητό να της αποδοθεί, δίνεται από τον παροχέα στην παράμετρο επιπέδου προφύλαξης.

Το επίπεδο προφύλαξης καθορίζει το ποσοστό της επισφαλούς περιοχής που χαρακτηρίζεται κρίσιμη (critical area), όπου δηλαδή η πιθανότητα συμφόρησης φτάνει σε κρίσιμα επίπεδα τόσο ώστε να επιβάλει τη λήψη προληπτικών μέτρων και να δικαιολογεί το κόστος από τις αντίστοιχες κυρώσεις. Το όριο εισόδου στην κρίσιμη περιοχή, έστω R_{CL} , καθορίζεται επί της επισφαλούς περιοχής R_{nom} έως R_{max} ανάλογα με το επίπεδο προφύλαξης PL ως $R_{CL} = R_{nom} + PL \cdot (R_{max} - R_{nom})$ (βλέπε Σχήμα B-23).



Σχήμα B-23: καθορισμός κρίσιμης περιοχής βάσει του επιπέδου προφύλαξης

Η λειτουργία δυναμικής διαχείρισης αποδοχής στην ομαλή κατάσταση φαίνεται στο Σχήμα B-24.

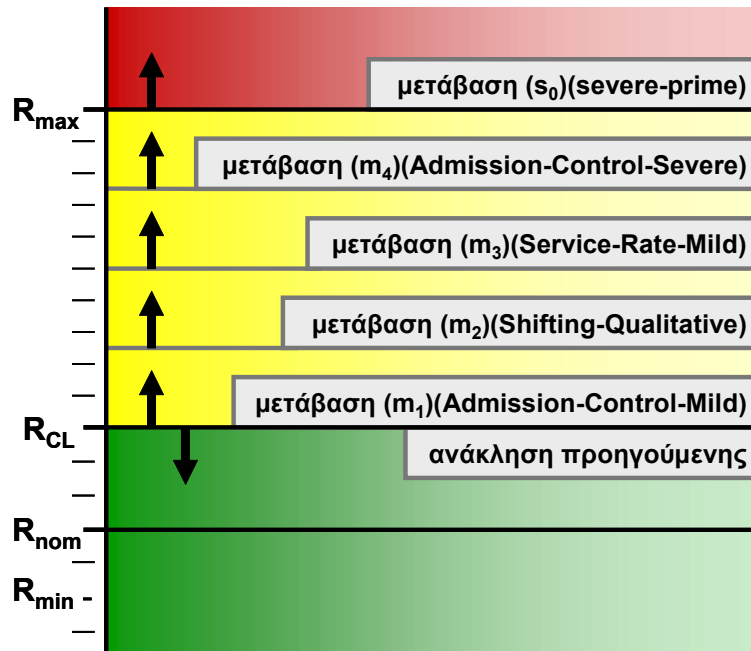
Όταν το LD ξεπεράσει το R_{CL} και εισέλθει στην κρίσιμη περιοχή, τότε λαμβάνονται προληπτικά ήπιες στρατηγικές για την προστασία του δικτύου από το ενδεχόμενο της συμφόρησης. Όσο περισσότερο αυξάνουν τα επίπεδα του φόρτου LD στην κρίσιμη περιοχή τόσο αυξάνει και η πιθανότητα συμφόρησης επιβάλλοντας την εφαρμογή ακόμα αυστηρότερων κυρώσεων.

Εφόσον δεν υπάρχει περιορισμός στη χρήση ολόκληρου του αποθέματος των soft-reserved μεριδίων των άλλων TTs, η πλέον αυστηρή στρατηγική s_0 που είναι επιτρεπτή στην ομαλή κατάσταση λαμβάνεται στο R_{max} . Η κρίσιμη περιοχή διαιρείται σε περιοχές καταλληλότητας όσες και οι εκάστοτε στρατηγικές της ήπιας ομάδας, εδώ m_1 έως m_4 . Κάθε τέτοια περιοχή οριοθετείται από το αντίστοιχο κάτω όριο της

$LD_{m_n}^b = R_{CL} + (n - 1) \left(\frac{R_{max} - R_{CL}}{num_m} \right)$, όπου num_m το πλήθος των ήπιων στρατηγικών, εδώ ίσο με 4. Όταν το

LD σε ανοδική πορεία ξεπεράσει για πρώτη φορά το κατώφλι $LD_{m_n}^b$ τότε εφαρμόζεται η στρατηγική m_n .

Έτσι, όταν το LD ξεπεράσει το R_{CL} και εισέλθει στην κρίσιμη περιοχή για πρώτη φορά, τότε λαμβάνεται η πρώτη ήπια στρατηγική m_1 , εάν μετά από χρόνο τουλάχιστον t_{min} εξακολουθεί να έχει ανοδική τάση και κάποτε ξεπεράσει και το $R_{CL} + \frac{R_{max} - R_{CL}}{4}$ τότε εφαρμόζεται η στρατηγική m_2 κ.ο.κ.



Σχήμα B-24: λειτουργία στην ομαλή κατάσταση

Από τα παραπάνω διαφαίνεται και ο τρόπος που το επίπεδο προφύλαξης επιδρά στη λειτουργία του συστήματος. Τελικά, η προφύλαξη μεταφράζεται σε αυστηρότητα των κυρώσεων που επιβάλλονται προληπτικά συναρτήσει του επιπέδου του φόρτου, το οποίο χρησιμεύει ως ενδεικτική παράμετρος για την πιθανότητα συμφόρησης. Το επίπεδο προφύλαξης αντιστοιχίζει αυστηρότητα κυρώσεων σε επίπεδο φόρτου με δύο τρόπους, ορίζοντας αφενός πότε, δηλαδή σε πόσο υψηλό ή χαμηλό επίπεδο φόρτου στην επισφαλή περιοχή, εφαρμόζεται κάθε προληπτική στρατηγική και αφετέρου έχοντας περαιτέρω αυξήσει ή ελαττώσει την αυστηρότητά της (βλέπε ενότητα B.2.2.5.4).

Το επίπεδο του φόρτου LD είναι το αποτέλεσμα αφενός της προσφερόμενης κίνησης από τις ενεργές ροές και αφετέρου των ρυθμίσεων της ισχύουσας στρατηγικής. Εάν, με την υιοθέτηση μιας στρατηγικής, σημειωθεί πτώση του LD , αυτή είναι πολύ πιθανό να οφείλεται στις νέες ρυθμίσεις και όχι σε μείωση της προσφερόμενης κίνησης. Η άρση των ρυθμίσεων και η επιστροφή στην προηγούμενη στρατηγική είναι κατ' αντιστοιχία πολύ πιθανό να επαναφέρει το LD σε υψηλά επίπεδα καθιστώντας εκ νέου αναγκαία τη λήψη της ίδιας στρατηγικής. Το φαινόμενο αυτό οδηγεί τον αλγόριθμο σε ταλαντώσεις μεταξύ δύο παρακείμενων στρατηγικών.

Αποσκοπώντας στην αποφυγή του φαινομένου των ταλαντώσεων και δεδομένου πως σε όλη την έκταση της κρίσιμης περιοχής εξ' ορισμού η πιθανότητα συμφόρησης θεωρείται υψηλή, η εκάστοτε ισχύουσα ήπια στρατηγική άρεται και αντικαθίσταται από την αμέσως προηγούμενη στη διατεταγμένη ακολουθία, μόνο όταν τα επίπεδα του φόρτου LD πέσουν κάτω από τα όρια της κρίσιμης περιοχής. Ή αλλιώς πως, αφού το LD ξεπεράσει μία φορά το κατώφλι $LD_{m_n}^b$ και υιοθετηθεί η στρατηγική m_n , αυτή παραμένει σε ισχύ όσο το LD κινείται στην περιοχή R_{CL} έως $LD_{m_{n+1}}^b$. Με άλλα λόγια το κατώφλι μιας στρατηγικής m_n υποδηλώνει

πως, εάν το LD βρεθεί πιο ψηλά από αυτό, τότε, αν όχι μια περισσότερο αυστηρή στρατηγική, κατ' ελάχιστο η στρατηγική m_n θα είναι σε ισχύ.

Όσο το LD , μετά την πάροδο χρόνου t_{min} από την τελευταία μεταβολή της στρατηγικής, παραμένει κάτω από τα όρια της κρίσιμης περιοχής και είναι σταθερό ή εξακολουθεί να μειώνεται, άρονται και οι επόμενες στη σειρά στρατηγικές μέχρι και την αποκατάσταση της βασικής στρατηγικής b_0 . Με αυτό τον τρόπο, εξασφαλίζεται πως το σύστημα δεν εμμένει σε κυρώσεις όταν ο φόρτος επανέρχεται σε χαμηλά επίπεδα εκτός της κρίσιμης περιοχής. Οσοδήποτε χαμηλά και αν βρίσκεται το LD ωστόσο, εάν η τάση LDT είναι ανοδική, η ισχύουσα στρατηγική παραμένει.

Γενικότερα, μεταβολή της στρατηγικής που οδηγεί σε επαύξηση της υπάρχουσας τάσης LDT , δηλαδή μεταβολή προς λιγότερο αυστηρή στρατηγική όταν η τάση είναι ανοδική και αντίστροφα, δεν θεωρείται συνετή επιλογή. Η υπάρχουσα τάση, χωρίς να έχει ολοκληρώσει την εξέλιξή της, ενισχύεται από τη νέα στρατηγική προς την κατεύθυνση της αναίρεσής της και της εκ νέου αντικατάστασής με την προηγουμένως ισχύουσα. Η μόνη περίπτωση που κάτι τέτοιο θεωρείται σκόπιμο είναι όταν είναι επιθυμητή η συγκράτηση του φόρτου LD σε συγκεκριμένα επίπεδα. Τότε, ακόμα και όταν η τάση LDT είναι πτωτική, εφόσον τα επίπεδα του φόρτου εξακολουθούν να είναι υψηλότερα από το επιδιωκόμενο όριο, υιοθετείται νέα, ακόμα αυστηρότερη στρατηγική συμβάλλοντας σε νέα μείωση του LD μέχρι και τη συμμόρφωσή του με το όριο.

B.2.2.5.6.3 Κατάσταση Αντίδρασης

Στην κατάσταση αντίδρασης, η κατάσταση διαθεσιμότητας του TT είναι red, η συμφόρηση είναι γεγονός και ο αντικειμενικός σκοπός είναι η άμεση επίλυσή της, κατανέμοντας τις κυρώσεις επί των ενεργών ροών με δίκαιο κατά το δυνατό τρόπο σε όλη την επικράτεια του δικτύου.

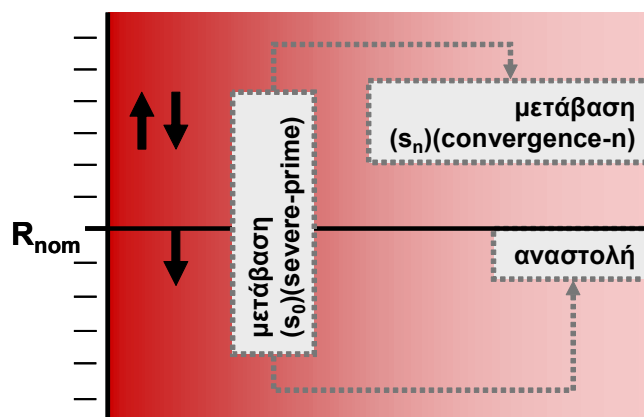
Η συμφόρηση είναι αποτέλεσμα της επικάλυψης μεταξύ διαφορετικών TTs και η επίλυσή της απαιτεί τη συντονισμένη δράση των στρατηγικών αποδοχής του κάθε TT. Στην πιο απλή περίπτωση, συμφόρηση συμβαίνει σε μία ζεύξη συνωστισμού (bottleneck link) στο δίκτυο κορμού και για ένα PHB, δηλαδή για μία QoS-κλάση, όταν η κίνηση που διέρχεται από τη ζεύξη για αυτό το PHB υπερβαίνει το R_{nom} που της αντιστοιχεί και τους τυχόν διαθέσιμους πόρους από τα άλλα PHBs που μοιράζονται τη φυσική χωρητικότητα της ζεύξης (βλέπε ενότητα B.1.2.2.5). Η απομόνωση και ο περιορισμός της συμφόρησης μόνο στο PHB με την αυξημένη ζήτηση και η διασφάλιση πως η κίνηση άλλων QoS-κλάσεων παραμένει ανεπηρέαστη, επιτυγχάνεται χάρη στις ρυθμίσεις αρχικοποίησης του traffic engineering αλλά και τις τυχόν δυναμικές ρυθμίσεις από τη λειτουργία δυναμικής διαχείρισης πόρων. Συνεπώς, τα TTs που επηρεάζονται, εφεξής αποκαλούμενα *πληγέντα* (affected), είναι εκείνα που κάποιο μονοπάτι τους περνάει από τη ζεύξη συνωστισμού και ανήκουν στην QoS-κλάση που υπερβαίνει το R_{nom} της σε αυτή τη ζεύξη.

Η κίνηση που διέρχεται από τη ζεύξη συνωστισμού για το PHB της συμφόρησης είναι η συνολική κίνηση από τα πληγέντα TTs, καθιστώντας τα έτσι υπεύθυνα για την πρόκληση της συμφόρησης και άρα και για την επίλυσή της. Η ειδοποίηση για την εκδήλωση ή την παύση της συμφόρησης διαδίδεται στα πληγέντα TTs από τις λειτουργίες παρακολούθησης με την έγερση ενός red ή green συναγερμού (alarm) μεταβάλλοντας αντιστοίχως και την κατάσταση των πληγέντων TTs. Με αυτό τον τρόπο, χάρη στην εμμέσως από κοινού παρακολούθηση της κάθε ζεύξης και προώθηση των αποτελεσμάτων, επιτυγχάνεται, χωρίς πληροφορία για τη φυσική ή τη λογική τοπολογία, ο συσχετισμός και ο συντονισμός μεταξύ των ανταγωνιζόμενων TTs κατά τη λειτουργία του συστήματος (runtime) και κατ' επέκταση η κινητοποίηση του μηχανισμού επίλυσης συμφόρησης σε κάθε ένα από τα πληγέντα και υπεύθυνα TTs, σε κάθε ένα από τους αντίστοιχα υπεύθυνους ακραίους κόμβους εισόδου.

Ο εναρμονισμός των ενεργειών της στρατηγικής καθενός από τα πληγέντα TTs για την από κοινού επίλυση της συμφόρησης στη ζεύξη συνωστισμού, επιτυγχάνεται χάρη στον λογικό καταμερισμό των πόρων σε απομονωμένα ανά TT μερίδια. Τα μερίδια αυτά υπολογίζονται κατά την αρχικοποίηση του συστήματος κεντρικά και βάσει της συνολικής φυσικής και λογικής τοπολογίας από την λειτουργία διαστασιοποίησης του δικτύου (βλέπε ενότητα B.1.2.2.4) και αποτυπώνονται στις ενδείξεις R_{nom} του RAB ανά TT (βλέπε ενότητα B.1.2.2.6). Εφόσον δεν υπάρχουν επικαλύψεις μεταξύ των κεντρικά προσδιοριζόμενων R_{nom} , διασφαλίζεται έτσι η σύγκλιση των ανεξάρτητων ενεργειών της στρατηγικής καθενός από τα πληγέντα TTs απαλλάσσοντας το σύστημα από την ανταλλαγή ειδικών μηνυμάτων. Αν δηλαδή κάθε TT συγκρατήσει τα επίπεδα του φόρτου LD που διοχετεύει στο δίκτυο κάτω του μεριδίου R_{nom} που του αντιστοιχεί, τότε λύνονται οι επικαλύψεις στους συνολικά διαθέσιμους πόρους της ζεύξης και η συμφόρηση υποχωρεί.

Η ανάγκη για άμεση επίλυση της συμφόρησης επιβάλλει την χωρίς δισταγμό επιβολή των μέγιστων κυρώσεων που δεν υποβαθμίζουν πέραν του αποδεκτού την ικανοποίηση, την επιτακτική υιοθέτηση δηλαδή της κύριας σφοδρής στρατηγικής s_0 . Εφόσον η απόκριση του συστήματος παραμένει η ίδια, εφόσον δηλαδή η συμφόρηση επιμένει και ο αλγόριθμος παραμένει στην κατάσταση αντίδρασης, η στρατηγική s_0 μπορεί μόνο να επαυξηθεί από στρατηγικές σύγκλισης αλλά όχι να αναιρεθεί και αντικατασταθεί από λιγότερο αυστηρές στρατηγικές. Με άλλα λόγια, στην κατάσταση αντίδρασης επιτρεπτές στρατηγικές είναι μόνο οι σφοδρές στρατηγικές.

Η λειτουργία δυναμικής διαχείρισης αποδοχής στην κατάσταση αντίδρασης φαίνεται στο Σχήμα B-25.



Σχήμα B-25: λειτουργία στην κατάσταση αντίδρασης

Το R_{nom} ορίζεται ως το δίκαιο μερίδιο των πόρων για το συγκεκριμένο TT. Υπολογίζεται βάσει των απαιτήσεων των προβλεπόμενων και των υπάρχοντων στην αρχή του τρέχοντος RPC εγγραφών υπηρεσιών, με μια μακροπρόθεσμη προοπτική (βλέπε ενότητα B.1.2.2.4). Ανεξάρτητα με τις προβλέψεις αφενός ως προς τις νέες εγγραφές που θα προκύψουν κατά τη διάρκεια του RPC και αφετέρου ως προς τις απαιτήσεις του συνολικού πληθυσμού των εγγραφών, η αμεροληψία του συστήματος κρίνεται στη μεταχείριση των υπηρεσιών που πραγματικά εγείρουν απαιτήσεις και για το πραγματικό ύψος αυτών των απαιτήσεων. Θεωρούμε συνεπώς πως η αμεροληψία αποτιμάται συγκρίνοντας την αυστηρότητα στην αποδοχή για κάθε TT ή αλλιώς τις κυρώσεις που θα υποστούν τελικά οι πραγματικές υπηρεσίες. Έτσι, δεδομένου πως για την επίλυση της συμφόρησης είναι απαραίτητο να μειωθεί η κίνηση που διοχετεύεται στο δίκτυο, δηλαδή να ενταθούν οι ισχύουσες κυρώσεις, το κριτήριο της αμεροληψίας επιβάλλει οι απαραίτητες κυρώσεις να καταναμηθούν με δίκαιο τρόπο σε όλες τις πληγείσες ενεργές υπηρεσίες ανά την επικράτεια του δικτύου, ανεξάρτητα με το φόρτο LD που εκείνη τη στιγμή τυχαίνει να παρουσιάζουν τα TTs από τα οποία

εξυπηρετούνται. Τελικά, ενώ το R_{nom} παρέχει το μέσο σύγκλισης του κατανεμημένου μηχανισμού επίλυσης της συμφόρησης, ωστόσο η δίκαιη κατανομή των κυρώσεων είναι το κριτήριο της αμεροληψίας.

Σε αυτή την κατεύθυνση ορίζουμε πως, με την είσοδο στην κατάσταση αντίδρασης, ανεξάρτητα από τα επίπεδα του φόρτου, ο αλγόριθμος είτε μεταβαίνει στην κύρια σφοδρή στρατηγική s_0 είτε, εάν βρίσκεται ήδη σε κάποια στρατηγική σύγκλισης s_n , παραμένει σε αυτή. Με άλλα λόγια, με την διάδοση του red συναγεμού, όλα τα πληγέντα TTs επιβάλουν τουλάχιστον τις κυρώσεις της κύριας σφοδρής στρατηγικής. Ένα TT είναι δυνατό να έχει ισχύουσα μια στρατηγική σύγκλισης εάν βρισκόταν πριν σε κατάσταση ομαλοποίησης η οποία προφανώς έληξε χωρίς να επιτύχει την ομαλή μετάβαση του συστήματος σε ομαλή κατάσταση. Σε αυτή την περίπτωση, σε αυτό το πρώτο βήμα δεν λαμβάνεται καμία περαιτέρω ενέργεια.

Εάν μετά από διάστημα t_{min} βρεθεί πως η κατάσταση συμφόρησης επιμένει και ο αλγόριθμος παραμένει σε κατάσταση αντίδρασης, τότε χρησιμοποιείται το κριτήριο σύγκλισης για την άμεση επίλυση της συμφόρησης σύμφωνα με το οποίο εκείνα τα TTs που διοχετεύουν κίνηση $LD > R_{nom}$ προχωρούν σε επόμενη στρατηγική σύγκλισης ενώ για όσα $LD \leq R_{nom}$ ο αλγόριθμος περνά σε αναστολή (stops) και παραμένει αδρανής για όσο ισχύει η συνθήκη $LD \leq R_{nom}$. Όταν η λειτουργία βασίζεται στο κριτήριο σύγκλισης τότε το μόνο μέγεθος που εξετάζεται είναι το επίπεδο του φόρτου LD ενώ η τάση του φόρτου LDT είναι αδιάφορη.

Η διαδικασία επαναλαμβάνεται μέχρι να πέσουν όλα τα TTs κάτω από R_{nom} οπότε είναι βέβαιο πως επιτυγχάνεται επίλυση της συμφόρησης. Ενώ η στρατηγική σύγκλισης διαμορφώνεται έτσι ώστε να επιφέρει άμεση αναχαίτιση του LD το αντίκτυπό της είναι αναμενόμενο να αποκλίνει από το στόχο οπότε και επιβάλλεται νέα διορθωτική στρατηγική σύγκλισης (βλέπε ενότητα B.2.2.5.4.2). Οι κυρώσεις που επιφέρει μια στρατηγική σύγκλισης αντιστοιχούν σε υποβάθμιση της ικανοποίησης πέραν του αποδεκτού almost-satisfied ορίου μέσω της μείωσης του ρυθμού εξυπηρέτησης κάτω από το SR_{AS} . Η υποβάθμιση αυτή δικαιολογείται εφόσον ο πρωταρχικός σκοπός είναι η επίλυση της συμφόρησης για την αποκατάσταση των χαρακτηριστικών ποιότητας μεταφοράς της QoS-κλάσης έστω και για το μικρότερο ρυθμό εξυπηρέτησης. Σημειώνεται πως επίλυση της συμφόρησης μπορεί να επιτευχθεί και από το πρώτο βήμα επιβολής της s_0 , ακόμα και με κάποια TTs να χρησιμοποιούν περισσότερους από R_{nom} πόρους εφόσον αυτοί είναι διαθέσιμοι από άλλα ανταγωνιζόμενα TTs.

Η σφοδρότητα των κυρώσεων που είναι αναγκαίο να επιβληθούν επί της προσφερόμενης κίνησης για να συμμορφωθεί με το κριτήριο σύγκλισης στο R_{nom} εξαρτάται προφανώς από την ένταση της προσφερόμενης κίνησης, που με τη σειρά της εξαρτάται από το επίπεδο ικανοποίησης SL και το επίπεδο προφύλαξης PL. Το επίπεδο ικανοποίησης επηρεάζει την ένταση του ρυθμού άφιξης εξουσιοδοτημένων αιτήσεων κλήσης, ενώ το επίπεδο προφύλαξης επηρεάζει το ποσοστό από αυτές τις αιτήσεις που θα γίνουν αποδεκτές, θα ενεργοποιηθούν και στη συνέχεια θα προσφέρουν κίνηση στο δίκτυο. Χαμηλό επίπεδο ικανοποίησης σε συνδυασμό με χαμηλό επίπεδο προφύλαξης έχουν ως αποτέλεσμα συχνότερη, περισσότερο επίμονη και πιο εκτεταμένη συμφόρηση και συνεπώς μεγαλύτερη υποβάθμιση της ποιότητας των ενεργών ροών, αφενός λόγω διάρκειας της συμφόρησης και αφετέρου λόγω σφοδρότητας των κυρώσεων για την επίλυσή της. Υψηλότερο επίπεδο προφύλαξης θα προστάτευε καλύτερα τις ενεργές ροές από τη συμφόρηση, μεταφέροντας το κόστος ικανοποίησης στη διαθεσιμότητα της υπηρεσίας. Υψηλότερο επίπεδο ικανοποίησης θα προστάτευε καλύτερα το βαθμό ικανοποίησης συνολικά, σε ποιότητα υπηρεσίας κατά τη χρήση αλλά και σε διαθεσιμότητα.

Επισημαίνουμε πως η κατάσταση αντίδρασης είναι η μόνη κατάσταση που η πολιτική του παροχέα δεν επεμβαίνει. Ο αλγόριθμος επιφορτίζεται με την αποκλειστικά μηχανιστική αντιμετώπιση του φαινομένου

της συμφόρησης που έχει προκύψει ως αποτέλεσμα της λειτουργίας στην ομαλή κατάσταση και στην κατάσταση ομαλοποίησης όπως έχει ρυθμιστεί από την πολιτική του παροχέα.

B.2.2.5.6.4 Κατάσταση Ομαλοποίησης

Στην κατάσταση ομαλοποίησης, η διαθεσιμότητα του TT έχει μόλις περάσει σε green, η συμφόρηση μόλις έχει επιτυχώς επιλυθεί κατά τη διάρκεια της προηγούμενης κατάστασης αντίδρασης και ο αντικειμενικός σκοπός εδώ είναι η ασφαλής αλλά και αποτελεσματική μετάβαση από την κατάσταση αντίδρασης στην ομαλή κατάσταση.

Η ασφάλεια αφορά στην αποφυγή της πρόσκαιρης άρσης των κυρώσεων που έχουν επιβληθεί κατά την κατάσταση αντίδρασης επιτυγχάνοντας την αναχαίτιση της προσφερόμενης κίνησης και την έξοδο από τη συμφόρηση. Ο περιορισμός της προσφερόμενης κίνησης που τελικά διοχετεύεται στο δίκτυο δεν ισοδυναμεί με την υποχώρηση των πηγών των χρηστών που διαμορφώνουν την προσφερόμενη κίνηση και άρα πρόσκαιρη άρση των κυρώσεων που την περιορίζουν θα οδηγήσει εκ νέου σε συμφόρηση. Όσο λοιπόν η προσφερόμενη κίνηση παραμένει υψηλή, παραμένει δικαιολογημένη και επιθυμητή και η επιβολή σφοδρών κυρώσεων.

Η αποτελεσματικότητα αφορά στην εμμονή σε απαράδεκτα ως προς το βαθμό ικανοποίησης σφοδρές κυρώσεις ενώ στο δίκτυο υπάρχουν διαθέσιμοι πόροι. Όταν τελικά η προσφερόμενη κίνηση υποχωρήσει τότε η άρση των κυρώσεων που περιορίζουν την προσφερόμενη κίνηση, αφενός επιτρέπεται αφού δεν οδηγεί πλέον σε συμφόρηση, και αφετέρου επιβάλλεται στα πλαίσια της μεγιστοποίησης της ικανοποίησης η οποία είναι κάτω του αποδεκτού όσο είναι σε ισχύ στρατηγικές σύγκλισης από την προηγούμενη κατάσταση αντίδρασης.

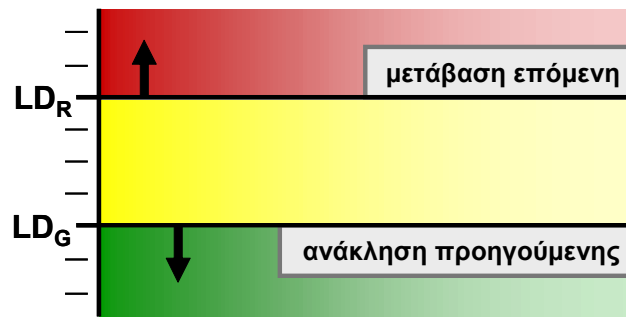
Διαπιστώνουμε πως, όπως στην ομαλή κατάσταση εκτιμώντας το ρίσκο επικείμενης συμφόρησης προβαίνουμε προληπτικά σε ήπιες κυρώσεις, έτσι στην κατάσταση ομαλοποίησης εκτιμώντας το ρίσκο επανάληψης της συμφόρησης εμμένουμε προληπτικά σε σφοδρές κυρώσεις. Με άλλα λόγια, στην ομαλή κατάσταση το ερώτημα είναι πόσο μπορούμε να καθυστερήσουμε την πιο αυστηρή στρατηγική, ενώ στην κατάσταση ομαλοποίησης πόσο μπορούμε να επισπεύσουμε την άρση της πιο αυστηρής στρατηγικής και το πέρασμα στην ομαλή κατάσταση θεωρώντας πως η προσφερόμενη κίνηση έχει υποχωρήσει και η συμφόρηση δεν θα επαναληφθεί.

Η μετάβαση στην ομαλή κατάσταση μπορεί να γίνει ως φυσική εξέλιξη ακολουθώντας το επίπεδο του φόρτου, ή μετά από επέμβαση του αλγορίθμου. Στις παρακάτω ενότητες διακρίνουμε μεταξύ της συμπεριφοράς του αλγορίθμου ως προς τη διαχείριση αποδοχής όσο παραμένει στην κατάσταση ομαλοποίησης οπότε και θεωρείται πως εξακολουθούν να ισχύουν οι συνθήκες που οδήγησαν σε συμφόρηση και ως προς τον μηχανισμό εξόδου από την κατάσταση ομαλοποίησης και μετάβασης στην ομαλή κατάσταση.

B.2.2.5.6.4.1 Διαχείριση αποδοχής

Κατά την έξοδο από την κατάσταση αντίδρασης και την είσοδο στην κατάσταση ομαλοποίησης το σύστημα είναι σε θέση να αποτιμήσει το ρίσκο συμφόρησης με περισσότερη ακρίβεια από ό,τι σε οποιαδήποτε άλλη περίπτωση. Έστω LD_R το επίπεδο του φόρτου όταν σημειώθηκε για τελευταία φορά συμφόρηση, το TT πέρασε σε red κατάσταση διαθεσιμότητας και ο αλγόριθμος εισήλθε σε κατάσταση αντίδρασης. Έστω αντίστοιχα LD_G το επίπεδο του φόρτου όταν εν συνεχεία υποχώρησε η συμφόρηση, το TT πέρασε σε green κατάσταση διαθεσιμότητας και ο αλγόριθμος εισήλθε στην κατάσταση ομαλοποίησης.

Η λειτουργία δυναμικής διαχείρισης αποδοχής στην κατάσταση ομαλοποίησης φαίνεται στο Σχήμα B-26.



Σχήμα B-26: λειτουργία στην κατάσταση ομαλοποίησης

Η απόκριση του συστήματος έχει καταδείξει πως όταν όλα τα πληγέντα TTs ξεπεράσουν το LD_R τότε το δίκτυο οδηγείται σε συμφόρηση. Συνεπώς, για να προστατευτεί το δίκτυο και να μην υποπέσει ξανά σε συμφόρηση κάθε TT οφείλει να συγκρατεί το φόρτο LD κάτω του ορίου LD_R .

Κατά τη διάρκεια της κατάστασης αντίδρασης, όλα τα TTs έχουν εφαρμόσει την κύρια σφοδρή στρατηγική, ανεξάρτητα με το επίπεδο του φόρτου τους, ενώ όσα TTs βρέθηκαν να υπερβαίνουν το R_{nom} εφάρμοσαν στρατηγικές σύγκλισης αναχαιτίζοντας την κίνηση πιθανώς κατά πολύ κάτω του R_{nom} . Αναμένεται με άλλα λόγια κατά την έξοδο από την κατάσταση αντίδρασης τα πληγέντα TTs να βρίσκονται κάποια αρκετά και κάποια μόλις κάτω του R_{nom} , αφήνοντας έτσι τους πόρους έως και το R_{nom} διαθέσιμους. Προκύπτει συνεπώς πως, μεταξύ της τιμής LD_G και της τιμής LD_R υπάρχει κάποιο περιθώριο στη διαθεσιμότητα των πόρων το οποίο επαυξάνεται αμέσως μετά την επίλυση της συμφόρησης, καθώς η προσφερόμενη κίνηση μειώνεται εξαιτίας της αποθάρρυνσης των χρηστών και του μηχανισμού ανίχνευσης συμφόρησης και υποχώρησης του TCP. Θεωρούμε έτσι πως η περιοχή που οριοθετείται από τα LD_G και LD_R , είναι κρίσιμη ωστόσο επιτρεπτή. Ενώ στη γενική περίπτωση $LD_G \leq LD_R$, ωστόσο είναι πιθανό για κάποιο TT η αύξηση σε προσφερόμενη κίνηση να υπήρξε μεγαλύτερη από την αύξηση σε σφοδρότητα των κυρώσεων κατά την κατάσταση αντίδρασης οπότε $LD_G > LD_R$. Σε αυτή την περίπτωση θέτουμε $LD_G = LD_R$ και η περιοχή μεταξύ τους συρρικνώνεται σε σημείο.

Στην κατάσταση ομαλοποίησης ο σκοπός είναι η μείωση των σφοδρών κυρώσεων. Έτσι, όταν ο φόρτος LD βρίσκεται κάτω από το όριο LD_G και παρουσιάζει πτωτική τάση LDT , δεδομένου πως το δίκτυο έχει δείξει πως μπορεί να προσφέρει χωρίς συμφόρηση τουλάχιστον το LD_G , η ισχύουσα στρατηγική αντικαθίσταται από την προηγούμενη λιγότερο αυστηρή στη διάταξη, η οποία διάταξη τώρα περιλαμβάνει τις στρατηγικές σύγκλισης όπως διαμορφώθηκαν κατά την κατάσταση αντίδρασης. Εάν το LD υπερβεί και πάλι το LD_G και για όσο μεταβάλλεται μέσα στην περιοχή έως και το LD_R δεν πραγματοποιείται καμία ενέργεια. Εάν ωστόσο ξεπεράσει και βρεθεί στην περιοχή άνω του LD_R και ανεξάρτητα από την τάση LDT που θα παρουσιάζει, τότε η ισχύουσα στρατηγική αντικαθίσταται από την επόμενη περισσότερο αυστηρή στη διάταξη, μέχρι και την πιο αυστηρή στρατηγική σύγκλισης που έχει ληφθεί κατά την κατάσταση αντίδρασης. Εάν ωστόσο ο φόρτος LD εξακολουθεί να υπερβαίνει το LD_R ο αλγόριθμος δεν προβαίνει σε νέα σφοδρότερη στρατηγική σύγκλισης για τη συγκράτησή του κάτω του LD_R . Αυτό συμβαίνει γιατί περαιτέρω αύξηση στη σφοδρότητα των κυρώσεων θεωρείται αδικαιολόγητη, δεδομένου πως οι ισχύουσες κυρώσεις είναι ήδη ιδιαίτερα σφοδρές, το σύστημα αποκρίνεται με κατάσταση green και η ζήτηση προφανώς είναι ιδιαίτερα αυξημένη καθώς παρατηρείται αύξηση πέρα του LD_R , που είχε σημειωθεί χωρίς την αυξημένη σφοδρότητα και πριν την αποθάρρυνση των χρηστών.

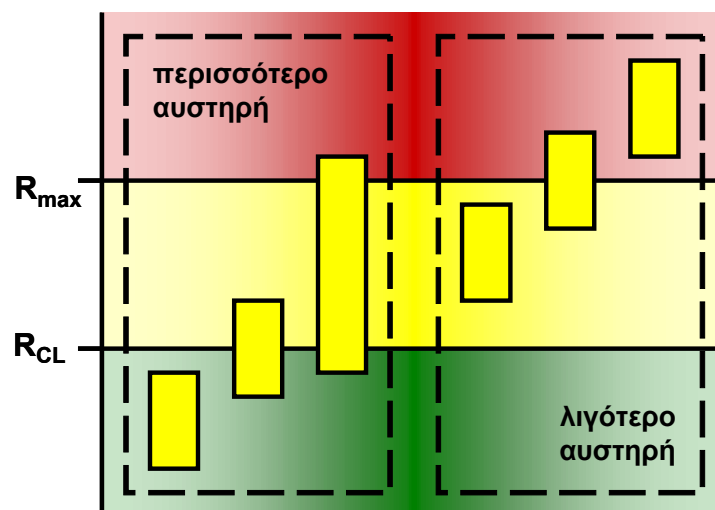
B.2.2.5.6.4.2 Μετάβαση στην ομαλή κατάσταση

Με την έξοδο από την κατάσταση αντίδρασης η αυστηρότητα της ισχύουσας στρατηγικής (τουλάχιστον κύρια σφοδρή στρατηγική ή σφοδρές στρατηγικές σύγκλισης) είναι μεγαλύτερη ή ίση από τη μέγιστη επιτρεπτή αυστηρότητα στην ομαλή κατάσταση (κύρια σφοδρή στρατηγική). Συνεπώς, προϋπόθεση για την μετάβαση στην ομαλή κατάσταση είναι η αναίρεση των στρατηγικών μέχρι και τα επιτρεπτά σε αυτή επίπεδα αυστηρότητας.

Η αναίρεση ή η επιβολή περισσότερο αυστηρών στρατηγικών γίνεται πάντα βάσει των αυξομειώσεων του φόρτου LD , είτε με τη λογική της κατάστασης ομαλοποίησης (βλέπε ενότητα B.2.2.5.6.4.1), είτε με τη λογική της ομαλής κατάστασης (βλέπε ενότητα B.2.2.5.6.2). Αναμένεται πως και με τη λογική της ομαλοποίησης εν καιρώ θα αναιρεθούν οι σφοδρές κυρώσεις που έχουν επιβληθεί κατά την κατάσταση αντίδρασης, επιτρέποντας στο σύστημα να μεταβεί στην ομαλή κατάσταση, οπότε και αναλαμβάνει η λογική της ομαλής κατάστασης.

Η ανάγκη για επέμβαση με σκοπό την επίσπευση της μετάβασης στην ομαλή κατάσταση ισχύει μόνο στην περίπτωση που η λογική στην κατάσταση ομαλοποίησης αποβαίνει σε πιο αυστηρές στρατηγικές από ό,τι θα συνέβαινε με τη λογική της ομαλής κατάστασης. Όταν η λειτουργία του αλγορίθμου στην κατάσταση ομαλοποίησης είναι λιγότερο αυστηρή τότε το σύστημα αφήνεται να επανέλθει στην ομαλή κατάσταση μόνο του, χωρίς περαιτέρω επέμβαση, όταν η μείωση στα επίπεδα του φόρτου επιτρέψει την αναίρεση των σφοδρών στρατηγικών.

Η αυστηρότητα ως προς τη λογική που διέπει την μετάβαση από αυστηρότερη σε ηπιότερη στρατηγική εξαρτάται από τις συνθήκες μετάβασης ως προς τις προϋποθέσεις του επιπέδου του φόρτου, ή αλλιώς από τον προσδιορισμό και την αντιστοίχιση περιοχών καταλληλότητας με συγκεκριμένες στρατηγικές. Τελικά, η αυστηρότητα της λογικής εκτιμάται βάσει της αντιστοιχίας φόρτου LD και αυστηρότητας στρατηγικής που χρησιμοποιεί η λογική.



Σχήμα B-27: αυστηρότητα σε κατάσταση ομαλοποίησης έναντι ομαλής κατάστασης

Χρησιμοποιώντας ως κριτήριο τη σχέση της κρίσιμης περιοχής όπως οριοθετείται από τα LD_G και LD_R στην κατάσταση ομαλοποίησης, με την κρίσιμη περιοχή όπως οριοθετείται από τα R_{CL} και R_{max} στην ομαλή κατάσταση διακρίνουμε έξι περιπτώσεις (βλέπε Σχήμα B-27). Κάθε μία από αυτές τις περιπτώσεις χαρακτηρίζεται ανάλογα με την αυστηρότητα της συνολικής συμπεριφοράς του αλγορίθμου στην

κατάσταση ομαλοποίησης συγκριτικά με την ομαλή κατάσταση, ως περισσότερο ή λιγότερο αυστηρή. Όταν η κρίσιμη περιοχή στην ομαλοποίηση ξεκινά πιο χαμηλά από την κρίσιμη περιοχή στην ομαλή κατάσταση, όταν δηλαδή το κάτω όριο LD_G είναι μικρότερο από το κάτω όριο R_{CL} , τότε το επίπεδο του φόρτου LD πρέπει να είναι χαμηλότερο από ό,τι στην ομαλή κατάσταση για να αντικατασταθεί η εκάστοτε ισχύουσα στρατηγική με λιγότερο αυστηρή στρατηγική. Όταν η κρίσιμη περιοχή στην ομαλοποίηση σταματά πιο χαμηλά από την κρίσιμη περιοχή στην ομαλή κατάσταση, όταν δηλαδή το άνω όριο LD_R είναι μικρότερο από το άνω όριο R_{max} , τότε η πιο σφοδρή δυνατή στρατηγική λαμβάνεται για μικρότερες τιμές του φόρτου LD στην κατάσταση ομαλοποίησης από την αντίστοιχη στρατηγική στην ομαλή κατάσταση. Και στις δύο αυτές περιπτώσεις η συμπεριφορά του αλγορίθμου είναι περισσότερο αυστηρή στην ομαλοποίηση.

Επισημαίνουμε πως, η κρίσιμη περιοχή της κατάστασης ομαλοποίησης καταδεικνύει και την τρέχουσα ζήτηση σε πόρους, την περίοδο που σημειώθηκε η συμφόρηση. Στις περιπτώσεις λιγότερο αυστηρής συμπεριφοράς, όπου $LD_G > R_{nom}$, δείχνει πως η ζήτηση είναι ιδιαίτερα υψηλή και έχει ήδη δεχθεί σφοδρές κυρώσεις καθώς στην κατάσταση αντίδρασης όσο το LD παραμένει μεγαλύτερο από R_{nom} υιοθετούνται όλο και σφοδρότερες στρατηγικές σύγκλισης. Επιζητώντας την αμερόληπτη μεταχείριση των υπηρεσιών, θεωρείται σκόπιμο να μην αυξηθούν περαιτέρω οι κυρώσεις, γι' αυτό και η λιγότερο αυστηρή συμπεριφορά του αλγορίθμου σε αυτές τις περιπτώσεις δικαιολογείται και αφήνεται χωρίς περαιτέρω περιορισμό.

Για τις περιπτώσεις που η κατάσταση ομαλοποίησης είναι λιγότερο αυστηρή από την ομαλή κατάσταση, η μετάβαση στην ομαλή κατάσταση λέγεται πως "κλειδώνει" όταν, μετά από την όποια διαδοχή στρατηγικών βάσει των διακυμάνσεων του φόρτου, μετά από αναίρεση φτάνει να υιοθετηθεί εκείνη η στρατηγική που το όριο της περιοχής επιβολής της στην ομαλή κατάσταση είναι το αμέσως μικρότερο από το LD_G . Έστω π.χ. $LD_{m_2}^b < LD_G \leq LD_{m_3}^b$, τότε η μετάβαση στην ομαλή κατάσταση "κλειδώνει" όταν, ακολουθώντας τις διακυμάνσεις του φόρτου με τη λογική της κατάστασης ομαλοποίησης, αναιρεθούν όλες οι στρατηγικές σύγκλισης που ενδεχομένως είχαν επιβληθεί κατά την κατάσταση αντίδρασης μέχρι και την στρατηγική m_3 οπότε και αυτή αντικατασταθεί με τη στρατηγική m_2 .

Αντίστοιχα, στις περιπτώσεις περισσότερο αυστηρής συμπεριφοράς, ενώ παρουσίασαν χαμηλή ζήτηση κατά τη συμφόρηση είναι πολύ πιθανό η ζήτηση να αυξηθεί περισσότερο προς τον προβλεπόμενο μέσο όρο, οπότε και να πρέπει να τους παραχωρηθεί μεγαλύτερο μέρος του μεριδίου τους χαλαρώνοντας την αυστηρότητα της λογικής στην κατάσταση ομαλοποίησης που περιορίζει τη χρήση πόρων άνω του LD_R και επιτρέποντας την υπέρβασή του σύμφωνα με τις αυξημένες απαιτήσεις της προσφερόμενης κίνησης με έχει δικαίωμα στους επιπλέον πόρους. Σε αυτή την περίπτωση ο αλγόριθμος επεμβαίνει και επισπεύδει τη μετάβαση στην ομαλή κατάσταση, οπότε και η αναίρεση ή επιβολή κυρώσεων διεξάγεται επί της κρίσιμης περιοχής και βάσει της λογικής της ομαλής κατάστασης. Αυτό για τις περιπτώσεις περισσότερο αυστηρής λογικής στην ομαλοποίηση, σημαίνει πως εάν ο φόρτος LD παρουσιάζει πτωτική τάση τότε αρκεί να είναι κάτω από R_{CL} και όχι κάτω από $LD_G < R_{CL}$ για να αναιρεθεί η ισχύουσα στρατηγική ενώ αν παρουσιάζει ανοδική τάση συνεχίζει και πάνω από το LD_R μέχρι και το όριο που επιτρέπεται για την ισχύουσα στρατηγική κατά την ομαλή κατάσταση.

Υπέρβαση του LD_R είναι ωστόσο πιθανό να οδηγήσει σε συμφόρηση και θα πρέπει να μην επιτρέπεται όσο η προσφερόμενη κίνηση δεν έχει αλλάξει από την τελευταία συμφόρηση και μπορεί δυναμικά να οδηγήσει ξανά σε συμφόρηση. Όπως είναι αδύνατο να προβλεφθεί με ακρίβεια η πιθανότητα του ενδεχομένου της συμφόρησης, είναι επίσης αδύνατο να προβλεφθεί με ακρίβεια μετά από πόσο χρόνο παύουν να ισχύουν οι συνθήκες που οδήγησαν σε συμφόρηση, ή αλλιώς πότε υποχωρεί αρκετά η προσφερόμενη κίνηση οπότε το σύστημα μπορεί με ασφάλεια να επιστρέψει στην ομαλή κατάσταση. Κατ' αναλογία με την ομαλή κατάσταση, υποτίμηση του χρόνου εκτόνωσης της συμφόρησης και πρόσκαιρη άρση των σφοδρών

κυρώσεων οδηγεί ξανά σε συμφόρηση ενώ υπερτίμηση του χρόνου εκτόνωσης και αδικαιολόγητη εμμονή στις σφοδρές κυρώσεις οδηγεί σε περαιτέρω υποβάθμιση της ικανοποίησης και απώλεια στην αξιοποίηση των πόρων. Το διάστημα που θεωρούμε πως παραμένουν οι συνθήκες που οδήγησαν σε συμφόρηση καθορίζει και την προφύλαξη που επιθυμούμε να εφαρμόσουμε έναντι του κόστους σε αξιοποίηση των πόρων το οποίο, αντίστοιχα με το επίπεδο προφύλαξης, προσδιορίζεται από την πολιτική του παροχέα στην παράμετρο *χρονικής διάρκειας συμφόρησης* (Congestion Duration Time – CDT).

Εάν ο χρόνος επέμβασης για την επίσπευση της μετάβασης στην ομαλή κατάσταση είναι ο ίδιος για όλα τα TTs, τότε το αντίκτυπο σε αύξηση του φόρτου στο δίκτυο συσσωρεύεται σε μία στιγμή, αυτή της ταυτόχρονης άρσης των κυρώσεων σε όλα τα TTs. Η συσσωρευμένη αύξηση του φόρτου είναι πολύ πιθανό να οδηγήσει ξανά σε συμφόρηση, και αφού περάσει η φάση αντίδρασης ξανά σε ταυτόχρονη επέμβαση κ.ο.κ. Με άλλα λόγια, κοινός χρόνος επέμβασης είναι δυνατό να συντονίσει την συμπεριφορά των TTs οδηγώντας το σύστημα σε ταλαντώσεις μεταξύ της κατάστασης red και της κατάστασης green. Είναι επομένως επιθυμητό, ο χρόνος που μεσολαβεί από την είσοδο στην κατάσταση ομαλοποίησης έως την επέμβαση για επίσπευση της μετάβασης στην ομαλή κατάσταση να διαφοροποιείται για κάθε TT και να διασπαρθεί κατά το δυνατόν με τυχαίο και μη εξαρτημένο τρόπο. Για αυτό το σκοπό χρησιμοποιείται ως κριτήριο διαφοροποίησης το συνολικό κόστος που έχει υποστεί κάθε TT από την αντίδραση στη συμφόρηση.

Τα TTs που ανήκουν στην περίπτωση της αυστηρότερης συμπεριφοράς στην ομαλοποίηση, διαφοροποιούνται περαιτέρω ανάλογα με το συνολικό κόστος που υπέστησαν από την αντίδραση στη συμφόρηση. Το συνολικό κόστος εξαρτάται από δύο παράγοντες, αφενός από τη σφοδρότητα των κυρώσεων που επεβλήθησαν και αφετέρου από το πλήθος των υπηρεσιών που βρέθηκαν ενεργές και υπέστησαν αυτές τις κυρώσεις. Η σφοδρότητα των κυρώσεων προσμετράται από τον *παράγοντα κυρώσεων* (Penalty Factor – PF) ενώ το πλήθος των ενεργών υπηρεσιών εκτιμάται μέσω των *καταβεβλημένων πόρων* (Contributed Resources – CR).

Δεδομένου πως όλα τα TTs έχουν εφαρμόσει την κύρια σφοδρή στρατηγική ο παράγοντας κυρώσεων διαφοροποιείται μόνο σε σχέση με την υποβάθμιση του ρυθμού εξυπηρέτησης κάτω του SR_{AS} . Εάν sp_{SR}^G είναι η τιμή της παραμέτρου ρυθμού εξυπηρέτησης (βλέπε ενότητα B.2.2.4.1.3) κατά την είσοδο στην κατάσταση διαθεσιμότητας green και στην ομαλοποίηση, τότε ορίζουμε $PF = \left| sp_{SR}^G \right|$.

Το πλήθος των ενεργών υπηρεσιών δεν είναι γνωστό. Ωστόσο, το πλάτος της κρίσιμης περιοχής LD_G έως LD_R αντιστοιχεί στους πόρους που έχασε το TT ως αποτέλεσμα του περιορισμού των ενεργών υπηρεσιών από τις σφοδρές στρατηγικές που επεβλήθησαν κατά τη διάρκεια της συμφόρησης. Οι πόροι που έχασε το TT κατά τη διάρκεια της κατάστασης αντίδρασης θεωρούνται ως η συνεισφορά του για την επίλυση της συμφόρησης και αποκαλούνται *καταβεβλημένοι πόροι* και υπολογίζονται ως $CR = \frac{LD_R - LD_G}{R_{\max}}$.

Έτσι, ο χρόνος επέμβασης για την επίσπευση της μετάβασης υπολογίζεται ως τυχαία ανά TT μεταβλητή, έστω T^n , χρησιμοποιώντας μια κατανομή με μέσο όρο έστω T_{avg}^n , ο οποίος υπολογίζεται βάσει των CDT, PF και CR.

B.2.2.5.6.5 Εξομάλυνση Ταλαντώσεων

Το φαινόμενο των ταλαντώσεων απορρέει από την ίδια τη φύση του βρόχου ανάδρασης. Η λογική της δυναμικής διαχείρισης αποδοχής καθοδηγείται από τη διαθεσιμότητα των πόρων στο δίκτυο και η διαθεσιμότητα των πόρων στο δίκτυο επηρεάζεται από τη λογική που καθορίζει τη στρατηγική αποδοχής.

Ταλαντώσεις συμβαίνουν καταρχήν μεταξύ των στρατηγικών. Η μεταβολή της ισχύουσας στρατηγικής ισοδυναμεί στη γενική περίπτωση με αντικατάσταση από την παρακείμενή της στην προδιαγεγραμμένη αυστηρά διατεταγμένη ακολουθία στρατηγικών. Χάρη σε αυτό τον περιορισμό καθίσταται εφικτή και η αναγνώριση ενός μοτίβου στην εναλλαγή τους που υπό προϋποθέσεις χαρακτηρίζεται ως ταλάντωση.

Ένας από τους βασικότερους λόγους που μπορεί να οδηγήσουν το σύστημα σε ταλαντώσεις μεταξύ των στρατηγικών είναι η συμπεριφορά των χρηστών και οι μηχανισμοί του TCP πρωτοκόλλου. Θεωρώντας για παράδειγμα πως το αντίκτυπο από την ανάληψη της ν-οστής στρατηγικής είναι τόσο έντονο που ο διοχετευόμενος φόρτος LD μειώνεται κατακόρυφα με αποτέλεσμα την αναίρεση της ν-οστής στρατηγικής και την αντικατάστασή της με την μικρότερη σε αυστηρότητα ($v-1$). Η μείωση του φόρτου ωστόσο πιθανώς δεν αντιστοιχεί σε πραγματική υποχώρηση της προσφερόμενης κίνησης αλλά είναι ενδεχομένως εξ ολοκλήρου αποτέλεσμα του μηχανισμού αποφυγής συμφόρησης (congestion avoidance) του TCP. Το TCP ανταποκρίνεται με μείωση της προσφερόμενης κίνησης στη μείωση του ρυθμού (throughput) κατά τη ν-οστή στρατηγική, ενώ τείνει να επαναφέρει τον αρχικό ρυθμό όταν διαγνώσει αποκατάσταση του ρυθμού από την αναίρεση της στρατηγικής. Ο αρχικός ρυθμός της προσφερόμενης κίνησης όμως θα προκαλέσει και πάλι την ανάληψη της ν-οστής στρατηγικής κ.ο.κ.

Πέρα από τις ταλαντώσεις μεταξύ στρατηγικών, ταλαντώσεις συμβαίνουν και στην πρόκληση και υποχώρηση της συμφόρησης, ή αλλιώς μεταξύ της τιμής red και green για την κατάσταση διαθεσιμότητας των πόρων του TT. Δεδομένου πως στην κατάσταση αντίδρασης λαμβάνονται σφοδρές κυρώσεις, όταν τελικά επιλυθεί η συμφόρηση, αργά ή γρήγορα η δυναμική διαχείριση αποδοχής θα επιχειρήσει να αναιρέσει αυτές τις κυρώσεις για κάθε πληγέν TT. Αν ωστόσο η προσφερόμενη κίνηση δεν έχει υποχωρήσει επαρκώς λόγω του τερματισμού ενεργών κλήσεων τότε θα επανέλθει στα αρχικά επίπεδα δημιουργώντας και πάλι συμφόρηση, οπότε αντιδραστικά θα επανέλθουν οι σφοδρές κυρώσεις κ.ο.κ.

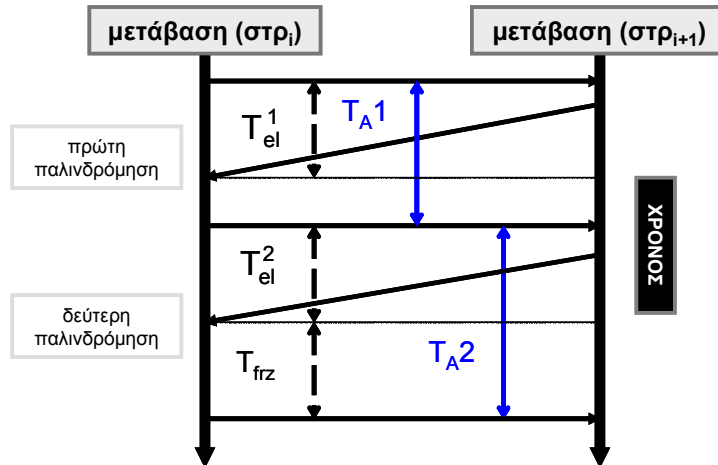
Γενικότερα στο κατανεμημένο περιβάλλον, η μεταβολή στη στρατηγική αποδοχής σε ένα ακραίο κόμβο επηρεάζει την κίνηση που διοχετεύεται στο δίκτυο, η οποία διασταυρώνεται στις ζεύξεις κορμού με την κίνηση από άλλους ακραίους κόμβους και προκαλεί ενδεχομένως αλλαγή της κατάστασης διαθεσιμότητας των αντίστοιχων TTs, πυροδοτώντας έτσι μεταβολή της στρατηγικής τους.

Ένας άλλος παράγοντας που μπορεί με ανεξέλεγκτο τρόπο να προκαλέσει ταλαντώσεις είναι ο συσχετισμός με τις αποφάσεις των άλλων δυναμικών λειτουργιών, δηλαδή των λειτουργιών δυναμικής διαχείρισης πόρων ανά κόμβο και δυναμικής διαχείρισης μονοπατιού ανά ακραίο κόμβο.

Επειδή οι συσχετίσεις και οι βρόχοι ελέγχου (control loops) που οδηγούν σε ταλαντώσεις είναι ιδιαίτερα δύσκολο και πολύπλοκο να μοντελοποιηθούν και να επιλυθούν με βέλτιστο τρόπο, γι' αυτό και οι μηχανισμοί αντιμετώπισης περιορίζονται στην αναγνώριση και την εξομάλυνση των ταλαντώσεων μέχρι να αλλάξουν οι συνθήκες των εξωτερικών, μη ελεγχόμενων παραγόντων που τις προξενούν. Διακρίνουμε δύο διαφορετικούς μηχανισμούς για τις ταλαντώσεις μεταξύ στρατηγικών και τις ταλαντώσεις στην κατάσταση διαθεσιμότητας εφόσον εφαρμόζονται σε διαφορετικές πλευρές και φάσεις του προβλήματος.

B.2.2.5.6.5.1 Εξομάλυνση Ταλαντώσεων Μεταξύ Στρατηγικών

Θεωρώντας δύο διαδοχικές στρατηγικές αποδοχής στην προδιαγεγραμμένη διατεταγμένη ακολουθία, έστω σ_{tr_i} και $\sigma_{tr_{i+1}}$, η αλληλουχία των μεταβολών ($\sigma_{tr_i} \rightarrow \sigma_{tr_{i+1}} \rightarrow \sigma_{tr_i}$), έστω πρώτη παλινδρόμηση (βλέπε Σχήμα B-28), δεν συνιστά υποψήφια ταλάντωση καθώς είναι η φυσική εξέλιξη ενός φαινομένου αύξησης και εν συνεχεία μείωσης του συνολικού φόρτου. Υποψήφια ταλάντωση συνιστά η αλληλουχία των μεταβολών ($\sigma_{tr_i} \rightarrow \sigma_{tr_{i+1}} \rightarrow \sigma_{tr_i} \rightarrow \sigma_{tr_{i+1}} \rightarrow \sigma_{tr_i}$), έστω δεύτερη παλινδρόμηση.



Σχήμα B-28: σενάριο παλινδρομήσεων μεταξύ στρατηγικών

Το κριτήριο για το χαρακτηρισμό διαδοχικών παλινδρομήσεων ως ταλάντωση είναι ο χρόνος εξέλιξης των φαινομένων. Εάν οι παλινδρομήσεις του φόρτου που οδηγούν στις παλινδρομήσεις μεταξύ των διαδοχικών στρατηγικών εξελίσσονται σε κάποιο επαναλαμβανόμενο χρονικό μοτίβο τότε θεωρούμε πως οι εξωτερικοί παράγοντες, δηλαδή οι ενεργές ροές που δημιουργούν την κίνηση, παραμένουν σταθεροί και πως το σύστημα δεν μπορεί να αντεπεξέλθει και να βρει κάποιο κατάλληλο σημείο ισορροπίας μεταξύ των δύο στρατηγικών. Ταλάντωση στο παράδειγμά μας παρουσιάζεται όταν ο χρόνος για την εκδήλωση της δεύτερης παλινδρόμησης, έστω T_{el}^2 , δεν διαφέρει σημαντικά από τον αντίστοιχο χρόνο της πρώτης παλινδρόμησης, έστω T_{el}^1 , ή αλλιώς εάν $|T_{el}^1 - T_{el}^2| < T_{el-\deltaelta}^{min}$, όπου $T_{el-\deltaelta}^{min}$ η ελάχιστη διαφορά μεταξύ αλληλάλληλων παλινδρομήσεων που θεωρείται πως καταδεικνύει διαφοροποίηση της συμπεριφοράς. Για την αντιμετώπιση του φαινομένου της ταλάντωσης, η μετάβαση στην στρατηγική σ_{tr_i} καθυστερείται κατά χρόνο ακινητοποίησης (freeze time), έστω T_{frz} .

Όσο οι ταλαντώσεις εξακολουθούν, όσο δηλαδή σημειώνονται παλινδρομήσεις και αυτές χαρακτηρίζονται ταλαντώσεις βάσει του κριτηρίου και χωρίς να υπολογίζεται η τεχνητή ακινητοποίηση, τόσο ο χρόνος ακινητοποίησης αυξάνει με εκθετικό τρόπο για την επιπλέον εξομάλυνση του φαινομένου. Επισημαίνουμε πως ο χρόνος ακινητοποίησης προστίθεται στον δεδομένο από τον αλγόριθμο ελάχιστο χρόνο μεταξύ διαδοχικών μεταβολών t_{min} (βλέπε ενότητα B.2.2.5.1). Από κάποιο αριθμό παλινδρομήσεων και μετά επιχειρείται επιστροφή σε φυσιολογική λειτουργία, και ο χρόνος ακινητοποίησης τίθεται και πάλι στην αρχική τιμή του. Ο αλγόριθμος που χρησιμοποιείται για τον υπολογισμό του χρόνου ακινητοποίησης βασίζεται στη μελέτη [JJS99].

B.2.2.5.6.5.2 Εξομάλυνση Ταλαντώσεων Κατάστασης Διαθεσιμότητας

Παλινδρομήσεις στην κατάσταση διαθεσιμότητας αναμένεται να συμβαίνουν όσο ο αλγόριθμος βρίσκεται στην κατάσταση ομαλοποίησης. Η επανάληψη της συμφόρησης στην κατάσταση ομαλοποίησης μπορεί να συμβαίνει είτε εξαιτίας της πρόσκαιρης αναίρεσης των σφοδρών κυρώσεων ενόσω η προσφερόμενη κίνηση δεν έχει ακόμα πραγματικά υποχωρήσει, είτε λόγω περαιτέρω αύξησης της έντασης της προσφερόμενης κίνησης. Η επανάληψη της συμφόρησης λόγω περαιτέρω αύξησης ωστόσο είναι μια φυσική εξέλιξη που ακολουθεί την εξέλιξη των εξωτερικών παραγόντων και δεν εμπίπτει στην περίπτωση της ταλάντωσης του συστήματος.

Η αναίρεση των κυρώσεων συμβαίνει είτε λόγω πτώσης του επιπέδου του φόρτου κάτω από LD_G , είτε μετά από την επίσπευση της μετάβασης στην ομαλή κατάσταση (βλέπε ενότητα B.2.2.5.6.4.2). Οι ταλαντώσεις αντιμετωπίζονται και στα δύο αυτά επίπεδα.

Με την είσοδο στην κατάσταση ομαλοποίησης για πρώτη φορά και πριν από την επιτυχή μετάβαση στην ομαλή κατάσταση, κάθε φορά που σημειώνεται συμφόρηση και το σύστημα παλινδρομεί μεταξύ κατάστασης αντίδρασης και κατάστασης ομαλοποίησης ακολουθώντας τις παλινδρομήσεις μεταξύ red και green της κατάστασης διαθεσιμότητας, αυξάνεται ο χρόνος επέμβασης για την επίσπευση, και ειδικότερα ο μέσος όρος της τυχαίας μεταβλητής T_{avg}^n , με την ίδια λογική που αυξάνεται ο χρόνος ακινητοποίησης T_{fix} στο μηχανισμό εξομάλυνσης ταλαντώσεων μεταξύ στρατηγικών. Επιπλέον, έστω LD_R^i και LD_G^i αντίστοιχα οι τιμές του φόρτου κατά τη μεταβολή σε red και green αντίστοιχα της *i*-οστής παλινδρόμησης, τότε ως LD_R κατά την κατάσταση ομαλοποίησης επιλέγεται πάντα το $\min(LD_R^i)$ και ως LD_G το $\min(LD_R^i) - \max(LD_R^i - LD_G^i)$ θέτοντας έτσι τα αυστηρότερα δυνατά όρια για την αναίρεση των κυρώσεων λόγω πτώσης του επιπέδου.

B.2.3 Ανακεφαλαίωση

Η διαχείριση εγγραφής είναι κεντρική λειτουργία.

Κάθε αίτηση εγγραφής μεταφράζεται και αντιστοιχίζεται βάσει του δομικού μοντέλου στα TTs που θα υποστηρίξουν την κίνηση που θα προκύψει από το αντίστοιχο συμβόλαιο και ο αλγόριθμος ελέγχου εισόδου εφαρμόζεται ανά TT. Ο αλγόριθμος αποφαινεται επί της μακροπρόθεσμης δυνατότητας του δικτύου, βάσει των διαθέσιμων πόρων του RAB, να εξυπηρετήσει ανά πάσα στιγμή τη συνολική κίνηση από τις ήδη εγγεγραμμένες και τη νέα υπηρεσία. Η συνολική ανά TT προσφερόμενη κίνηση ανά πάσα στιγμή, είναι στατιστικό μέγεθος και εξάγεται βάσει parameter-based μοντέλων.

Η συσχέτιση των TTs που ανταγωνίζονται για τους πόρους μιας ζεύξης κορμού δεν είναι γνωστή με αποτέλεσμα το αντίκτυπο της αποδοχής μιας νέας υπηρεσίας να μπορεί να υπολογιστεί μόνο για τα άμεσα εμπλεκόμενα TTs. ***Η λειτουργική πολιτική του παροχέα αποτυπώνεται στην παράμετρο επιπέδου ικανοποίησης SL που ο παροχέας θεωρεί επαρκές να παρέχει στις εγγεγραμμένες υπηρεσίες.*** Η ρύθμιση του επιπέδου ικανοποίησης δεν προεξοφλεί το βαθμό ικανοποίησης που θα λάβουν τελικά οι υπηρεσίες, ο οποίος αποκλίνει ανάλογα με τα στατιστικά λάθη των μοντέλων και την προβλεψιμότητα στη συμπεριφορά των χρηστών. Η χρησιμότητα του επιπέδου ικανοποίησης έγκειται στη δυνατότητα εμπειρικής ρύθμισης της συμπεριφοράς του αλγορίθμου. ***Η πιθανότητα υποβάθμισης της ποιότητας είναι αντιστρόφως ανάλογη προς την τιμή του επιπέδου ικανοποίησης.***

Στη φάση αρχικοποίησης υπολογίζεται από τους πόρους του RAB και βάσει του επιπέδου ικανοποίησης, η επιτρεπτή περιοχή που θα παραχωρηθεί προς χρήση στη συνολική προβλεπόμενη κίνηση. Ενώ η επιτρεπτή

περιοχή είναι σταθερή, η συνολική προσφερόμενη κίνηση μεταβάλλεται ανάλογα με τις νέες αιτήσεις που γίνονται αποδεκτές αλλά και τη διακοπή ή λήξη συμβολαίων εγγραφής. Μία νέα αίτηση γίνεται αποδεκτή εάν η εκάστοτε συνολική προσφερόμενη κίνηση δεν υπερβαίνει την επιτρεπτή περιοχή, σε αντίθετη περίπτωση το σύστημα προχωρά σε διαπραγματεύσεις.

Η λειτουργία διαχείρισης κλήσης είναι καταναμημένη σύμφωνα με το feedback-based μοντέλο. Το στιγμιότυπο διαχείρισης κλήσης σε έναν ακραίο κόμβο ελέγχει εκείνα τα TTs στα οποία διοχετεύει κίνηση, ή αλλιώς για τα οποία αποτελεί κόμβο εισόδου.

Ο έλεγχος εισόδου επιδρά αφενός στην αποδοχή κίνησης από τις ενεργές ροές και αφετέρου στην αποδοχή νέων ενεργών ροών. Ονομάζουμε *στρατηγική αποδοχής* την τοποθέτηση της διαδικασίας λήψης αποφάσεων ως προς κάθε παράγοντα που επηρεάζει την κίνηση που διοχετεύεται στο δίκτυο και **διακρίνουμε μεταξύ των διαδικασιών λήψης και υλοποίησης των αποφάσεων οι οποίες ανατίθενται αντίστοιχα στις λειτουργίες δυναμικής διαχείρισης αποδοχής και εφαρμογής στρατηγικής αποδοχής.** Η λειτουργία δυναμικής διαχείρισης αποδοχής, καθοδηγούμενη από τις μετρήσεις, διαθέτει τη λογική να αποφανθεί την κατάλληλη στρατηγική αποδοχής την οποία μεταθέτει προς υλοποίηση στη λειτουργία εφαρμογής στρατηγικής αποδοχής. Η λειτουργία εφαρμογής στρατηγικής αποδοχής επικοινωνεί με το χρήστη για το χειρισμό των αιτήσεων κλήσης και με τον ακραίο κόμβο για την ενεργοποίηση των υπηρεσιών.

Η ενδεδειγμένη στρατηγική αποδοχής είναι η περισσότερο παραχωρητική στρατηγική που δεν οδηγεί σε υπερφόρτωση του δικτύου. Το ενδεχόμενο επικάλυψης στη χρήση κοινόχρηστων μεριδίων και συνεπώς υπερφόρτωσης και συμφόρησης, δεν είναι δυνατό να προβλεφθεί με ακρίβεια. **Ο βαθμός στον οποίο η προληπτική προφύλαξη του συστήματος είναι αφενός δικαιολογημένη και αφετέρου επιθυμητή καθορίζεται από τη λειτουργική πολιτική του παροχέα μέσω του επιπέδου προφύλαξης PL.**

Το επίπεδο ικανοποίησης SL είναι αυτό που καθορίζει τον εν δυνάμει μέγιστο βαθμό ικανοποίησης των υπηρεσιών. Στα περιθώρια που οριοθετούνται από το επίπεδο ικανοποίησης, το επίπεδο προφύλαξης PL καθορίζει τον τύπο της υποβάθμισης της ικανοποίησης, είτε σε ποιότητα υπηρεσίας λόγω συμφόρησης κατά τη χρήση, είτε σε διαθεσιμότητα λόγω αποκλεισμού από τη χρήση με την απόρριψη των αιτήσεων κλήσης. Εάν το επίπεδο προφύλαξης τεθεί σε αναντιστοιχία με τα επίπεδα επάρκειας των πόρων, τότε παρατηρείται απόκλιση από τον μέγιστο δυνατό βαθμό ικανοποίησης.

Η υλοποίηση της στρατηγικής προϋποθέτει την αντιστοίχιση μεταξύ υπηρεσιών και ανά TT στρατηγικών αποδοχής. Η εξυπηρέτηση κάθε υπηρεσίας πρέπει να συμβαδίζει με τη στρατηγική αποδοχής του κάθε TT στο οποίο διοχετεύει κίνηση και αντίστροφα, η μεταβολή στη στρατηγική αποδοχής για ένα TT θα πρέπει να εφαρμόζεται σε κάθε υπηρεσία που το χρησιμοποιεί. Η έγκριση για την αποδοχή μιας υπηρεσίας μετατίθεται και προωθείται σε κάθε εμπλεκόμενο ακραίο εισόδο. Η λειτουργία διαχείρισης κλήσης επιφορτίζεται τέλος με τη μετάφραση και αντιστοίχιση των κανόνων ενεργοποίησης του συμβολαίου σε ρυθμίσεις των μηχανισμών του επιπέδου δεδομένων ώστε η ροή κίνησης να αποκτήσει πρόσβαση στην υπηρεσία.

Ανάλογα με την απόκριση του συστήματος ως προς την κατάσταση συμφόρησης, ο αλγόριθμος μεταβαίνει στην ομαλή κατάσταση, την κατάσταση αντίδρασης και την κατάσταση ομαλοποίησης. Στην ομαλή κατάσταση, η τροποποίηση της στρατηγικής αποδοχής είναι προαιρετική, καθοδηγείται από το επίπεδο του φόρτου τοπικά και εξαρτάται από το επίπεδο προφύλαξης. Στην κατάσταση αντίδρασης είναι επιτακτικό να ληφθούν οι απαραίτητες ενέργειες ώστε να επιλυθεί η συμφόρηση. Η κατάσταση ομαλοποίησης μεσολαβεί για την ασφαλή μετάβαση από την κατάσταση αντίδρασης στην ομαλή κατάσταση όταν αποφανθεί την υποχώρηση της προσφερόμενης κίνησης και κατ' επέκταση της λανθάνουσας συμφόρησης.

Θεωρούμε ένα πεπερασμένο σύνολο προκαθορισμένων στρατηγικών αποδοχής, απόλυτα διατεταγμένων μεταξύ τους ως προς την παραχωρητικότητα τους. Σε κάθε κατάσταση αρμόζουν και αντιστοιχίζονται διαφορετικές ομάδες στρατηγικών και διαφορετική λογική μετάβασης μεταξύ των επιμέρους στρατηγικών. Για την προληπτική προστασία του δικτύου προβλέπεται η εφαρμογή στρατηγικών, το κόστος των οποίων θεωρείται από το επίπεδο προφύλαξης δικαιολογημένο προκειμένου να αποφευχθεί μεγαλύτερο κόστος λόγω συμφόρησης στο δίκτυο.

Λόγω του βρόγχου ανάδραση το φαινόμενο των ταλαντώσεων είναι έμφυτο στο βασιζόμενο σε ανάδραση μοντέλο. *Διακρίνουμε ταλαντώσεις μεταξύ παρακείμενων στη διατεταγμένη ακολουθία στρατηγικών αποδοχής και ταλαντώσεις στην κατάσταση συμφόρησης του δικτύου. Κατάλληλοι μηχανισμοί εντοπισμού και εξομάλυνσης των ταλαντώσεων εφαρμόζονται στην κάθε περίπτωση.*

ΜΕΡΟΣ C. ΑΞΙΟΛΟΓΗΣΗ

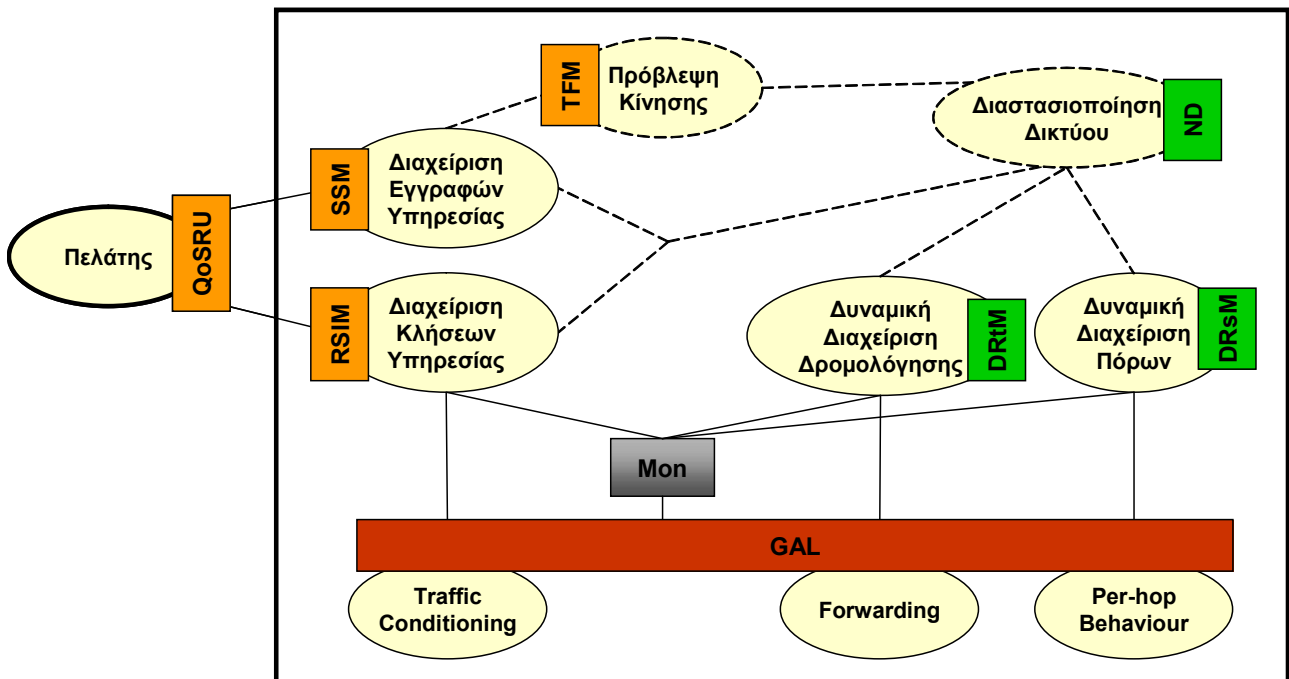
C.1 Υλοποίηση και Πρακτική Εφαρμογή

C.1.1 Μοντέλο Οντοτήτων Λογισμικού

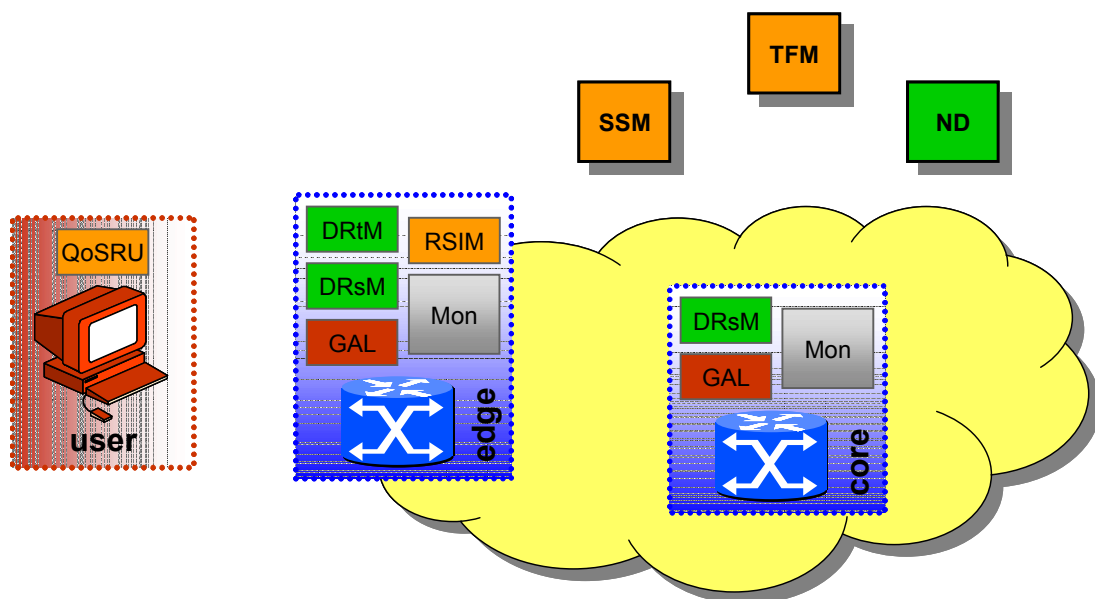
Οι οντότητες λογισμικού που υλοποιούν το λειτουργικό μοντέλο που έχουμε θεωρήσει για την ολοκληρωμένη παροχή ποιότητας υπηρεσίας (βλέπε ενότητα B.1.2.2.1) είναι οι εξής (βλέπε Σχήμα C-1):

- *Traffic Forecast Module (TFM)* : υλοποιεί τη λειτουργία πρόβλεψης κίνησης (βλέπε ενότητα B.1.2.2.3). Είναι κεντρική οντότητα, ένα στιγμιότυπο αντιστοιχεί στην επικράτεια του δικτύου.
- *Network Dimensioning (ND)* : υλοποιεί τη λειτουργία διαστασιοποίησης δικτύου (βλέπε ενότητα B.1.2.2.4). Είναι κεντρική οντότητα, ένα στιγμιότυπο αντιστοιχεί στην επικράτεια του δικτύου.
- *Service Subscription Module (SSM)* : υλοποιεί τη λειτουργία διαχείρισης εγγραφής (βλέπε ενότητα B.2.1). Είναι κεντρική οντότητα, ένα στιγμιότυπο αντιστοιχεί στην επικράτεια του δικτύου.
- *Router Service Invocation Module (RSIM)* : υλοποιεί τη λειτουργία διαχείρισης κλήσης (βλέπε ενότητα B.2.2). Είναι κατανεμημένη οντότητα, ένα στιγμιότυπο αντιστοιχεί σε κάθε ακραίο κόμβο της επικράτειας του δικτύου.
- *Dynamic Resource Management (DRsM)* και *Dynamic Route Management (DRtM)* : υλοποιούν τις λειτουργίες δυναμικής διαχείρισης πόρων και δυναμικής διαχείρισης μονοπατιού αντιστοίχως (βλέπε ενότητα B.1.2.2.5). Είναι κατανεμημένες οντότητες, η μεν DRtM αντιστοιχεί σε κάθε κόμβο ακραίο κόμβο και η δε DRsM σε κάθε κόμβο, ακραίο ή κόμβο κορμού.
- *Generic Adaptation Layer (GAL)* : υλοποιεί τις λειτουργίες του επιπέδου δεδομένων, δηλαδή τις λειτουργίες Traffic Conditioning, Forwarding και Per Hop Behaviour (βλέπε ενότητα B.1.2.2.1). Είναι κατανεμημένη οντότητα, αντιστοιχεί σε κάθε κόμβο και ουσιαστικά καλύπτει τις εγγενείς δυνατότητες ενός δρομολογητή που υποστηρίζει τεχνολογία DiffServ και MPLS με μια διεπιφάνεια σχεδιασμένη με γενικό τρόπο και ανεξάρτητη από την υποκείμενη υλοποίηση του κάθε κατασκευαστή, απλοποιώντας με αυτό τον τρόπο την εφαρμογή των λειτουργιών ελέγχου σε υλικό από διαφορετικό κατασκευαστή και κατ' επέκταση την αναβάθμιση του συστήματος με την αναβάθμιση του εξοπλισμού.
- *Monitoring (Mon)* : υλοποιεί τις λειτουργίες παρακολούθησης (βλέπε ενότητα B.1.2.2.1). Η αρχιτεκτονική των λειτουργιών παρακολούθησης είναι πολύπλοκη [TEQ-MON] και περιλαμβάνει οντότητες για τη συλλογή των στατιστικών τοπικά σε κάθε κόμβο, για την αποστολή δοκιμαστικών πακέτων από άκρο σε άκρο στους ακραίους κόμβους, για την επεξεργασία και τη διαμόρφωση αναφορών κεντρικά κ.λπ. Εδώ, περιορίζουμε την περιγραφή μας χρησιμοποιώντας τη σύμβαση η οντότητα Mon κατανεμημένη ανά κόμβο υλοποιεί όλες τις ως άνω λειτουργίες και παρέχει τις σχετικές υπηρεσίες στις άλλες οντότητες.
- *QoS Request Utility (QoSRU)* : υλοποιεί τις εξειδικευμένες λειτουργίες που είναι απαραίτητες στην πλευρά του πελάτη/χρήστη για να δημιουργεί, να αποστέλλει και να διαπραγματεύεται αιτήσεις εγγραφής και κλήσης μιας υπηρεσίας με τις οντότητες SSM και RSIM αντίστοιχα, πριν να είναι σε θέση να χρησιμοποιήσει την υπηρεσία.

Η αντιστοιχία των οντοτήτων λογισμικού στις φυσικές οντότητες παρουσιάζεται στο Σχήμα C-2.



Σχήμα C-1: οντότητες λογισμικού λειτουργικού μοντέλου

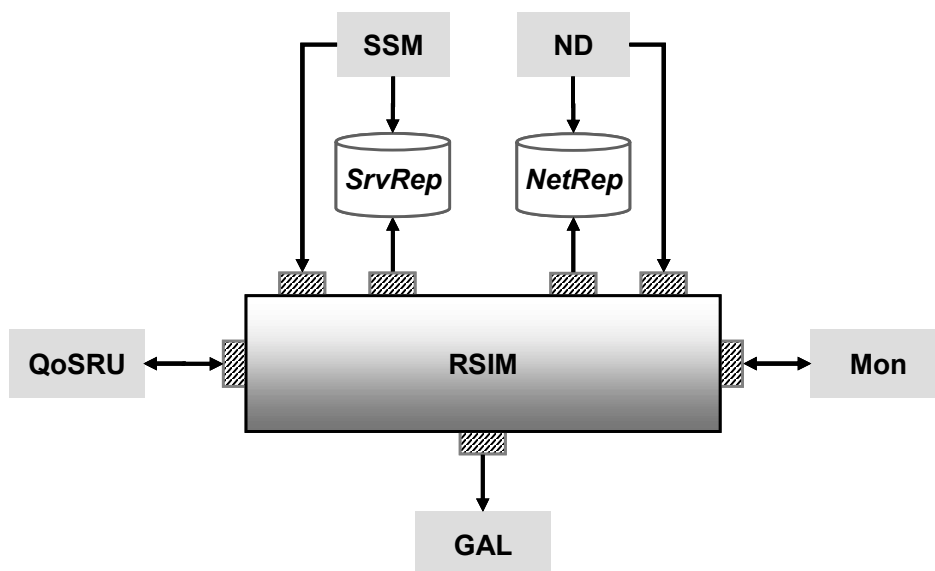


Σχήμα C-2: φυσική αντιστοιχία οντοτήτων λογισμικού

C.1.2 Router Service Invocation Module

C.1.2.1 Interfaces

Κατ' αντιστοιχία με την λειτουργία διαχείρισης κλήσης (βλέπε Σχήμα B-18), η οντότητα RSIM επικοινωνεί με τις οντότητες SSM, ND, QoSRU, Mon και GAL (βλέπε Σχήμα C-3).



Σχήμα C-3: RSIM interactions

ΑΠΟ	ΠΡΟΣ	ΜΕΘΟΔΟΣ
ND	NetRep	<i>store_RAM</i>
ND	RSIM	<i>new_RPC</i>
RSIM	NetRep	<i>get_RAM</i>
RSIM	Mon	<i>start_monitor_TT</i>
RSIM	Mon	<i>stop_monitor_TT</i>
Mon	RSIM	<i>handle_load_update</i>
Mon	RSIM	<i>handle_net_status_alarm</i>
QoSru	RSIM	<i>invocation_request</i>
RSIM	QoSru	<i>invocation_request_reply</i>
QoSru	RSIM	<i>invocation_termination</i>
SSM	SrvRep	<i>add_subscription</i>
SSM	SrvRep	<i>remove_subscription</i>
SSM	RSIM	<i>new_permanent_subscription</i>
SSM	RSIM	<i>subscription_removed</i>
RSIM	SrvRep	<i>get_subscription_info</i>
RSIM	GAL	<i>add_traffic_conditioning_block</i>
RSIM	GAL	<i>modify_traffic_conditioning_block</i>
RSIM	GAL	<i>remove_traffic_conditioning_block</i>

Πίνακας C-1: μέθοδοι των interfaces της οντότητας RSIM

Η πληροφορία που αφορά τις εγγεγραμμένες υπηρεσίες αποθηκεύεται στην αποθήκη υπηρεσιών (Service Repository – SrvRep), ενώ η πληροφορία που αφορά τη φυσική και τη λογική τοπολογία του δικτύου αποθηκεύεται στην αποθήκη δικτύου (Network Repository – NetRep). Οι αλληλεπιδράσεις μεταξύ των

οντοτήτων και οι αντίστοιχες μέθοδοι (operations) των interfaces (βλέπε Πίνακας C-1) περιγράφονται ακολούθως.

C.1.2.1.1 *RSIM και ND*

Η οντότητα ND κατά τη φάση αρχικοποίησης του RPC υπολογίζει τη νέα λογική τοπολογία και το καινούριο RAM και το αποθηκεύει στο NetRep καλώντας την *store_RAM*. Στη συνέχεια ειδοποιεί όλες τις ενδιαφερόμενες οντότητες για την έναρξη του νέου RPC ώστε να ενημερωθούν για τη νέα λογική τοπολογία και το νέο RAM καλώντας σε καθεμία από αυτές τη *new_RPC*. Η οντότητα RSIM ανακτά την πληροφορία για το νέο RAM καλώντας την *get_RAM* στο NetRep.

C.1.2.1.2 *RSIM και Mon*

Η οντότητα RSIM θέτει υπό παρακολούθηση κάθε TT που διαχειρίζεται καλώντας την *start_monitor_TT* της οντότητας Mon και τερματίζει την παρακολούθησή του καλώντας αντίστοιχα την *stop_monitor_TT*. Καθ' όλη τη διάρκεια παρακολούθησης ενός TT η Mon ενημερώνει την RSIM για τα επίπεδα του φόρτου *LD* που διοχετεύεται από τον τοπικό ακραίο κόμβο στο TT καλώντας περιοδικά την *handle_load_update*, και σε περίπτωση αλλαγής της κατάστασης διαθεσιμότητας του TT καλώντας την *handle_net_status_alarm*.

C.1.2.1.3 *RSIM και QoSRU*

Ο χρήστης μέσω της QoSRU επικοινωνεί με την οντότητα RSIM για την κλήση υπηρεσιών ρητής κλήσης καλώντας την *invocation_request* και για τον τερματισμό της την *invocation_termination* αντίστοιχα. Η έκβαση της διαδικασίας ελέγχου εισόδου ως προς την αποδοχή ή την απόρριψη της αίτησης κλήσης υπηρεσίας διαβιβάζεται από την οντότητα RSIM στο χρήστη καλώντας την *invocation_request_reply* της QoSRU.

C.1.2.1.4 *RSIM και SSM*

Η οντότητα SSM ενημερώνει το SrvRep για τη σύναψη ενός νέου συμβολαίου εγγραφής ή για την απόσυρση ενός υπάρχοντος συμβολαίου καλώντας τις *add_subscription* και *remove_subscription* αντίστοιχα. Σε περίπτωση σύναψης ενός νέου συμβολαίου για μια υπηρεσία μόνιμης ενεργοποίησης (βλέπε ενότητα B.1.2.1.1.2), η SSM ενημερώνει την RSIM καλώντας την *new_permanent_subscription* ώστε να αναλάβει την αυτόματη ενεργοποίηση βάσει του χρονοδιαγράμματος της υπηρεσίας. Σε περίπτωση απόσυρσης ενός νέου συμβολαίου, η SSM ενημερώνει την RSIM καλώντας την *subscription_removed* ώστε να τερματίσει κάθε τυχόν ενεργή κλήση της υπηρεσίας. Η οντότητα RSIM ανακτά κάθε απαραίτητη πληροφορία για την εξουσιοδότηση και την ενεργοποίηση μιας υπηρεσίας καλώντας την *get_subscription_info* στο SrvRep.

C.1.2.1.5 *RSIM και GAL*

Η οντότητα RSIM ενεργοποιεί μια υπηρεσία, διαμορφώνει δηλαδή τις ρυθμίσεις για την εξυπηρέτηση της κίνησης μιας υπηρεσίας στο επίπεδο δεδομένων, καλώντας την *add_traffic_conditioning_block* της οντότητας GAL. Αντίστοιχα, ο τερματισμός μιας ενεργής υπηρεσίας πραγματοποιείται με την άρση των ρυθμίσεων καλώντας την *remove_traffic_conditioning_block* της οντότητας GAL. Τέλος, η

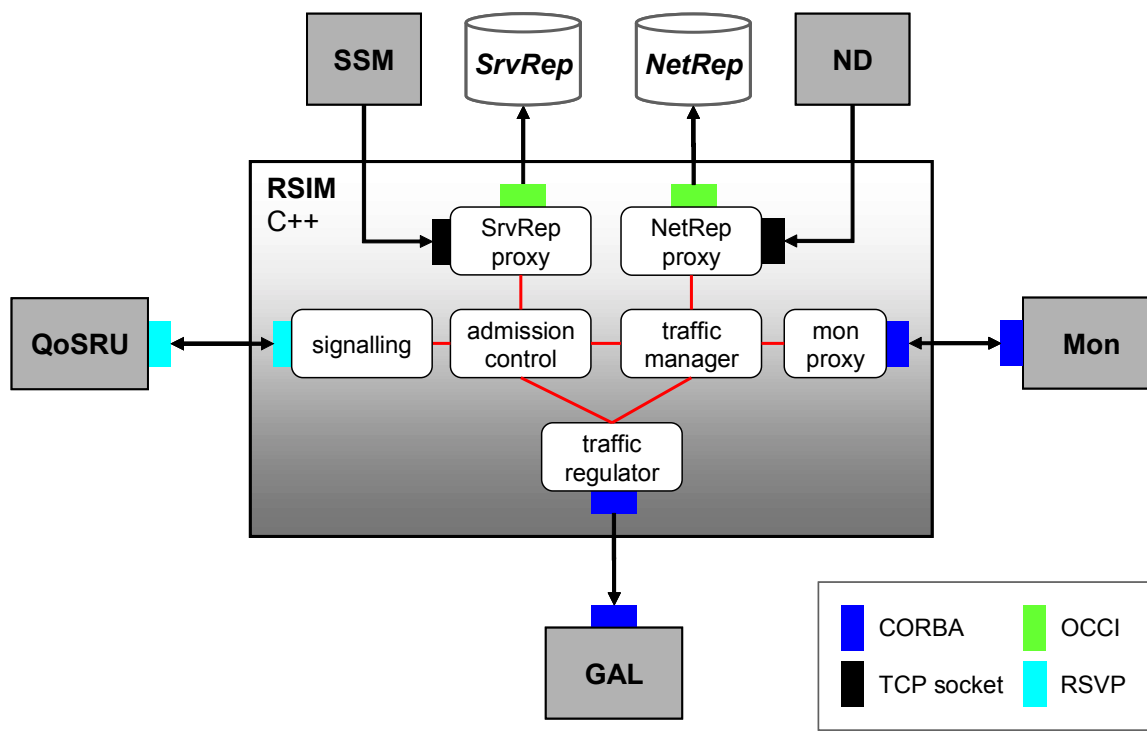
ποιότητα που παρέχεται στην υπηρεσία μεταβάλλεται ως αποτέλεσμα μεταβολής της στρατηγικής αποδοχής με την μετατροπή των ρυθμίσεων καλώντας την `modify_traffic_conditioning_block`.

C.1.2.2 Επιμερισμός υπολειτουργιών

Ήδη από τις προδιαγραφές της λειτουργίας διαχείρισης κλήσης έχουμε διακρίνει μεταξύ της λειτουργίας εφαρμογής στρατηγικής αποδοχής και της λειτουργίας δυναμικής διαχείρισης αποδοχής (βλέπε ενότητα B.2.2.3).

C.1.2.2.1 Λειτουργία δυναμικής διαχείρισης αποδοχής

Η λειτουργία δυναμικής διαχείρισης αποδοχής περαιτέρω επιμερίζεται στις λειτουργίες αντιπροσώπου παρακολούθησης (mon proxy), αντιπροσώπου αποθήκης δικτύου (NetRep proxy) και διαχειριστή της κίνησης (traffic manager) (βλέπε Σχήμα C-4).



Σχήμα C-4: υλοποίηση RSIM

Η λειτουργία mon proxy υλοποιεί το interface με τις λειτουργίες παρακολούθησης Mon (βλέπε ενότητα C.1.2.1.2), περιβάλλοντας τις δυνατότητες, τις προδιαγραφές και το κανάλι επικοινωνίας με την εκάστοτε υλοποίηση του Mon, εξάγοντας έτσι μια ανεξάρτητη και προσαρμοσμένη στις ανάγκες του αλγορίθμου θεώρηση των λειτουργιών παρακολούθησης.

Η λειτουργία NetRep proxy υλοποιεί το interface με τη λειτουργία διαστασιοποίησης του δικτύου (βλέπε ενότητα C.1.2.1.1). Το NetRep proxy είναι υπεύθυνο για τη διαμόρφωση και ενημέρωση της πληροφορίας που είναι απαραίτητη στον αλγόριθμο όσον αφορά τη φυσική και λογική τοπολογία του δικτύου και τους διαθέσιμους πόρους του RAM. Ο αλγόριθμος αναζητά την απαραίτητη πληροφορία καλώντας το NetRep proxy, το οποίο γνωρίζει εάν αυτή η πληροφορία υπάρχει σε τοπικό αντίγραφο, εάν χρειάζεται να την ανακτήσει καλώντας την απομακρυσμένη οντότητα του NetRep, τότε πρέπει να την ενημερώσει κ.λπ. Αντίστροφα το NetRep proxy με την έναρξη νέου RPC προκαλεί επανεκκίνηση του αλγορίθμου χρησιμοποιώντας το νέο RAM.

Η λειτουργία traffic manager ουσιαστικά υλοποιεί τον ανά TT αλγόριθμο δυναμικής διαχείρισης αποδοχής (βλέπε ενότητα B.2.2.5.6). Βάσει του RAM όπως προκύπτει από το NetRep proxy και των μετρήσεων όπως προκύπτουν από το mon proxy αποφαινεται την εκάστοτε στρατηγική αποδοχής την οποία και αναθέτει προς εφαρμογή στις υπολειτουργίες admission control και traffic regulator της λειτουργίας εφαρμογής στρατηγικής αποδοχής.

C.1.2.2.2 Λειτουργία εφαρμογής στρατηγικής αποδοχής

Η λειτουργία εφαρμογής στρατηγικής αποδοχής επιμερίζεται στις λειτουργίες *αντιπροσώπου αποθήκης υπηρεσιών* (SrvRep proxy), *σηματοδοσίας* (signalling), *ελέγχου αποδοχής* (admission control) και *ρυθμιστή της κίνησης* (traffic regulator) (βλέπε Σχήμα C-4).

Η λειτουργία SrvRep proxy υλοποιεί το interface με τη λειτουργία διαχείρισης εγγραφής (βλέπε ενότητα C.1.2.1.4). Το SrvRep proxy παρέχει τοπικά τη δυνατότητα αναζήτησης και ανεύρεσης κάθε πληροφορίας σχετικής με τις εγγεγραμμένες υπηρεσίες, σύμφωνα με τα περιεχόμενα του συμβολαίου εγγραφής ήδη αναλυμένα βάσει του δομικού μοντέλου από τη λειτουργία διαχείρισης εγγραφής. Η επικοινωνία με το απομακρυσμένο SrvRep και η δημιουργία τοπικών αντιγράφων για την αποτελεσματικότερη διαχείριση της πληροφορίας είναι θέματα υλοποίησης του SrvRep. Επιπλέον, το SrvRep proxy πυροδοτεί τον τερματισμό τυχόν ενεργών κλήσεων κατά την κατάργηση ενός συμβολαίου υπηρεσίας και αναλαμβάνει την έγκαιρη βάσει χρονοδιαγράμματος ειδοποίηση για την ενεργοποίηση των υπηρεσιών μόνιμης ενεργοποίησης.

Η λειτουργία signalling υλοποιεί τη μηχανή καταστάσεων (state machine) και την κωδικοποίηση των μηνυμάτων για την ενεργοποίηση μιας κλήσης στο κανάλι επικοινωνίας με την οντότητα QoSRU του χρήστη αλλά και με άλλα στιγμιότυπα της οντότητας RSIM σε άλλους ακραίους κόμβους. Όποια τεχνολογία και όποιες συμβάσεις και αν χρησιμοποιούνται, ο αλγόριθμος τροφοδοτείται με τις ειδικά προδιαγεγραμμένες πρωτογενείς λειτουργίες (primitives) και τις αντίστοιχες παραμέτρους.

Η λειτουργία admission control υλοποιεί τελικά τη διαδικασία αποδοχής ή απόρριψης κλήσεων υπηρεσιών ρητής κλήσης (βλέπε ενότητες B.2.2.4.1.1 και B.2.2.4.3). Τα μηνύματα που φτάνουν από τους χρήστες μεταφράζονται μέσω του signalling σε αιτήσεις κλήσης και διαβιβάζονται στο admission control. Εκεί, αφού ανακτηθεί η απαραίτητη πληροφορία από το SrvRep proxy, η αίτηση κλήσης περνά εξουσιοδότηση βάσει συμβολαίου και για κάθε TT στο οποίο αντιστοιχεί ελέγχονται οι κατάλληλοι παράμετροι ελέγχου αποδοχής, όπως έχουν ρυθμιστεί από το traffic manager. Εάν είναι απαραίτητη η επικοινωνία με άλλες ομότιμες οντότητες RSIM για την αποδοχή της αίτησης αυτή πραγματοποιείται μέσω του signalling. Μετά την αποδοχή μιας κλήσης, η υπηρεσία ενεργοποιείται μέσω του traffic regulator. Αντίστοιχα, ο τερματισμός μιας υπηρεσίας διαβιβάζεται μέσω του signalling στο admission control και από εκεί καλείται ο traffic regulator για την άρση των σχετικών ρυθμίσεων.

Η λειτουργία traffic regulator υλοποιεί τις ρυθμίσεις της κίνησης σύμφωνα με τις επιταγές των προδιαγεγραμμένων απαιτήσεων βάσει των συμβολαίων των υπηρεσιών (βλέπε ενότητα B.2.2.4.2) και των ρυθμίσεων ποιότητας και ρυθμού εξυπηρέτησης της ισχύουσας στρατηγικής αποδοχής (βλέπε ενότητες B.2.2.4.1.2 και B.2.2.4.1.3), όπως διαμορφώνεται κάθε φορά από το traffic manager. Έτσι, η μετάφραση και η αντιστοίχιση των υπηρεσιών που περνούν από το admission control σε ρυθμίσεις της κίνησης γίνεται από τον traffic regulator, όπως επίσης και η υλοποίηση των δυναμικών μεταβολών των στρατηγικών αποδοχής επί των υπηρεσιών όπως αποφασίζονται ανά TT από το traffic manager. Η υλοποίηση των ρυθμίσεων πραγματοποιείται με την εφαρμογή τους στη λειτουργία GAL.

C.1.2.3 Τεχνολογίες

Η υλοποίηση της οντότητας RSIM έγινε στα πλαίσια του πρωτοτύπου (prototype) του συστήματος TEQUILA [TEQ-ARCH] και χρησιμοποιήθηκαν οι τεχνολογίες και οι προδιαγραφές που υιοθετήθηκαν σε αυτό.

Η ίδια η οντότητα RSIM (βλέπε Σχήμα C-4), υλοποιήθηκε σε γλώσσα C++ σε περιβάλλον Linux με κέλυφος Linux kernel έκδοση 2.4.9³ και χρησιμοποιήθηκε μεταγλωττιστής GCC έκδοση 2.95.3⁴.

Η οντότητα GAL υλοποιεί ένα CORBA⁵ interface, το οποίο εξομοιώνει το πρωτόκολλο COPS-PR [COPS-PR]. Βάσει του COPS-PR το Policy Enforcement Point (PEP) που εδώ εδρεύει στο GAL υλοποιεί τις αποφάσεις του Policy Decision Point (PDP) που εδώ είναι κάθε φορά ο χρήστης του interface, σε αυτή την περίπτωση η οντότητα RSIM. Το περιεχόμενο και η σύνταξη των αποφάσεων είναι προδιαγεγραμμένα σε σχήματα XML⁶. Το CORBA interface διαθέτει μία μοναδική μέθοδο, την *enforce_decisions*, η οποία λαμβάνει ως παράμετρο το XML κείμενο που περιέχει τις αποφάσεις και επιστρέφει ένα XML κείμενο με τις ανά απόφαση αναφορές για την περάτωση της εφαρμογής, σύμφωνα με τις προδιαγραφές του COPS-PR. Για τη αυτοματοποίηση της διαμόρφωσης του XML κειμένου των αποφάσεων και της ανάγνωσης του XML κειμένου των αναφορών χρησιμοποιείται από το traffic regulator στην οντότητα RSIM η Xerces-C++ βιβλιοθήκη συντακτικής ανάλυσης XML έκδοση 1.6.0⁷. Η υλοποίηση της CORBA που χρησιμοποιείται είναι η ORBacus ORB⁸ έκδοση 4.1.0.

Η οντότητα Mon υλοποιεί ένα CORBA interface, για τον προσδιορισμό *αντικειμένων παρακολούθησης* (monitor) και τη δημιουργία *εργασιών παρακολούθησης* (monitor jobs). Από τις προδιαγραφές της οντότητας Mon προκύπτει πως αγνοεί τον τύπο TT, ενώ ο πλησιέστερος τύπος που αναγνωρίζει είναι το μονοπάτι. Η λειτουργία mon proxy αναλαμβάνει την αντιστοίχιση ενός TT σε αντικείμενα και εργασίες παρακολούθησης ανά μονοπάτι που το εξυπηρετεί, καθώς και τη συσχέτιση των αναφορών από κάθε μονοπάτι υπό παρακολούθηση για την εξαγωγή των συνολικών αναφορών ανά TT. Η αποστολή των αναφορών υλοποιείται στο Mon χρησιμοποιώντας την *υπηρεσία ειδοποίησης* (notification service) της CORBA. Με τη δημιουργία των εργασιών παρακολούθησης η διεργασία της οντότητας RSIM εγγράφεται στο *κανάλι γεγονότων* (event channel) στο οποίο εκπέμπονται από το Mon οι αναφορές που αφορούν όλες τις εργασίες για ένα αντικείμενο ως γεγονότα τα οποία, φιλτραρισμένα ανάλογα με το ενδιαφέρον κάθε εγγεγραμμένης διεργασίας ακροατή, καλούν τις κατάλληλες μεθόδους διαχείρισης γεγονότων (event handling), που εδώ επαυξάνονται και διακρίνονται από το mon proxy του RSIM σε *handle_load_update* και *handle_net_status_alarm* (βλέπε C.1.2.1.2). Η υλοποίηση της υπηρεσίας ειδοποίησης της CORBA που χρησιμοποιείται είναι η ORBacus Notify έκδοση 2.0.0⁹.

³ βλέπε <http://www.kernel.org/>

⁴ βλέπε <http://gcc.gnu.org/>

⁵ βλέπε <http://www.corba.org/>

⁶ βλέπε <http://www.w3.org/XML/>

⁷ βλέπε <http://xml.apache.org/xerces-c/>

⁸ βλέπε <http://www.orbacus.com/>

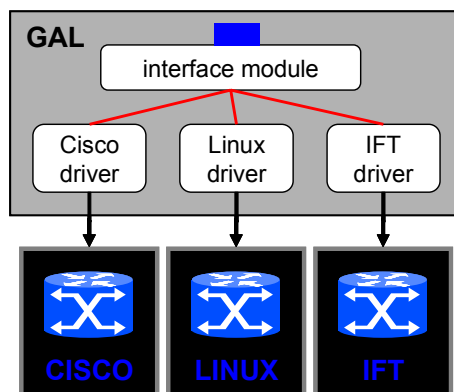
⁹ βλέπε <http://www.ionacom/products/orbacus/notify.htm>

Τα SrvRep και NetRep είναι σχήματα σε μια βάση δεδομένων. Η υλοποίηση της βάσης δεδομένων έγινε σε Oracle 9.0.1¹⁰. Για την ανάκτηση και την προσθήκη δεδομένων από μια απομακρυσμένη C++ διεργασία παρέχεται η βιβλιοθήκη OCCI (Oracle C++ Call Interface), η οποία έχει ενσωματωθεί και χρησιμοποιείται στη διεργασία RSIM για την επικοινωνία με τη βάση δεδομένων που υλοποιεί τα SrvRep και NetRep. Οι λειτουργίες SrvRep proxy και NetRep proxy γνωρίζουν το σχήμα των πινάκων της βάσης και σχηματίζουν τις κατάλληλες SQL εντολές οι οποίες μεταβιβάζονται μέσω του OCCI, εκτελούνται στη συνέχεια στη βάση και τα αποτελέσματα επιστρέφονται στη διεργασία ως αποτέλεσμα της κλήσης της τοπικής συνάρτησης (function call). Οι υλοποιήσεις των οντοτήτων SSM και ND έχουν γίνει σε Oracle PL/SQL¹¹, είναι συνεπώς ασύμβατες με κάθε άλλο τρόπο επικοινωνίας. Έτσι, για την επικοινωνία με τις οντότητες SSM και ND χρησιμοποιείται απλό TCP socket, πάνω από το οποίο περνούν τα μηνύματα από τον SSM και τον ND προς τον RSIM. Η συντακτική ανάλυση (parsing) της ακολουθίας χαρακτήρων που περνά από το TCP socket για την αναγνώριση των μηνυμάτων και των παραμέτρων τους βάσει των ειδικών συμβάσεων που έχουν συμφωνηθεί είναι και αυτή μέρος των λειτουργιών SrvRep proxy και NetRep proxy.

Το πρωτόκολλο σηματοδότησης θεωρείται πως είναι εμβόλιμο (piggybacked) στο πρωτόκολλο RSVP (Resource ReSerVation Protocol) [RSVP-FS]. Η λειτουργία signalling της οντότητας RSIM χρησιμοποιεί και επαυξάνει την υλοποίηση του RSVP δαίμονα για Linux έκδοση 4.2a4¹².

C.1.2.4 Υλοποίηση σε στοιχεία δικτύου

Πέραν της υλοποίησης του ίδιου του αλγορίθμου στην οντότητα RSIM όπως περιγράφεται στις παραπάνω ενότητες, θεωρείται απαραίτητη η τεκμηρίωση της δυνατότητας εφαρμογής του στα στοιχεία του δικτύου (network elements).



Σχήμα C-5: υλοποίηση GAL

Ολόκληρο το σύστημα TEQUILA εφαρμόζεται χωρίς άλλη μεταβολή πάνω από κάθε στοιχείο δικτύου που υποστηρίζεται από το GAL, καθώς αυτό αποκρύπτει τις ιδιομορφίες υλικού από διαφορετικούς κατασκευαστές. Το GAL υλοποιήθηκε για εμπορικούς δρομολογητές της εταιρείας Cisco, για πειραματικούς δρομολογητές υλοποιημένους με λογισμικό του λειτουργικού συστήματος Linux, και για πειραματικούς υβριδικούς δρομολογητές της εταιρείας France Telecom, επονομαζόμενους Internet Fast Translator (IFT)

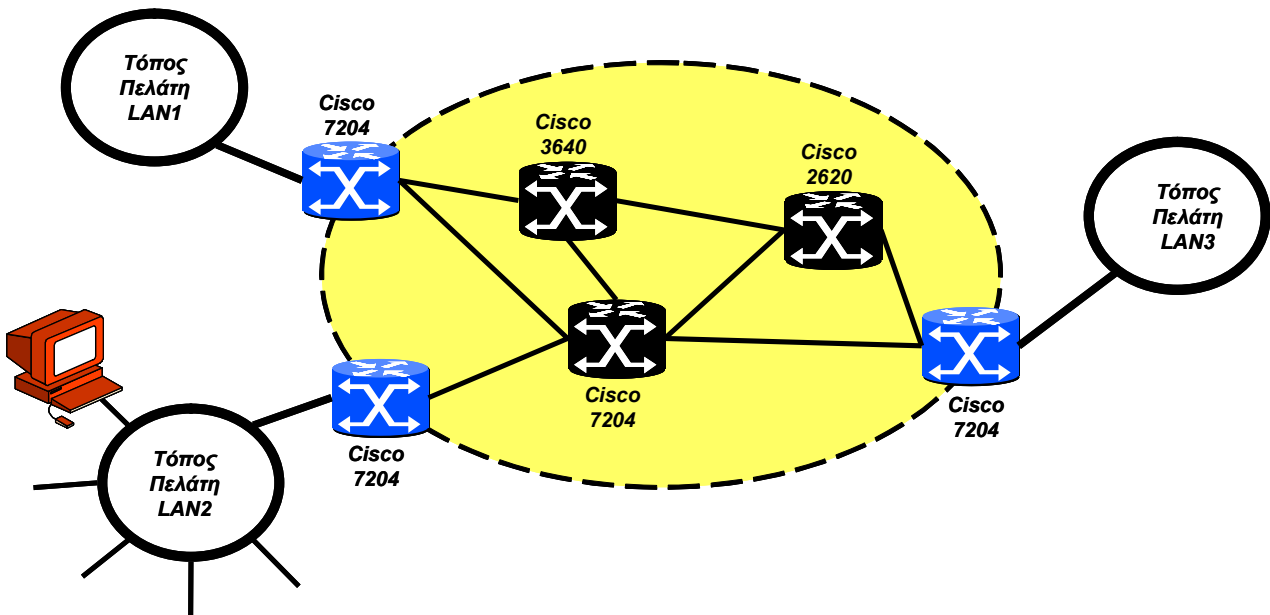
¹⁰ βλέπε <http://www.oracle.com/>

¹¹ βλέπε http://otn.oracle.com/tech/pl_sql/index.html

¹² βλέπε <http://dsmpis.atlantis.rug.ac.be/>

δρομολογητές, από τους οποίους μέρος των λειτουργιών αναπτύχθηκε πειραματικά σε υλικό και μέρος σε λογισμικό Linux (βλέπε Σχήμα C-5).

Ειδικότερα εδώ εξετάζουμε την εφαρμογή σε δρομολογητές Cisco. Η υλοποίηση του GAL απαιτεί ο υποκείμενος δρομολογητής να υποστηρίζει λογισμικό έκδοσης IOS 12.2¹³. Οι λειτουργίες του Traffic Conditioning υλοποιούνται χρησιμοποιώντας access list για τα φίλτρα ταξινόμησης και rate limit για το μετρητή, τη σημείωση και τη διαγραφή (βλέπε Σχήμα B-20). Ολόκληρο το σύστημα TEQUILA υλοποιήθηκε και εφαρμόστηκε σε διαφορετικά δικτυακά περιβάλλοντα δοκιμών (testbeds) αποτελούμενα από δρομολογητές Cisco ή Linux (π.χ. βλέπε Σχήμα C-6).



Σχήμα C-6: δικτυακό περιβάλλον εφαρμογής επί δρομολογητών Cisco

¹³ βλέπε <http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/index.htm>

C.2 Θεωρητική Μελέτη Πολυπλοκότητας

Η πολυπλοκότητα του αλγορίθμου σχετίζεται άμεσα με τη δυνατότητα εφαρμογής του σε μεγάλη κλίμακα.

Μερικές από τις βασικές αρχές που ακολουθήσαμε κατά το σχεδιασμό του συστήματος, του περιβάλλοντος και του ίδιου του αλγορίθμου, συνεισφέρουν στην εξασφάλιση της σταθερότητας και στη μείωση της πολυπλοκότητας έτσι ώστε να καθίσταται δυνατή η εφαρμογή της προτεινόμενης λύσης σε συστήματα μεγάλης κλίμακας. Τέτοιες αρχές είναι:

- *Συσσωμάτωση στο δομικό μοντέλο.* Χάρη στην επιλογή της τεχνολογίας DiffServ το μοντέλο των δομών πληροφορίας (βλέπε ενότητα B.1.2.1.3) που χρησιμοποιείται, τόσο από τις λειτουργίες traffic engineering του δικτύου, όσο και από τις λειτουργίες service management, επιτρέπει την αναπαράσταση των δυνατοτήτων του δικτύου και το λογισμό επί αυτών βάσει κλάσεων κίνησης, εδώ των QoS-κλάσεων, που προκύπτουν από τη συσσωμάτωση πολλών ξεχωριστών ροών. Η πολυπλοκότητα της διαχείρισης των κλάσεων είναι προφανώς μειωμένη σε σχέση με την πολυπλοκότητα της διαχείρισης των ροών.
- *Επιμερισμός των λειτουργιών.* Οι λειτουργίες που αφορούν στον έλεγχο εισόδου κίνησης επιμερίζονται και ανατίθενται σε διαφορετικές οντότητες και σε διαφορετική χρονική κλίμακα (βλέπε ενότητα B.1.2.2.2). Οι πολύπλοκες λειτουργίες βέλτιστης κατανομής των πόρων εκτελούνται στο παρασκήνιο κατά τη φάση αρχικοποίησης του συστήματος και τα αποτελέσματά τους εφοδιάζουν τις δυναμικές λειτουργίες για την άμεση εξυπηρέτηση της πραγματικής κίνησης. Η λειτουργία του ελέγχου εισόδου κίνησης επιμερίζεται περαιτέρω σε λειτουργία διαχείρισης εγγραφής και λειτουργία διαχείρισης κλήσης, μειώνοντας ακόμα περισσότερο τα καθήκοντα και τη συναφή πολυπλοκότητα του ελέγχου εισόδου κίνησης για την ενεργοποίηση μιας ροής. Οι αλληλεπιδράσεις μεταξύ των διαφορετικών λειτουργιών έχουν σχεδιαστεί με σαφήνεια και αποτελεσματικά χάρη και στην απλότητα του δομικού μοντέλου.
- *Κατανομή των λειτουργιών.* Η δυναμική λειτουργία ελέγχου εισόδου κίνησης κατανέμεται σε διαφορετικά στιγμιότυπα ανά ακραίο κόμβο που αναλαμβάνουν τη διαχείριση του υποσυνόλου της κίνησης που διοχετεύεται από τον κόμβο που τους αντιστοιχεί προς το δίκτυο. Ο συντονισμός των ενεργειών μεταξύ των ανταγωνιζόμενων στιγμιότυπων επιτυγχάνεται χάρη στην λήψη κοινών αναφορών παρακολούθησης όπως διαδίδονται από τις λειτουργίες παρακολούθησης. Κάθε στιγμιότυπο διαθέτει την απαραίτητη πληροφορία και αυτοδυναμία για να αποφανθεί για τους πόρους του δικτύου από τον ακραίο κόμβο. Ανταλλαγή μηνυμάτων συμβαίνει μόνο για την εξυπηρέτηση κλήσεων υπηρεσιών που εμπλέκουν περισσότερες από μία από άκρο σε άκρο συνδέσεις μονής κατεύθυνσης.
- *Συγκράτηση της δυναμικής ενημέρωσης.* Οι μετρήσεις που απαιτούνται από τον έλεγχο εισόδου κίνησης βασίζονται σε περιοδική ενημέρωση του τοπικού φόρτου και σε γνωστοποίηση κατά τη μεταβολή της δυαδικής ένδειξης της από άκρο σε άκρο κατάστασης του δικτύου. Επιπλέον, όπου αυτό είναι εφικτό, οι μετρήσεις περνούν από φίλτρα εξομάλυνσης, όπου η τιμή του εκάστοτε μεγέθους υπολογίζεται και βάσει του προηγούμενου δείγματος, επιτυγχάνοντας με αυτό τον τρόπο εξομάλυνση και απορρόφηση περιστασιακών ακραίων τιμών.

Στις ενότητες που ακολουθούν αναλύουμε την πολυπλοκότητα της λειτουργίας διαχείρισης κλήσης [MCG03,TEQ-SCST]. Εξετάζουμε την πολυπλοκότητα σε δύο επίπεδα, στο επίπεδο λήψης αποφάσεων και στο επίπεδο εφαρμογής των αποφάσεων. Η ανάλυση στηρίζεται στη θεώρηση της χειρότερης περίπτωσης (worst-case). Η δυνατότητα εφαρμογής μιας λύσης κρίνεται ανάλογα με το πεδίο εφαρμογής, θέτοντας έτσι διαφορετικές απαιτήσεις για την απόκριση του συστήματος στη φάση κλήσης της υπηρεσίας, διαφορετικές

στη φάση εγγραφής της υπηρεσίας και διαφορετικές για τις λειτουργίες που εκτελούνται στο παρασκήνιο για την προετοιμασία ενός νέου RPC.

Η πολυπλοκότητα επηρεάζεται από τους εξής παράγοντες:

- (1) αριθμό ακραίων κόμβων,
- (2) αριθμό QoS-κλάσεων,
- (3) αριθμό των εγγεγραμμένων υπηρεσιών και
- (4) αριθμό των SLSs που περιλαμβάνονται συνολικά στις εγγεγραμμένες υπηρεσίες.

Σημειώνεται πως εφόσον ένα SLS αντιστοιχεί σε ένα μέλος σύνδεσης από άκρο σε άκρο, ο αριθμός των SLSs ανά εγγεγραμμένη υπηρεσία είναι φραγμένος από το τετράγωνο του αριθμού των ακραίων κόμβων. Επιπλέον, καθώς ένα TT αντιστοιχεί σε μια σύνδεση μονής φοράς μεταξύ δύο ακραίων κόμβων και για μια QoS-κλάση, ο αριθμός των TTs στην επικράτεια του δικτύου περιορίζεται από το τετράγωνο του αριθμού των ακραίων κόμβων επί τον αριθμό των QoS-κλάσεων. Αντίστοιχα περιορίζεται και ο αριθμός των TTs που αντιστοιχίζονται και εξυπηρετούν την κίνηση από μια υπηρεσία.

Η λειτουργία διαχείρισης κλήσης εξετάζεται μέσω των δύο ξεχωριστών διαδικασιών που πυροδοτούνται από εξωτερικούς παράγοντες, της διαδικασίας εξυπηρέτησης αίτησης κλήσης και της διαδικασίας δυναμικής προσαρμογής βάσει των μηνυμάτων παρακολούθησης.

C.2.1 Εξυπηρέτηση κλήσης

Η εξυπηρέτηση κλήσης αναλύεται στα παρακάτω στάδια.

- *Στάδιο εξακρίβωσης γνησιότητας και εξουσιοδότησης βάσει συμβολαίου εγγραφής*

Αυτό το στάδιο περιλαμβάνει εύρεση και σύγκριση των τιμών των διαφόρων πεδίων της αίτησης κλήσης με τις αντίστοιχες του συμβολαίου εγγραφής. Καθώς τα στοιχεία των εγγεγραμμένων υπηρεσιών αποθηκεύονται σε μια κεντρική απομακρυσμένη βάση δεδομένων, η πρόσβαση σε απομακρυσμένα δεδομένα επιφέρει αντίστοιχη επιβάρυνση. Στην υλοποίησή μας διατηρούμε αντίγραφο της πληροφορίας για κάθε εγγραφή που έχει έστω και μία ενεργή κλήση ενώ, ανάλογα με το κόστος της απομακρυσμένης πρόσβασης, άλλες υλοποιήσεις είναι δυνατό να διατηρούν περισσότερο κάποια αντίγραφα στην ενδιάμεση μνήμη (cache memory). Κατά την εξουσιοδότηση γίνεται έλεγχος και των δικαιωμάτων ενεργοποίησης έναντι των ήδη ενεργών κλήσεων μιας υπηρεσίας, το οποίο όμως χρησιμοποιεί τοπικά διαθέσιμη πληροφορία.

- *Στάδιο εξουσιοδότησης βάσει διαθεσιμότητας πόρων*

Για κάθε TT που αντιστοιχεί στην υπηρεσία το στάδιο εξουσιοδότησης βάσει διαθεσιμότητας πόρων περιλαμβάνει την εξέταση της πιθανότητας αποδοχής όπως ορίζονται στην ισχύουσα για κάθε TT στρατηγική αποδοχής και σύμφωνα με το σχήμα βάσει της λογικής RED (βλέπε ενότητα B.2.2.4.1.1).

- *Έλεγχος εισόδου κίνησης σε ομότιμους εμπλεκόμενους κόμβους εισόδου*

Η έγκριση για την αποδοχή της κλήσης και η εφαρμογή των κατάλληλων ρυθμίσεων στο επίπεδο δεδομένων γίνεται σε κάθε εμπλεκόμενο ακραίο κόμβο εισόδου, ή αλλιώς για κάθε SLS. Αυτό το στάδιο εισάγει επιβάρυνση λόγω της προώθησης της αίτησης κλήσης από τον κόμβο που αρχικά τη λαμβάνει από το χρήστη παράλληλα προς κάθε άλλο εμπλεκόμενο κόμβο, την καθυστέρηση λόγω εξουσιοδότησης βάσει διαθεσιμότητας πόρων στους κόμβους και την αποστολή της απάντησης πίσω στον αρχικό κόμβο.

Αφού γίνει αποδεκτή μια αίτηση κλήσης η εφαρμογή της απόφασης σημαίνει την εφαρμογή των κατάλληλων ρυθμίσεων στο επίπεδο δεδομένων σε κάθε ακραίο κόμβο εισόδου κίνησης που θα εξυπηρετηθεί από την υπηρεσία. Ο αριθμός και η πολυπλοκότητα των ρυθμίσεων σχετίζονται με τον αριθμό των SLSs ανά υπηρεσία. Ανάλογα με το υλικό κάθε κατασκευαστή είναι πιθανό να τίθενται περιορισμοί στον αριθμό των ρυθμίσεων που μπορούν να υποστηριχτούν ταυτόχρονα, όπως π.χ. των αριθμό των φίλτρων ταξινόμησης για κάθε ενεργή ροή.

Ο χρόνος απόκρισης της διαδικασίας εξυπηρέτησης κλήσης είναι σημαντικός για την απόδοση του σχήματος και θεωρείται πως πρέπει να περιορίζεται σε κλίμακα δευτερολέπτων. Εκτός από τους περιορισμούς που τίθενται από το χρόνο μετάβασης με επιστροφή (ground trip time) κατά τη σηματοδότηση μεταξύ των ακραίων κόμβων και την καθυστέρηση από το υλικό του δρομολογητή για την εφαρμογή των ρυθμίσεων, η επιβάρυνση που εισάγεται στη διαδικασία από τον αλγόριθμο είναι αμελητέα καθώς συνοψίζεται στην εξέταση της τιμής της πιθανότητας όπως ορίζεται από την παράλληλη λειτουργία δυναμικής διαχείρισης αποδοχής.

C.2.2 Δυναμική προσαρμογή

Η διαδικασία δυναμικής προσαρμογής καλεί τη λειτουργία δυναμικής διαχείρισης αποδοχής και πυροδοτείται από τις αναφορές παρακολούθησης. Οι αναφορές παρακολούθησης λαμβάνονται ασύγχρονα κατά την εκδήλωση ενός γεγονότος όπως η υπέρβαση ενός κατωφλιού ή η ανίχνευση συμφόρησης. Ενώ ένα γεγονός όπως η συμφόρηση σε μια ζεύξη κορμού μπορεί να επηρεάζει πολλά διαφορετικά TTs, σε διαφορετικούς αλλά και στον ίδιο ακραίο κόμβο, θεωρούμε πως κάθε TT λαμβάνει ανεξάρτητη αναφορά του γεγονότος. Στην υλοποίησή μας, αυτό γίνεται αποτελεσματικά χάρη στην υπηρεσία ειδοποίησης που αναλαμβάνει την αποστολή αντιγράφων του μοναδικά παραγόμενου γεγονότος σε κάθε ενδιαφερόμενη οντότητα.

Σε κάθε TT ο αλγόριθμος δυναμικής διαχείρισης αποδοχής αποφαινεται τη νέα στρατηγική εξετάζοντας καταρχήν την τρέχουσα κατάσταση του αλγορίθμου, την κατάσταση συμφόρησης του TT και τα επίπεδα του φόρτου σε σχέση με προκαθορισμένα κατά την αρχικοποίηση όρια. Συνεπώς, η πολυπλοκότητα του αλγορίθμου ως προς τη λήψη αποφάσεων εξαντλείται στην εξέταση μηδαμινής και τοπικά διαθέσιμης πληροφορίας.

Η άμεση περάτωση της διαδικασίας θεωρείται σημαντική, καθώς εκτελεί τις διορθωτικές ρυθμίσεις που έχουν κριθεί απαραίτητες για την εύρυθμη λειτουργία του δικτύου. Η εφαρμογή της νέας στρατηγικής περιλαμβάνει την αλλαγή των τιμών των παραμέτρων που ρυθμίζουν τον έλεγχο αποδοχής κλήσης και τη μεταβολή των ρυθμίσεων ποιότητας και ρυθμού εξυπηρέτησης στο επίπεδο δεδομένων για όσες ενεργές ροές επηρεάζονται από τη μεταβολή της στρατηγικής αποδοχής για το συγκεκριμένο TT. Καθώς οι μεταβολές στο επίπεδο δεδομένων αφενός επιβαρύνουν σημαντικά τον κόμβο και αφετέρου, ανάλογα με την υλοποίηση από τον κατασκευαστή του υποκείμενου δρομολογητή, ενδεχομένως διαταράσσουν την ποιότητα που λαμβάνουν οι ενεργές ροές, είναι σημαντικό να εξετάσουμε και να περιορίσουμε τη συχνότητά τους. Σχετικά μέτρα έχουν ληφθεί με τη συγκράτηση της δυναμικής ενημέρωσης, τον καθορισμό μιας ακολουθίας προδιαγεγραμμένων στρατηγικών και την επιβολή ενός ελάχιστου χρονικού διαστήματος t_{\min} μεταξύ διαδοχικών μεταβολών. Με τη ρύθμιση αυτών των παραγόντων θεωρείται πως μπορεί να περιοριστεί η πολυπλοκότητα εφαρμογής της νέας στρατηγικής σε αποδεκτά επίπεδα.

Το αντίκτυπο της εφαρμογής της νέας στρατηγικής συνδέεται άμεσα και με τον αριθμό των ενεργών ροών που επηρεάζει. Είναι δυνατό να οριστούν υποσύνολα και οι προδιαγεγραμμένες στρατηγικές να διαιρεθούν

ανάλογα με τον αριθμό των υποσυνόλων. Αυτό επηρεάζει ωστόσο και το αντίκτυπο της νέας στρατηγικής στην κατάσταση του δικτύου. Τα σχετικά θέματα συνιστούν ανοικτό πεδίο προς περαιτέρω έρευνα.

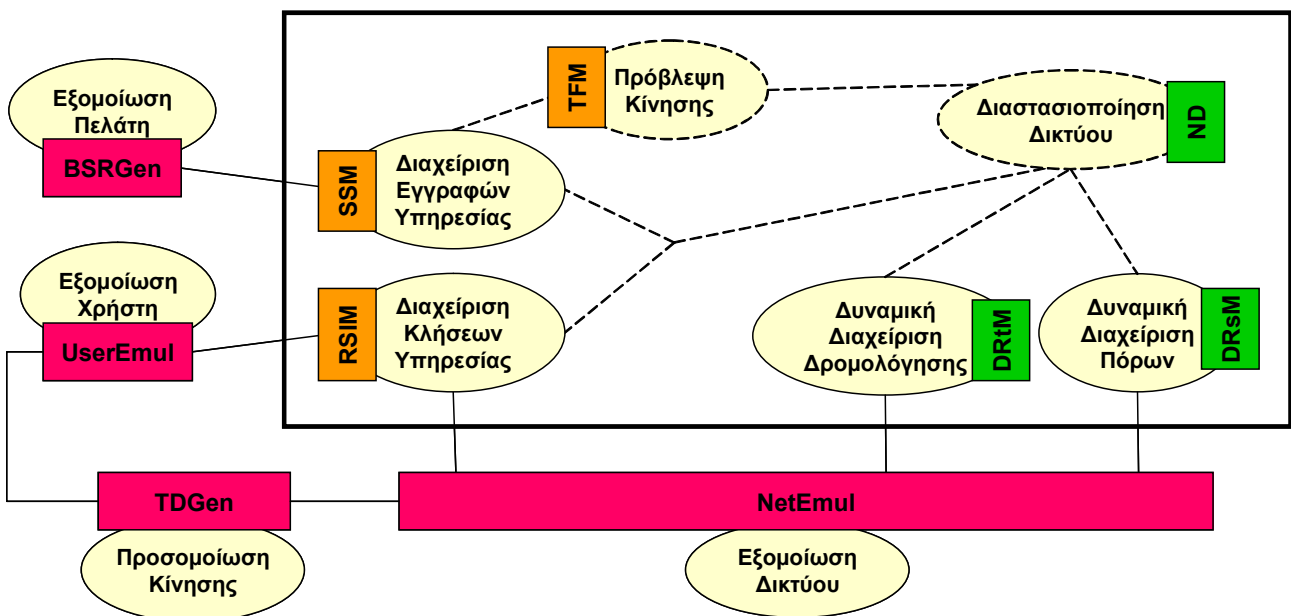
Μία σημαντική παράμετρος που κρίνει και τη δυνατότητα εφαρμογής σε μεγάλη κλίμακα είναι οι απαιτήσεις επεξεργασίας και πληροφορίας των λειτουργιών παρακολούθησης που απαιτούνται από τη δυναμική προσαρμογή. Ειδικότερα εδώ απαιτείται συλλογή και επεξεργασία των πρωτογενών δεδομένων αναφορικά με τον φόρτο τοπικά με τη χρήση παθητικών μετρήσεων. Οι μετρήσεις που απαιτούνται για την ανίχνευση της συμφόρησης είναι δυνατό να υλοποιηθούν με ενεργές μετρήσεις από άκρο σε άκρο ή, για την αποφυγή της εκθετικά αυξανόμενης πολυπλοκότητας, με παθητικές μετρήσεις ανά κόμβο (hop-by-hop), συλλογή και συσχέτιση (correlation) των αποτελεσμάτων για την σύνταξη των κατάλληλων αναφορών από άκρο σε άκρο. Οι λεπτομέρειες του βέλτιστου σχεδιασμού των λειτουργιών παρακολούθησης είναι έξω από τα πλαίσια της παρούσας εργασίας.

C.3 Πειράματα

C.3.1 Πειραματικό Πλαίσιο

Για τη διεξαγωγή εκτεταμένων πειραμάτων είναι απαραίτητη η δυνατότητα δοκιμής διαφορετικών φυσικών τοπολογιών και η δυνατότητα δοκιμής διαφορετικών συνθηκών κίνησης.

Καθώς ο εξοπλισμός σε δρομολογητές που αποτελούν τα δικτυακά περιβάλλοντα δοκιμής (testbeds) είναι αναγκαστικά περιορισμένος, αναπτύξαμε λειτουργίες εξομοίωσης δικτύου (network emulation) και εξομοίωσης χρηστών (user emulation) καθιστώντας δυνατό τον καθορισμό του δικτυακού περιβάλλοντος δοκιμής χωρίς τους περιορισμούς του υλικού αλλά με την ευελιξία που παρέχει το λογισμικό και ανάλογα με τις ανάγκες του εκάστοτε επιθυμητού πειράματος. Επιπλέον, για τη δυνατότητα δοκιμής διαφορετικών συνθηκών κίνησης δημιουργήσαμε κατάλληλες λειτουργίες προσομοίωσης της κίνησης (traffic demand simulation functions). Οι λειτουργίες αυτές συνθέτουν την πλατφόρμα δοκιμών (βλέπε Σχήμα C-7).



Σχήμα C-7: οντότητες πλατφόρμας δοκιμών

Διακρίνουμε τις λειτουργίες εξομοίωσης δικτύου, εξομοίωσης πελάτη, εξομοίωσης χρήστη και προσομοίωσης κίνησης.

C.3.1.1 Λειτουργίες Πλατφόρμας Δοκιμών

C.3.1.1.1 Προσομοίωση Κίνησης

Η προσφερόμενη κίνηση είναι αποτέλεσμα των χαρακτηριστικών κίνησης των ροών, οι οποίες αποκτούν πρόσβαση στους πόρους του δικτύου μετά από κλήσεις υπηρεσιών, που αντιστοιχούν σε συμβόλαια εγγραφής υπηρεσιών. Συνεπώς, η ρύθμιση της έντασης και του τύπου της προσφερόμενης κίνησης αφορά και τα τρία αυτά στάδια ζήτησης. Διακρίνουμε έτσι μεταξύ της προσομοίωσης αιτήσεων εγγραφών, προσομοίωσης αιτήσεων κλήσεων και προσομοίωσης πηγής κίνησης (traffic source).

Θεωρούμε καταρχήν διακριτούς τύπους υπηρεσιών. Ένας τύπος υπηρεσίας ορίζεται από τη μέθοδο ενεργοποίησης της υπηρεσίας και τα τεχνικά χαρακτηριστικά της υπηρεσίας όπως αποτυπώνονται στα πεδία χαρακτηριστικών κίνησης και εγγυήσεων ποιότητας των SLSs του συμβολαίου εγγραφής (βλέπε ενότητα B.1.2.1.2).

C.3.1.1.1.1 Προσομοίωση Εγγραφών

Ένα συμβόλαιο εγγραφής έχει μακροπρόθεσμη ισχύ. Θεωρούμε έτσι πως, στη διάρκεια των πειραμάτων μας, δεν συμβαίνουν αυξομειώσεις των συμβολαίων εγγραφής, αντίθετα ο πληθυσμός τους παραμένει σταθερός. Η προσομοίωση αιτήσεων εγγραφών εξαντλείται στον προσδιορισμό των αιτήσεων που θα τροφοδοτηθούν άπαξ προς διαπραγμάτευση στη διαδικασία διαχείρισης εγγραφής, βάσει των οποίων θα προκύψουν και τα συμβόλαια εγγραφής. Η διαδικασία προσομοίωσης των αιτήσεων και εξομοίωσης της λειτουργίας διαπραγμάτευσης συμβολαίου του πελάτη συγχωνεύονται σε μία οντότητα που ονομάζεται *παραγωγή πολλαπλών αιτήσεων υπηρεσίας* (Bulk Service Request Generator – BSRGen). Η οντότητα BSRGen τροφοδοτείται με το σύνολο των πελατών και των τόπων τους (customer sites), τους τύπους των υπηρεσιών, και την επιθυμητή ζήτηση που εκφράζεται από τον αριθμό των αιτήσεων που είναι επιθυμητό να παραχθούν, και τα ποσοστά αιτήσεων ανά τύπο υπηρεσιών. Με αυτά τα στοιχεία, κατανέμει τη ζήτηση στους διαφορετικούς χρήστες και παράγει τις κατάλληλες αιτήσεις εγγραφής. Οι αιτήσεις εγγραφής τροφοδοτούνται στην οντότητα SSM οπότε και κάποιες θα γίνουν αποδεκτές ενώ κάποιες άλλες ενδέχεται να απορριφθούν. Οι αποδεκτές αιτήσεις εγγραφής αποτελούν και τα συμβόλαια εγγραφής επί των οποίων γίνεται στη συνέχεια η προσομοίωση αιτήσεων κλήσεων.

C.3.1.1.1.2 Προσομοίωση Κλήσεων

Η προσομοίωση αιτήσεων κλήσης παράγει τις αιτήσεις κλήσης για τις υπηρεσίες ρητής κλήσης. Μια υπηρεσία καλείται πολλές φορές μέσα στο χρόνο και συνεπώς και κατά τη διάρκεια του πειράματος. Επιπλέον, για την ίδια υπηρεσία είναι δυνατό να γίνονται περισσότερες από μία ταυτόχρονες κλήσεις ανάλογα με τις προδιαγραφές του βαθμού διαθεσιμότητας του συμβολαίου εγγραφής. Η προσομοίωση αιτήσεων κλήσης παράγει τις αιτήσεις κλήσης κάθε εγγεγραμμένης υπηρεσίας και ανάλογα με τη μέθοδο ενεργοποίησης της υπηρεσίας για όλη τη διάρκεια του πειράματος. Έτσι, σε κάθε συμβόλαιο υπηρεσίας μόνιμης ενεργοποίησης (βλέπε ενότητα B.1.2.1.1.2) αντιστοιχεί μία κλήση υπηρεσίας που ξεκινά στο χρόνο μηδέν του πειράματος και παύει να ισχύει στο τέλος του πειράματος. Σε κάθε συμβόλαιο υπηρεσίας ενεργοποίησης ανά αίτηση αντιστοιχίζεται ένας αριθμός χρηστών έστω n_{users} . Κάθε ένας από τους χρήστες της υπηρεσίας δημιουργεί αιτήσεις κλήσης της υπηρεσίας ανά χρόνο έστω t_{inter} οι οποίες διαρκούν έστω t_{hold} οπότε και γίνεται αίτηση τερματισμού της κλήσης. Οι χρόνοι t_{inter} και t_{hold} είναι τυχαίες μεταβλητές και παίρνουν τιμές από τη γενικευμένη εκθετική κατανομή (generalised exponential) [KX89] με μέσο όρο αντίστοιχα t_{inter}^{avg} και t_{hold}^{avg} . Με τον ίδιο τρόπο δημιουργούνται και οι αιτήσεις κλήσης για κάθε συμβόλαιο υπηρεσίας ευέλικτης ενεργοποίησης με τη διαφορά πως το 100% του ρυθμού εξυπηρέτησης διανέμεται στους χρήστες n_{users} , οπότε η αίτηση κλήσης καθενός ενεργοποιεί το $1/n_{users}$ ποσοστό ρυθμού εξυπηρέτησης. Οι παράμετροι n_{users} , t_{inter}^{avg} και t_{hold}^{avg} ορίζονται ανά τύπο υπηρεσίας.

C.3.1.1.1.3 Προσομοίωση Προσφερόμενης Κίνησης

Η προσομοίωση πηγής κίνησης παράγει ανά αίτηση κλήσης τις περιόδους λειτουργίας και σιωπής της πηγής βάσει του μοντέλου ON/OFF. Σύμφωνα με το μοντέλο πηγής κίνησης ON/OFF, μια πηγή που

χαρακτηρίζεται από μέγιστο ρυθμό ροής έστω Γ_{\max} , θεωρείται πως ανά πάσα στιγμή μπορεί να βρίσκεται σε μία από τις δύο πιθανές καταστάσεις, είτε να παράγει κίνηση με Γ_{\max} (κατάσταση ON), είτε να μην παράγει καθόλου κίνηση (κατάσταση OFF), ή όπως αλλιώς αποκαλείται να διέρχεται μια περίοδο σιωπής. Η αναλογία της διάρκειας των περιόδων που η πηγή βρίσκεται σε κατάσταση ON και OFF καθορίζει και τον μέσο όρο του ρυθμού ροής της κίνησης. Ο χρόνος διάρκειας κάθε περιόδου θεωρούμε πως είναι τυχαία μεταβλητή με μέσο όρο έστω t_{ON}^{avg} και t_{OFF}^{avg} και είναι δυνατό να υπολογίζεται είτε από τη γενικευμένη εκθετική κατανομή, είτε από κατανομή Pareto. Επίσης, είναι δυνατό να χρησιμοποιείται το αποκαλούμενο μοντέλο σταθερού ρυθμού (constant rate) αντί για το ON/OFF μοντέλο, σύμφωνα με το οποίο η πηγή παράγει συνεχώς και αδιαλείπτως για όσο διαρκεί η κλήση (t_{hold}), κίνηση σταθερού ρυθμού, έστω $\Gamma_{avg} < \Gamma_{\max}$. Ο τύπος του μοντέλου constant ή ON/OFF, η κατανομή γενικευμένη εκθετική ή Pareto, και οι παράμετροι t_{ON}^{avg} και t_{OFF}^{avg} ή Γ_{avg} ορίζονται ανά τύπο υπηρεσίας, ενώ το Γ_{\max} εξάγεται από το πεδίο χαρακτηριστικών της κίνησης του συμβολαίου κάθε υπηρεσίας όπως έχει ήδη οριστεί ανά τύπο υπηρεσίας.

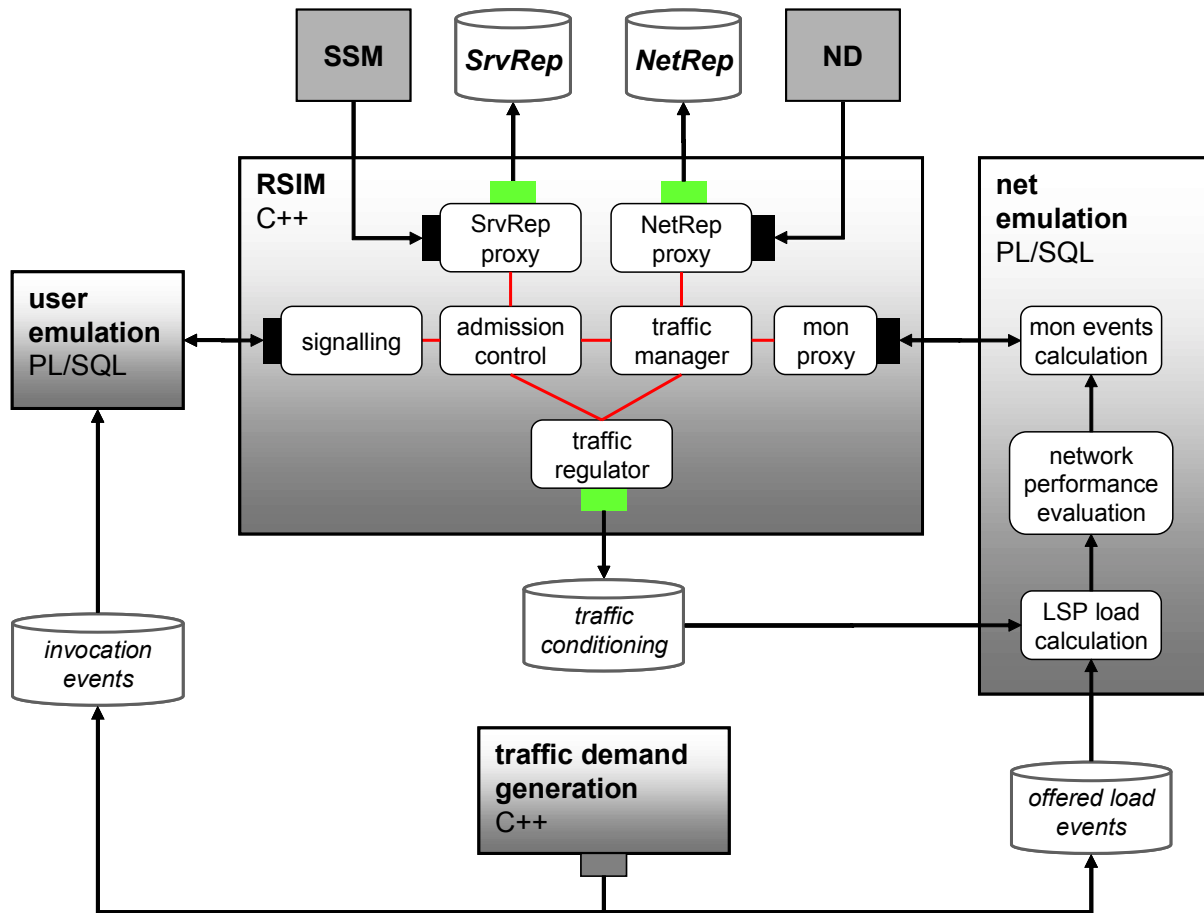
C.3.1.1.2 Εξομοίωση Δικτύου

Η λειτουργία εξομοίωσης δικτύου αρχικοποιείται με τον ορισμό της φυσικής τοπολογίας. Η φυσική τοπολογία περιγράφεται από τους ακραίους κόμβους και τους κόμβους κορμού, τις ζεύξεις μεταξύ τους και τη χωρητικότητα αυτών, τις IP διευθύνσεις των interfaces κ.λπ.

Κατά τη λειτουργία της πλατφόρμας δοκιμών, η λειτουργία εξομοίωσης τροφοδοτείται με τις ρυθμίσεις των λειτουργιών διαχείρισης του δικτύου και την προσφερόμενη κίνηση από τη λειτουργία προσομοίωσης κίνησης. Όπως και στο πρωτότυπο σύστημα (βλέπε ενότητα B.1.2.2.1, Σχήμα B-7), οι λειτουργίες διαχείρισης είναι η δυναμική διαχείριση πόρων που ρυθμίζει την κατανομή των πόρων ανά PHB σε κάθε ζεύξη, η δυναμική διαχείριση δρομολόγησης που καθορίζει τα μονοπάτια και ρυθμίζει τη δρομολόγηση των ροών σε αυτά, και η δυναμική διαχείριση κλήσεων υπηρεσίας που ρυθμίζει την πρόσβαση των ροών στους πόρους του δικτύου. Αντίθετα με το πρωτότυπο σύστημα, οι ρυθμίσεις εδώ δεν γίνονται επί των λειτουργιών επιπέδου δεδομένων Traffic Conditioning, Forwarding και Per-Hop Behaviour, αλλά επί της λειτουργίας εξομοίωσης του δικτύου. Από την προσφερόμενη κίνηση όπως προκύπτει από τη λειτουργία προσομοίωσης κίνησης, η λειτουργία εξομοίωσης δικτύου υπολογίζει την κίνηση που διοχετεύεται στο δίκτυο βάσει των ρυθμίσεων της δυναμικής διαχείρισης κλήσεων υπηρεσίας, την κατανομή της κίνησης στους κόμβους του δικτύου από τις ρυθμίσεις της δυναμικής διαχείρισης δρομολόγησης και τελικά την ποιότητα που λαμβάνει η κίνηση από την επάρκεια των πόρων σε κάθε ζεύξη βάσει των ρυθμίσεων της δυναμικής διαχείρισης πόρων και της κίνησης που διέρχεται από την κάθε ζεύξη. Από τους παραπάνω υπολογισμούς εξάγονται οι τιμές των μεγεθών που χρησιμοποιούνται στις μετρήσεις, βάσει των οποίων παράγονται τα περιοδικά ή ασύγχρονα γεγονότα, και τροφοδοτούνται πίσω στις λειτουργίες διαχείρισης σαν αποτελέσματα των λειτουργιών παρακολούθησης.

Η λειτουργία εξομοίωσης χρήστη αναλαμβάνει τη δημιουργία και αποστολή των αιτήσεων κλήσης και τερματισμού κλήσης σύμφωνα με τα αποτελέσματα της προσομοίωσης κλήσης. Για κάθε αίτηση κλήσης που ενεργοποιείται επιτυχώς από τη λειτουργία διαχείρισης κλήσης η προσφερόμενη κίνηση όπως προκύπτει από τα αποτελέσματα της προσομοίωσης κίνησης τελικά τροφοδοτείται στη λειτουργία εξομοίωσης δικτύου.

C.3.1.2 Υλοποίηση Πλατφόρμας Δοκιμών



Σχήμα C-8: υλοποίηση RSIM και πλατφόρμας δοκιμών

C.3.1.3 Στήσιμο Πειραμάτων

Το στήσιμο ενός πειράματος περιλαμβάνει τα εξής στάδια:

Στάδιο 1ο: Καθορισμός τύπων υπηρεσιών.

Στάδιο 2ο: Καθορισμός φυσικής τοπολογίας.

Στάδιο 3ο: Υπολογισμός αιτήσεων εγγραφής υπηρεσιών (BSRGen).

Στάδιο 4ο: Εκτέλεση λειτουργίας διαχείρισης κλήσης, υπολογισμός των εγγεγραμμένων υπηρεσιών. Εκτέλεση λειτουργιών αρχικοποίησης RPC, υπολογισμός του RAM για τις εγγεγραμμένες υπηρεσίες.

Στάδιο 5ο: Υπολογισμός αιτήσεων κλήσης και προσφερόμενης κίνησης (TDGen).

C.3.1.4 Διεξαγωγή Πειραμάτων

Σε κάθε διακριτή στιγμή πειράματος (simulation time unit – STU) τα βήματα διεξαγωγής του πειράματος έχουν ως εξής:

Βήμα 1ο: Οι νέες κλήσεις υπηρεσίας και οι αιτήσεις τερματισμού που έχουν προγραμματιστεί για αυτή τη STU τροφοδοτούνται από την οντότητα UserEmul στην οντότητα RSIM. Οι πίνακες traffic

conditioning ενημερώνονται κατάλληλα με νέες ρυθμίσεις για όσες νέες κλήσεις γίνονται αποδεκτές και με άρση των ρυθμίσεων για όσες κλήσεις τερματίζονται.

Βήμα 2ο: Η προσφερόμενη κίνηση ανά ενεργή κλήση εξάγεται από τις καταχωρήσεις στον πίνακα των offered load events και χρησιμοποιώντας τις ρυθμίσεις του traffic conditioning υπολογίζεται ο συνολικός φόρτος ανά μονοπάτι (load calculation).

Βήμα 3ο: Με βάση το φόρτο ανά μονοπάτι και τις ρυθμίσεις κατανομής των πόρων ανά ζεύξη υπολογίζεται η από άκρο σε άκρο ποιότητα υπηρεσίας και εξάγεται η κατάσταση διαθεσιμότητας ανά TT (network performance evaluation).

Βήμα 4ο: Διαμορφώνονται οι απαραίτητες αναφορές και ειδοποιήσεις παρακολούθησης και τροφοδοτούνται στην οντότητα RSIM. Εάν οι αναφορές παρακολούθησης πυροδοτήσουν μεταβολές στη στρατηγική αποδοχή που επηρεάζουν τις ρυθμίσεις ποιότητας και ρυθμού εξυπηρέτησης τότε ο πίνακας ρυθμίσεων traffic conditioning ενημερώνεται με τις αντίστοιχες μεταβολές.

C.3.2 Πειράματα

Στα πειράματα που διεξήχθησαν θεωρήσαμε πως δεν υπάρχει δυναμικό traffic engineering. Χρησιμοποιήθηκαν οι ίδιοι τύποι υπηρεσιών (Στάδιο 1ο), μία φυσική τοπολογία (Στάδιο 2ο) και ένα σύνολο από αιτήσεις εγγραφής (Στάδιο 3ο).

Τα πειράματα διαφοροποιούνται ως προς:

- (α) Το επίπεδο ικανοποίησης SL (βλέπε ενότητα B.2.1.3) με το οποίο ρυθμίζεται να λειτουργεί ο SSM στο 4ο Στάδιο και το οποίο επηρεάζει το πλήθος των εγγεγραμμένων υπηρεσιών που θα προκύψουν. Περιορίζουμε το συνεχές φάσμα των επιτρεπτών τιμών του επιπέδου ικανοποίησης $[-1,1]$ σε πέντε διακριτές τιμές για τα πειράματά μας, την τιμή μέγιστης ικανοποίησης $SL=1$, υψηλής ενδιάμεσης ικανοποίησης $SL=0,6$, μέσης ενδιάμεσης ικανοποίησης $SL=0,3$, χαμηλής ενδιάμεσης ικανοποίησης $SL=0,1$ και ελάχιστης ικανοποίησης $SL=-1$.
- (β) Την ένταση της προσφερόμενης κίνησης όπως προκύπτει από τη ρύθμιση της διαδικασίας προσομοίωσης κίνησης στο 5ο Στάδιο.
- (γ) Το επίπεδο προφύλαξης PL (βλέπε ενότητα B.2.2.3) με το οποίο ρυθμίζεται να λειτουργεί ο RSIM κατά τη διεξαγωγή του πειράματος. Περιορίζουμε το συνεχές φάσμα των επιτρεπτών τιμών του επιπέδου προφύλαξης $[0,1]$ σε τρεις διακριτές τιμές για τα πειράματά μας, την τιμή μέγιστης προφύλαξης $PL=0$, ενδιάμεσης προφύλαξης $PL=0,5$ και ελάχιστης προφύλαξης $PL=1$.

Τα πειράματα οργανώνονται σε τρία επίπεδα, κάθε μεμονωμένη δοκιμή (test) ανήκει σε μια υπόθεση δοκιμών (test case) που με τη σειρά της ανήκει σε μια σειρά δοκιμών (test suite). Διακρίνουμε δύο βασικές σειρές δοκιμών, η πρώτη, έστω TS-R, εστιάζει αποκλειστικά στη μελέτη της συμπεριφοράς της οντότητας RSIM ή αλλιώς του ελέγχου εισόδου σε επίπεδο κλήσης και η δεύτερη, έστω TS-S στη μελέτη της επίδρασης της οντότητας SSM στη συμπεριφορά της οντότητας RSIM, δηλαδή του επίδρασης ελέγχου εισόδου σε επίπεδο εγγραφής στον έλεγχο εισόδου σε επίπεδο κλήσης.

Η κίνηση που δεν υφίσταται συμφόρηση και εξυπηρετείται με την προδιαγεγραμμένη σύμφωνα με τα χαρακτηριστικά της QoS-κλάσης ποιότητα αποκαλείται στο εξής green κίνηση σε αναλογία με την κατάσταση διαθεσιμότητας του TT. Αντίστοιχα, η κίνηση που υφίσταται συμφόρηση και κατ' επέκταση παραβίαση της ποιότητας, αποκαλείται στο εξής red κίνηση. Όταν η οντότητα RSIM αποδέχεται

περισσότερη κίνηση από αυτή που το δίκτυο είναι σε θέση να υποστηρίξει και προκαλείται συμφόρηση τότε αυτόματα η κίνηση που διέρχεται από τις ζεύξεις συνωστισμού μετατρέπεται σε red κίνηση. Αντίθετα, όταν απορρίπτεται κίνηση που το δίκτυο μπορεί να υποστηρίξει, ουσιαστικά απορρίπτεται green κίνηση. Είναι φανερό πως, ο βασικός στόχος διασφάλισης της ποιότητας με την αποφυγή της συμφόρησης και ταυτόχρονα μεγιστοποίησης της αξιοποίησης των πόρων με την μεγιστοποίηση της αποδοχής μεταφράζεται στη μεγιστοποίηση της green κίνησης.

Τα μεγέθη που χρησιμοποιούνται στα διαγράμματα και στην ανάλυση των αποτελεσμάτων είναι:

- *προσφερόμενη κίνηση* (offered traffic) : η κίνηση που παράγεται από τις ενεργές ροές που αντιστοιχούν σε επιτυχώς ενεργοποιημένες αιτήσεις κλήσης υπηρεσιών.
- *διοχετευόμενη κίνηση* (injected traffic) : εκείνο το μέρος από την προσφερόμενη κίνηση που περνά την αστυνόμευση του ρυθμού εξυπηρέτησης και διοχετεύεται τελικά στο δίκτυο.
- *κίνηση green* (green traffic) : εκείνο το μέρος από τη διοχετευόμενη κίνηση που εξυπηρετείται από TT σε κατάσταση green, δηλαδή δεν αντιμετωπίζει συμφόρηση και λαμβάνει την απαιτούμενη ποιότητα υπηρεσίας.
- *κίνηση red* (red traffic) : εκείνο το μέρος από τη διοχετευόμενη κίνηση που εξυπηρετείται από TT σε κατάσταση red, δηλαδή αντιμετωπίζει συμφόρηση και λαμβάνει υποβαθμισμένη, σε σχέση με την απαιτούμενη, ποιότητα υπηρεσίας.
- *εφικτή κίνηση* (sustainable traffic) : η κίνηση που δυνητικά μπορεί να υποστηριχθεί από το δίκτυο χωρίς να προκληθεί συμφόρηση για τη συγκεκριμένη QoS-κλάση, δεδομένης της διοχετευόμενης κίνησης από τις άλλες QoS-κλάσεις. Η εφικτή κίνηση ισοδυναμεί με το υπόλοιπο που απομένει από τους συνολικούς διαθέσιμους πόρους αφού αφαιρεθούν οι πόροι που απασχολούνται από τις άλλες QoS-κλάσεις. Είναι προφανές πως σε περιόδους συμφόρησης η εφικτή κίνηση είναι τουλάχιστον το ανά QoS-κλάση δίκαιο μερίδιο R_{nom} , το οποίο και αποκαλείται *ελάχιστη εφικτή κίνηση σε κατάσταση συμφόρησης* (minimum sustainable traffic at congestion).

Ορίζουμε ως κανονικοποιημένο κέρδος του ελέγχου εισόδου κίνησης τον εκάστοτε λόγο της green κίνησης που επιτυγχάνεται με την εφαρμογή της οντότητας RSIM προς την green κίνηση που επιτυγχάνεται κάτω από τις ίδιες συνθήκες χωρίς την εφαρμογή της οντότητας RSIM. Στο εξής, το κανονικοποιημένο κέρδος θα αναφέρεται και απλά ως κέρδος. Το κόστος από την εφαρμογή του ελέγχου εισόδου κίνησης αντιστοιχεί στο κόστος στην ικανοποίηση των υπηρεσιών και εκτιμάται για την απόρριψη κλήσεων από το ποσοστό των απορριφθέντων προς το σύνολο των αιτήσεων κλήσης και για τη μείωση του ρυθμού εξυπηρέτησης από το λόγο του πραγματικού ρυθμού εξυπηρέτησης καθ' όλη τη διάρκεια μιας ενεργής υπηρεσίας προς τον συμφωνημένο στο συμβόλαιο ρυθμό εξυπηρέτησης.

C.3.2.1 Βασικό Στήσιμο Πειραμάτων

C.3.2.1.1 Τύποι Υπηρεσιών

Θεωρούμε πως το δίκτυο υποστηρίζει τρεις QoS-κλάσεις που αντιστοιχούν και υλοποιούνται από τα EF, AF1 και AF3 OAs αντιστοίχως. Ορίζουμε τους εξής τύπους υπηρεσιών (βλέπε Σχήμα C-9):

- *πλασματικού ιδιωτικού δικτύου μόνιμης ενεργοποίησης* (permanent VPN) : αναμένεται να εξυπηρετούν κίνηση μικτή από διαφορετικές εφαρμογές και απαιτούν την υψηλότερη ποιότητα που τους αποδίδεται

από την EF QoS-κλάση. Ο ρυθμός πρόσβασης σε ένα συμβόλαιο permanent VPN αναμένεται να κυμαίνεται μεταξύ 1000 και 2000 kbps σε κάθε κατεύθυνση.

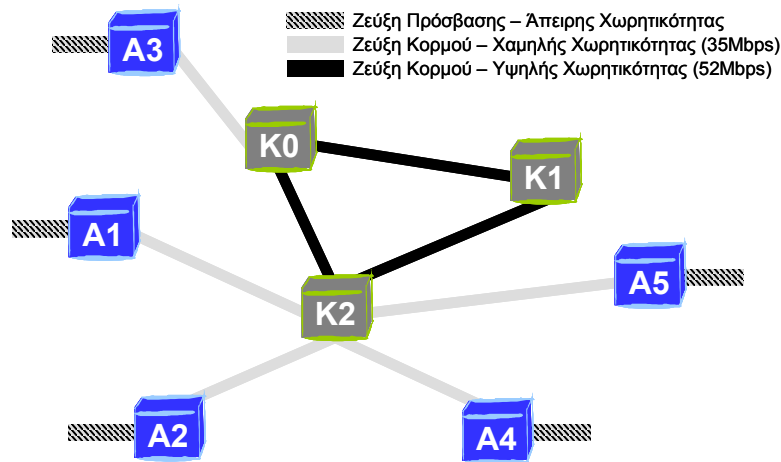
- *πλασματικού ιδιωτικού δικτύου ευέλικτης ενεργοποίησης (flexible VPN)* : αναμένεται να εξυπηρετούν κίνηση για μεταφορά δεδομένων και δεν απαιτούν υψηλή ποιότητα οπότε αντιστοιχίζονται στην AF3 QoS-κλάση. Ο ρυθμός πρόσβασης σε ένα συμβόλαιο flexible VPN συμφωνείται στα 500 kbps σε κάθε κατεύθυνση μεταξύ των διαφορετικών χρηστών που συνδέονται στο πλασματικό δίκτυο, από τα οποία κάθε κλήση ενεργοποιεί με ευέλικτη ενεργοποίηση ένα μόνο μέρος, όσο χρειάζεται κάθε φορά για τη νέα σύνοδο (session).
- *διμερής ευέλικτης ενεργοποίησης (flexible peer-to-peer)* : αναμένεται να εξυπηρετούν κίνηση μικτή από διαφορετικές εφαρμογές και απαιτούν μια ενδιάμεση ποιότητα οπότε και αντιστοιχίζονται στην AF1 QoS-κλάση. Ο ρυθμός πρόσβασης σε ένα συμβόλαιο flexible peer-to-peer αναμένεται να κυμαίνεται μεταξύ 500 και 1000 kbps σε κάθε κατεύθυνση.
- *διμερής ενεργοποίησης ανά αίτηση (on-demand peer-to-peer)* : αναμένεται να εξυπηρετούν κίνηση από εφαρμογές τύπου τηλεφωνίας μέσω διαδικτύου (Voice over IP – VoIP) οι οποίες απαιτούν μια ενδιάμεση ποιότητα οπότε και αντιστοιχίζονται στην AF1 QoS-κλάση. Ο ρυθμός πρόσβασης σε κάθε κλήση ενός συμβολαίου on-demand peer-to-peer είναι συμμετρικός στα 64 kbps, όσο χρειάζεται το κανάλι φωνής.

Τύποι Υπηρεσιών		ΟΑ		προσομοίωση εγγραφών					
				αριθμός εγγραφών	ποσοστό	ρυθμός καθόδου		ρυθμός ανόδου	
						min	max	min	max
Permanent VPN	EF	455	20%	1000	2000	1000	2000		
Flexible VPN	AF3		25%	500	500	500	500		
Flexible Peer-to-Peer	AF1		15%	500	1000	500	1000		
On-Demand Peer-to-Peer	AF1		30%	64	64	64	64		

Σχήμα C-9: τύποι υπηρεσιών και ρυθμίσεις προσομοίωσης εγγραφών

C.3.2.1.2 Φυσική Τοπολογία

Το δίκτυο που χρησιμοποιήθηκε (βλέπε Σχήμα C-10) αποτελείται από πέντε ακραίους κόμβους A1 έως A5 και τρεις κόμβους κορμού K0 έως K2. Στους ακραίους κόμβους συνδέονται οι πελάτες μέσω ζεύξεων πρόσβασης (access links) έστω άπειρης χωρητικότητας. Οι ακραίοι κόμβοι συνδέονται με τους κόμβους κορμού μέσω ζεύξεων χαμηλής χωρητικότητας (35 Mbps) ενώ οι συνδέσεις των κόμβων κορμού γίνονται μέσω ζεύξεων υψηλής χωρητικότητας (52 Mbps).



Σχήμα C-10: φυσική τοπολογία δοκιμών

C.3.2.1.3 Αιτήσεις Εγγραφής

Οι ρυθμίσεις που χρησιμοποιούνται για την προσομοίωση της παραγωγής αιτήσεων εγγραφών από την οντότητα προσομοίωσης εγγραφών (BSRGen) αφορούν στην ανά τύπο υπηρεσίας ελάχιστη και μέγιστη τιμή του ρυθμού πρόσβασης, τον συνολικό αριθμό των αιτήσεων εγγραφών που πρόκειται να παραχθούν και την ανά τύπο υπηρεσιών αναλογία. Οι τιμές που χρησιμοποιήθηκαν στο στήσιμο των πειραμάτων παρουσιάζονται στο Σχήμα C-9.

C.3.2.2 Σειρά Δοκιμών TS-R

C.3.2.2.1 Σχεδιασμός

Η συμπεριφορά της οντότητας RSIM ρυθμίζεται και συνεπώς εξαρτάται από την παράμετρο λειτουργικής πολιτικής του επιπέδου προφύλαξης PL, και πιο συγκεκριμένα από την αντιστοιχία του επιπέδου προφύλαξης με τα επίπεδα επάρκειας των πόρων.

Η επάρκεια των πόρων καθορίζεται από τη σχέση της έντασης της προσφερόμενης κίνησης με τους διαθέσιμους πόρους που παραχωρούνται για την εξυπηρέτησή της. Δεδομένου πως οι πεπερασμένοι πόροι του φυσικού εξοπλισμού κατανέμονται ανά TT βάσει της προβλεπόμενης κίνησης, η επάρκεια των πόρων εξαρτάται από τη σχέση της πραγματικής προσφερόμενης κίνησης με τις αντίστοιχες προβλέψεις, τόσο ως προς την ένταση, όσο και ως προς την κατανομή της. Συνεπώς, οι δύο παράγοντες που καθορίζουν την επάρκεια των πόρων είναι αφενός η απόκλιση των προβλέψεων από την πραγματική προσφερόμενη κίνηση και αφετέρου η ομοιογένεια της απόκλισης μεταξύ των TTs. Με άλλα λόγια, η προσφερόμενη κίνηση μπορεί να ξεπερνά τις προβλέψεις τόσο ώστε οι φυσικοί πόροι να μην επαρκούν, ή/και να κατανέμεται διαφορετικά από τις προβλέψεις και άρα η κατανομή των φυσικών πόρων ανά TT να μην αντιστοιχεί στις πραγματικές ανάγκες. Η συμπεριφορά της οντότητας RSIM στην πραγματική προσφερόμενη κίνηση καθορίζεται αφενός από το απόθεμα των διαθέσιμων πόρων του RAM που αντανακλά τις προβλέψεις ανά TT, και αφετέρου από το επίπεδο προφύλαξης.

Ορίζουμε ως *συντελεστή πληρότητας* την ένταση της προσφερόμενης κίνησης δια του μη επικαλυπτόμενου δίκαιου μεριδίου διαθέσιμων πόρων R_{nom} . Ως προσφερόμενη κίνηση θεωρείται η κίνηση που παράγεται για κάθε αίτηση κλήσης, ανεξάρτητα αν η αίτηση γίνεται αποδεκτή ή απορρίπτεται. Ο συνολικός συντελεστής πληρότητας αντανακλά την επάρκεια των πόρων, ενώ, δεδομένου συντελεστή υπέρβασης μεγαλύτερου της

μονάδας, η σχετική διαφορά μεταξύ των συντελεστών πληρότητας κάθε QoS-κλάσης αντανακλά την ομοιογένεια στην απόκλιση από τις προβλέψεις.

Ο συντελεστής πληρότητας, για δεδομένο εξοπλισμό σε φυσικούς πόρους, εξαρτάται από το πλήθος των εγγεγραμμένων υπηρεσιών, την ένταση παραγωγής αιτήσεων κλήσης ανά εγγεγραμμένη υπηρεσία και την ένταση παραγωγής κίνησης ανά κλήση. Θεωρούμε δεδομένο για όλα τα πειράματα πληθυσμό εγγεγραμμένων υπηρεσιών που προκύπτει ρυθμίζοντας τον παράγοντα (α) με ελάχιστη ικανοποίηση, δηλαδή επίπεδο ικανοποίησης $SL=-1$. Διακρίνουμε τέσσερις υποθέσεις δοκιμών, έστω TC-R-1 έως TC-R-4, ανάλογα με το συνολικό συντελεστή πληρότητας (χαμηλή για TC-R-[1,2] και υψηλή για TC-R-[3,4]) και την ομοιογένεια των συντελεστών πληρότητας ανά QoS-κλάση (ομοιογενής κατανομή για TC-R-[1,3] και ανομοιογενής για TC-R-[2,4]) (βλέπε Πίνακας C-2). Ως βάρος υπολοίπου θεωρούμε το σχετικό βάρος βάσει του οποίου μοιράζεται μεταξύ των QoS-κλάσεων όση από τη συνολική ένταση της προσφερόμενης κίνησης απομένει αφού καταναμηθεί η ελάχιστη πληρότητα.

	πληρότητα						
	συνολική C_{TOT}	ελάχιστη C_{MIN}			βάρος υπολοίπου C_{RM-W}		
		EF	AF1	AF3	EF	AF1	AF3
TC-R-1	0.6				1	1	1
TC-R-2	0.7	0.1		0.1		1	
TC-R-3	1.25	0.4				1	1
TC-R-4	1.5	0.5				2	1

Πίνακας C-2: υποθέσεις ακολουθίας δοκιμών TS-R

Στην TC-R-1 ρυθμίζουμε τη συνολική προσφερόμενη κίνηση να κυμαίνεται κατά μέσο όρο στο 0.6 των διαθέσιμων πόρων, συμμετρικά για όλες τις QoS-κλάσεις. Επιδιώκουμε έτσι να μετρήσουμε το κόστος του ελέγχου εισόδου κίνησης σε δίκτυα με υπερεπάρκεια πόρων, η οποία κατανέμεται συμμετρικά μεταξύ των QoS-κλάσεων σε σχέση με τις προβλέψεις.

Στην TC-R-2 ρυθμίζουμε τη συνολική προσφερόμενη κίνηση να κυμαίνεται κατά μέσο όρο στο 0.7 των διαθέσιμων πόρων, αλλά ασύμμετρα κατανεμημένη μεταξύ των QoS-κλάσεων. Ειδικότερα, η προσφερόμενη κίνηση για τις κλάσεις EF και AF3 έχει υπερτιμηθεί κατά τις προβλέψεις και στην πραγματικότητα ανέρχεται μόλις στο 0.1 του R_{nom} των διαθέσιμων πόρων που της έχουν παραχωρηθεί, ενώ η AF1 έχει υποτιμηθεί και στην πραγματικότητα φτάνει ακόμα και να ξεπεράσει το R_{nom} τόσο ώστε να επιτευχθεί ο μέσος όρος στο 0.7. Επιδιώκουμε έτσι να μετρήσουμε το κόστος του ελέγχου εισόδου κίνησης σε δίκτυα με υπερεπάρκεια πόρων, όπου οι προβλέψεις είναι λανθασμένες, η προσφερόμενη κίνηση ασύμμετρα κατανεμημένη με αποτέλεσμα κάποιες QoS-κλάσεις να υπερβαίνουν τους πόρους που τους έχουν παραχωρηθεί, ενώ άλλες να αφήνουν πόρους ανεκμετάλλετους. Εδώ μελετάται και η δυνατότητα της οντότητας RSIM να απορροφήσει την απόκλιση από τις προβλέψεις εφόσον υπάρχει το περιθώριο επάρκειας των φυσικών πόρων.

Στην TC-R-3 ρυθμίζουμε τη συνολική προσφερόμενη κίνηση να κυμαίνεται κατά μέσο όρο στο 1.25 των διαθέσιμων πόρων, κρατώντας την EF κάτω του προβλεπόμενου και κατανέμοντας συμμετρικά την προσφερόμενη κίνησης μεταξύ των AF1 και AF3 QoS-κλάσεων. Επιδιώκουμε έτσι να μετρήσουμε για διαφορετικές τιμές του επιπέδου προφύλαξης την απόδοση της οντότητας RSIM στη διαχείριση υψηλού

φόρτου όταν οι προβλέψεις και η συναφής κατανομή των πόρων σε μερίδια συμπίπτουν με την πραγματική αναλογία μεταξύ των διαφορετικών QoS-κλάσεων.

Στην TC-R-4 θέτουμε πολύ υψηλή προσφερόμενη κίνηση, στο 1.5 κατά μέσο όρο των διαθέσιμων πόρων, και ασύμμετρα κατανεμημένη έτσι ώστε η EF να χρησιμοποιεί το μισό των προβλεπόμενων πόρων, και η AF1 και AF3 να μοιράζονται το υπόλοιπο της υπέρβασης των προβλεπόμενων πόρων σε αναλογία 2 και 1 αντίστοιχα. Εδώ επιδιώκουμε να μετρήσουμε την απόδοση της οντότητας RSIM στη διαχείριση υψηλού φόρτου με ασύμμετρα κατανεμημένη απόκλιση από τις προβλέψεις.

	επίπεδο προφύλαξης			
	T-0	T-1	T-2	T-3
TC-R-1				0
TC-R-2	✓	0.5	0.2	0
TC-R-3	✓	1	0.5	0
TC-R-4	✓	1	0.5	0

Πίνακας C-3: πειράματα ανά υπόθεση ακολουθίας δοκιμών TS-R

Σε κάθε μία υπόθεση δοκιμών, για τις ίδιες ακριβώς συνθήκες στησίματος του πειράματος και με σταθερούς τους παράγοντες (α) και (β) διεξάγονται στη γενική περίπτωση τέσσερα ξεχωριστά πειράματα, T-0 έως T-3 (βλέπε Πίνακας C-3). Στο πείραμα T-0 δεν εφαρμόζεται η οντότητα RSIM, οπότε όλη η προσφερόμενη κίνηση γίνεται αποδεκτή και τα αποτελέσματα που λαμβάνουμε χρησιμεύουν ως βάση για τον υπολογισμό του κέρδους ελέγχου εισόδου κίνησης στα πειράματα T-1 έως T-3. Στα πειράματα T-1 έως T-3 εφαρμόζεται η οντότητα RSIM, η προσφερόμενη κίνηση ανακόπτεται βάσει του επιπέδου προφύλαξης, που σε κάθε πείραμα τίθεται σε μία από τις τρεις διακριτές τιμές του παράγοντα (γ), σε ελάχιστη προφύλαξη PL=1 για το πείραμα T-1, ενδιάμεση προφύλαξη PL=0,5 για το πείραμα T-2 και μέγιστη προφύλαξη PL=0 για το πείραμα T-3.

Εξαίρεση αποτελούν οι υποθέσεις δοκιμών με χαμηλή πληρότητα. Ειδικότερα, για την TC-R-1 εκτελείται μόνο το πείραμα T-3 καθώς είναι δεδομένο πως δεν θα υπάρχει συμφόρηση, άρα δεν θα υπάρχει red κίνηση και άρα δεν θα υπάρχει κέρδος από την οντότητα RSIM οπότε μελετάται μόνο το υψηλότερο δυνατό κόστος που επέρχεται με τη ρύθμιση του επιπέδου προφύλαξης στη πιο συντηρητική τιμή του PL=0. Για την TC-R-2, δεν χρησιμοποιείται η χαμηλή τιμή του επιπέδου προφύλαξης καθώς και πάλι σε συνδυασμό με την ιδιαίτερα χαμηλή πληρότητα έχει ως αποτέλεσμα μηδενικό κέρδος και μηδενικό κόστος. Αντ' αυτής χρησιμοποιούνται δύο ενδιάμεσες τιμές, στα PL=0.5 και PL=0.2 για τη λήψη αποτελεσμάτων με μεγαλύτερη λεπτομέρεια.

Κάθε πείραμα εκτελείται για 1500 μονάδες χρόνου προσομοίωσης από τις οποίες στην ανάλυση των αποτελεσμάτων δεν λαμβάνονται υπόψη οι πρώτες 500 οπότε και θεωρείται πως το σύστημα δεν έχει ισορροπήσει ακόμα. Επιπλέον, το ελάχιστο χρονικό διάστημα μεταβολής της στρατηγικής αποδοχής του αλγορίθμου της οντότητας RSIM t_{min} ορίζεται στις 3 μονάδες χρόνου προσομοίωσης για όλα τα πειράματα (βλέπε ενότητα B.2.2.5.1).

Αναλυτικά, οι ρυθμίσεις της προσομοίωσης παρουσιάζονται στα παρακάτω σχήματα.

		Προσομοίωση Προσφερόμενης Κίνησης									
		Κλήσεις					Κίνηση				
		παράγοντας διόγκωσης	αριθμός χρηστών	κατανομή	μέσος χρόνος διάρκειας	μέσος χρόνος μεταξύ κλήσεων	διακύμανση	κατανομή	μέσος χρόνος OFF	μέσος χρόνος ON	διακύμανση
Τύποι Υπηρεσιών	OA										
Permanent VPN	EF	0.46					Pareto	30	3	N/A	
Flexible VPN	AF3	1.95	10	Γεν. Εκθ.	100	78.2	5	Σταθερή			
Flexible Peer-to-Peer	AF1	0.65	8	Γεν. Εκθ.	1200	4015	5	Pareto	30	39	N/A
On-Demand Peer-to-Peer	AF1	0.60	11	Γεν. Εκθ.	240	600	5	Γεν. Εκθ.	15	45	8

Σχήμα C-11: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-1

		Προσομοίωση Προσφερόμενης Κίνησης									
		Κλήσεις					Κίνηση				
		παράγοντας διόγκωσης	αριθμός χρηστών	κατανομή	μέσος χρόνος διάρκειας	μέσος χρόνος μεταξύ κλήσεων	διακύμανση	κατανομή	μέσος χρόνος OFF	μέσος χρόνος ON	διακύμανση
Τύποι Υπηρεσιών	OA										
Permanent VPN	EF	0.09					Pareto	30	0.5	N/A	
Flexible VPN	AF3	0.60	10	Γεν. Εκθ.	50	183	5	Σταθερή			
Flexible Peer-to-Peer	AF1	1.40	8	Γεν. Εκθ.	1200	1543	5	Pareto	30	53	N/A
On-Demand Peer-to-Peer	AF1	1.08	19	Γεν. Εκθ.	240	600	5	Γεν. Εκθ.	15	45	8

Σχήμα C-12: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-2

		Προσομοίωση Προσφερόμενης Κίνησης									
		Κλήσεις					Κίνηση				
		παράγοντας διόγκωσης	αριθμός χρηστών	κατανομή	μέσος χρόνος διάρκειας	μέσος χρόνος μεταξύ κλήσεων	διακύμανση	κατανομή	μέσος χρόνος OFF	μέσος χρόνος ON	διακύμανση
Τύποι Υπηρεσιών	OA										
Permanent VPN	EF	0.33					Pareto	30	2.1	N/A	
Flexible VPN	AF3	3.17	10	Γεν. Εκθ.	600	173	5	Σταθερή			
Flexible Peer-to-Peer	AF1	3.20	8	Γεν. Εκθ.	1200	338	5	Pareto	30	137	N/A
On-Demand Peer-to-Peer	AF1	1.08	19	Γεν. Εκθ.	240	600	5	Γεν. Εκθ.	15	45	8

Σχήμα C-13: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-3

		Προσομοίωση Προσφερόμενης Κίνησης									
		Κλήσεις					Κίνηση				
		παράγοντας διόγκωσης	αριθμός χρηστών	κατανομή	μέσος χρόνος διάρκειας	μέσος χρόνος μεταξύ κλήσεων	διακύμανση	κατανομή	μέσος χρόνος OFF	μέσος χρόνος ON	διακύμανση
Τύποι Υπηρεσιών	OA										
Permanent VPN	EF	0.43					Pareto	30	2.84	N/A	
Flexible VPN	AF3	2.20	10	Γεν. Εκθ.	600	382	5	Σταθερή			
Flexible Peer-to-Peer	AF1	4.90	8	Γεν. Εκθ.	1200	12.2	5	Pareto	30	2970	N/A
On-Demand Peer-to-Peer	AF1	1.08	19	Γεν. Εκθ.	240	600	5	Γεν. Εκθ.	15	45	8

Σχήμα C-14: ρυθμίσεις προσομοίωσης προσφερόμενης κίνησης υπόθεσης δοκιμών TS-R-4

C.3.2.2.2 Αποτελέσματα

C.3.2.2.2.1 TC-R-1

Από τα αποτελέσματα της TC-R-1 (βλέπε Σχήμα C-15) διαπιστώνουμε πως, ακόμα και με τη μεγαλύτερη δυνατή προφύλαξη, ακόμα δηλαδή και εάν θεωρητικά επιβάλλονται προληπτικές κυρώσεις, σε ένα δίκτυο που έχει υπερεπάρκεια πόρων και η πληρότητα είναι χαμηλή, ο αλγόριθμος μένει ουσιαστικά ανενεργός και δεν επιφέρει κανένα κόστος σε απόρριψη κλήσεων ή σε μείωση ρυθμού εξυπηρέτησης.

Από τα αναλυτικά αποτελέσματα διαπιστώνουμε ακόμα πως, παρά την υπερεπάρκεια, παρατηρούνται κάποια αμελητέα ποσοστά κίνησης red. Αυτό οφείλεται στην απουσία λειτουργιών δυναμικού traffic engineering. Υπενθυμίζουμε πως τα μεγέθη που χρησιμοποιούμε μας δίνουν τις αθροιστικές τιμές για τη συνολική κίνηση ανά QoS-κλάση σε όλη την επικράτεια του δικτύου. Ωστόσο, αυτή η κίνηση απαρτίζεται

από μεμονωμένα TTs από άκρο σε άκρο τα οποία, εφόσον δεν γίνεται εξισορρόπηση του φόρτου, τυχαίνει σε κάποιο σημείο συνωστισμού να προκαλέσουν συμφόρηση ενώ στο υπόλοιπο δίκτυο η πληρότητα παραμένει χαμηλή.

C.3.2.2.2.2 TC-R-2

Στην υπόθεση TC-R-2 η προσφερόμενη κίνηση στο πείραμα T-0 (βλέπε Σχήμα C-16) χωρίς άλλη επέμβαση και έλεγχο εισόδου καλύπτεται και ικανοποιείται απόλυτα, η red κίνηση είναι μηδενική. Συνεπώς, οποιεσδήποτε προληπτικές κυρώσεις μπορεί να προκύψουν με την εφαρμογή του αλγορίθμου δεν μπορεί παρά να αποφέρουν μείωση του κέρδους.

Παρατηρούμε πράγματι πως για την ελάχιστη προφύλαξη στο πείραμα T-1 (βλέπε Σχήμα C-17) δεν εφαρμόζεται καμία κύρωση προληπτικά και έτσι τα αποτελέσματα ταυτίζονται με αυτά του πειράματος T-0, το κέρδος όπως και το κόστος μηδενικό. Ωστόσο, στην ενδιάμεση και ακόμα περισσότερο στη μέγιστη προφύλαξη των πειραμάτων T-1 και T-2 αντίστοιχα (βλέπε Σχήμα C-18 και Σχήμα C-19), η υπερβολική σε σχέση με την πραγματική πιθανότητα συμφόρησης τιμή του επιπέδου προφύλαξης έχει ως αποτέλεσμα, πέρα από το κόστος σε ικανοποίηση υπηρεσιών, αρνητικό κέρδος σε αξιοποίηση πόρων και συνεπώς κακή απόδοση του συστήματος. Όπως και σε κάθε άλλη περίπτωση έτσι και στην υπόθεση TC-R-2 όσο αυξάνει το επίπεδο προφύλαξης τόσο αυξάνεται το κόστος. Εδώ επιπλέον, όσο αυξάνεται το επίπεδο προφύλαξης τόσο περισσότερο μειώνεται το κέρδος (βλέπε Σχήμα C-20).

Παρατηρούμε επιπλέον πως, εφόσον οι πόροι επαρκούν, η ανομοιογένεια στην κατανομή της πληρότητας αφομοιώνεται επιτυχώς από το δίκτυο χάρη στην υπόθεση μηχανισμού χρονοπρογραμματισμού μη σταθερού έργου. Έτσι, όταν η προφύλαξη είναι χαμηλή εφόσον το δίκτυο αφομοιώνει την υπέρβαση του δίκαιου μεριδίου ο αλγόριθμος δεν λαμβάνει προληπτικές κυρώσεις και δεν επιφέρει κανένα κόστος ικανοποίησης. Εάν ωστόσο η προφύλαξη είναι αυξημένη, τότε για εκείνες τις QoS-κλάσεις που υπερβαίνουν το δίκαιο μεριδίό τους λαμβάνονται προληπτικές κυρώσεις οι οποίες ωστόσο δεν δικαιολογούνται καθώς η εφικτή κίνηση είναι πολύ μεγαλύτερη.

C.3.2.2.2.3 TC-R-3

Στην υπόθεση TC-R-3 η πληρότητα υπερβαίνει τους διαθέσιμους πόρους με αποτέλεσμα η προσφερόμενη και εξ ολοκλήρου διοχετευόμενη στο δίκτυο κίνηση του πειράματος T-0 (βλέπε Σχήμα C-21) να μετατρέπεται πολύ γρήγορα σε red κίνηση. Συνεπώς, αναμένουμε εδώ η προληπτική ή αντιδραστική επέμβαση από τον αλγόριθμο να έχει θετικές συνέπειες και να επιφέρει κάποιο κέρδος.

Πράγματι διαπιστώνουμε πως ήδη για την ελάχιστη προφύλαξη στο πείραμα T-1 (Σχήμα C-22) το κέρδος εκτινάσσεται περίπου σε πέντε φορές υψηλότερο από αυτό του πειράματος T-0 (βλέπε Σχήμα C-25). Είναι φανερό πως, ακόμα και με την ελάχιστη προφύλαξη, εφόσον η πληρότητα είναι ιδιαίτερα υψηλή, το σύστημα γρήγορα θα οδηγηθεί σε κατάσταση αντίδρασης επιβάλλοντας γρήγορα αρκετά συγκρατημένες στρατηγικές αποδοχής.

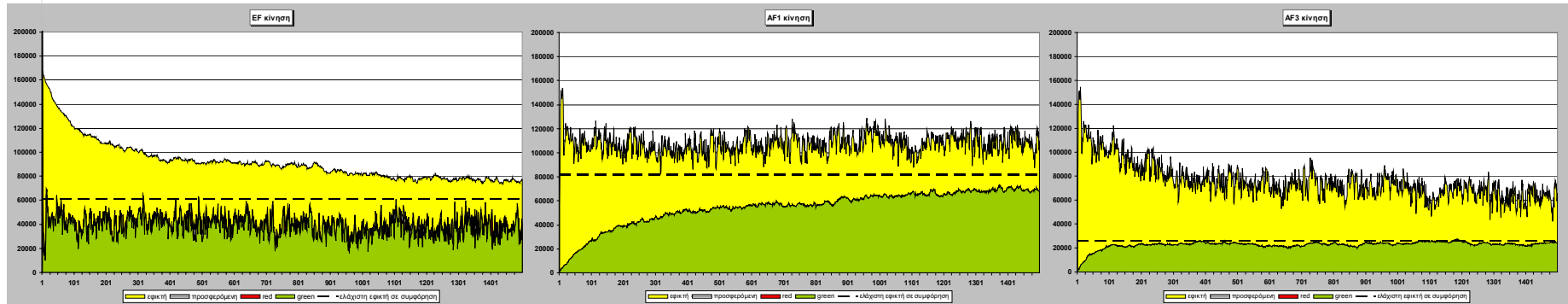
Καθώς αυξάνεται το επίπεδο προφύλαξης (βλέπε Σχήμα C-23 και Σχήμα C-24) παρατηρούμε πως το κέρδος, αφού φτάσει κάποιο μέγιστο σημείο, κάνει καμπή και αρχίζει να μειώνεται όσο αυξάνει η προφύλαξη. Υπάρχει δηλαδή μια τιμή του επιπέδου προφύλαξης η οποία αποφέρει και το μέγιστο κέρδος στην αξιοποίηση των πόρων, μεγαλύτερη προφύλαξη έχει ως συνέπεια την απόρριψη προσφερόμενης κίνησης η οποία είναι εφικτό να εξυπηρετηθεί από τους διαθέσιμους πόρους με την απαραίτητη ποιότητα. Η απόρριψη με άλλα λόγια εφικτής green κίνησης έχει ως αποτέλεσμα μείωση του κέρδους σε σχέση με την εν δυνάμει μέγιστη τιμή του.

Παράλληλα παρατηρούμε πως, όσο μεγαλώνει το επίπεδο προφύλαξης, τόσο το κόστος ικανοποίησης μετατοπίζεται, η μείωση του ρυθμού εξυπηρέτησης γίνεται ηπιότερη και πιο έντονη η απόρριψη κλήσεων. Αυτός είναι ακριβώς ο τρόπος που λειτουργεί ο έλεγχος εισόδου κίνησης και τον οποίο ρυθμίζει η λειτουργική πολιτική του παροχέα.

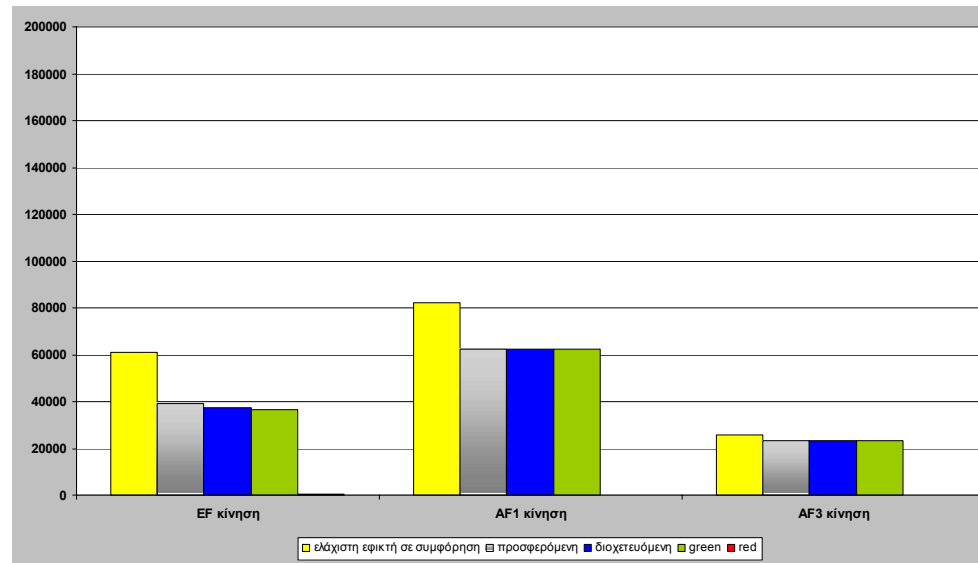
Επισημαίνουμε πως, η αύξηση του κέρδους που αντιστοιχεί σε αύξηση της αξιοποίησης των πόρων δεν είναι πάντα επιθυμητή αλλά κρίνεται υπό το πρίσμα των δύο εκφάνσεων του κόστους. Είναι έτσι δυνατό, ο παροχέας να επιλέξει μεγαλύτερη προφύλαξη από αυτή που μεγιστοποιεί το κέρδος εάν είναι επιθυμητό να συγκρατηθεί σε κάποια επίπεδα η μείωση του ρυθμού εξυπηρέτησης. Το κόστος σε απόρριψη των κλήσεων που συνοδεύει τη συγκράτηση του κόστους σε μείωση του ρυθμού εξυπηρέτησης εξαρτάται τελικά από το πλήθος των εξουσιοδοτημένων κλήσεων, που με τη σειρά του εξαρτάται από το πλήθος των εγγεγραμμένων υπηρεσιών, το οποίο ρυθμίζεται βάσει του επιπέδου ικανοποίησης με το οποίο λειτουργεί ο έλεγχος εισόδου σε επίπεδο εγγραφής. Διαπιστώνουμε με άλλα λόγια πως ο βαθμός ικανοποίησης των υπηρεσιών ρυθμίζεται καταρχήν σε επίπεδο εγγραφής, ενώ σε επίπεδο κλήσης ρυθμίζεται το είδος του κόστους ικανοποίησης, ενώ τυχόν πρόκληση σφοδρότερου κόστους μπορεί να προκληθεί εάν το επίπεδο προφύλαξης έχει τεθεί σε τιμή που δεν αντιστοιχεί στην επάρκεια των πόρων.

C.3.2.2.2.4 TC-R-4

Τα αποτελέσματα που λαμβάνουμε στην υπόθεση TC-R-4 (βλέπε Σχήμα C-26 έως Σχήμα C-30), πέρα από τις αρχές που διαπιστώσαμε ήδη από την υπόθεση TC-R-3, δείχνουν ακόμα πως, όπως είναι αναμενόμενο, η μεγαλύτερη πληρότητα επιφέρει και υψηλότερο κόστος ικανοποίησης και μικρότερα περιθώρια στην αύξηση του κέρδους. Επιπλέον, η ανομοιογένεια στην κατανομή δεν επηρεάζει τη συμπεριφορά του αλγορίθμου καθώς η υψηλή πληρότητα καθιστά τη λειτουργία του κατά κόρο αντιδραστική παρά προληπτική. Η ασυμφωνία της πραγματικής προσφερόμενης κίνησης με την προβλεπόμενη επηρεάζει την προληπτική λειτουργία του αλγορίθμου οπότε εδώ δεν εκδηλώνεται.

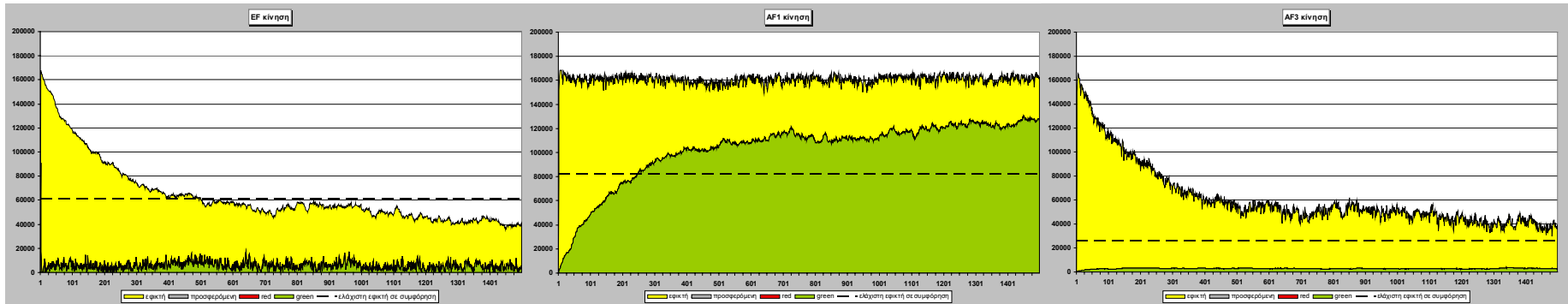


(α) φόρτος ανά QoS-κλάση στο χρόνο

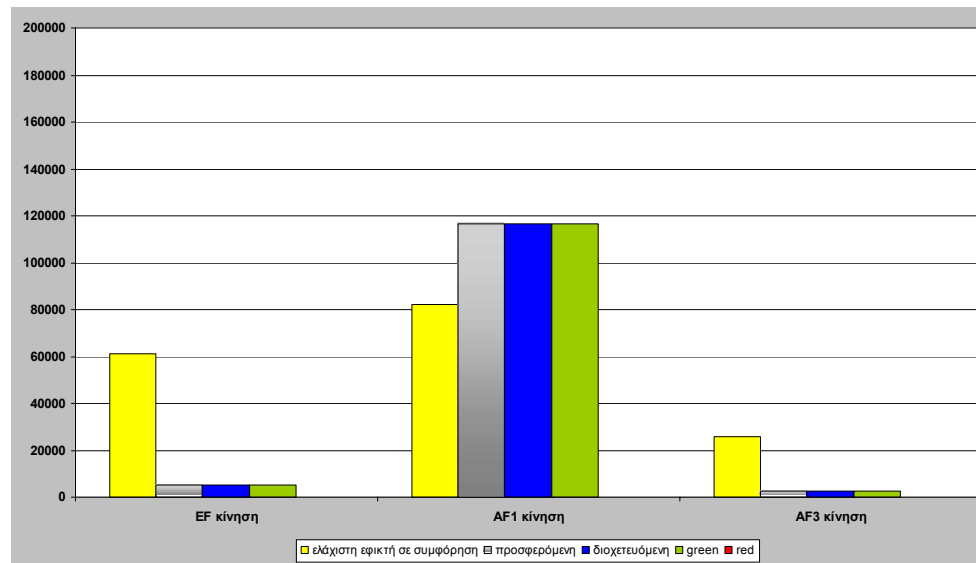


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-15: αποτελέσματα πειράματος [TC-R-1].[T-3]

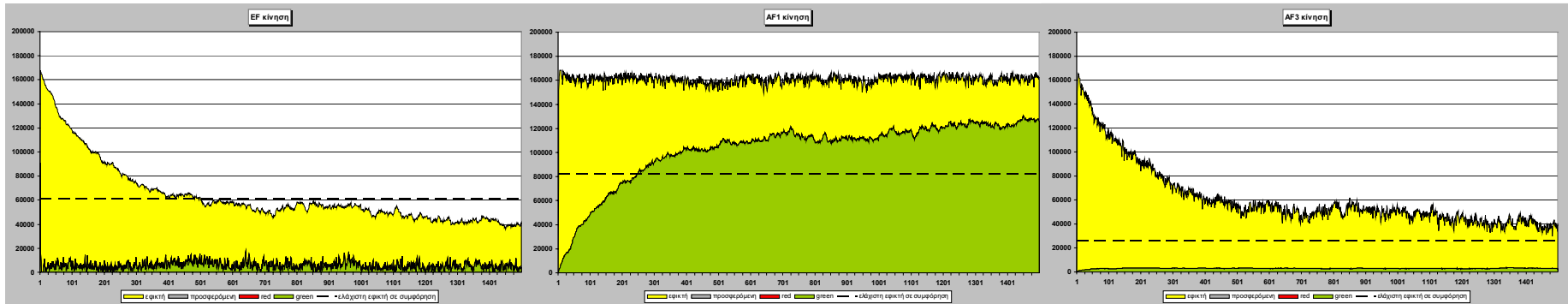


(α) φόρτος ανά QoS-κλάση στο χρόνο

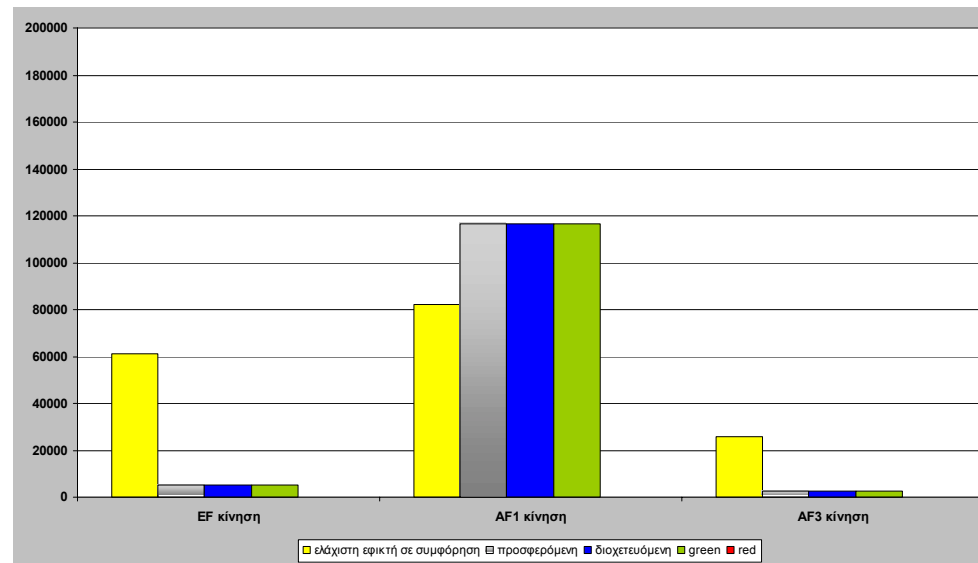


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-16: αποτελέσματα πειράματος [TC-R-2].[T-0]

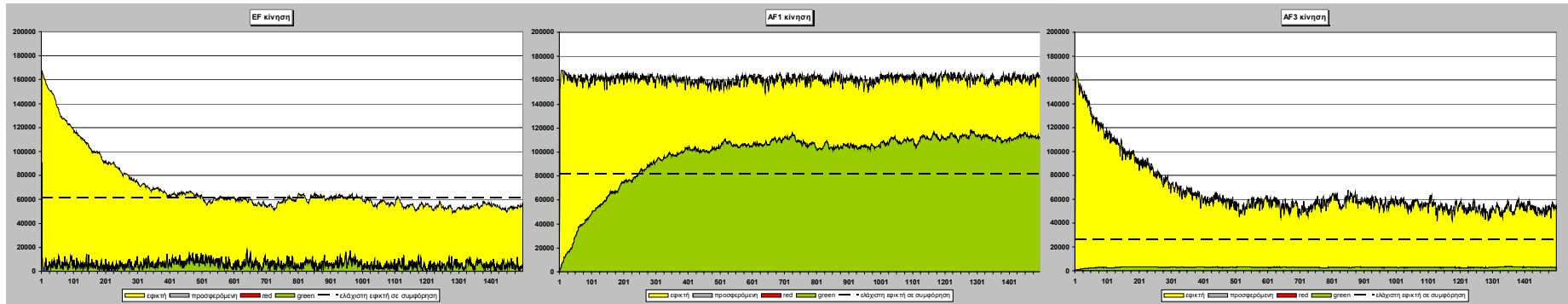


(α) φόρτος ανά QoS-κλάση στο χρόνο

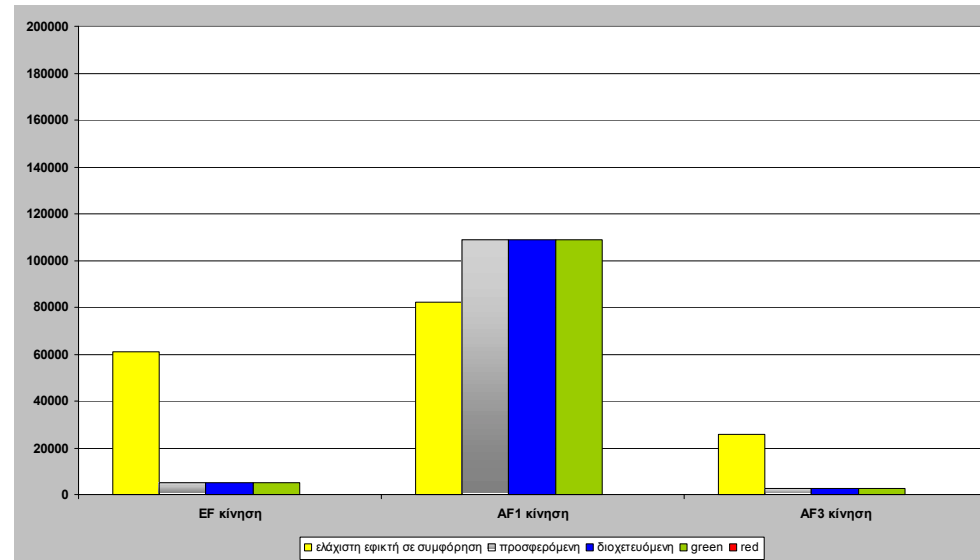


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-17: αποτελέσματα πειράματος [TC-R-2].[T-1]

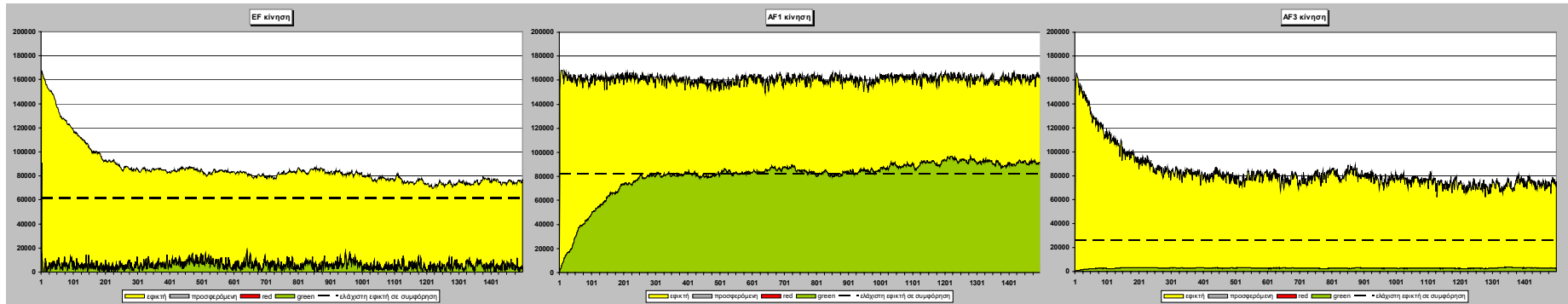


(α) φόρτος ανά QoS-κλάση στο χρόνο

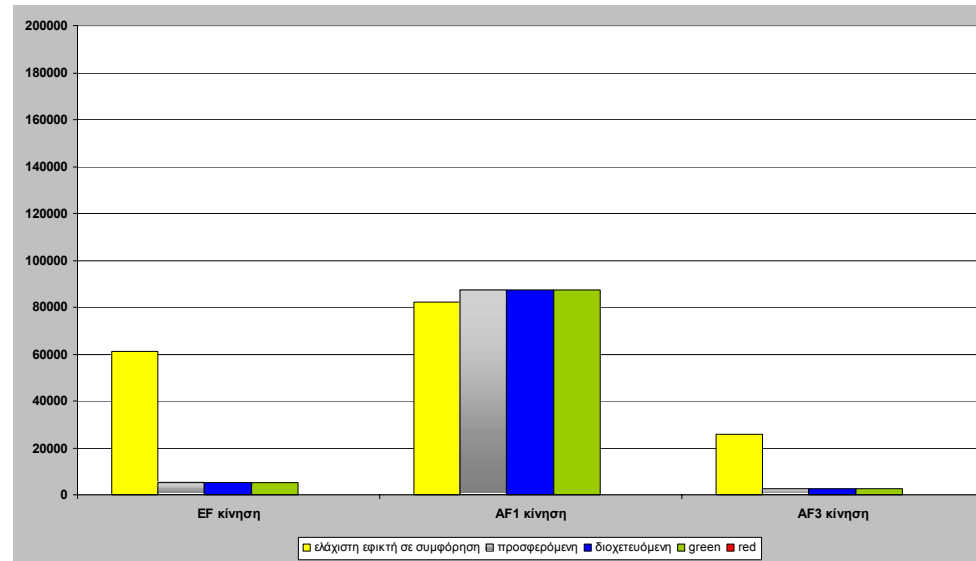


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-18: αποτελέσματα πειράματος [TC-R-2].[T-2]

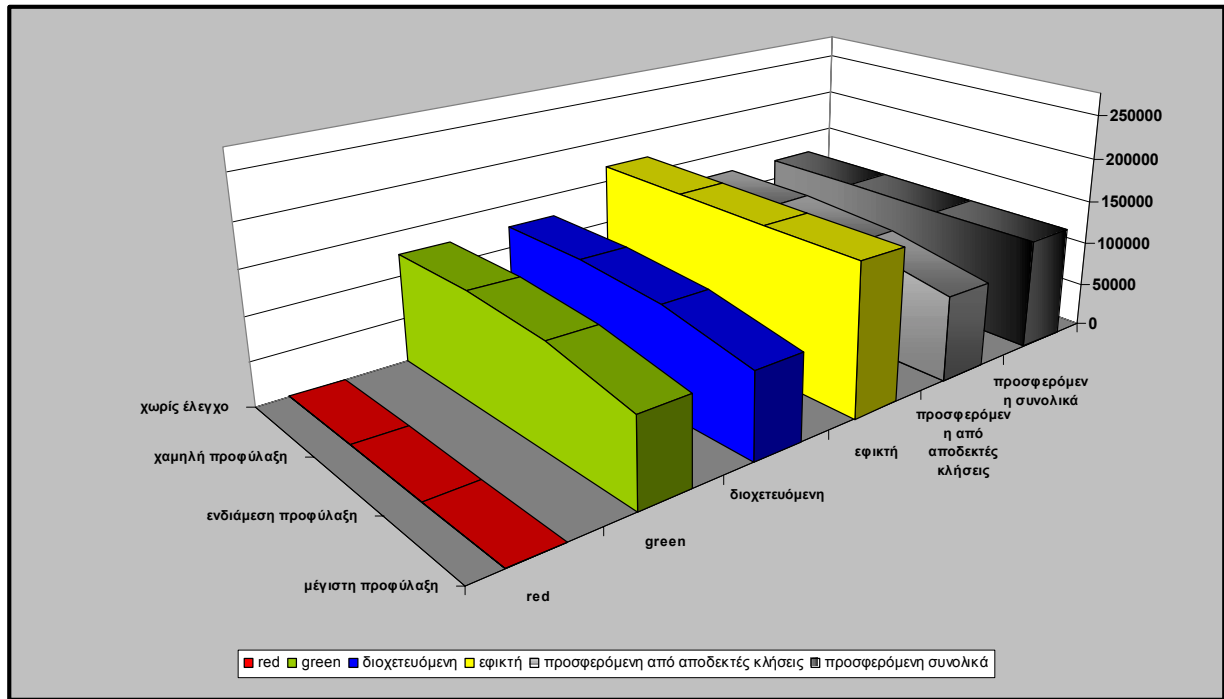


(α) φόρτος ανά QoS-κλάση στο χρόνο

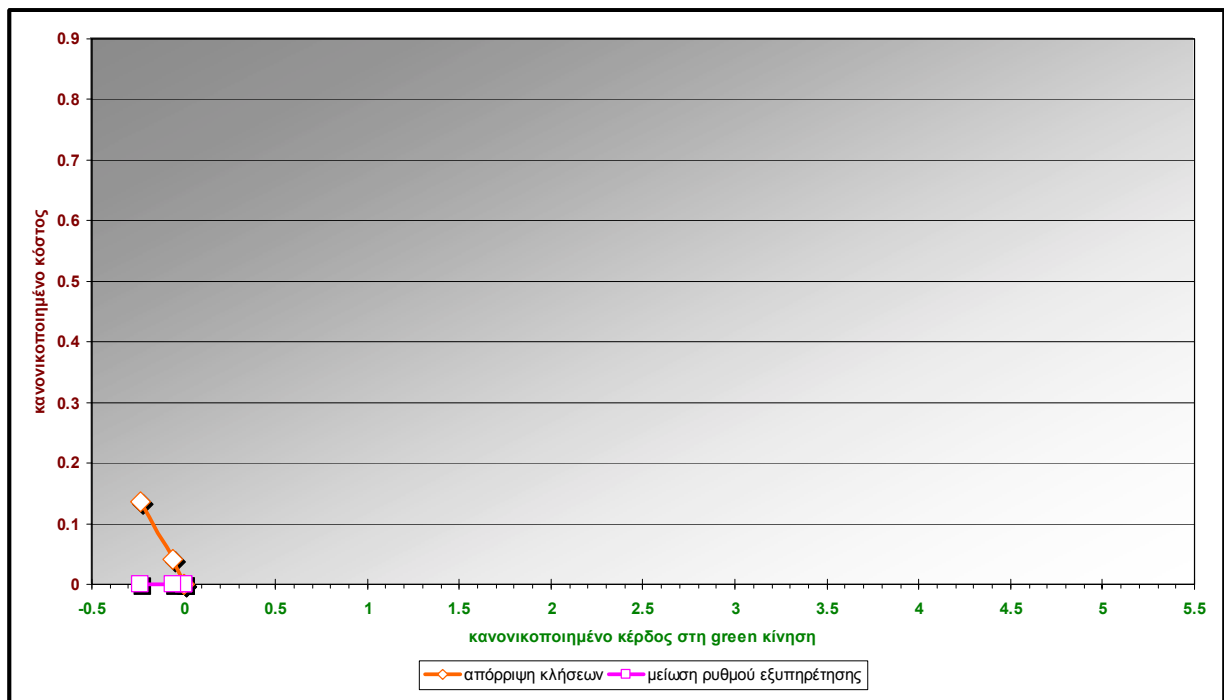


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-19: αποτελέσματα πειράματος [TC-R-2].[T-3]

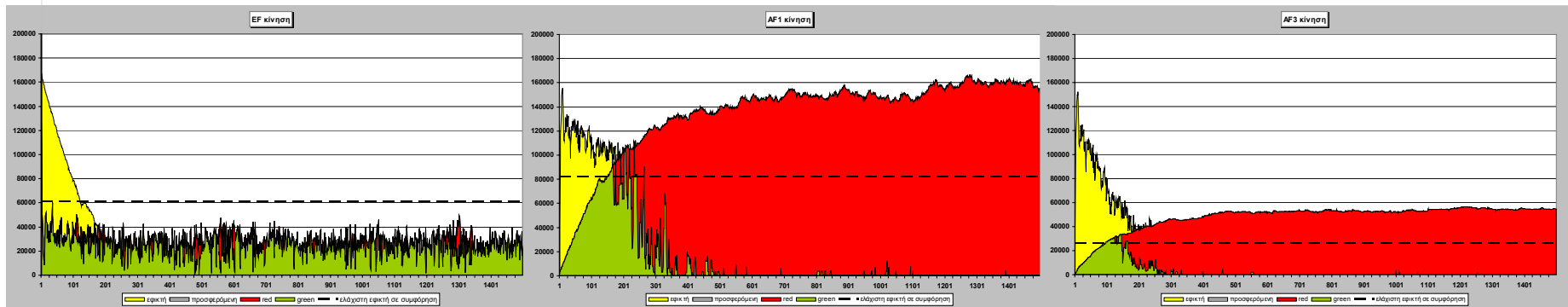


(α) συνολικός φόρτος κατά μέσο όρο ανά πείραμα (μετατόπιση 500 μονάδων χρόνου)

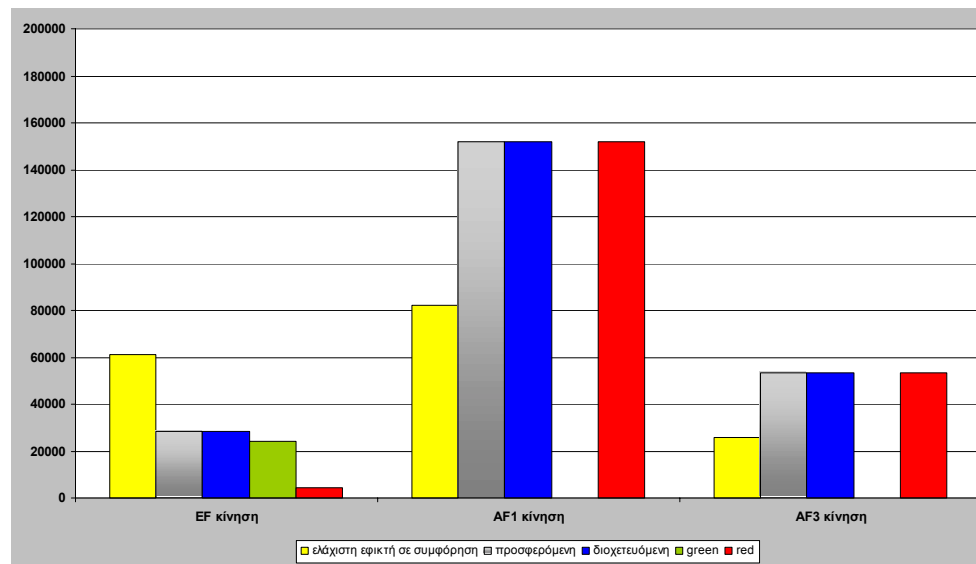


(β) κανονικοποιημένο κέρδος και κόστος ανά πείραμα

Σχήμα C-20: συγκριτικά αποτελέσματα πειραμάτων υπόθεσης δοκιμών TC-R-2

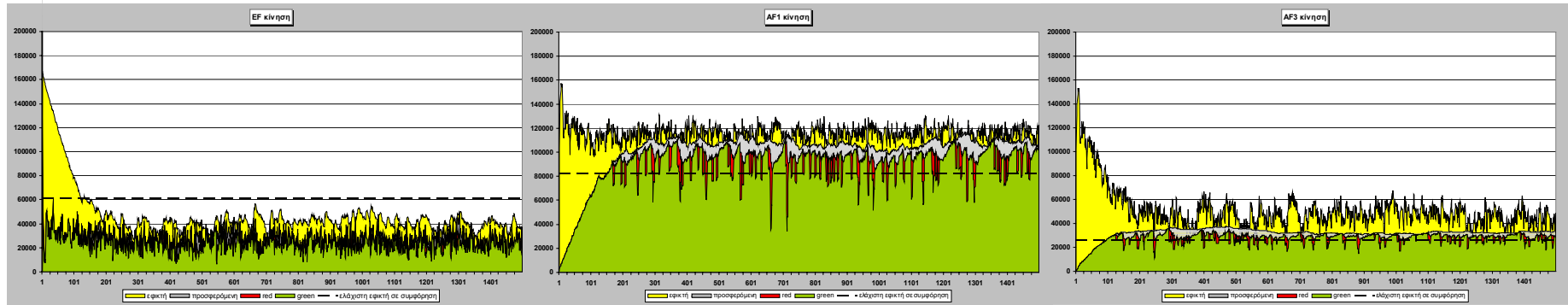


(α) φόρτος ανά QoS-κλάση στο χρόνο

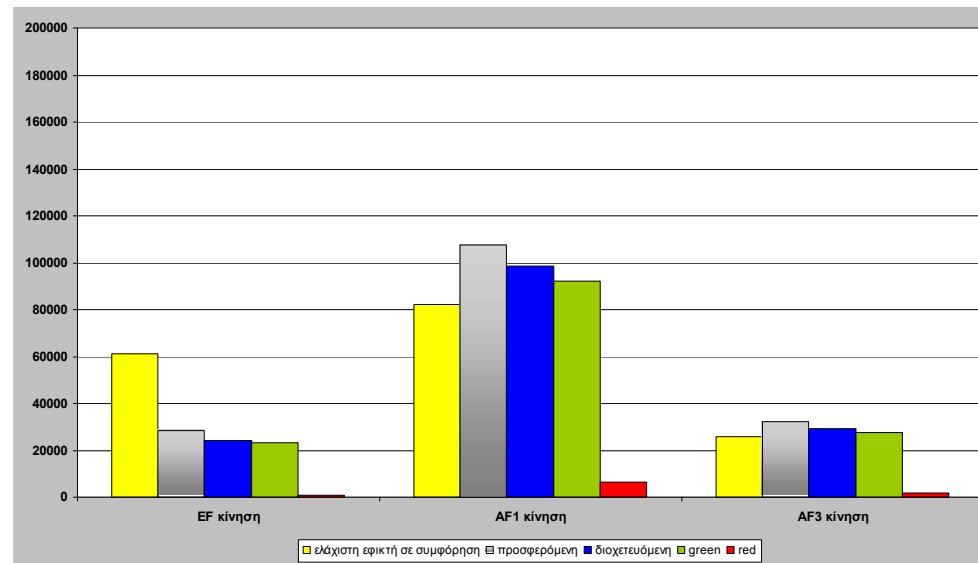


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-21: αποτελέσματα πειράματος [TC-R-3].[T-0]

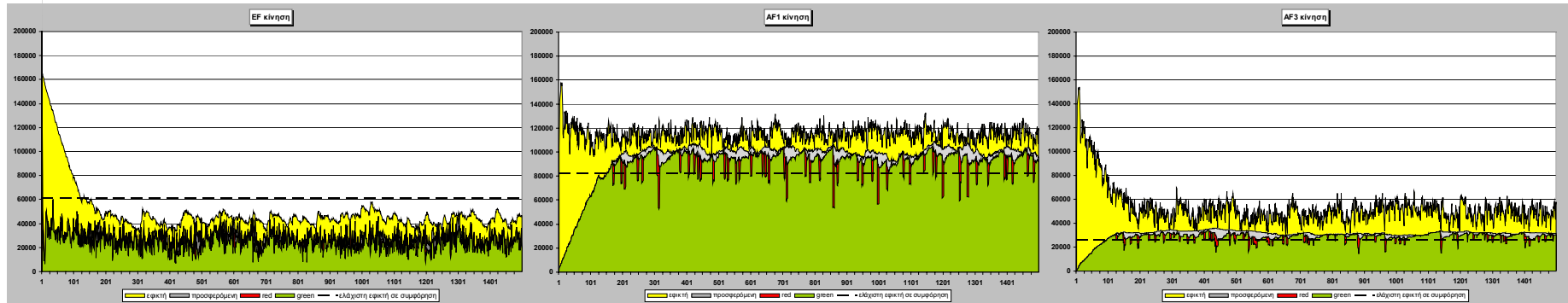


(α) φόρτος ανά QoS-κλάση στο χρόνο

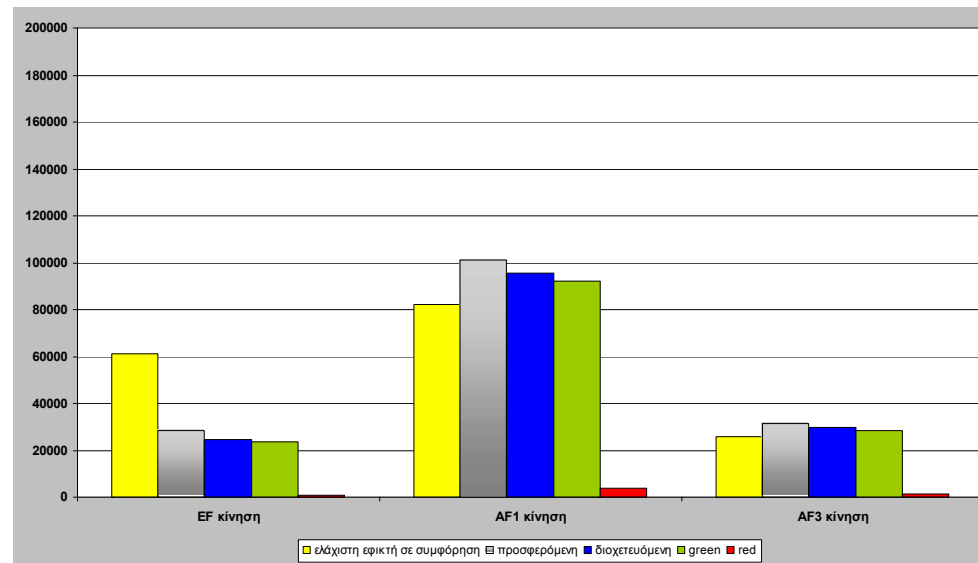


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-22: αποτελέσματα πειράματος [TC-R-3].[T-1]

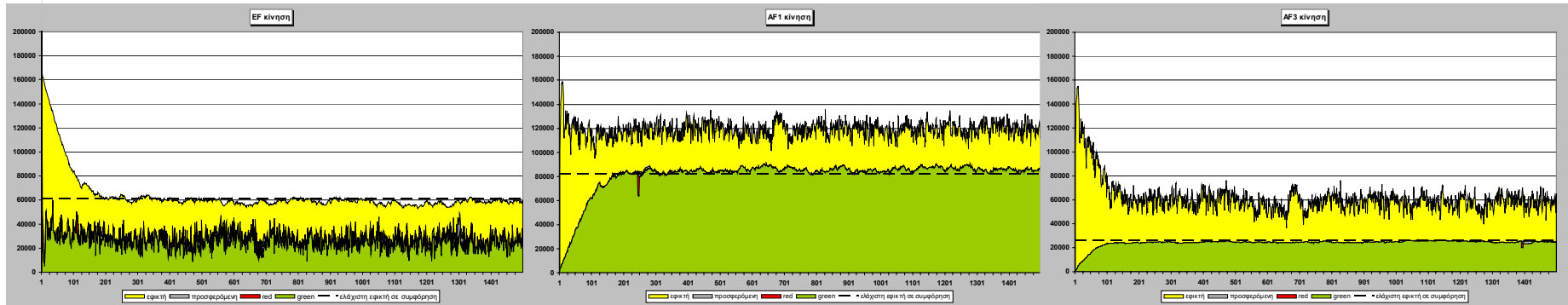


(α) φόρτος ανά QoS-κλάση στο χρόνο

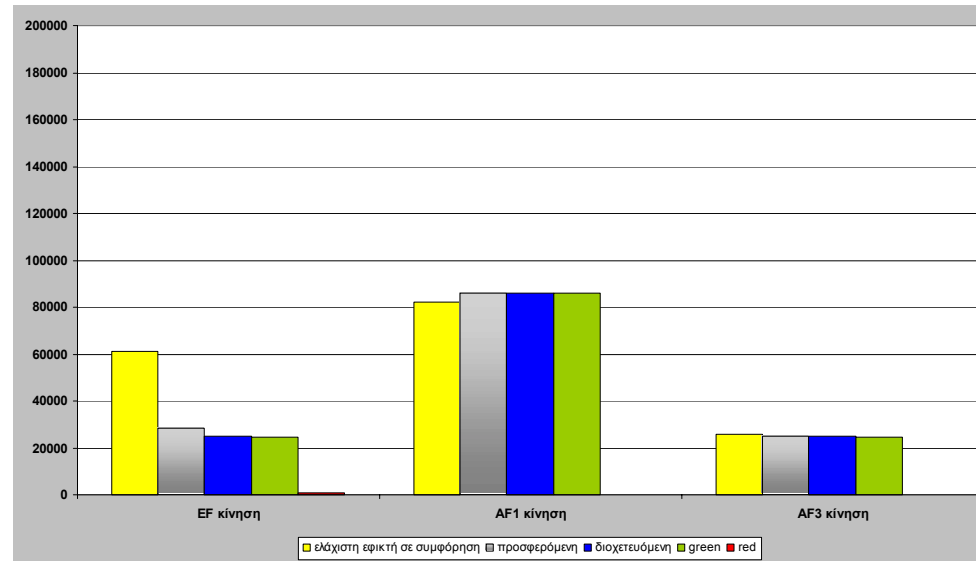


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-23: αποτελέσματα πειράματος [TC-R-3].[T-2]

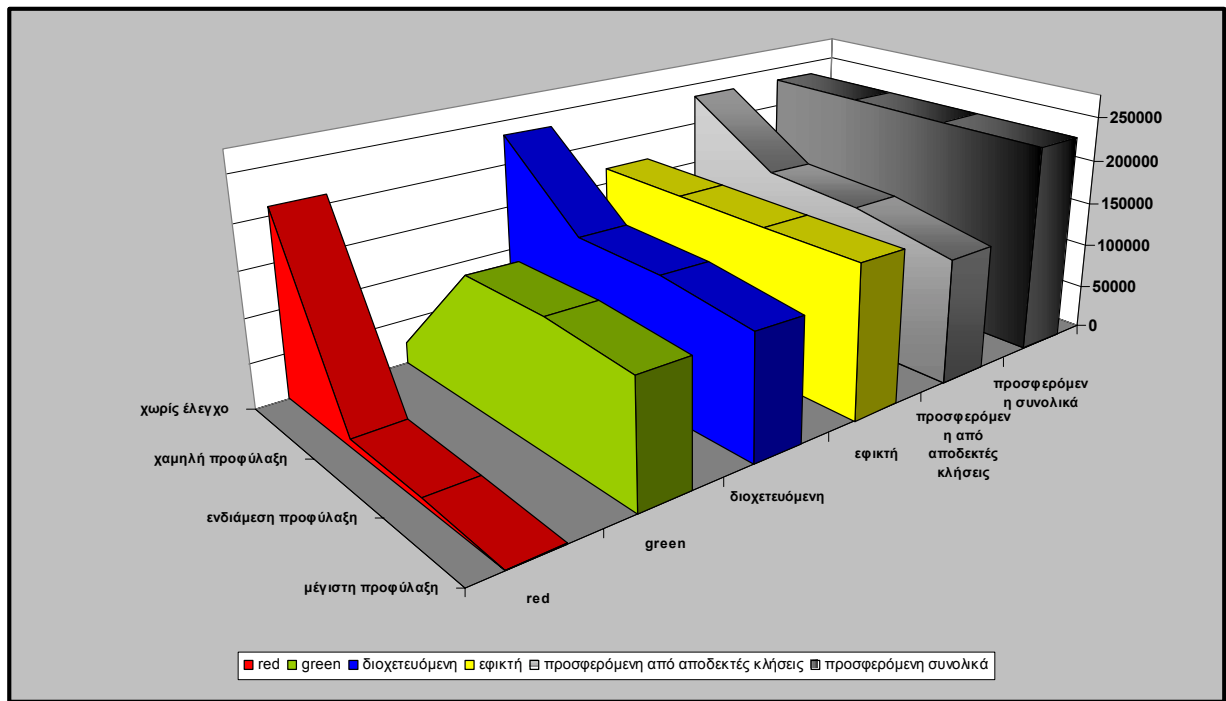


(α) φόρτος ανά QoS-κλάση στο χρόνο

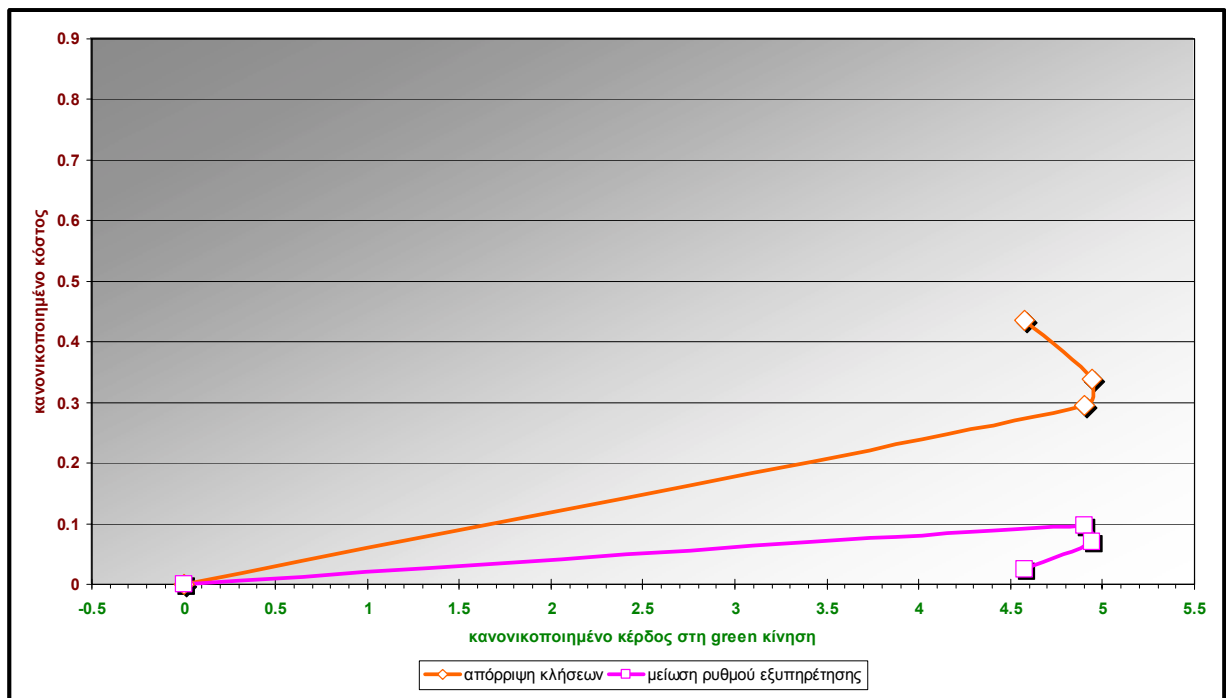


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-24: αποτελέσματα πειράματος [TC-R-3].[T-3]

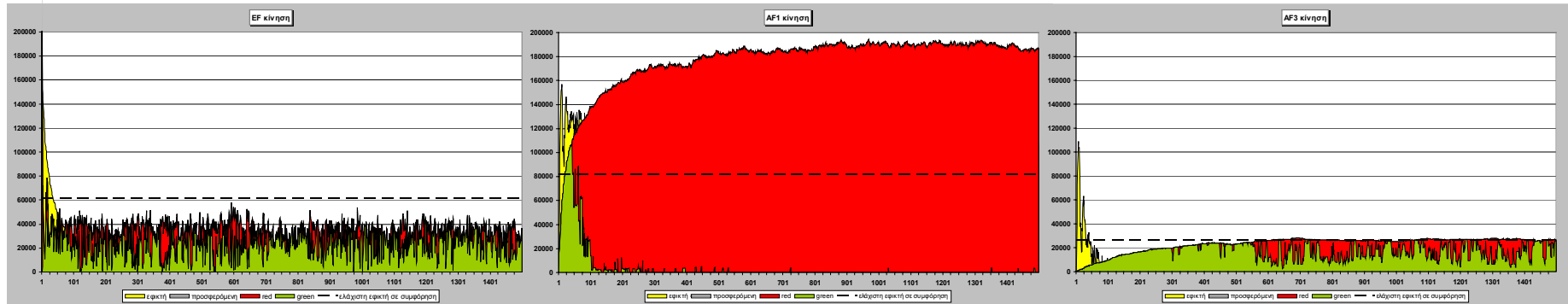


(α) συνολικός φόρτος κατά μέσο όρο ανά πείραμα (μετατόπιση 500 μονάδων χρόνου)

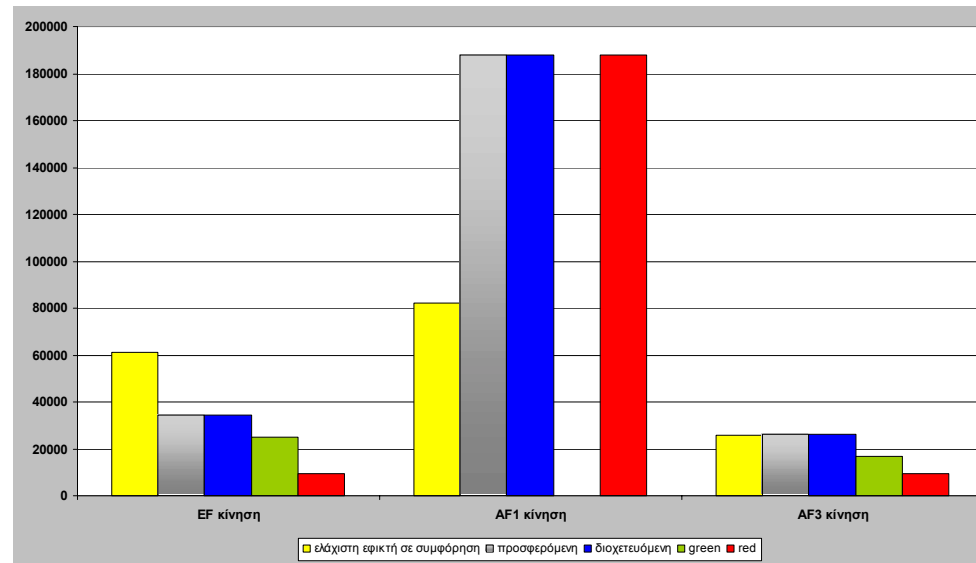


(β) κανονικοποιημένο κέρδος και κόστος ανά πείραμα

Σχήμα C-25: συγκριτικά αποτελέσματα πειραμάτων υπόθεσης δοκιμών TC-R-3

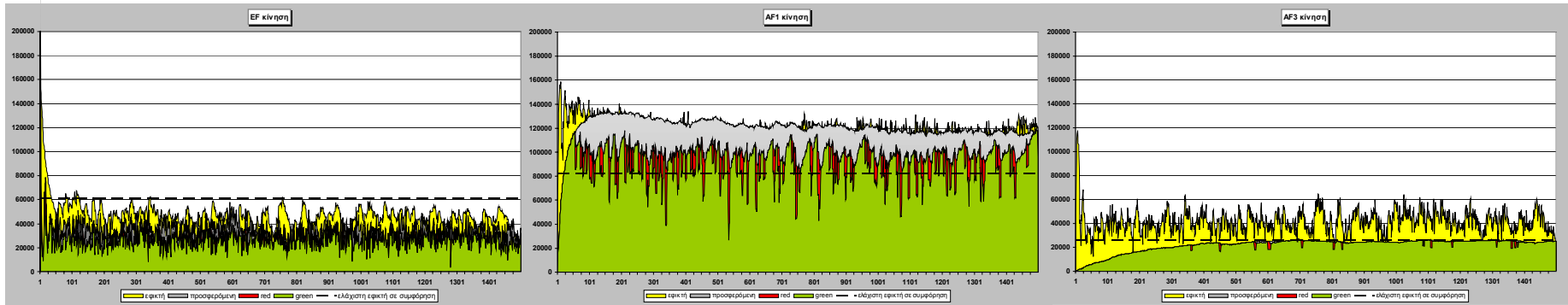


(α) φόρτος ανά QoS-κλάση στο χρόνο

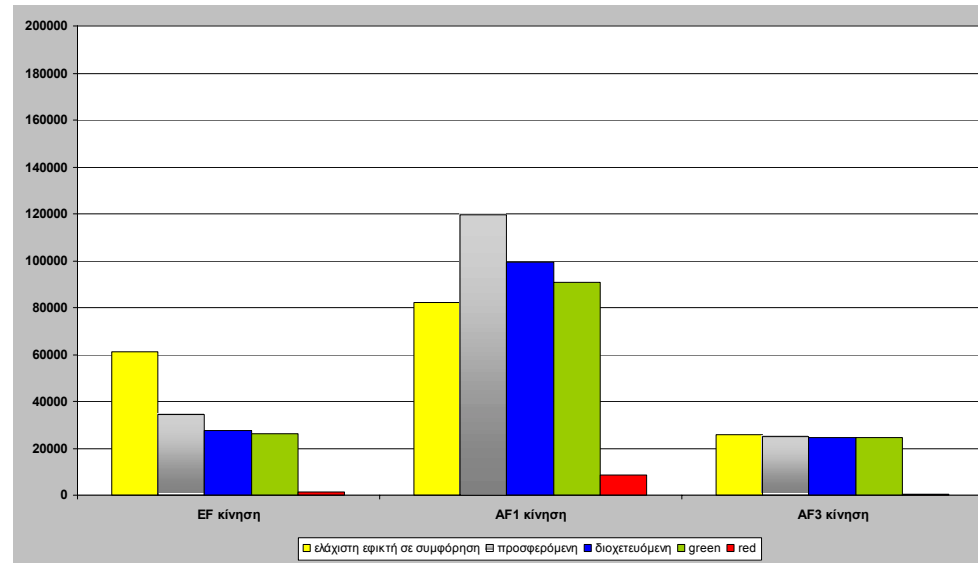


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-26: αποτελέσματα πειράματος [TC-R-4].[T-0]

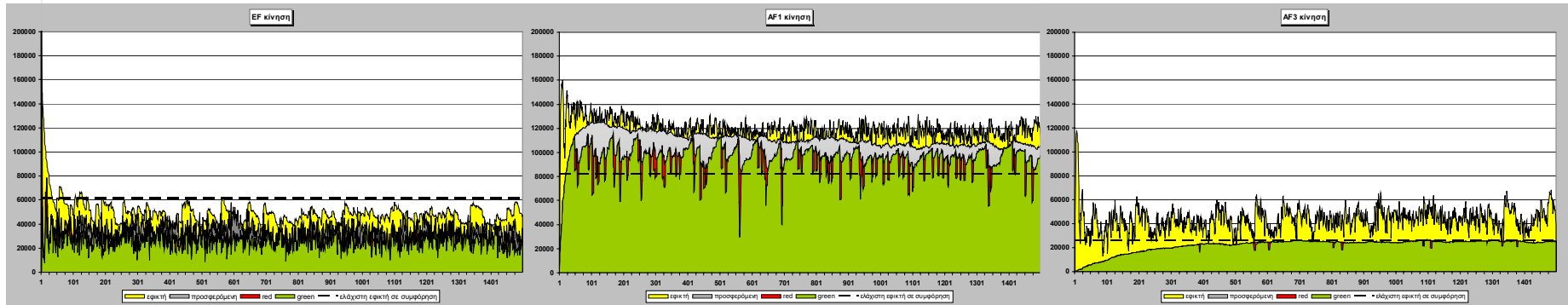


(α) φόρτος ανά QoS-κλάση στο χρόνο

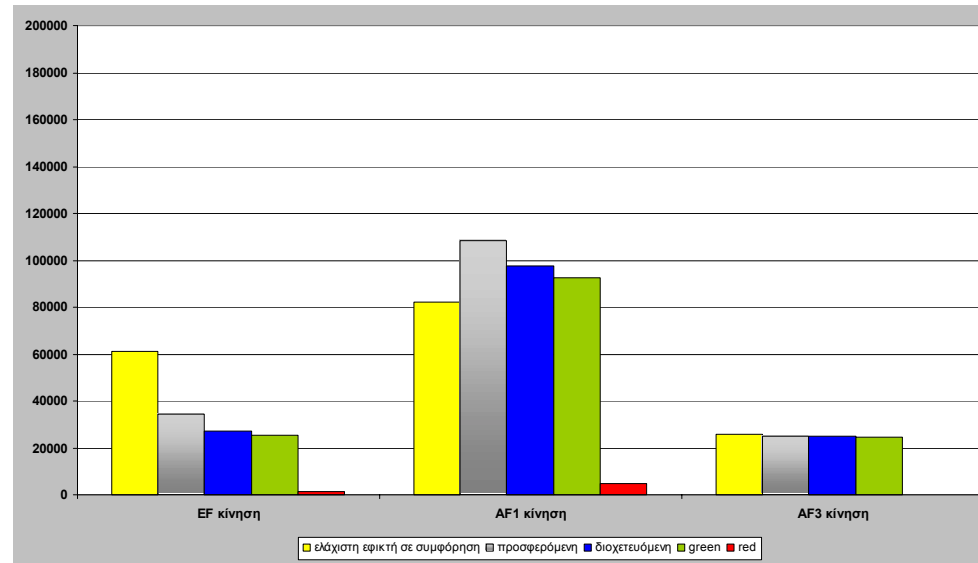


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-27: αποτελέσματα πειράματος [TC-R-4].[T-1]

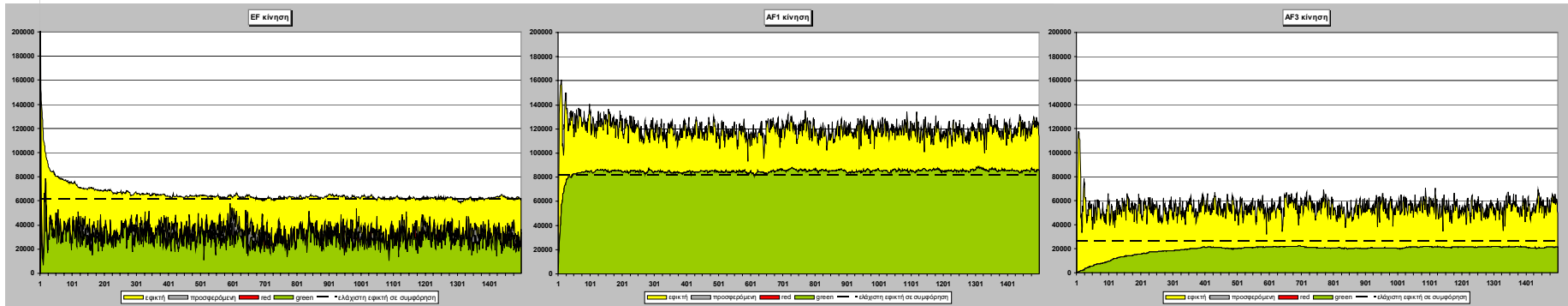


(α) φόρτος ανά QoS-κλάση στο χρόνο

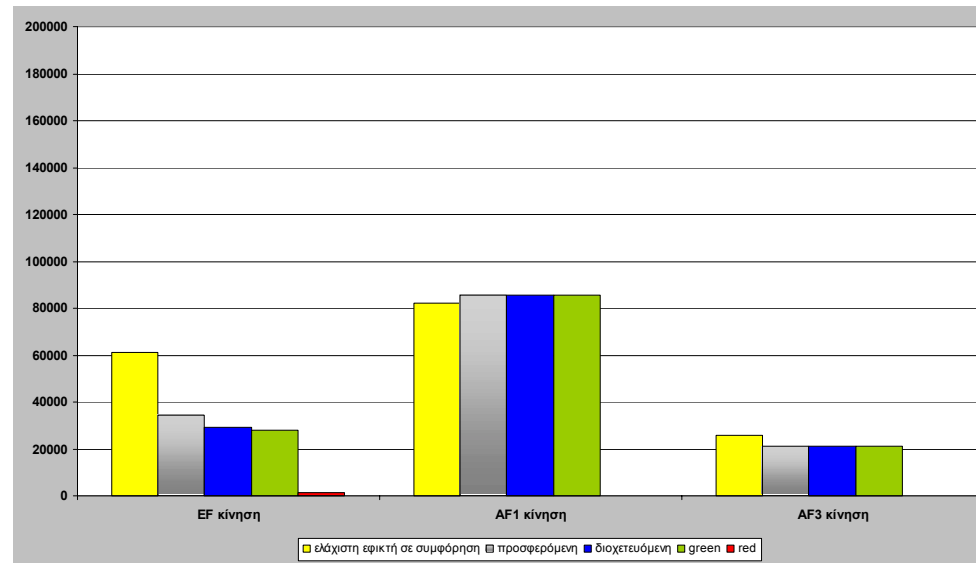


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-28: αποτελέσματα πειράματος [TC-R-4].[T-2]

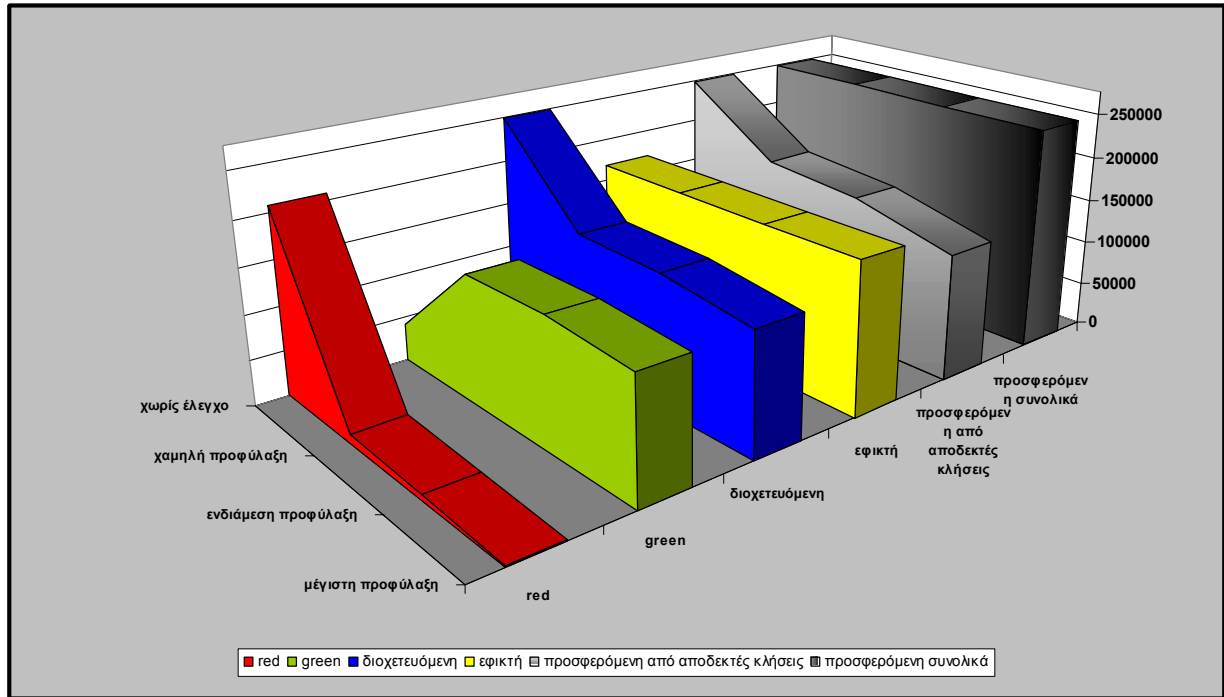


(α) φόρτος ανά QoS-κλάση στο χρόνο

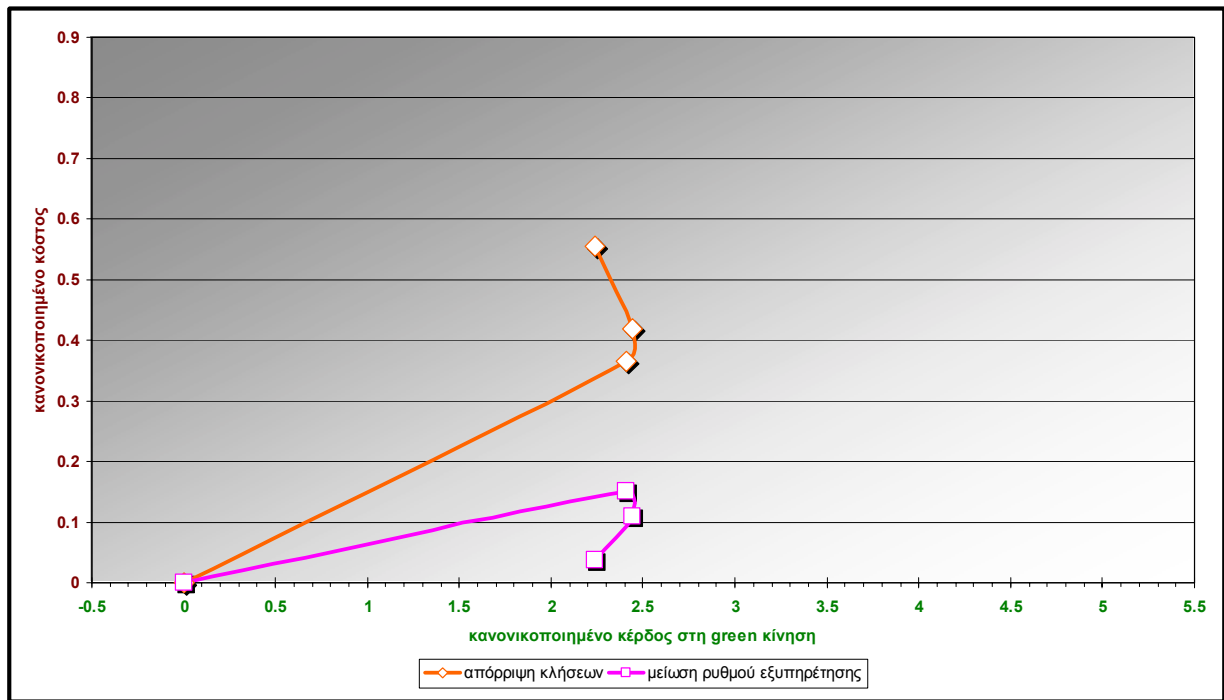


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-29: αποτελέσματα πειράματος [TC-R-4].[T-3]



(α) συνολικός φόρτος κατά μέσο όρο ανά πείραμα (μετατόπιση 500 μονάδων χρόνου)



(β) κανονικοποιημένο κέρδος και κόστος ανά πείραμα

Σχήμα C-30: συγκριτικά αποτελέσματα πειραμάτων υπόθεσης δοκιμών TC-R-4

C.3.2.3 Σειρά Δοκιμών TS-S

C.3.2.3.1 Σχεδιασμός

Η συμπεριφορά της οντότητας RSIM εξαρτάται επιπλέον από τον αριθμό των εγγεγραμμένων υπηρεσιών που θα παράγουν εξουσιοδοτημένες προς ενεργοποίηση κλήσεις, ο οποίος εξαρτάται από τη συμπεριφορά της οντότητας SSM, που με τη σειρά της ρυθμίζεται από την παράμετρο λειτουργικής πολιτικής του επιπέδου ικανοποίησης SL, δηλαδή τον (α) παράγοντα διαφοροποίησης των πειραμάτων.

Όταν η πληρότητα είναι ήδη αρκετά χαμηλή, το κέρδος και το κόστος που σχετίζεται με την εφαρμογή της οντότητας RSIM είναι μηδενικό. Εφόσον με την εφαρμογή της οντότητας SSM η πληρότητα δεν μπορεί παρά να μειωθεί ακόμα περισσότερο, η επίδρασή της στη συμπεριφορά της οντότητας RSIM θα είναι ανύπαρκτη. Είναι συνεπώς προτιμότερο, η μελέτη της επίδρασης της οντότητας SSM στη συμπεριφορά της οντότητας RSIM να χρησιμοποιεί υψηλή και συγκεκριμένα αρκετά μεγάλη έτσι ώστε να παρατηρείται κάποια επίδραση σε μεγάλο φάσμα από το πεδίο τιμών του επιπέδου ικανοποίησης SL που διαφοροποιεί και την επίδραση.

Για να απομονώσουμε τη μελέτη μας στην επίδραση της οντότητας SSM και εφόσον έχουμε μελετήσει ξεχωριστά υποθέσεις διαφορετικής πληρότητας, χρησιμοποιούμε για όλα τα πειράματα της σειράς TS-S ένα σταθερό σενάριο πληρότητας, δηλαδή σταθερό τον παράγοντα διαφοροποίησης των πειραμάτων (β). Συγκεκριμένα, το σενάριο που χρησιμοποιείται είναι αυτό της υψηλής ανομοιογενώς κατανεμημένης πληρότητας που χρησιμοποιήθηκε στην υπόθεση δοκιμών TC-R-4 (βλέπε Πίνακας C-2). Αυτό το σενάριο εκτελέστηκε για πληθυσμό εγγεγραμμένων υπηρεσιών που έχει προκύψει ουσιαστικά χωρίς καμία απόρριψη αίτησης εγγραφής ή αλλιώς με επίπεδο ικανοποίησης SL=-1. Θεωρούμε έτσι πως η υπόθεση δοκιμών TC-R-4 ισοδυναμεί με τη βασική υπόθεση δοκιμών TC-S-0, τα αποτελέσματα από τα πειράματα της οποίας χρησιμοποιούνται ως βάση για την αξιολόγηση υποθέσεων δοκιμών με διαφορετικές τιμές για το επίπεδο ικανοποίησης. Διακρίνουμε περαιτέρω τέσσερις υποθέσεις δοκιμών TC-S-1 έως TC-S-4, για καθεμία από τις καθορισμένες διακριτές τιμές του επιπέδου ικανοποίησης που χρησιμοποιούνται στον παράγοντα (α) διαφοροποίησης των πειραμάτων (βλέπε Πίνακας C-4).

			επίπεδο προφύλαξης	
			T-1	T-2
TC-S-0	επίπεδο ικανοποίησης	-1	1	0
TC-S-1		0.1	1	0
TC-S-2		0.3	1	0
TC-S-3		0.6	1	0
TC-S-4		1		0

Πίνακας C-4: πειράματα ανά υπόθεση ακολουθίας δοκιμών TS-S

Για κάθε υπόθεση δοκιμών, οι αιτήσεις εγγραφής τροφοδοτούνται στην οντότητα SSM, η οποία, ρυθμισμένη με διαφορετικό επίπεδο ικανοποίησης, αποδέχεται διαφορετικό πληθυσμό αιτήσεων εγγραφών. Για κάθε τέτοια υπόθεση εκτελούνται δύο πειράματα T-1 και T-2, που αντιστοιχούν στα T-1 και T-3 των υποθέσεων της σειράς δοκιμών TS-R, με την οντότητα RSIM ρυθμισμένη να λειτουργεί με την ελάχιστη και τη μέγιστη προφύλαξη αντίστοιχα. Εξάιρεση αποτελεί η υπόθεση δοκιμών TC-S-4, όπου το πείραμα T-1 δεν εκτελείται

καθώς η πληρότητα είναι τόσο χαμηλή που, με την ελάχιστη προφύλαξη, η οντότητα RSIM παραμένει τελικά ανενεργή και χωρίς καμία επίδραση στα αποτελέσματα.

Με τα πειράματα αυτά επιδιώκουμε να μελετήσουμε τη φύση και το μέγεθος της επίδρασης του επιπέδου ικανοποίησης στην απόδοση της οντότητας RSIM. Συνεπώς, τα αποτελέσματα χρησιμεύουν για τη σύγκριση του κέρδους και του κόστους με τη λειτουργία της οντότητας RSIM σε διαφορετικές τιμές του επιπέδου ικανοποίησης.

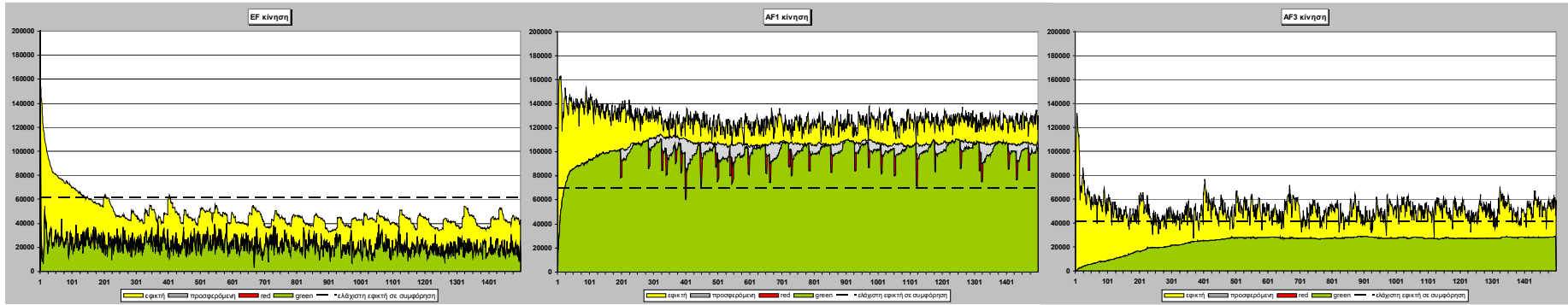
C.3.2.3.2 Αποτελέσματα

Τα αποτελέσματα των πειραμάτων της σειράς δοκιμών TS-S (βλέπε Σχήμα C-31 έως Σχήμα C-37) ομαδοποιούνται ανάλογα με το επίπεδο προφύλαξης στις κατηγορίες ελάχιστης και μέγιστης προφύλαξης των πειραμάτων T-1 και T-2 αντίστοιχα κάθε υπόθεσης δοκιμής. Η απόδοση του αλγορίθμου για κάθε κατηγορία αποτιμάται συναρτήσει του επιπέδου ικανοποίησης της κάθε υπόθεσης δοκιμής (βλέπε Σχήμα C-38).

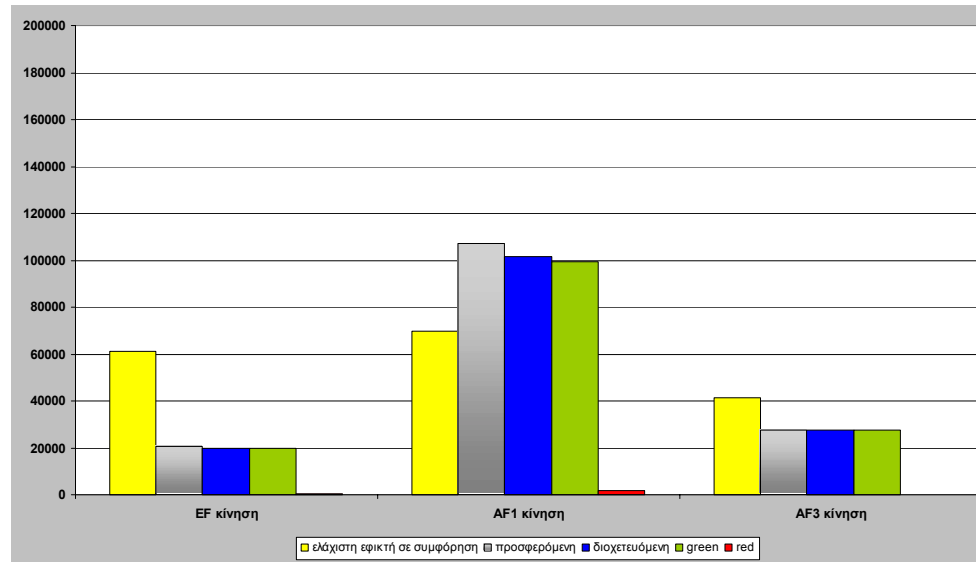
Παρατηρούμε πως πράγματι, σε κάθε περίπτωση, η αύξηση του επιπέδου ικανοποίησης έχει ως αποτέλεσμα τη μείωση του κόστους ικανοποίησης, σε οποιαδήποτε αναλογία και αν κατανέμεται αυτό μεταξύ μείωσης ρυθμού εξυπηρέτησης και απόρριψης κλήσεων.

Ειδικότερα, στην περίπτωση της ελάχιστης προφύλαξης, το κόστος γρήγορα μηδενίζεται. Επιπλέον, εφόσον όσο μειώνεται η πληρότητα, η ελάχιστη προφύλαξη γίνεται περισσότερο κατάλληλη ρύθμιση, παρατηρείται ακόμα και κάποια αύξηση του κέρδους, μέχρι που η πληρότητα μειώνεται σε επίπεδα μικρότερα από αυτά της εφικτής χωρητικότητας του δικτύου οπότε το σχετικό κέρδος περνάει σε αρνητικές τιμές. Η ελάχιστη αρνητική τιμή του κέρδους στη μέγιστη τιμή του επιπέδου ικανοποίησης ισοδυναμεί με την απώλεια από τη χρήση της λογικής ανάθεσης αιχμής (peak allocation, βλέπε ενότητα A.1.4.3.2), δηλαδή την απώλεια από την απόρριψη συμβολαίων που χάρη στο κέρδος πολυπλεξίας (multiplexing gain) είναι εφικτό να εξυπηρετηθούν.

Στην περίπτωση της μέγιστης προφύλαξης παρατηρούμε επιπλέον πως το κόστος σε απόρριψη κλήσεων παραμένει υψηλό ακόμα και για υψηλές τιμές του επιπέδου ικανοποίησης. Αυτό συμβαίνει λόγω της ανομοιογένειας της κατανομής της πληρότητας του σεναρίου TC-R-4 που χρησιμοποιούμε. Με τη μέγιστη προφύλαξη ο αλγόριθμος με την ελάχιστη υπέρβαση του δίκαιου μεριδίου λαμβάνει προληπτικά μέτρα, η ήπια στρατηγική που εφαρμόζεται πρώτη αυξάνει ειδικά την πιθανότητα απόρριψης κλήσεων. Με άλλα λόγια, ενώ οι πόροι στο δίκτυο υπερεπαρκούν για τη συνολική κίνηση, το λάθος των προβλέψεων ως προς την κατανομή της μεταξύ των QoS-κλάσεων έχει ως αποτέλεσμα την αντίστοιχα εσφαλμένη κατανομή των δίκαιων μεριδίων, στα οποία βασίζεται ο αλγόριθμος όταν είναι ρυθμισμένος να λειτουργεί προληπτικά με τη μέγιστη προφύλαξη. Επιβεβαιώνεται δηλαδή πως, με τη μέγιστη προφύλαξη το σύστημα δεν αφήνεται να λειτουργήσει αντιδραστικά βάσει ανάδρασης και δεν μπορεί έτσι να απορροφήσει το λάθος των προβλέψεων, με υψηλό κόστος σε ικανοποίηση και προφανώς και επιπλέον μείωση του κέρδους. Αξίζει ωστόσο να σημειώσουμε πως, με τη μέγιστη τιμή επιπέδου ικανοποίησης και τη μέγιστη προφύλαξη, εκείνες οι λιγότερες από τις εφικτές κλήσεις που θα γίνουν αποδεκτές λαμβάνουν εγγυημένη ποιότητα.

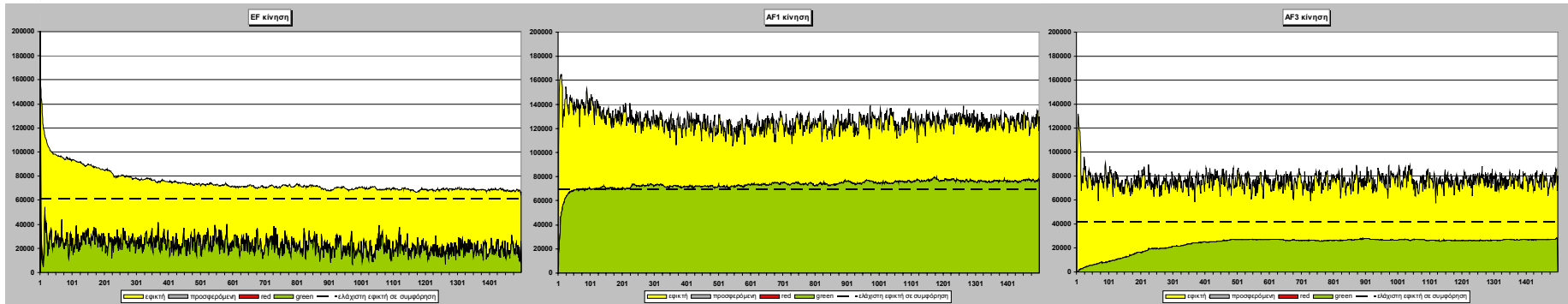


(α) φόρτος ανά QoS-κλάση στο χρόνο

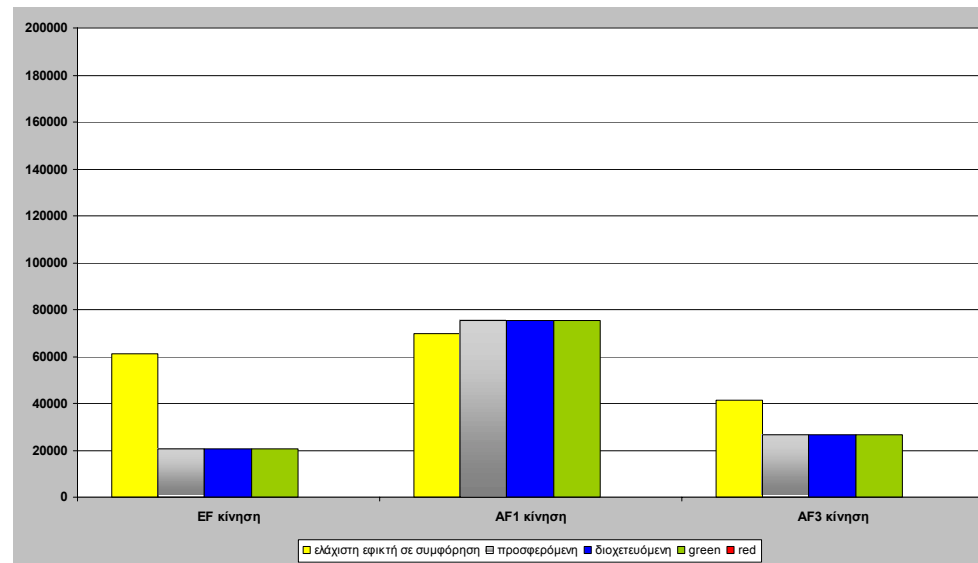


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-31: αποτελέσματα πειράματος [TC-S-1].[T-1]

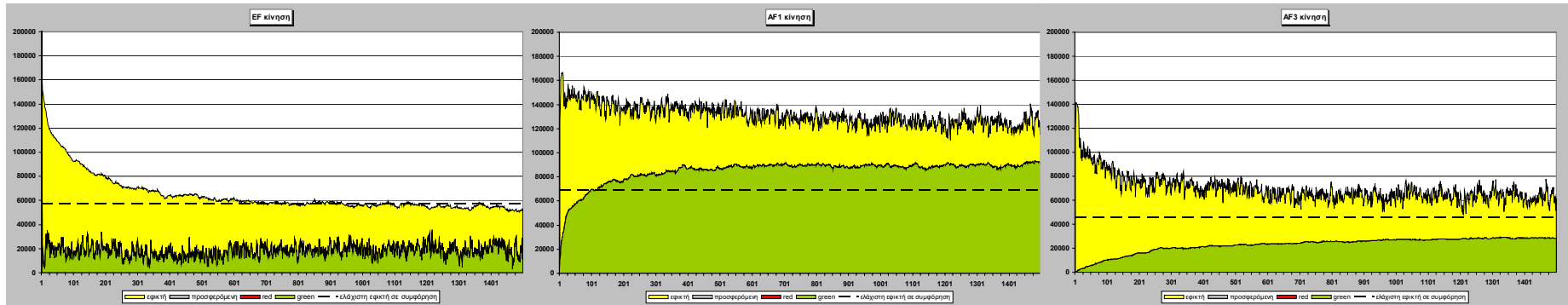


(α) φόρτος ανά QoS-κλάση στο χρόνο

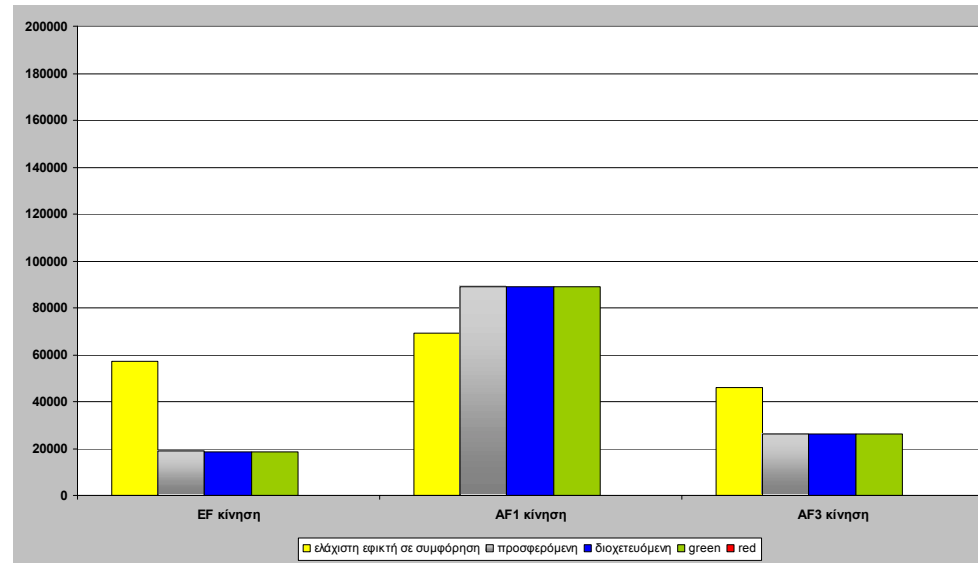


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-32: αποτελέσματα πειράματος [TC-S-1].[T-2]

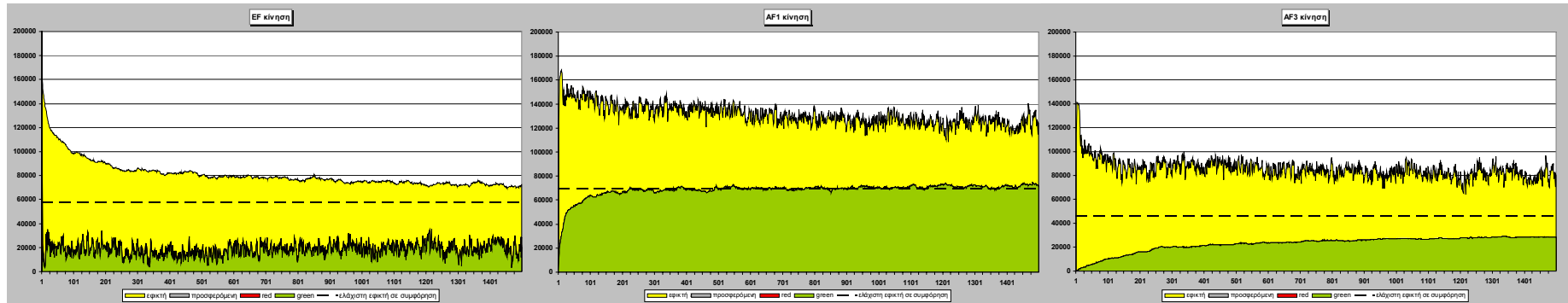


(α) φόρτος ανά QoS-κλάση στο χρόνο

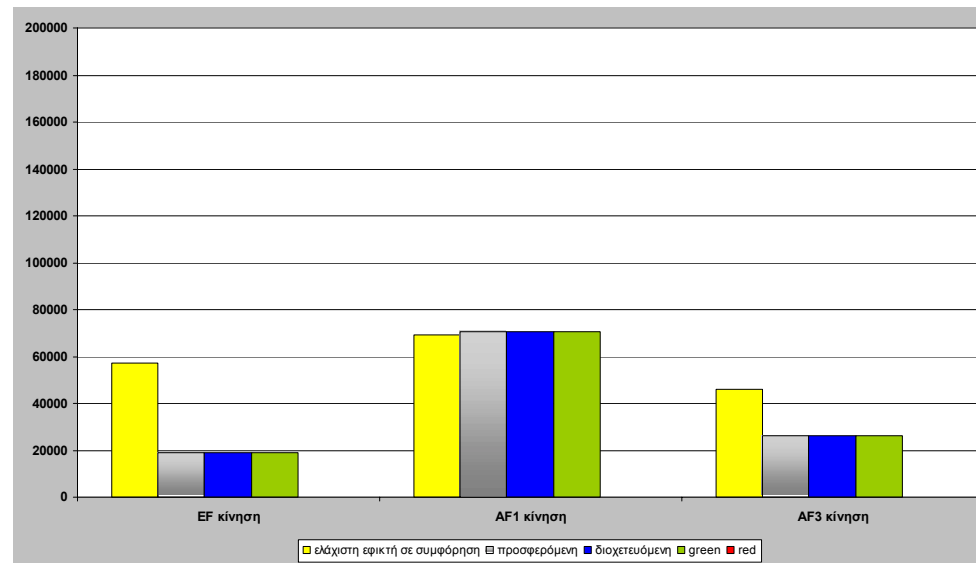


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-33: αποτελέσματα πειράματος [TC-S-2].[T-1]

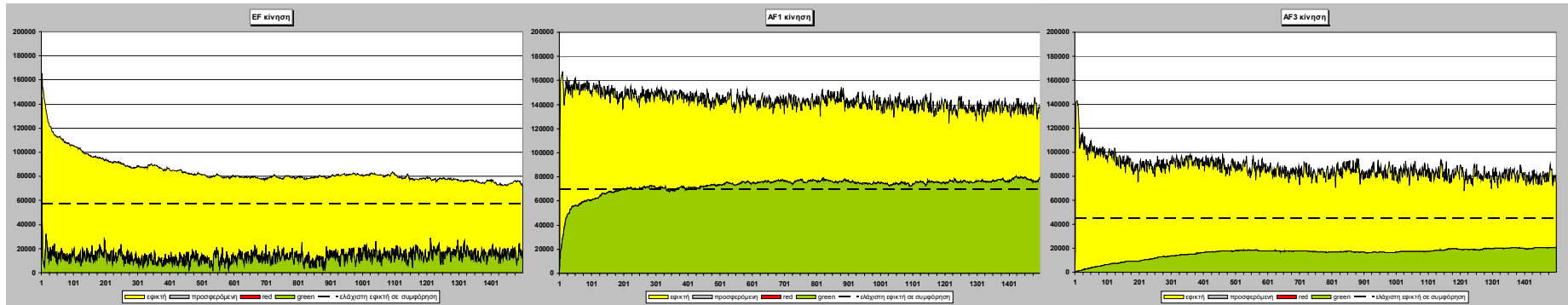


(α) φόρτος ανά QoS-κλάση στο χρόνο

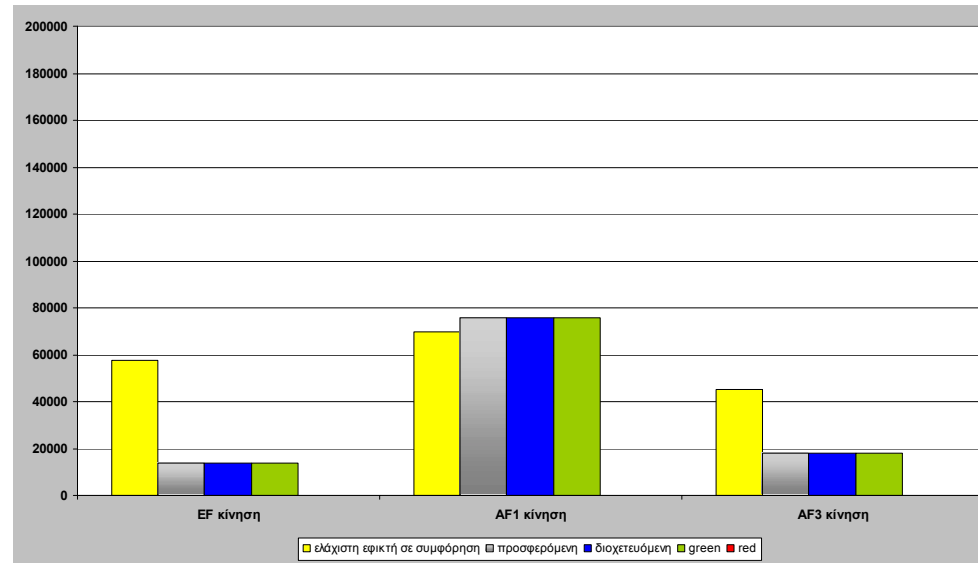


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-34: αποτελέσματα πειράματος [TC-S-2].[T-2]

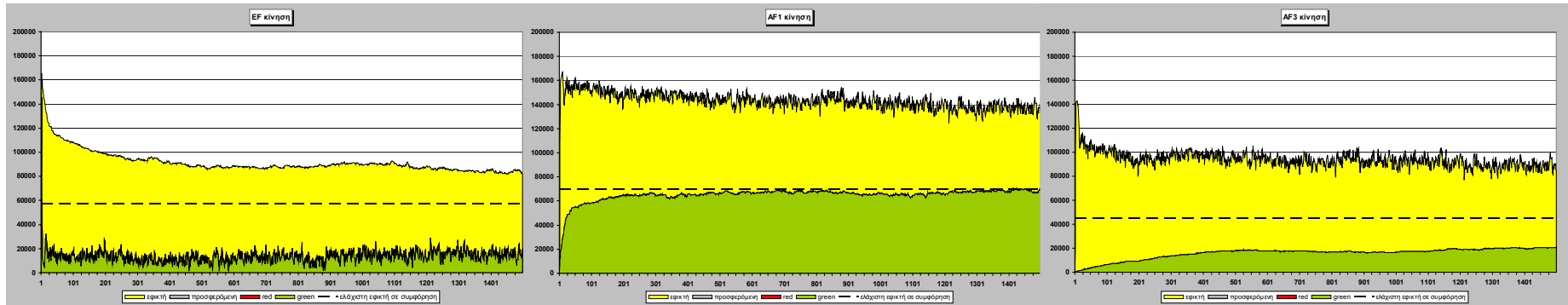


(α) φόρτος ανά QoS-κλάση στο χρόνο

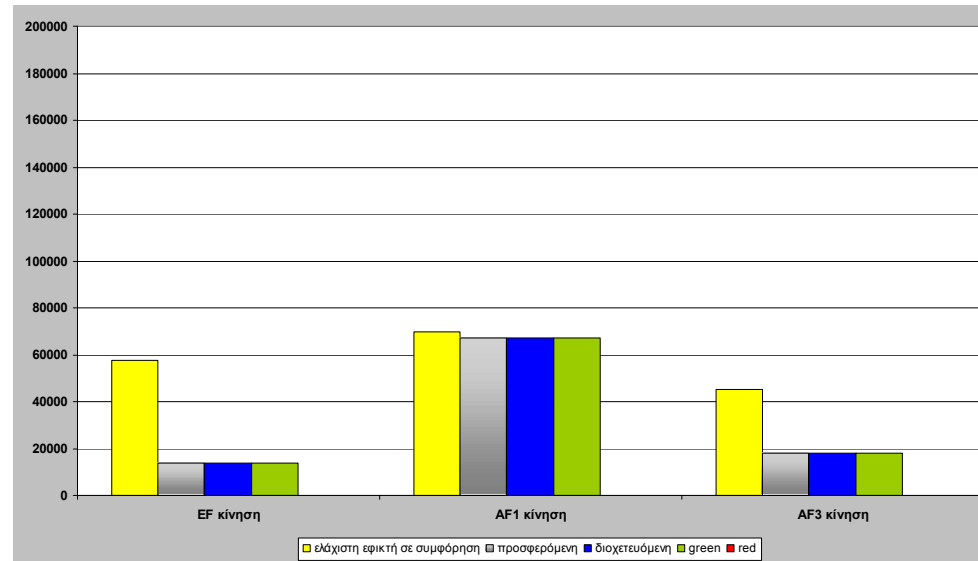


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-35: αποτελέσματα πειράματος [TC-S-3].[T-1]

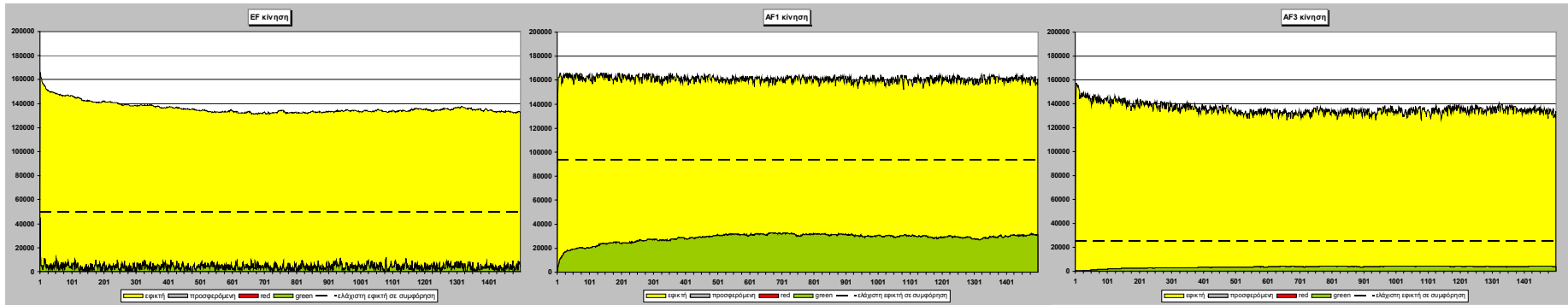


(α) φόρτος ανά QoS-κλάση στο χρόνο

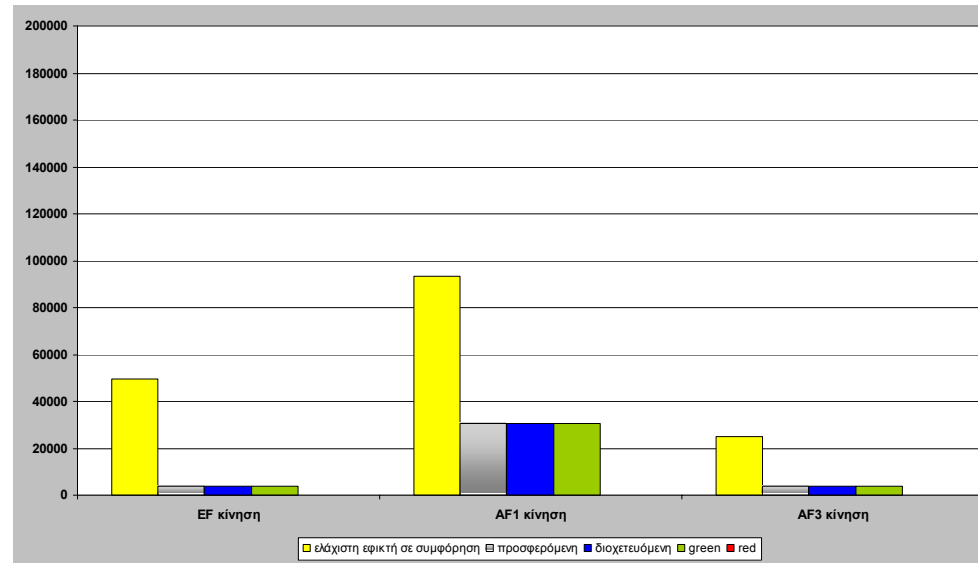


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-36: αποτελέσματα πειράματος [TC-S-3].[T-2]

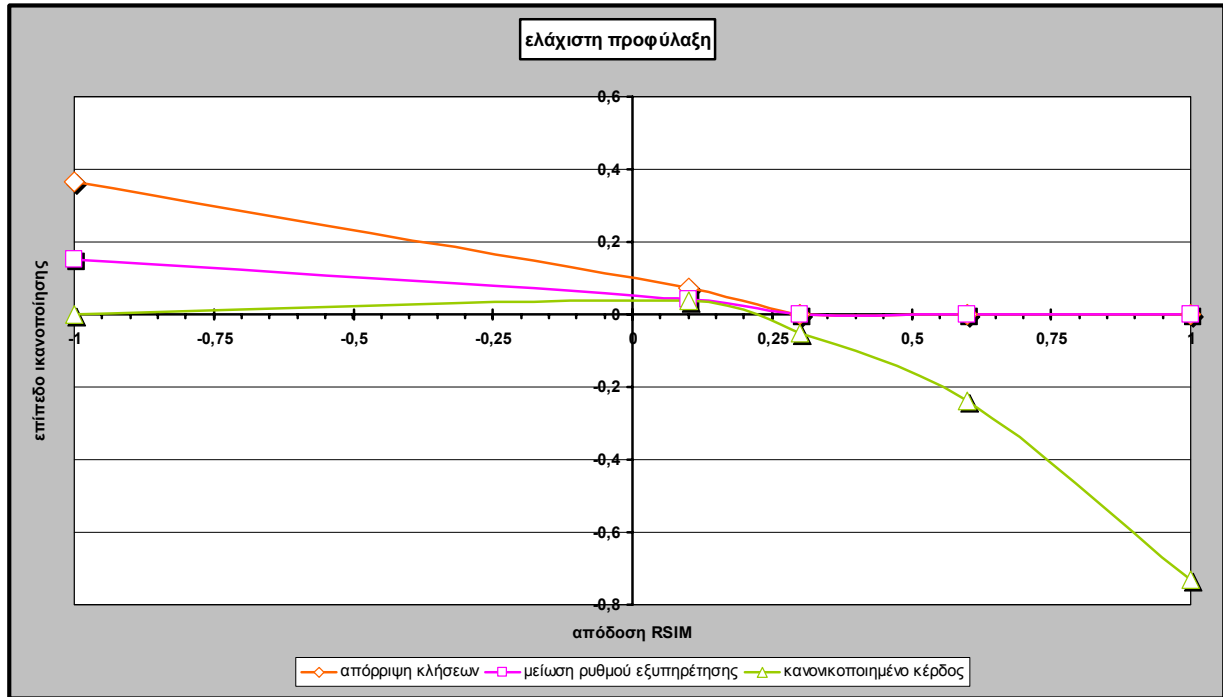


(α) φόρτος ανά QoS-κλάση στο χρόνο

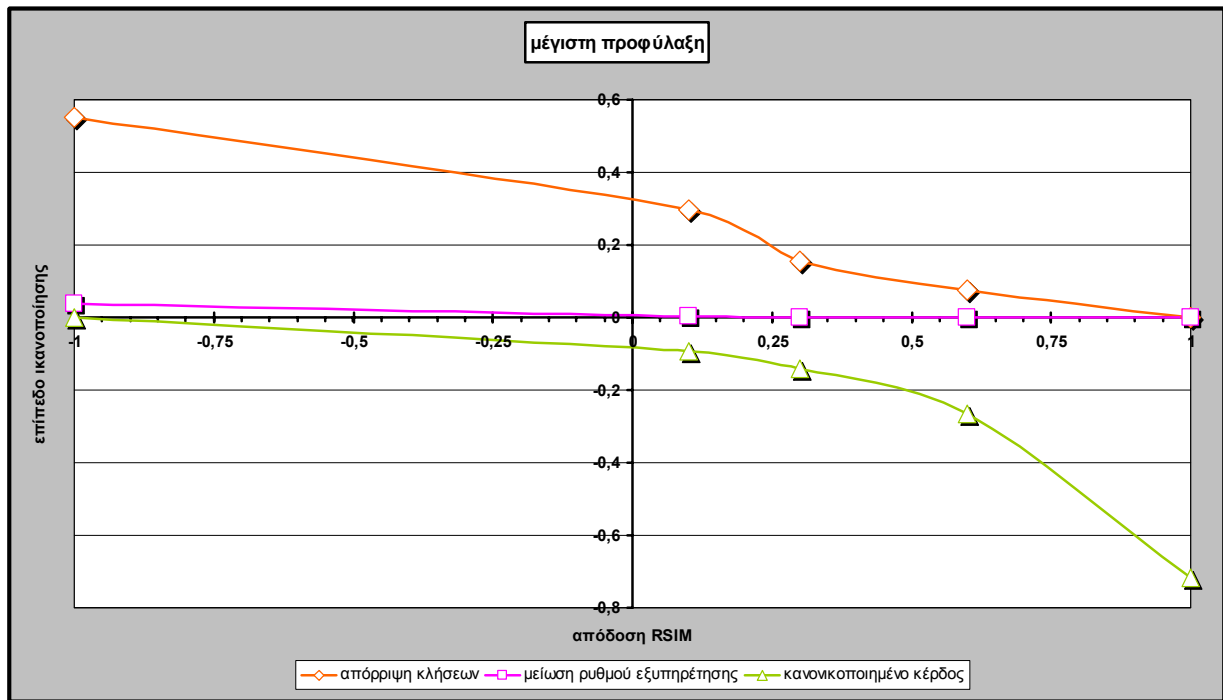


(β) φόρτος ανά QoS-κλάση κατά μέσο όρο (μετατόπιση 500 μονάδων χρόνου)

Σχήμα C-37: αποτελέσματα πειράματος [TC-S-4].[T-2]



(α) κανονικοποιημένο κέρδος και κόστος στην ελάχιστη προφύλαξη συναρτήσεϊ του επιπέδου ικανοποίησης



(β) κανονικοποιημένο κέρδος και κόστος στη μέγιστη προφύλαξη συναρτήσεϊ του επιπέδου ικανοποίησης

Σχήμα C-38: συγκριτικά αποτελέσματα πειραμάτων σειράς δοκιμών TS-S

C.4 Συμπεράσματα

Στην παρούσα εργασία μετά από εκτενή εξέταση των μελετών ελέγχου εισόδου κίνησης για την παροχή ποιότητας σε δίκτυα DiffServ και τον προσδιορισμό των πλεονεκτημάτων και των μειονεκτημάτων της κάθε προσέγγισης αναπτύξαμε ένα νέο μοντέλο που αποκαλούμε βασιζόμενο σε ανάδραση. Η προτεινόμενη λύση εντάχθηκε σε ένα περιβάλλον πλαίσιο παροχής υπηρεσίας και αναγνωρίστηκαν οι αλληλεπιδράσεις με τις εξωτερικές οντότητες. Αναπτύχθηκε ένα πρωτότυπο υλοποίησης και εφαρμόστηκε επί εμπορικών και πειραματικών δρομολογητών. Επιπλέον αναπτύχθηκε μια πειραματική πλατφόρμα εξομίωσης του δικτύου η οποία μας επέτρεψε τη διεξαγωγή εκτεταμένων πειραμάτων για την αξιολόγηση της συμπεριφοράς και της απόδοσης του αλγορίθμου.

Ακολούθως, συνοψίζονται η καινοτομία του μοντέλου, η αξιολόγηση της προτεινόμενης λύσης και ορισμένα από τα θέματα προς περαιτέρω έρευνα που προέκυψαν κατά την εκπόνηση της παρούσας εργασίας.

C.4.1 Καινοτομία

Η καινοτομία της προτεινόμενης λύσης έγκειται σε τρία σημεία, (α) στην ανάπτυξη του νέου μοντέλου βασιζόμενου σε ανάδραση, (β) στο διαχωρισμό του ελέγχου εισόδου στο επίπεδο εγγραφής και στο επίπεδο κλήσης και χρήσης της υπηρεσίας και (γ) στη ρύθμιση της συμπεριφοράς του αλγορίθμου από την πολιτική του παροχέα με τρόπο ανεξάρτητο των υποκείμενων μοντέλων.

- *Μοντέλο βασιζόμενο σε ανάδραση*

Μετά από εκτενή εξέταση των μελετών ελέγχου εισόδου κίνησης για την παροχή ποιότητας σε δίκτυα DiffServ, τοποθετηθήκαμε στο κατανεμημένο μοντέλο αμιγώς ακραίων κόμβων με τη χρήση μετρήσεων. Οι υπάρχουσες μελέτες διακρίνονται περαιτέρω σε δύο κατηγορίες. Στην πρώτη προσέγγιση, η από άκρο σε άκρο χωρητικότητα θεωρείται γνωστή και διαμορφώνεται από τον α ριγοί ή το δυναμικό και συντονισμένο διαμερισμό των πόρων του δικτύου σε απομονωμένα μη επικαλυπτόμενα μερίδια. Μετρήσεις του φόρτου που διοχετεύεται τοπικά προς το δίκτυο επί της γνωστής χωρητικότητας χρησιμοποιούνται για την εκτίμηση της διαθεσιμότητας των πόρων. Στη δεύτερη προσέγγιση η από άκρο σε άκρο χωρητικότητα είναι άγνωστη, οι πόροι δεν διαμερίζονται, το δίκτυο αναπαριστάται ως μαύρο κουτί και η διαθεσιμότητα των πόρων εκτιμάται άμεσα χρησιμοποιώντας απευθείας μετρήσεις της ποιότητας από άκρο σε άκρο. Διακρίνοντας τις αδυναμίες των δύο διαφορετικών προσεγγίσεων προτείνουμε μια νέα υβριδική measurement-based μέθοδο που συνδυάζει χαρακτηριστικά των γνωστών προσεγγίσεων και αποκαλούμε μοντέλο βασιζόμενο σε ανάδραση.

- *Έλεγχος εισόδου σε δύο επίπεδα*

Υιοθετώντας την ευρέως διαδεδομένη επιχειρησιακή πρακτική της διαδικασίας εγγραφής προ της χρήσης της υπηρεσίας, εισάγουμε και προσδιορίζουμε τη λογική ελέγχου αποδοχής σε δύο επίπεδα, στο επίπεδο εγγραφής όπου ο αλγόριθμος καλείται να αποφανθεί επί της μακροπρόθεσμης διαθεσιμότητας των πόρων για την ικανοποίηση του υπό διαπραγμάτευση συμβολαίου και στο επίπεδο κλήσης όπου ο αλγόριθμος καλείται να αποφανθεί επί της τρέχουσας διαθεσιμότητας των πόρων για την άμεση ικανοποίηση της υπό εξέταση εξουσιοδοτημένης αίτησης κλήσης. Με αυτό τον τρόπο επαυξάνεται και σε ένα δεύτερο επίπεδο ο έλεγχος για την παροχή ποιότητας.

- *Ρύθμιση με τρόπο ανεξάρτητο των υποκείμενων μοντέλων*

Η παραμετροποίηση και ρύθμιση της συμπεριφοράς του ελέγχου εισόδου δεν είναι καινοτομία στα μοντέλα που χρησιμοποιούν μετρήσεις. Η καινοτομία εδώ έγκειται στην αποσύνδεση των παραμέτρων ρύθμισης με μαθηματικά προβλέψιμη πιθανότητα παραβίασης της ποιότητας η οποία έχει αποδειχθεί ανέφικτη στις υπάρχουσες μελέτες. Αντίθετα, προτείνεται η ρύθμιση της συμπεριφοράς του αλγορίθμου με σχετικό τρόπο ανάλογα με την αντίληψη του παροχέα για το τι σημαίνει επαρκής ικανοποίηση των υπηρεσιών, έχοντας αφομοιώσει και εμμέσως συνεκτιμήσει τις ιδιομορφίες του εκάστοτε περιβάλλοντος στο οποίο εφαρμόζεται ο αλγόριθμος, και συνεπώς με τρόπο ανεξάρτητο των υποκείμενων μαθηματικών μοντέλων. Ο μαθηματικός λογισμός επί της επαρκούς ικανοποίησης χρησιμοποιεί ορισμούς που βασίζονται σε μια ρυθμοκεντρική θεώρηση της ικανοποίησης.

C.4.2 Αξιολόγηση

Το θεμελιώδες πρόβλημα του ελέγχου εισόδου κίνησης συνοψίζεται στην αποτροπή της υποβάθμισης της ποιότητας των ενεργών υπηρεσιών, ενώ παράλληλα μεγιστοποιείται η διαθεσιμότητα της υπηρεσίας και συνεπώς η αξιοποίηση των πόρων. Η απόδοση της προτεινόμενης λύσης ως προς το θεμελιώδες αυτό πρόβλημα υπολογίζεται συνεκτιμώντας το κόστος σε υποβάθμιση της ικανοποίησης τόσο ως προς την ποιότητα κατά τη χρήση όσο και ως προς τη διαθεσιμότητα της υπηρεσίας.

Από τα πειράματα που διεξήχθησαν διαπιστώνουμε καταρχήν πως η εφαρμογή του αλγορίθμου αυξάνει συνολικά την απόδοση του συστήματος και δύναται να ελέγξει και να περιορίσει τη συμφόρηση, είτε προληπτικά, είτε αντιδραστικά. Επιπλέον, η εφαρμογή ελέγχου αποδοχής στο επίπεδο εγγραφής, πράγματι επαυξάνει τα περιθώρια απόδοσης του συστήματος. Τέλος, η απόδοση του συστήματος διαφοροποιείται ποιοτικά κατά τις επιταγές των ρυθμίσεων του παροχέα, ρυθμίζοντας το κόστος σε διαθεσιμότητα έναντι του κόστους σε ποιότητα κατά τη χρήση.

Σύμφωνα με τα κριτήρια αξιολόγησης των προτεινόμενων λύσεων για το πρόβλημα του ελέγχου εισόδου κίνησης διαπιστώνουμε επιπλέον τα παρακάτω.

- *Κριτήριο αποτελεσματικότητας ως προς το δίκτυο*

Ως προς τη διαδικασία της λήψης αποφάσεων η λειτουργία του ελέγχου εισόδου κίνησης δεν απαιτεί κανένα είδους συντονισμό μεταξύ των κατανεμημένων στους ακραίους κόμβους στιγμιότυπων της. Η μόνη επιβάρυνση του δικτύου από μηνύματα που ανταλλάσσονται για τον έλεγχο εισόδου κίνησης προκύπτει αποκλειστικά από τη διαδικασία αποδοχής κλήσης που λαμβάνεται και υλοποιείται με συντονισμένο τρόπο σε κάθε ακραίο κόμβο που εμπλέκεται στην ενεργοποίηση της εκάστοτε κλήσης.

Οι μετρήσεις που απαιτούνται και η ανταλλαγή μηνυμάτων μεταξύ της λειτουργίας ελέγχου εισόδου και των λειτουργιών παρακολούθησης είναι προσεκτικά σχεδιασμένες ως ιδιαίτερο στοιχείο της προτεινόμενης λύσης και περιορίζονται σε τοπικές μετρήσεις του φόρτου που διοχετεύεται στο δίκτυο από τον υποκείμενο ακραίο κόμβο και σε αποστολή γνωστοποιήσεων κατά την ανίχνευση συμφόρησης και συνεπώς υποβάθμισης της ποιότητας από άκρο σε άκρο. Επισημαίνουμε πως η αποτελεσματικότητα της συλλογής των μετρήσεων, επεξεργασίας των πρωτογενών δεδομένων και σύνθεσης αναφορών κατατάσσεται στις απαιτήσεις των λειτουργιών παρακολούθησης και είναι έξω από τα πλαίσια της παρούσας εργασίας.

- *Κριτήριο αποτελεσματικότητας ως προς τους χρήστες*

Ο χρόνος απόκρισης του συστήματος στους χρήστες που καλούν μια υπηρεσία εξαρτάται από την απόκριση της διαδικασίας ελέγχου αποδοχής σε κάθε κόμβο και από την κατάληξη του συντονισμού μεταξύ των κατανεμημένων διαδικασιών σε όλους τους εμπλεκόμενους κόμβους. Η διαδικασία ελέγχου αποδοχής στην

προτεινόμενη λύση συνοψίζεται στον έλεγχο της πιθανότητας αποδοχής και συνεπώς η απόκρισή της είναι άμεση. Επισημαίνουμε πως η ρύθμιση της πιθανότητας αποδοχής είναι το αντικείμενο της παράλληλης διαδικασίας δυναμικής διαχείρισης αποδοχής. Ο σχεδιασμός του συστήματος σηματοδοσίας που καθορίζει και την αποτελεσματικότητα του συντονισμού μεταξύ των κατανεμημένων διαδικασιών είναι έξω από τα πλαίσια της παρούσας εργασίας.

- *Κριτήριο αποτελεσματικότητας αλγορίθμου*

Ανεξάρτητα από το μέρος της λειτουργίας αυτόματου ελέγχου εξουσιοδότησης των κλήσεων εφαρμογής και ενεργοποίησης της υπηρεσίας με την υλοποίηση των κατάλληλων ρυθμίσεων στο δρομολογητή, η πληροφορία και η επεξεργασία που απαιτεί ο αλγόριθμος είναι μηδαμινή. Αυτό εξασφαλίζεται από το δομικό μοντέλο που χρησιμοποιείται και την ελαχιστοποίηση των αλληλεπιδράσεων με τις εξωτερικές οντότητες.

- *Κριτήριο δυνατότητας επίλυσης της συμφόρησης*

Βεβαιώνεται από το σχεδιασμό της προτεινόμενης λύσης πως ο αλγόριθμος διαθέτει τη δυνατότητα να επιλύει τη συμφόρηση με άμεσο και αποτελεσματικό τρόπο, αναχαιτίζοντας την ένταση της κίνησης από τις ενεργές ροές όσο είναι απαραίτητο μόνο και επιτρέποντας έτσι την άμεση επίλυση της συμφόρησης με τις λιγότερες δυνατές κυρώσεις.

- *Κριτήριο αμεροληψίας*

Ο αλγόριθμος σχεδιάστηκε λαμβάνοντας υπόψη το κριτήριο της αμεροληψίας σε κάθε πλευρά της ικανοποίησης των υπηρεσιών, τόσο ως προς τη διαθεσιμότητα της υπηρεσίας όσο και ως προς την ποιότητα που λαμβάνει κατά τη χρήση. Επισημαίνουμε πως η αμεροληψία υπόκειται στην υποκειμενική κατά τον παροχέα αξιολόγηση της κάθε υπηρεσίας ανάλογα με τις επιχειρησιακές επιδιώξεις του.

- *Κριτήριο ευστάθειας και ευρωστίας*

Η ευστάθεια του συστήματος αποτέλεσε αντικείμενο ιδιαίτερης έρευνας στην προτεινόμενη λύση. Είναι δεδομένο πως σε ένα σύστημα βασιζόμενο σε ανάδραση, με μη ελεγχόμενους παράγοντες όπως η συμπεριφορά των χρηστών και ιδιαίτερα πολύπλοκο από την αλληλεπίδραση πολλών άλλων δυναμικών και κατανεμημένων λειτουργιών, η ευστάθεια είναι ιδιαίτερα κρίσιμο σημείο. Για αυτό το λόγο σχεδιάστηκαν και εφαρμόστηκαν ιδιαίτερη μηχανισμοί αντίχενωσης και εξομάλυνσης των ταλαντώσεων, όπου ταλαντώσεις θεωρούμε την επαναλαμβανόμενη παλινδρόμηση τόσο στην κατάσταση και τις ενέργειες του αλγορίθμου (του μηχανισμού ελέγχου), όσο και στην κατάσταση συμφόρησης του δικτύου (του αντικειμένου ελέγχου).

Η ευρωστία του συστήματος και η αντιμετώπιση όλων των πιθανών περιπτώσεων εξασφαλίζεται κατά το σχεδιασμό του αλγορίθμου ως προς την αντιμετώπιση κάθε ενδεχομένου μεταβολής των παραγόντων που επηρεάζουν τη συμπεριφορά του.

- *Κριτήριο πρακτικής εφαρμογής*

Στο σχεδιασμό της προτεινόμενης λύσης θεωρήσαμε μόνο τις ήδη προδιαγεγραμμένες και ευρέως υλοποιημένες λειτουργίες του επιπέδου δεδομένων και χωρίς να απαιτεί νέες ειδικές λειτουργίες. Η πρακτική εφαρμογή του βεβαιώνεται μέσω της υλοποίησής του σε υπάρχοντα περιβάλλοντα δικτύου επί εμπορικών δρομολογητών.

C.4.3 Περαιτέρω Έρευνα

Από την ίδια εκτίμηση της παρούσας εργασίας αναγνωρίζουμε ορισμένα θέματα που χρήζουν περισσότερο εκτεταμένης προσοχής αλλά και δυνατότητες επέκτασης σε ευρύτερο φάσμα εφαρμογής της προτεινόμενης λύσης.

Ειδικότερα, κρίνεται σκόπιμη η διεξαγωγή περισσότερο εκτεταμένων πειραμάτων για την εξέταση της σταθερότητας του συστήματος. Η αλληλεπίδραση με τις δυναμικές λειτουργίες του traffic engineering πρέπει να μελετηθούν, να αξιολογηθεί η συμπεριφορά των μηχανισμών εξομάλυνσης των ταλαντώσεων, τόσο ως προς τις παραμέτρους που τη ρυθμίζουν όσο και ως προς τις συνθήκες του περιβάλλοντος που την επηρεάζουν.

Σε μια πιο εκτεταμένη εξέταση της συμπεριφοράς του αλγορίθμου δυναμικής διαχείρισης αποδοχής είναι δυνατό να μελετηθεί το αντίκτυπο άλλων λειτουργικών παραμέτρων, όπως το ελάχιστο χρονικό διάστημα μεταξύ διαδοχικών μεταβολών στη στρατηγική αποδοχής.

Τροποποιήσεις που πιστεύεται πως θα δώσουν περισσότερα στοιχεία για την απόδοση του συστήματος αφορούν στη δοκιμή με διαφορετικές διατεταγμένες ακολουθίες προδιαγεγραμμένων στρατηγικών αποδοχής, ειδικότερα ως προς την ομάδα των ήπιων στρατηγικών. Εκτός από το είδος, τους παράγοντες και την αυστηρότητα της κάθε στρατηγικής είναι δυνατό αλλάζει και το πλήθος τους που, εφόσον εφαρμόζεται στην ίδια κρίσιμη περιοχή των διαθέσιμων πόρων, θα σημαίνει και περισσότερο ή λιγότερο συχνές μεταβολές στρατηγικής. Ενδιαφέρον παρουσιάζει και η δυνατότητα καθορισμού περιορισμένου υποσυνόλου των υπηρεσιών στις οποίες θα εφαρμόζεται η εκάστοτε στρατηγική. Τα κριτήρια ομαδοποίησης των υπηρεσιών σε αυτά τα υποσύνολα μπορεί να είναι τυχαία επιλογή, βάσει της προτεραιότητας της υπηρεσίας που καθορίζεται από τον παροχέα, βάσει μιας παραμέτρου στην οποία θα καταγράφεται ο βαθμός ικανοποίησης της κάθε εγγεγραμμένης υπηρεσίας σε προηγούμενες ενεργοποιήσεις της, κ.ά. Ενδιαφέρον παρουσιάζει η μελέτη του αντίκτυπου κάθε στρατηγικής ανάλογα τον παράγοντα που μεταβάλλει, το μέγεθος της μεταβολής του παράγοντα, το πλήθος των υπηρεσιών στο οποίο εφαρμόζεται.

Ενδιαφέρον ακόμα παρουσιάζει η μελέτη της συσχέτισης ενεργειών και αντίκτυπου με το μονοπάτι που ακολουθούν οι υπηρεσίες, το οποίο, μεταξύ των εναλλακτικών που προβλέπονται για το συγκεκριμένο TT, είναι κατά πάσα πιθανότητα διαφορετικό για κάθε υπηρεσία. Η πληροφορία αυτή θεωρείται χρήσιμη τόσο για την καλύτερη κατανόηση και ρύθμιση της συμπεριφοράς του αλγορίθμου δυναμικής διαχείρισης αποδοχής όσο και για την εφαρμογή της στρατηγικής αποδοχής. Είναι δυνατό κάποιος από τους παράγοντες της στρατηγικής όπως ο ρυθμός εξυπηρέτησης να αστυνομεύεται στη ζεύξη εξόδου προς το δίκτυο για το σύνολο της κίνησης του TT και όχι στη ζεύξη εισόδου από το δίκτυο πρόσβασης, περιορίζοντας τις μεταβολές στο επίπεδο δεδομένων και συνεπώς τη διατάραξη των ενεργών ροών και μειώνοντας την επιφόρτιση του δρομολογητή σε επεξεργασία.

Έχοντας ορίσει τις παραμέτρους κέρδους και κόστους σε απόδοση, θεωρούμε πως η ανάθεση της βέλτιστης τιμής των παραμέτρων λειτουργικής πολιτικής του παροχέα είναι δυνατό να γίνεται αυτόματα, με τον καθορισμό των εκάστοτε επιχειρησιακών κανόνων (business rules) υπολογισμού του κέρδους του παροχέα. Η σύνταξη και η μετάφραση αυτών των κανόνων στις υπάρχουσες παραμέτρους αποτελεί μια πολύ ενδιαφέρουσα επέκταση.

Τέλος, η πλέον ενδιαφέρουσα όσο και απαιτητική επέκταση της παρούσας εργασίας αφορά στην εφαρμογή της σε ένα περιβάλλον διαφορετικών επικρατειών (inter-domain) και στη μελέτη των θεμάτων που προκύπτουν για την ενεργοποίηση μιας κλήσης μεταξύ δύο τερματικών κόμβων μέσω πολλαπλών επικρατειών.

ΜΕΡΟΣ D. ΒΙΒΛΙΟΓΡΑΦΙΑ

- [AVJ01] A. Asgari, S. Van den Berghe, C. Jacquenet, P. Trimintzios, R. Egan, D. Goderis, L. Georgiades, E. Mykoniati, P. Georgatsos, D. Griffin, "A Monitoring and Measurement Architecture for Traffic Engineered IP Networks", Proceedings of IEEE/IFIP/IEE International Symposium on Telecommunications (IST2001), Tehran, Iran, 1-3 September 2001.
- [BBFP01] G. Bianchi, N. Blefari-Melazzi, M. Femminella, F. Pugini, "Performance Evaluation of a Measurement-Based Algorithm for Distributed Admission Control in a DiffServ Framework", *IEEE Globecom 2001*, November 2001.
- [BJS00] L. Breslau, S. Jamin, S. Shenker, "Comments on the Performance of Measurement-Based Admission Control Algorithm", in proc. of *IEEE Infocom 2000*, March 2000.
- [BN03] S. Bhatnagar, B. Nath, "Distributed Admission Control to Support Guaranteed Services in Core-Stateless Networks", in proc. of *IEEE Infocom*, April 2003.
- [CB01] B. Choi, R. Bettati, "Endpoint Admission Control: Network Based Approach", in proc. of IEEE ICDCS, pp. 227-235, April 2001.
- [CKK01] C. Cetinkaya, V. Kanodia, E. Knightly, "Scalable Services via Egress Admission Control", *IEEE Transactions on Multimedia: Special Issue on Multimedia over IP*, vol. 3, no. 1, pp. 69-81, March 2001.
- [COPS-PR] K. Chan, J. Seligson, D. Durham, S. Gai, K. McCloghrie, S. Herzog, F. Reichmeyer, R. Yavatkar, A. Smith, "COPS Usage for Policy Provisioning (COPS-PR)", *RFC 3084*, March 2001.
- [CSKJ00] C.-N. Chuah, L. Subramanian, R.H. Katz, A.D. Joseph, "QoS Provisioning Using A Clearing House Architecture", *International Workshop on Quality of Service*, pp. 115-124, June 2000.
- [DKS90] A. Demers, S. Keshav and S. Shenker, "Analysis and Simulation of a Fair Queueing Algorithm", *Internetworking: Research and Experience*, vol. 1, 1990, pp. 3-26.
- [DS-AF] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group", *RFC 2597*, June 1999.
- [DS-ARCH] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", *RFC 2475*, December 1998.
- [DS-EF] B. Davie, A. Charny, J.C.R. Bennet, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu, D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", *RFC 3246*, March 2002.
- [DS-EFS] A. Charny, J. Bennet, K. Benson, J. Boudec, A. Chiu, W. Courtney, S. Davari, V. Firoiu, C. Kalmanek, K. Ramakrishnan, "Supplemental Information for the New Definition of the EF PHB (Expedited Forwarding Per-Hop Behavior)", *RFC 3247*, March 2002.
- [DS-FIELD] K. Nichols, S. Blake, F. Baker, D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", *RFC 2474*, December 1998.
- [DS-PDB] K. Nichols, B. Carpenter, "Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification", *RFC 3086*, April 2001.

- [DS-RMOD] Y. Bernet, S. Blake, D. Grossman, A. Smith, "An Informal Management Model for Diffserv Routers", *RFC 3290*, May 2002.
- [DS-TB] K. Nichols, V. Jacobson, L. Zhang, "A two-bit differentiated services architecture for the internet", *RFC 2638*, July 1999.
- [DS-TERM] D. Grossman, "New Terminology and Clarifications for Diffserv", *RFC 3260*, April 2002.
- [DS-WG] <http://www.ietf.org/html.charters/OLD/diffserv-charter.html>
- [DSGLBZ98] B. Devalla, A. Sahoo, Y. Guan, C. Li, R. Bettati, W. Zhao, "Adaptive Connection Admission Control for Mission Critical Real-Time Communication Networks", in Proc. of *IEEE Milcom*, October 1998.
- [F94] S. Floyd, "TCP and Explicit Congestion Notification", *ACM Computer Communication Review*, vol. 24, no. 5, pp. 10-23, October 1994.
- [FJ93] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance", *IEEE/ACM Transactions on Networking*, vol.1 no.4, pp. 397-413, August 1993, <http://www.aciri.org/floyd/papers.html>
- [FJ95] S. Floyd and V. Jacobson, "Link-sharing and Resource Management Models for Packet Networks", *IEEE/ACM Transactions on Networking*, 3(4), August 1995, <http://www.aciri.org/floyd/papers.html>
- [FKSS97] W. Feng, D. Kandlur, D. Saha, K. Shin, "Understanding TCP Dynamics in an Integrated Services Internet", *NOSSDAV '97*, May 1997.
- [GRIP-TR] G. Bianchi, N. Blefari-Melazzi, M. Femminella, F. Pugini, "GRIP: QoS Support over a Stateless IP Domain by Means of Localized Measurements and Decisions", technical report, work in progress, http://drake.diei.unipg.it/netweb/GRIP_tech_rep.pdf.
- [GVT02] D. Goderis, S. Van den Bosch, Y. T'Joens, P. Georgatsos, D. Griffin, G. Pavlou, P. Trimintzios, G. Memenios, E. Mykoniati, C. Jacquenet, "A Scalable Service-Centric IP Quality of Service Architecture for Next Generation Networks", *Proceedings of IEEE/IFIP Network Operations and Management Symposium (NOMS2002)*, Florence, Italy, pp. 139-154, IEEE, April 2002.
- [IPv4] J. Postel, "Internet Protocol", *STD 5, RFC 791*, September 1981.
- [IPv6] S. Deering, R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", *RFC 2460*, December 1998.
- [IS-ARCH] R. Braden, D. Clark, S. Shenker, "Integrated Services in the Internet Architecture: An Overview", *RFC 1633*, June 1994.
- [IS-RAGGR] F. Baker, C. Iturralde, F. Le Faucheur, B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", *RFC 3175*, September 2001.
- [IS-CLS] J. Wroclawski, "Specification of the Controlled-Load Network Element Service", *RFC 2211*, September 1997.
- [IS-GS] S. Shenker, C. Partridge, R. Guerin, "Specification of Guaranteed Quality of Service", *RFC 2212*, September 1997.
- [IS-RSVP] J. Wroclawski, "The Use of RSVP with IETF Integrated Services", *RFC 2210*, September 1997.

- [IS-WG] <http://www.ietf.org/html.charters/OLD/intserv-charter.html>
- [JJS99] H. Jamjoom, S. Jamin, K. Shin, "Self-Organizing Network Services", Technical Report, CSE-TR-407-99, University of Michigan, 1999.
- [K97] S. Keshav, "An Engineering Approach to Computer Networking, ATM Networks, the Internet, and the Telephone Networks", Addison-Wesley, 1997, ISBN 0-201-63442-2.
- [KHAWG98] S. Kalyanaraman, D. Harrison, S. Arora, K. Wanglee, G. Guarriello, "A One-bit Feedback Enhanced Differentiated Services Architecture", draft-shivkuma-ecn-diffserv-01.txt, March 1998.
- [KS99] E. Knightly, N. Shroff, "Admission Control for Statistical QoS: Theory and Practice", *IEEE Network*, vol. 13, no. 2, pp. 20-29, March 1999.
- [KX89] D.D. Kouvatsos, N.P. Xenios, "MEM for Arbitrary Queueing Networks with Multiple General Servers and Repetitive-Service Blocking", *Performance Evaluation*, vol. 10, no. 3, pp. 169-195, 1989.
- [MCG03] E. Mykoniati, C. Charalampous, P. Georgatsos, T. Damilatis, D. Goderis, P. Trimintzios, G. Pavlou, D. Griffin, "Admission Control for Providing QoS in DiffServ IP Networks: the TEQUILA Approach", *IEEE Communications Magazine*, special issue on QoS Advances: the European IST Premium Projects, IEEE, vol. 41, no. 1, pp. 38-45, January 2003.
- [MHOC01] H.A. Mantar, J. Hwang, I.T. Okumus, S.J. Chapin, "Edge-to-edge Resource Provisioning and Admission Control in DiffServ Networks", *IEEE SoftCom*, October 2001.
- [MPCC00] R. Mortier, I. Pratt, C. Clark, S. Crosby, "Implicit Admission Control", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, pp. 2629-2639, December 2000.
- [MPLS] E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol Label Switching Architecture", *RFC 3031*, January 2001.
- [MR99] L. Massoulié, J. Roberts, "Arguments in favour of admission control for TCP flows", *Proc. of ITC-16: Teletraffic Engineering in a Competitive World*, eds. D. Smith, P. Key, vol. 3a of *Teletraffic Science and Engineering*, Elsevier, pp. 33-44, June 1999.
- [NortelQoS] Nortel Networks, "Introduction to Quality of Service (QoS)" White Paper, <http://www.nortelnetworks.com/products/library/collateral/56058.25-09-01.pdf>.
- [PSZG02] B. Pang, H. Shao, W. Zhu, W. Gao, "An Admission Control Scheme to Provide End-to-End Statistical QoS Provision in IP Networks", 21st IEEE International Performance, Computing, and Communications Conference (IPCCC 2002), April 2002.
- [RSVP-AS] A. Mankin, F. Baker, B. Braden, S. Bradner, M. O'Dell, A. Romanow, A. Weinrib, L. Zhang, "Resource ReSerVation Protocol (RSVP) – Version 1 Applicability Statement, Some Guidelines on Deployment", *RFC 2208*, September 1997.
- [RSVP-FS] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification", *RFC 2205*, September 1997.
- [RSVP-WG] <http://www.ietf.org/html.charters/rap-charter.html>
- [S95] S. Shenker, "Fundamental Design Issues for the Future Internet", *IEEE JSAC*, vol. 13, no. 7, pp. 1176-1188, September 1995.

- [SLS-T] D. Goderis, Y. T' Joens, C. Jacquenet, G. Memenios, G. Pavlou, R. Egan, D. Griffin, P. Georgatsos, L. Georgiadis, P. Van Heuven, "Service Level Specification Semantics, Parameters and Negotiation Requirements", *Internet Draft*, <draft-tequila-sls-01.txt>, June 2001.
- [SZ99] I. Stoica, H. Zhang, "Providing Guaranteed Services Without per Flow Management", in proc. of *ACM Sigcomm*, pp. 81-94, September 1999.
- [TEQ-SCST] T. Damilatis (ed.), "D3.4: Final System Evaluation (Part C) – Scalability and Stability Analysis", TEQUILA Consortium Deliverable, CEC no. 304/Algonet/b1, <http://www.ist-tequila.org/deliverables/D3-4c.pdf>, October 2002.
- [TEQ-ARCH] T. Damilatis (ed.), "D3.4: Final System Evaluation (Part B) – Final Architecture, Protocol and Algorithm Specification", TEQUILA Consortium Deliverable, CEC no. 304/Algonet/b1, <http://www.ist-tequila.org/deliverables/D3-4b.pdf>, October 2002.
- [TEQ-MON] D. Griffin (ed.), "D2.3: Status Report on Implementation (Part III2) – Design of the monitoring sub-system and modelling of monitoring services", TEQUILA Consortium Deliverable, CEC no. 203/UCL/b1, <http://www.ist-tequila.org/deliverables/D2-3.zip>, October 2001.
- [TEQ-RES] T. Damilatis (ed.), "D3.4: Final System Evaluation (Part A) – Tests and Results", TEQUILA Consortium Deliverable, CEC no. 304/Algonet/b1, <http://www.ist-tequila.org/deliverables/D3-4a.pdf>, October 2002.
- [TEQ-SRV] D. Griffin (ed.), "D2.3: Status Report on Implementation (Part III1) – Design issues and solutions for the common services of the TEQUILA system", TEQUILA Consortium Deliverable, CEC no. 203/UCL/b1, <http://www.ist-tequila.org/deliverables/D2-3.zip>, October 2001.
- [TPF03] P. Trimintzios, G. Pavlou, P. Flegkas, P. Georgatsos, A. Asgari, E. Mykoniati, "Service-driven Traffic Engineering for Intra-domain Quality of Service Management", *IEEE Network*, special issue on Network Management of Multi-service, Multimedia, IP-based Networks, vol. 17, no. 3, pp. 29-36, IEEE, May/June 2003.
- [WXBZ01a] S. Wang, D. Xuan, R. Bettati, W. Zhao, "Providing Absolute Differentiated Services for Real-Time Applications in Static-Priority Scheduling Networks", in proc. of *IEEE Infocom*, April 2001.
- [WXBZ01b] S. Wang, D. Xuan, R. Bettati, W. Zhao, "Differentiated Services with Statistical Real-Time Guarantees in Static-Priority Scheduling Networks", in proc. of *IEEE RTSS*, December 2001.
- [XLBCZ00] D. Xuan, C. Li, R. Bettati, J. Chen, W. Zhao, "Utilization-Based Admission Control for Real-Time Applications", *IEEE International Conference on Parallel Processing*, August 2000.
- [ZDH00] Z.-L. Zhang, Z. Duan, Y.T. Hou, "Virtual time reference system: A unifying scheduling framework for scalable support of guaranteed services", *IEEE JSAC*, vol. 18, no. 12, pp. 2684-2695, December 2000.
- [ZDH01] Z.-L. Zhang, Z. Duan, Y.T. Hou, "On Scalable Design of Bandwidth Brokers", *IEICE Transactions on Communications*, vol. E84-B, no. 8, pp. 2011-2025, August 2001.

- [ZDGH00] Z.-L. Zhang, Z. Duan, L. Gao, Y.T. Hou, "Decoupling QoS Control from Core Routers: A Novel Bandwidth Broker Architecture for Scalable Support of Guranteed Services", proc. *ACM Sigcomm*, pp. 71-83, August 2000.