

## Κατανεμημένα συστήματα

Παπαφώτη Μαρία Α.Μ: 45

12/5/03

2. Answer the following questions based on the survey paper:

D. S. Milojevic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, and Z. Xu

"Peer-to-Peer Computing", HP Technical Report, HPL-2002-57

(online copy at: [http://www.cs.uoi.gr/~pitoura/courses/ds03\\_gr/projects.html](http://www.cs.uoi.gr/~pitoura/courses/ds03_gr/projects.html))

### QUESTIONS

(cover sec 1-5)

1. What is the definition of a p2p system given by the authors in sec 1? Compare it with at least one of the definitions surveyed in the last paragraph of pg 2.

Σύμφωνα με τους συγγραφείς ένα p2p σύστημα είναι ένας τρόπος να υλοποιήσουμε συστήματα, εφαρμογές και αλγορίθμους με έναν τρόπο όλο και περισσότερο μη κεντρικοποιημένο. Σε ένα τέτοιο σύστημα στόχος είναι ο διαμοιρασμός των πόρων ανάμεσα στους διαφόρους peers.

Ο ορισμός αυτός πλησιάζει περισσότερο τον ορισμό που δίνει το Intel p2p working group το οποίο ορίζει ως p2p σύστημα το σύστημα εκείνο όπου επιτυγχάνεται διαμοιρασμός των πόρων και των υπηρεσιών του συστήματος με άμεση επικοινωνία μεταξύ των συστημάτων.

3. What is a hierarchical and what is a flat client-server model?

Ένα flat client-server μοντέλο είναι το μοντέλο εκείνο του συστήματος όπου υπάρχει ένας server ο οποίος εξυπηρετεί της αιτήσεις κάποιων client γύρω από αυτόν. Αντίθετα σε ένα ιεραρχικό client-server μοντέλο τα διάφορα μηχανήματα είναι οργανωμένα σε μια ιεραρχία και τα μηχανήματα ενός επιπέδου λειτουργούν ως servers για τα μηχανήματα του αμέσως κατώτερου επιπέδου.

4. What is a super peer?

Super peer είναι ένας κόμβος του p2p συστήματος ο οποίος έχει περισσότερες δυνατότητες από τους υπόλοιπους κόμβους ο οποίος συγκεντρώνει πολλές από τις αιτήσεις που γίνονται στους διάφορους κόμβους με σκοπό την εξοικονόμηση bandwidth.

5. What is the difference between a compute-intensive and a componentized application? How does this relate to vertical and horizontal distribution?

Οι εφαρμογές που χαρακτηρίζονται ως compute-intensive τρέχουν την ίδια διεργασία σε πολλούς διαφορετικούς peers, αντίθετα οι εφαρμογές που χαρακτηρίζονται ως componentized οι οποίες τρέχουν διαφορετικά μέρη μιας διεργασίας σε διαφορετικούς peers.

## 6. What is according to the authors the main challenge of communication in p2p?

Σύμφωνα με τους συγγραφείς το βασικό χαρακτηριστικό της p2p επικοινωνίας είναι το γεγονός ότι τέτοια συστήματα ξεπερνούν τις δυσκολίες εκείνες που δημιουργεί σε άλλους τύπους δικτύων η ύπαρξη δυναμικών αλλαγών στο πλήθος, τη διασύνδεση και τις δυνατότητες των κόμβων. Αυτό επιτυγχάνεται λόγω του ότι την επικοινωνία κυρίως διαχειριζόμαστε στο επίπεδο εφαρμογών.

## 7. What is the most common solution to reliability across p2p systems?

Η πιο διαδεδομένη λύση στο πρόβλημα της διαθεσιμότητας των p2p συστημάτων είναι η τεχνική του πλεονασμού (redundancy).

## 8. What are the advantages/disadvantages of the centralized directory, the flooded requests, and the document routing models?

- Centralized directory μειονεκτήματα:
  - Δεν είναι πλήρως καταναμημένο
  - Περιορισμένη κλιμάκωση
- Centralized directory πλεονεκτήματα:
  - Στην πράξη ισχυρό και αποτελεσματικό (χρησιμοποιήθηκε στο Napster)
- Flooded requests μειονεκτήματα:
  - Απαιτεί μεγάλο bandwidth
  - Περιορισμένη κλιμάκωση
- Flooded requests πλεονεκτήματα:
  - Αποτελεσματικό σε μικρά δίκτυα
- Document routing μειονεκτήματα:
  - Πρέπει το όνομα του έγγραφου που θα σταλεί να είναι γνωστό εκ των προτέρων
  - Ίσως οδηγηθούμε σε αποκοπή μέρους του δικτύου από το υπόλοιπο
- Document routing πλεονεκτήματα:
  - Πολύ αποτελεσματικό για μεγάλα δίκτυα

## 9. In the centralized directory approach, after the best peer is located, the file exchange occurs directly between it and the requesting peer. What are the advantages/disadvantages of this?

Πλεονεκτήματα:

Η επικοινωνία δεν φορτώνει πλέον το δίκτυο

Μειονεκτήματα:

Αν ο peer που εντοπίστηκε παύσει να είναι διαθέσιμος τότε έχουμε αποτυχία

## 10. What can be considered as a closure mechanism in Gnutella?

Στη Gnutella κάθε χρήστης πρέπει να ξέρει την IP διεύθυνση ενός τουλάχιστον ακόμα χρήστη. Στη συνέχεια προωθεί την ερώτηση σε αυτόν και αυτός στους γνωστούς του χωρίς να χρειάζεται να επικοινωνήσει ο τελευταίος με τον πρώτο.

11. What are the factors that affect scalability, give one example for each.

- Η ποσότητα των κεντρικοποιημένων λειτουργιών. Στη χειρότερη περίπτωση κεντρικοποιημένο σύστημα δεν έχω κλιμάκωση.
- Το πλήθος των καταστάσεων που πρέπει να διαχειριστεί το σύστημα.
- Ο παραλληλισμός των λειτουργιών.
- Ο τρόπος προγραμματισμού του συστήματος.

12. Given the ad-hoc nature of connectivity in p2p, comment on what type of (message-oriented) communication (i.e., synchronous/asynchronous, transient/persistent) would be more appropriate.

Η ad-hoc φύση της συνδεσιμότητας των p2p δικτύων συνεπάγεται ότι ο αριθμός των κόμβων του δικτύου καθώς και οι συνδέσεις μεταξύ τους δεν είναι σταθερές. Πιο κατάλληλη επομένως θα ήταν μια ασύγχρονη, persistent επικοινωνία μεταξύ των κόμβων του δικτύου έτσι ώστε το μήνυμα να μην χάνεται αν ο κόμβος προορισμού είναι ανενεργός τη στιγμή που φτάνει αλλά και το σύστημα να μην μπλοκάρει τη λειτουργία του περιμένοντας μηνύματα από κόμβους που πιθανώς έχουν βγει από το δίκτυο ή έχουν αποσυνδεθεί από τον κόμβο και δεν θα στείλουν ποτέ.

13. pg 17, 1st column, last par "The geographical distribution of the peers help to reduce congestion on both peers and the network". Explain.

Η γεωγραφική κατανομή των κόμβων που έχουν αντίγραφα των δεδομένων (προκύπτουν με βάση το replication) μας βοηθάει να αποφύγουμε τη συμφόρηση τόσο στο δίκτυο αφού με το να εντοπίζουμε τα δεδομένα σχετικά κοντά περιορίζεται ο αριθμός των μηνυμάτων που στέλνουμε στο δίκτυο ενώ ταυτόχρονα περιορίζουμε και τη συμφόρηση στους κόμβους του δικτύου αφού κάθε κόμβος εξυπηρετεί μονό εκείνους τους κόμβους που βρίσκονται γεωγραφικά κοντά.

14. What is the goal of caching in p2p? What are the advantages/disadvantages of caching the reply at all nodes in the return path? Can you think of any alternatives? Is this possible in Gnutella?

Η τεχνική του caching μειώνει το μήκος του μονοπατιού που χρειάζεται για να μεταφέρουμε ένα αντικείμενο και επομένως μειώνει και τον αριθμό των μηνυμάτων στο δίκτυο. Η μείωση αυτή του φόρτου του δικτύου είναι πολύ σημαντική στα p2p συστήματα όπου οι επικοινωνίες επιφέρουν σημαντικά προβλήματα επίδοσης. Το μειονέκτημα της μεθόδου είναι η πολλαπλή διατήρηση της πληροφορίας. Εναλλακτικά θα μπορούσαμε να αποθηκεύουμε την πληροφορία σε μέρος του μονοπατιού ή ακόμη καλύτερα να αποθηκεύουμε σε ολόκληρο το μονοπάτι πληροφορία σχετική με τη θέση του αντικειμένου που ζητάμε τότε βέβαια θα έχουμε πρόβλημα αν ο κόμβος που είχε την πληροφορία αποσυνδεθεί από το δίκτυο. Για αυτό τον λόγο δεν θα ήταν καλή ιδέα να εφαρμόσουμε την τελευταία τεχνική στο Gnutella.

17. Why is the fault tolerance problem a greater challenge in collaborative p2p systems than in file sharing p2p systems?

Στα collaborative p2p συστήματα έχω πολλούς κόμβους οι οποίοι μετακινούνται και οι οποίοι δεν είναι συνέχεια στο δίκτυο. Αυτή η μετακίνηση αλλά και η αποσύνδεση των κόμβων κάνει το δίκτυο λιγότερο αξιόπιστο και για αυτό πρέπει να δίνεται μεγαλύτερη προσοχή στην αντοχή σε σφάλματα τέτοιων συστημάτων.