

Μελέτη Αποδοτικών Τρόπων Διαχείρισης Profiles σε Συστήματα Publish/Subscribe

Μαρίνα Δρόσου Μυρτώ Ντέτσικα Γρηγόριος Τζώρτζης

{mdrosou, mntetsik, gtzortzi} @ cs.uoi.gr

Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

Περίληψη

Η πρόοδος που έχει συντελεστεί στον τομέα των p2p συστημάτων έχει οδηγήσει στην ανάπτυξη εφαρμογών Publish/Subscribe στις οποίες οι χρήστες ενημερώνονται αυτόματα για τα δεδομένα που τους ενδιαφέρουν. Στο άρθρο αυτό μελετούμε ένα τέτοιο σύστημα στο οποίο οι publishers χαρακτηρίζονται από profiles που περιγράφουν τα δεδομένα που παράγουν. Προτείνουμε κάποιες μορφές για τα profiles αυτά καθώς και κάποιους τρόπους ανανέωσής τους. Στη συνέχεια μελετούμε πειραματικά την απόδοση όσων προτείνουμε και συγκρίνουμε τα αποτελέσματα.

1 ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια τα p2p συστήματα έχουν γίνει ιδιαίτερα δημοφιλή καθώς αποτελούν ένα εύκολο και οικονομικό μέσο διανομής πόρων και ανταλλαγής δεδομένων. Για τον λόγο αυτό έχουν γίνει αντικείμενο εντατικής έρευνας με σκοπό την παροχή καλύτερων υπηρεσιών στους χρήστες. Η έρευνα έχει οδηγήσει στην ανάπτυξη δύο κύριων κατηγοριών p2p συστημάτων ανάλογα με την ύπαρξη ή όχι συγκεκριμένης τοπολογίας στο δίκτυο, τα δομημένα (όπως τα CAN (1), CHORD (2), BATON (3)) και τα αδόμητα (όπως η Gnutella (4) και το BitTorrent (5)).

Κοινό χαρακτηριστικό των περισσότερων p2p συστημάτων είναι ότι οι χρήστες

χρειάζεται να κάνουν μία νέα αναζήτηση στο δίκτυο κάθε φορά που θέλουν να ανακτήσουν δεδομένα. Μία διαφορετική προσέγγιση είναι τα συστήματα Publish/Subscribe, τα οποία επιτρέπουν στους χρήστες να δηλώνουν την προτίμησή τους για κάποιες κατηγορίες δεδομένων και στη συνέχεια τους ενημερώνουν για την εισαγωγή στο δίκτυο νέων δεδομένων αυτών των κατηγοριών.

Τα συστήματα Publish/Subscribe στηρίζονται στην ύπαρξη ενός Distributed Hash Table (DHT) και για τον λόγο αυτό ανήκουν στην κατηγορία των δομημένων p2p συστημάτων.

Τα συστήματα αυτά περιγράφονται αναλυτικότερα στην Ενότητα 2. Επιγραμματικά, έχουν προταθεί (6) δύο πρότυπα σχεδίασης για συστήματα Publish/Subscribe, το πρό-

τυπο Store-Sub και το Store-Pub. Η βασική διαφορά ανάμεσα στα δύο αυτά πρότυπα είναι ότι στο Store-Sub αποθηκεύεται στο δίκτυο πληροφορία σχετική με τα ενδιαφέροντα των χρηστών (subscriptions) ενώ στο Store-Pub αποθηκεύεται πληροφορία για τα δεδομένα που παράγουν οι κόμβοι (profiles). Επιπλέον, το Store-Pub δεν εγγυάται την ανάκτηση όλων των νέων δεδομένων από τους χρήστες που ενδιαφέρονται για αυτά.

Στο άρθρο αυτό επικεντρωνόμαστε στο πρότυπο Store-Pub. Προτείνουμε κάποιες μορφές για τα profiles καθώς και ορισμένους τρόπους ανανέωσης αυτών. Επίσης, πραγματοποιούμε πειράματα για την μελέτη διαφόρων συνδυασμών μορφών profiles/τρόπων ανανέωσης. Συγκρίνουμε την απόδοσή τους τόσο με βάση τον φόρτο που δημιουργούν στο δίκτυο όσο και το βαθμό στον οποίο επιτυγχάνουν να ενημερώσουν τους χρήστες για τα νέα δεδομένα.

Το υπόλοιπο του άρθρου διαρθρώνεται ως εξής: στην Ενότητα 2 περιγράφουμε αναλυτικότερα τα συστήματα Publish/Subscribe και τα δύο πρότυπα σχεδίασης Store-Sub και Store-Pub. Στην Ενότητα 3 περιγράφουμε το μοντέλο του συστήματος που χρησιμοποιήσαμε στα πειράματά μας. Στην Ενότητα 4 παρουσιάζουμε και συγκρίνουμε τα πειραματικά μας αποτελέσματα. Τέλος, η Ενότητα 5 περιέχει τα συμπεράσματα της μελέτης αυτής.

2 ΣΥΣΤΗΜΑΤΑ PUBLISH/SUBSCRIBE

Στα περισσότερα p2p συστήματα οι χρήστες αναζητούν δεδομένα χρησιμοποιώντας λέξεις-κλειδιά. Τα αποτελέσματα που λαμβάνουν αντιστοιχούν σε δεδομένα που βρίσκονται εκείνη τη στιγμή στο δίκτυο.

Για να λάβουν δεδομένα που εισέρχονται αργότερα στο δίκτυο θα πρέπει να επαναλάβουν την αναζήτηση. Τα συστήματα Publish/Subscribe προσφέρουν έναν εναλλακτικό τρόπο ανάκτησης δεδομένων με βάση τον οποίο οι χρήστες δηλώνουν τα ενδιαφέροντά τους και το σύστημα αναλαμβάνει να τους ενημερώσει για νέα δεδομένα σχετικά με αυτά.

Ένας κόμβος που συμμετέχει σε ένα σύστημα publish/subscribe μπορεί να είναι publisher ή/και subscriber. Οι publishers είναι υπεύθυνοι για την προσθήκη νέων δεδομένων στο δίκτυο. Οι subscribers λαμβάνουν κάποια από αυτά με βάση τα ενδιαφέροντα που έχουν δηλώσει. Με βάση το (6) μπορούμε να διακρίνουμε δύο σενάρια χρήσης, τα user-to-user και publisher-to-user. Στο πρώτο οι κόμβοι είναι συνήθως και publishers και subscribers ταυτόχρονα και καθένας δημοσιεύει ένα μικρό αριθμό δεδομένων. Στο δεύτερο έχουμε ένα μικρό αριθμό publishers που δημοσιεύουν πολλά δεδομένα και ένα μεγάλο αριθμό subscribers που ενδιαφέρονται για αυτά.

Έχουν προταθεί δύο μέθοδοι σχεδίασης για συστήματα Publish/Subscribe, οι Store-Sub και Store-Pub, τις οποίες περιγράφουμε στις Ενότητες 2.1 και 2.2.

2.1 Μέθοδος Store-Sub

Σύμφωνα με τη μέθοδο αυτή οι subscribers περιγράφουν τα ενδιαφέροντά τους μέσω κάποιων δομών που ονομάζονται subscriptions. Συνήθως αυτές αποτελούνται από ένα σύνολο λέξεων-κλειδιών. Στη συνέχεια αποθηκεύουν αυτές τις subscriptions στο δίκτυο. Κάθε φορά που ένας publisher παράγει ένα νέο δεδομένο ανακτά από το δίκτυο τις subscriptions που περιέχουν λέξεις-κλειδιά που περιγράφουν το νέο δεδο-

μένο και στη συνέχεια αποστέλλει το δεδομένο στους αντίστοιχους subscribers.

2.2 Μέθοδος Store-Pub

Σύμφωνα με τη μέθοδο αυτή κάθε publisher δημιουργεί ένα profile. Το profile είναι μία δομή η οποία περιγράφει το περιεχόμενο των άρθρων που έχει δημοσιεύσει στο παρελθόν ο publisher. Επομένως το profile αποτελεί και μία ένδειξη για το περιεχόμενο των άρθρων που θα δημοσιεύσει στο μέλλον. Βασισμένοι σε αυτό το γεγονός, οι publishers αποθηκεύουν στο δίκτυο το profile τους και στη συνέχεια το ανανεώνουν περιοδικά. Οι subscribers αναζητούν στο δίκτυο profiles που ανταποκρίνονται στα ενδιαφέροντά τους και στη συνέχεια κάνουν register στους αντίστοιχους publishers. Οι publishers αποστέλλουν τα νέα δεδομένα που παράγουν στους subscribers που έχουν κάνει register σε αυτούς. Οι subscribers ελέγχουν περιοδικά το δίκτυο για νέα profiles που τους αφορούν. Επίσης, αν μεταβληθούν τα ενδιαφέροντά τους, κάνουν unregister από τους αντίστοιχους publishers.

Σε αντίθεση με την μέθοδο Store-Sub, στη μέθοδο Store-Pub υπάρχει περίπτωση κάποιοι subscribers να μη λάβουν όλα τα σχετικά με τα ενδιαφέροντά τους δεδομένα λόγω της ύπαρξης των profiles που αποτελούν μία εκτίμηση του τι παράγει κάθε publisher και όχι μία ακριβή πρόβλεψη.

3 ΜΟΝΤΕΛΟ ΣΥΣΤΗΜΑΤΟΣ

Στα πλαίσια της εργασίας αυτής υλοποιήσαμε ένα σύστημα Publish/Subscribe βασισμένο στη μέθοδο Store-Pub.

Στο μοντέλο μας θεωρούμε ότι κάθε δεδομένο που παράγεται περιγράφεται από κάποιες λέξεις-κλειδιά τις οποίες καλούμε

terms. Θεωρούμε ένα Distributed Hash Table (DHT) στο οποίο κατακερματίζονται τα terms και αποθηκεύονται τα profiles. Λόγω της ύπαρξης του DHT, το σύστημά μας είναι ένα δομημένο p2p σύστημα. Το DHT παρέχει μία λειτουργία $nodeID \leftarrow lookup(term)$.

Υπάρχουν συνολικά $|P|$ publishers οι οποίοι διαθέτουν τρεις λειτουργίες:

- Publish_data: Οι publishers παράγουν κάθε φορά ένα τυχαίο αριθμό από δεδομένα, το περιεχόμενο των οποίων περιγράφεται από terms που ακολουθούν μία zipf κατανομή.
- Update_profile: Μέσω της λειτουργίας αυτής οι publishers αποθηκεύουν το profile τους στο δίκτυο.
- Send_data: Μέσω της λειτουργίας αυτής οι publishers στέλνουν τα δεδομένα που παράγουν στους subscribers που έχουν κάνει register σε αυτούς.

Μελετήσαμε τρεις διαφορετικές στρατηγικές για την ανανέωση των profiles. Πρώτον, την άμεση ανανέωση του profile κάθε φορά που παράγονται νέα δεδομένα. Δεύτερον, την περιοδική ανανέωση και τέλος την ανανέωση όταν το profile διαφέρει αισθητά από το προηγούμενο, δηλαδή όταν έχουν παραχθεί αρκετά δεδομένα που περιγράφονται από κάποιο term.

Επίσης μελετήσαμε και δύο διαφορετικές μορφές για τα profiles. Τα δυαδικά, τα οποία παρέχουν πληροφορία για το αν ο συγκεκριμένος publisher έχει παράγει δεδομένα για κάποιο term ή όχι και τα στατιστικά, τα οποία δίνουν και μία ποσοτική περιγραφή για το τι έχει παραχθεί έως τώρα.

Στα πλαίσια της πειραματικής μας μελέτης υλοποιήσαμε και συγκρίναμε πέντε διαφορετικά σενάρια-συνδυασμούς της μορφής

των profiles και του τρόπου ανανέωσής τους. Αυτοί είναι οι:

- Άμεση ανανέωση / δυαδικά profiles
- Περιοδική ανανέωση / δυαδικά profiles
- Άμεση ανανέωση / στατιστικά profiles
- Περιοδική ανανέωση / στατιστικά profiles
- Ανανέωση λόγω μεταβολής / στατιστικά profiles

Ο συνδυασμός ανανέωσης λόγω μεταβολής με δυαδικά profiles δεν έχει νόημα καθώς με τα δυαδικά profiles δεν υπάρχει η έννοια της ποσοτικής μεταβολής.

Εκτός από τους publishers στο σύστημα υπάρχουν και $|S|$ subscribers οι οποίοι διαθέτουν τις εξής τρεις λειτουργίες:

- `Generate_subscription`: Μέσω της λειτουργίας αυτής οι subscribers παράγουν έναν τυχαίο αριθμό από terms τα οποία αντιστοιχούν στα ενδιαφέροντά τους και ακολουθούν μία zipf κατανομή.
- `Retrieve_profiles`: Μέσω της λειτουργίας αυτής οι subscribers αναζητούν στο δίκτυο profiles που ανταποκρίνονται στα ενδιαφέροντά τους και στη συνέχεια κάνουν register στους αντίστοιχους publishers.
- `Unregister`: Μέσω της λειτουργίας αυτής οι subscribers κάνουν unregister από κάποιους publishers σύμφωνα με τις μεταβολές των ενδιαφερόντων τους.

Επειδή τα profiles που βρίσκονται στο δίκτυο ανανεώνονται συχνά και νέα profiles προστίθενται σε αυτό, οι subscribers αναζητούν επαναληπτικά κατάλληλα profiles

στο δίκτυο χωρίς να έχουν απαραίτητα αλλάξει τα ενδιαφέροντά τους.

4 ΠΕΙΡΑΜΑΤΙΚΗ ΜΕΛΕΤΗ

Κύριος στόχος αυτής της εργασίας είναι η πειραματική μελέτη της απόδοσης της μεθόδου Store-Pub. Για το λόγο αυτό υλοποιήσαμε ένα σύστημα Publish/Subscribe αποτελούμενο από publishers και subscribers. Υποθέσαμε επίσης την ύπαρξη ενός DHT και υλοποιήσαμε τις λειτουργίες αποθήκευσης και ανάκτησης των profiles σε αυτό.

Μελετήσαμε τον φόρτο που προκαλούν στο δίκτυο τα πέντε σενάρια που αναφέρθηκαν στην Ενότητα 3 με βάση τον αριθμό μηνυμάτων που ανταλλάσσονται. Επίσης, συγκρίναμε τα πέντε σενάρια και με βάση τον αριθμό των δεδομένων που παράγονται από τους publishers και δεν αποστέλλονται σε subscribers που ενδιαφέρονται για αυτά καθώς η μέθοδος Store-Pub δεν εγγυάται την ενημέρωση κάθε subscriber για όλα τα σχετικά με τα ενδιαφέροντά του δεδομένα.

4.1 Παράμετροι πειραματικής μελέτης

Στα πειράματά μας χρησιμοποιήσαμε ένα σύστημα αποτελούμενο από $|N| = 520$ κόμβους εκ των οποίων οι $|P| = 20$ είναι publishers και οι $|S| = 500$ είναι subscribers. Η αναλογία αυτή αντιστοιχεί στο σενάριο publisher-to-user (όπως αυτό αναφέρεται στο (6)), ένα σενάριο που ταιριάζει καλύτερα στη μέθοδο Store-Pub.

Τα ενδιαφέροντα των subscribers καθώς και τα profiles των publishers περιγράφονται από $|T| = 100$ terms η δημοφιλία των οποίων ακολουθεί zipf κατανομή με παράμετρο $\vartheta = 1.0$.

Κάθε subscription αποτελείται από ένα term για λόγους απλότητας και κάθε subscriber μπορεί να έχει ταυτόχρονα παραπάνω από μία subscription.

Τα profiles μπορούν να είναι είτε δυαδικά είτε στατιστικά. Στη δυαδική τους μορφή περιέχουν για κάθε term την τιμή 0 ή 1 ανάλογα με το αν ο αντίστοιχος publisher έχει παράγει στο παρελθόν δεδομένα για το συγκεκριμένο term ή όχι. Στη στατιστική τους μορφή περιέχουν τη συχνότητα εμφάνισης του συγκεκριμένου term στα δεδομένα που έχει παράγει ο publisher έως τώρα. Οι δύο αυτές μορφές φαίνονται στο Σχήμα 1.

Δυαδική:

0	1	1	0	0	1	...	1	0	0
---	---	---	---	---	---	-----	---	---	---

Στατιστική:

0	10	14	0	0	8	...	26	0	0
---	----	----	---	---	---	-----	----	---	---

Σχήμα 1: Μορφή profiles.

Η ανανέωση των profiles μπορεί να γίνει με τρεις διαφορετικούς τρόπους:

- Άμεσα, μετά από την παραγωγή νέων δεδομένων.
- Περιοδικά, κάθε 3 φορές που παράγει νέα δεδομένα.

- Κάθε φορά που παράγονται τουλάχιστον 4 επιπλέον δεδομένα για οποιοδήποτε term.

Στην υλοποίησή μας οι subscribers αναζητούν κατάλληλα profiles στο δίκτυο κάθε φορά που ενεργοποιούνται και ανανεώνουν τα ενδιαφέροντά τους κάθε 3^η φορά που ενεργοποιούνται. Επιπλέον, στην περίπτωση των στατιστικών profiles, τα profiles ανακτούνται μόνο εάν τα δεδομένα που αντιστοιχούν στο term για το οποίο ενδιαφέρεται ο subscriber αποτελούν τουλάχιστον το 5% των δεδομένων που περιέχουν.

Κάθε εκτέλεση του πειράματος αποτελείται από 1000 rounds και κάθε round αποτελείται από 10 γεγονότα. Το κάθε γεγονός είναι η ενεργοποίηση ενός τυχαίου publisher ή subscriber. Η πιθανότητα να επιλεγεί publisher (αντίστοιχα subscriber) σε κάθε γεγονός είναι ίση με 50%. Το ποιος publisher (αντίστοιχα subscriber) θα επιλεγεί στη συνέχεια καθορίζεται μέσω μιας ομοιόμορφης κατανομής. Επαναλάβαμε το κάθε πείραμα 10 φορές και ως τελικό αποτέλεσμα λάβαμε το μέσο όρο αυτών των επαναλήψεων.

Όνομα	Περιγραφή	Τιμή
$ N $	Αριθμός κόμβων	520
$ P $	Αριθμός publishers	20
$ S $	Αριθμός subscribers	500
$ T $	Αριθμός terms	100
ϑ	Παράμετρος zipf κατανομής	1.0
UPDATE_PUB_INTERVAL	Περίοδος ανανέωσης profiles	3
UPDATE_SUB_INTERVAL	Περίοδος ανανέωσης ενδιαφερόντων	3
TERM_PERCENTAGE	Ελάχιστο ποσοστό αποδοχής στατιστικού profile	5%
ROUNDS	Αριθμός rounds	1000
RUNS_PER_ROUND	Αριθμός γεγονότων ανά round	10

Πίνακας 1: Παράμετροι πειραμάτων.

Οι τιμές των παραμέτρων που χρησιμοποιήθηκαν στα πειράματά μας φαίνονται στον Πίνακα 1.

4.2 Τύποι μηνυμάτων

Τα είδη των μηνυμάτων που ανταλλάσσονται κατά τη διάρκεια των πειραμάτων μπορούν να χωριστούν στις κατηγορίες που φαίνονται στον Πίνακα 2.

Τύπος	Περιγραφή
ERASE_OLD_PROFILE	Διαγραφή profile από το DHT
STORE_NEW_PROFILE	Αποθήκευση profile στο DHT
SENT_DATA	Αποστολή δεδομένου σε subscriber
LOOK_UP	Αναζήτηση κόμβου στο DHT
REGISTER	Μήνυμα register προς κάποιον publisher
UNREGISTER	Μήνυμα unregister προς κάποιον publisher

Πίνακας 2: Τύποι μηνυμάτων.

Τα μηνύματα τύπου LOOK_UP θεωρούμε ότι έχουν κόστος $O(\log N)$, καθώς αυτός είναι ο χρόνος αναζήτησης που επιτυγχάνουν αρκετά DHTs, όπως π.χ. το CHORD (2). Τα υπόλοιπα μηνύματα ανταλλάσσονται απευθείας μεταξύ των κόμβων και για το λόγο αυτό θεωρούμε ότι το κόστος τους είναι $O(1)$.

4.3 Μη-ανακτημένα δεδομένα

Όπως αναφέραμε, η μέθοδος Store-Pub δεν εγγυάται την ενημέρωση κάθε subscriber για όλα τα σχετικά με τα ενδιαφέροντά του δεδομένα. Για αυτόν τον λόγο, κάθε φορά που ένας publisher παράγει κάποιο νέο δεδομένο (που αντιστοιχεί σε κάποιο term) προβαίνει σε δύο ενέργειες.

Πρώτον, αποστέλλει στους subscribers που έχουν κάνει register σε αυτόν για το συγκεκρι-

μενόμενο term το νέο δεδομένο. Δεύτερον, ελέγχει αν υπάρχουν subscribers οι οποίοι τη συγκεκριμένη στιγμή ενδιαφέρονται για αυτό το term, αλλά δεν έχουν κάνει register. Αυτό μπορεί να συμβεί λόγω της περιοδικής ανανέωσης των profiles ή λόγω του ότι ο publisher έχει παράγει για το συγκεκριμένο term δεδομένα που αντιστοιχούν σε λιγότερο από 5% των συνολικών.

Αυτός ο έλεγχος δεν αποτελεί μέρος ενός πραγματικού συστήματος Publish/Subscribe, αλλά υλοποιήθηκε προκειμένου να μετρηθεί ο όγκος των μη-ανακτημένων δεδομένων.

4.4 Πειραματικά αποτελέσματα

Στην ενότητα αυτή παρουσιάζουμε τα κυριότερα από τα πειραματικά μας αποτελέσματα. Στις γραφικές παραστάσεις που ακολουθούν, ο άξονας-y παρουσιάζεται σε λογαριθμική κλίμακα.

Στο Σχήμα 2 φαίνεται ο αριθμός των STORE_NEW_PROFILE μηνυμάτων που αντιστοιχεί σε κάθε ένα από τα πέντε σενάρια που μελετούμε. Παρατηρούμε ότι τα δύο σενάρια με άμεση ανανέωση των profiles παράγουν αρκετά περισσότερα μηνύματα κατά την αποθήκευση των profiles στο δίκτυο. Αυτό συμβαίνει επειδή οι publishers ανανεώνουν τα profiles τους κάθε φορά που ενεργοποιούνται. Τα σενάρια με περιοδική ανανέωση των profiles παράγουν λιγότερα μηνύματα επειδή οι publishers τα ανανεώνουν κάθε 3^η φορά που ενεργοποιούνται. Παρατηρούμε ότι σε αυτήν την περίπτωση τα μηνύματα που παράγονται είναι περίπου το 1/3 της προηγούμενης, επομένως η διαφορά αυτή είναι ανάλογη της περιόδου ανανέωσης των profiles, κάτι που είναι αναμενόμενο. Τα μηνύματα

αυτού του τύπου που παράγονται με το σενάριο της ανανέωσης λόγω μεταβολής εξαρτώνται από το κριτήριο που έχουμε ορίσει για τη μεταβολή αυτή. Από τις μετρήσεις αυτές συμπεραίνουμε ότι για τις παραμέτρους του πειράματός μας η περίοδος ανανέωσης είναι περίπου ίση με 2 ενεργοποιήσεις. Επίσης διαπιστώνουμε ότι η μορφή των profiles δεν επηρεάζει το πλήθος των μηνυμάτων, κάτι αναμενόμενο. Παρόμοια αποτελέσματα με πριν έχουμε και για τα ERASE_OLD_PROFILE μηνύματα για τα οποία η γραφική παράσταση παραλείπεται, καθώς όταν ανανεώνουμε τα profiles διαγράφουμε τα παλιά.

Στο Σχήμα 3 παρουσιάζεται ο αριθμός των LOOK_UP μηνυμάτων. Τέτοια μηνύματα παράγονται κατά την αποθήκευση και ανάκτηση των profiles στο δίκτυο. Για την ανάκτηση από τους subscribers και για τα πέντε σενάρια το πλήθος των LOOK_UP μηνυμάτων είναι σταθερό καθώς εξαρτάται μόνο από τα ενδιαφέροντα που παράγουν οι subscribers. Συνεπώς η διαφοροποίηση μεταξύ των σεναρίων οφείλεται στην αποθήκευση των profiles και στο πόσο συχνά γίνεται αυτό. Για το λόγο αυτό, όπως επιβεβαιώνεται και από τη γραφική παράσταση, η άμεση ανανέωση οδηγεί σε περισσότερα LOOK_UP μηνύματα.

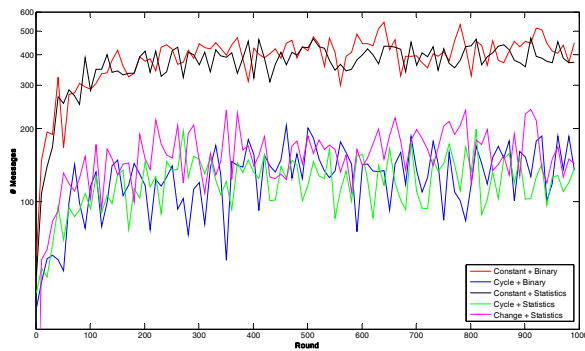
Η μορφή των profiles παίζει καθοριστικό ρόλο στο πλήθος των REGISTER μηνυμάτων όπως φαίνεται στο Σχήμα 4. Στην περίπτωση των δυαδικών profiles οι subscribers κάνουν register σε κάθε publisher με profile σχετικό με τα ενδιαφέροντά τους. Αντίθετα, η χρήση στατιστικών profiles οδηγεί σε απόρριψη από τους subscribers των publishers που έχουν μικρό ποσοστό δεδομένων για κάποια terms (στο πείραμα

το ποσοστό είναι 5%). Συνεπώς γίνονται λιγότερα register το οποίο επαληθεύεται από το Σχήμα 4.

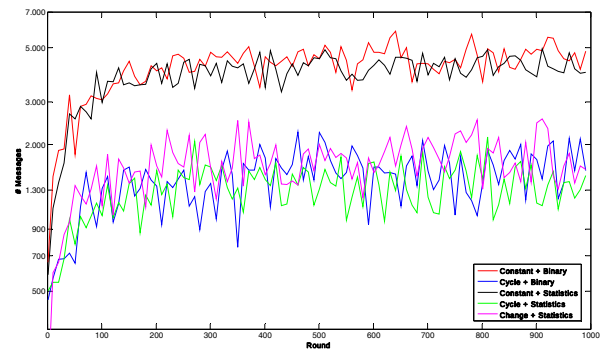
Τα μηνύματα τύπου UNREGISTER εξαρτώνται από το πόσο συχνά μεταβάλλουν οι subscribers τα ενδιαφέροντά τους και σε πόσους publishers έχουν κάνει ήδη register. Η πρώτη συνθήκη είναι ίδια για όλα τα σενάρια. Επομένως τα σενάρια με δυαδικά profiles παράγουν περισσότερα μηνύματα UNREGISTER λόγω των περισσότερων μηνυμάτων REGISTER, όπως είδαμε προηγουμένως. Λόγω αυτής της εξάρτησης η γραφική παράσταση παραλείπεται.

Στο επόμενο πείραμα παρουσιάζονται τα SENT_DATA μηνύματα (Σχήμα 5). Κάθε τέτοιο μήνυμα αντιστοιχεί στην αποστολή ενός δεδομένου σε έναν subscriber. Ο αριθμός τους εξαρτάται σε μεγαλύτερο βαθμό από τη μορφή των profiles παρά από τη συχνότητα ανανέωσης. Στην περίπτωση των δυαδικών profiles οι subscribers δεν απορρίπτουν κανένα profile σχετικό με τα ενδιαφέροντά τους από αυτά που βρίσκουν στο δίκτυο. Επομένως λαμβάνουν δεδομένα από περισσότερους publishers και για αυτό το λόγο τα μηνύματα SENT_DATA είναι περισσότερα.

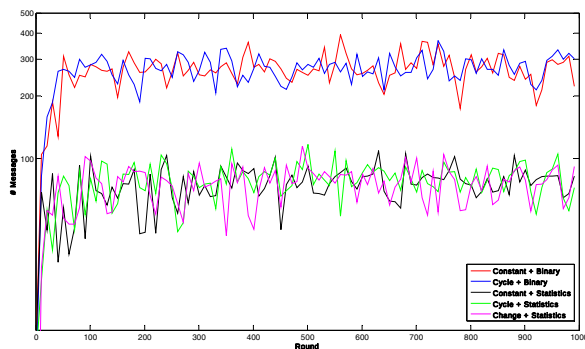
Στο Σχήμα 6 φαίνονται τα συνολικά μηνύματα που αποστέλλονται σε κάθε σενάριο. Όπως παρατηρούμε τα σενάρια που στηρίζονται στην άμεση ανανέωση παράγουν γενικά περισσότερα μηνύματα. Αυτό οφείλεται στο γεγονός ότι τα profiles ανανεώνονται συνεχώς, κάτι που οδηγεί στην αύξηση των μηνυμάτων STORE_NEW_PROFILE και LOOK_UP, όπως είδαμε και πριν. Τα υπόλοιπα σενάρια, τα οποία στηρίζονται στην περιοδική ανανέωση των profiles, παράγουν αντίστοιχα λιγότερα μηνύματα.



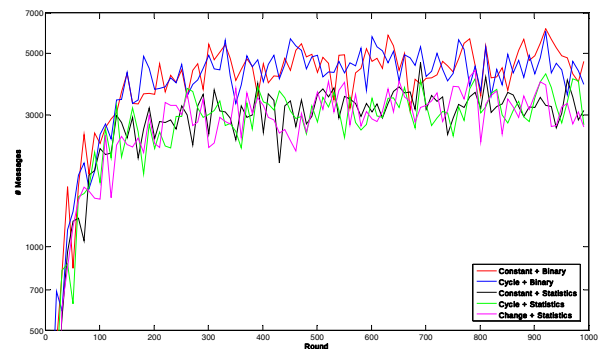
Σχήμα 2: Μηνύματα *STORE_NEW_PROFILE*.



Σχήμα 3: Μηνύματα *LOOK_UP*.



Σχήμα 4: Μηνύματα *REGISTER*.



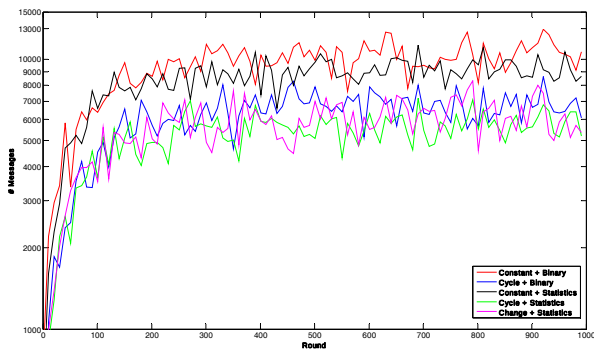
Σχήμα 5: Μηνύματα *SENT_DATA*.

Τόσο ανάμεσα στα σενάρια της άμεσης ανανέωσης όσο και ανάμεσα στα σενάρια της περιοδικής ανανέωσης περισσότερα μηνύματα παράγουν αυτά που χρησιμοποιούν δυαδικά profiles. Αυτό είναι κάτι που αναμέναμε λόγω των περισσότερων μηνυμάτων τύπου REGISTER/UNREGISTER και SENT_DATA που έχουμε σε αυτήν την περίπτωση.

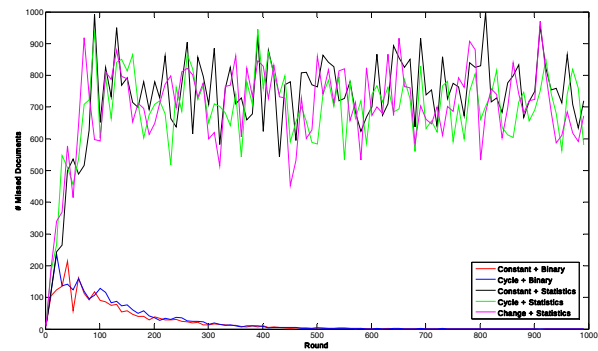
Τέλος, πραγματοποιήσαμε και κάποια πειράματα για τον αριθμό των μη ανακτημένων δεδομένων. Τα αποτελέσματα φαίνονται στο Σχήμα 7 (στο οποίο ο άξονας-y είναι σε γραμμική κλίμακα).

Η διαφορά μεταξύ των δυαδικών και των στατιστικών profiles όσο αφορά τα μη ανακτημένα δεδομένα είναι εμφανής. Αυτή

οφείλεται στο ότι με τη χρήση των στατιστικών profiles οι subscribers δεν επιλέγουν τελικά όλα τα σχετικά με τα ενδιαφέροντά τους profiles που βρίσκουν στο δίκτυο. Κατά συνέπεια δεν κάνουν register σε ορισμένους publishers οι οποίοι στη συνέχεια τυγχάνει να παράγουν δεδομένα που τους ενδιαφέρουν. Στην περίπτωση των δυαδικών profiles ο αριθμός των μη ανακτημένων δεδομένων τείνει μετά από κάποια rounds στο μηδέν. Αυτό συμβαίνει επειδή από τη στιγμή που κάποιος publisher παράγει έστω και ένα δεδομένο για κάποιο συγκεκριμένο term, όλοι οι subscribers που ενδιαφέρονται για αυτό το term θα κάνουν register σε αυτόν. Επομένως, ακόμα και αν ο publisher αυτός παράγει συνολικά λίγα δεδομένα για το term αυτό, κανένα από αυτά δε θα χαθεί.



Σχήμα 6: Αριθμός συνολικών μηνυμάτων.



Σχήμα 7: Αριθμός μη-ανακτημένων δεδομένων.

5 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Στο άρθρο αυτό περιγράψαμε τα συστήματα Publish/Subscribe και ιδιαίτερα τη μέθοδο Store-Pub. Στόχος της μελέτης μας ήταν αφενός το να προτείνουμε κάποιες εναλλακτικές μορφές για τα profiles και τρόπους ανανέωσης αυτών και αφετέρου το να αξιολογήσουμε την απόδοση των παραπάνω.

Με βάση την πειραματική μας μελέτη μπορούμε να πούμε πως η άμεση ανανέωση των profiles οδηγεί στην παραγωγή περισσότερων μηνυμάτων άρα και σε μεγαλύτερο φόρτο στο δίκτυο. Για το λόγο αυτό είναι προτιμότερη η χρήση της περιοδικής ανανέωσης ώστε να μειωθεί ο φόρτος αυτός.

Όσον αφορά τη μορφή των profiles, τα δυαδικά υπερτερούν έναντι των στατιστικών ως προς τον αριθμό των μη-ανακτημένων δεδομένων. Όμως δεν κάνουν καμία διάκριση μεταξύ των publishers που παράγουν πολλά δεδομένα για κάποιο συγκεκριμένο term και αυτών που παράγουν λιγότερα. Η διάκριση αυτή είναι χρήσιμη όταν οι subscribers διαθέτουν περιορισμένους πόρους για την αποθήκευση και διαχείριση των registrations τους, καθώς σε αυτήν την πε-

ρίπτωση είναι σημαντικό να επιλέγονται οι publishers που παράγουν περισσότερα δεδομένα για το term που ενδιαφέρονται. Επομένως στις περιπτώσεις αυτές είναι προτιμότερη η χρήση των στατιστικών profiles.

Αν δεν υπάρχουν τέτοιοι περιορισμοί στο σύστημά μας, μία καλή επιλογή είναι η χρήση περιοδικής ανανέωσης σε συνδυασμό με δυαδικά profiles καθώς επιτυγχάνει σχετικά χαμηλό φόρτο στο δίκτυο αλλά και χαμηλότερο αριθμό μη-ανακτημένων δεδομένων.

Ακόμα, ένα ενδιαφέρον σενάριο είναι αυτό της ανανέωσης λόγω μεταβολής. Αν μπορούμε να παρατηρήσουμε το δίκτυό μας και ιδιαίτερα το ρυθμό παραγωγής νέων δεδομένων από τους publishers, θα μπορούσαμε να καταλήξουμε σε ένα κριτήριο ανανέωσης βασισμένο στο ρυθμό αυτό, το οποίο θα μεταβαλλόταν δυναμικά. Με τον τρόπο αυτό θα μπορούσαμε να επιτύχουμε έναν καλό συμβιβασμό για το δίκτυό μας μεταξύ του συνολικού αριθμού μηνυμάτων και των μη-ανακτημένων δεδομένων.

Η εύρεση ενός τέτοιου κριτηρίου καθώς και ο τρόπος με τον οποίο μπορούμε να παρατηρούμε το δίκτυο για τον καθορισμό του είναι ένα ενδιαφέρον ζήτημα για μελλοντική έρευνα.

Ένα άλλο ενδιαφέρον ζήτημα είναι η πληροφορία που περιλαμβάνεται στα profiles. Πιο συγκεκριμένα, θα ήταν επιθυμητό τα profiles να διαθέτουν ένα είδος μνήμης, δηλαδή να δίνουν περισσότερη βαρύτητα στα δεδομένα που ο publisher παράγαγε πιο πρόσφατα.

ΑΝΑΦΟΡΕΣ

1. **S. Ratnasamy, P. Francis, M. Handley, R. Karp, S. Shenker.** *A Scalable Content-Addressable Network*. ACM SIGCOMM. 2001. pp. 161-171.

2. **I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, H. Balakrishnan.** *CHORD: A scalable peer-to-peer lookup service for internet applications*. 2001.

3. **H.V. Jagadish, B. C. Ooi, Q. H. Vu.** *BATON: A Balanced Tree Structure for Peer-to-Peer Network*. VLDB. 2005. pp. 661-672.

4. *Gnutella*. [Online]
<http://gnutella.wego.com>.

5. **Cohen, B.** *Incentives Build Robustness in BitTorrent*. 2003.

6. **M. Bender, S. Michel, S. Parkitny, G. Weikum.** *A Comperative Study of Pub/Sub Methods in Structured P2P Networks*.