

Πέτσιος Στέφανος – Κων/νος

A.M. #47

Οι απαντήσεις του 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

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.

**A.** *"The term "peer-to-peer" (P2P) refers to a class of systems and applications that employ distributed resources to perform a critical function in a decentralized manner."*

Vs

**B.** *"The sharing of computer resources and services by direct exchange between systems"*

**C.** *"The use of devices on the internet periphery in a non client capacity"*

**D.** *"P2P is a class of applications that takes advantage of resources – storage, cycles, content, human presence – available at the edges of the Internet. Because accessing these decentralized resources means operating in an environment of unstable connectivity and Unpredictable IP addresses, P2P nodes must operate outside the DNS system and have significant or total autonomy from central servers"*

Ο Α ορισμός είναι ο πιο επιτυχημένος κατά την άποψή μου. Ο Β είναι αρκετά γενικός χωρίς να θέτει κάποιους περιορισμούς σαν τον decentralized manner ο οποίος είναι βασικός σε ένα p2p δίκτυο. Ο C περιορίζεται μόνο στην αναφορά devices ενώ στην πραγματικότητα οι εφαρμογές είναι αυτές που καθιστούν τα μηχανήματα ως μέρη ενός p2p δικτύου. Ο D είναι ο ορισμός ενός p2p που είναι σχεδιασμένο για το internet.

2. In Fig 2 (pg 3), the authors compare some aspects of the client-server and the p2p computing models. List and explain these aspects.

Στο σχήμα αυτό είναι εμφανείς οι διαφορές ενός p2p δίκτυο με ένα client – server αλλά ταυτόχρονα παρατηρούμε την αντιστοιχία του ενός μοντέλου με το άλλο.

Managed – Self organized

Στο client server έχουμε μία εκ των προτέρων οργάνωση και παραμένει για πάντα η ίδια, αντίθετα στο p2p έχουμε μια αυτό-οργάνωση του δικτύου μας.

Configured – ad-hoc

Στα p2p οι κόμβοι που συμμετέχουν στο δίκτυο έχουν κάποια σχετική ελευθερία στο να «έρχονται» και να «φεύγουν» από το δίκτυο ενώ στο client server εν γένη είναι εκ το προτέρων γνωστό τι θα συμβεί εάν κάποιος κόμβος αποχωρίσει ή φύγει. (??? Δεν ξέρω ακριβώς)

Lookup – discover

Στα client server ο server ψάχνει στα δεδομένα του την απάντηση και την δίνει στον client ενώ στο p2p μας ενδιαφέρει να ανακαλύψουμε αρχικά ποιος κόμβος έχει αυτό που επιθυμεί ο client.

Hierarchy – mesh

Στο p2p έχουμε κάποιο ακαθόριστο πλέγμα που έχει το δίκτυό μας ενώ το client server ακολουθεί ένα ιεραρχικό μοντέλο με ρίζα τον server.

Static – mobile

Στο client server τα πάντα είναι στατικά: από τον server, τις τοποθεσίες των client και του server μέχρι και την δομή του. Αντίθετα στο p2p τα πάντα μπορεί να είναι μέχρι ένα βαθμό μεταβλητά. (Ίσως ο όρος εδώ να αναφέρεται μόνο στις θέσεις των κόμβων)

Server dependencies – Independent lifetime

Στο client server μοντέλο είναι αναμενόμενο στο δίκτυο όλοι οι περιορισμοί για τον χρόνο λειτουργίας να είναι ευαίσθητες από την λειτουργία του server κάτι που δεν συμβαίνει στο p2p δίκτυο.

IP centric – also non IP & DNS based – custom naming

Εν γένη το client server βασίζεται στις αρχές του λειτουργίας του internet: τα DNS και IP. Στα p2p υπάρχουν εσωτερικοί μηχανισμοί για την δρομολόγηση.

## RPC – async

Το client-server μοντέλο ακολουθεί σύγχρονες επικοινωνίες ενώ το p2p μπορεί να λειτουργήσει και σε ασύγχρονο περιβάλλον.

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

Ένα flat μοντέλο βασίζεται σε ενός επιπέδου επικοινωνία του server με τους clients, δηλαδή η απόσταση του client με τον server θα είναι πάντα ένα hop, ενώ του hierarchical μοντέλου λόγω της ιεραρχικής δομής η απόσταση του client από τον server δεν είναι σταθερή και εξαρτάται από την θέση των δύο κόμβων στο ιεραρχικό δένδρο.

### 4. What is a super peer?

Με τον όρο super peer αναφέρονται σε ένα σύνολο από κόμβους οι οποίοι λειτουργούν σαν ένας. Για παράδειγμα ένα δίκτυο 1000 κόμβων θα είναι λογικά ισοδύναμο με το δίκτυο με 100 super peer (εάν ο καθένας έχει 10 κόμβους από το προηγούμενο δίκτυο). Ο κάθε super peer έχει την ιδιότητα να μαζεύει όλες τις αιτήσεις.

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

Η κύρια διαφορά ανάμεσα στα παραπάνω είναι ότι στις compute-intensive εφαρμογές στον κάθε κόμβο εκτελούμε με διαφορετικές παραμέτρους ενώ στις componentized εφαρμογές ο κώδικας από κόμβο σε κόμβο είναι διαφορετικός.

Η compute-intensive χρησιμοποιεί οριζόντιο διαχωρισμό ενώ η componentized κάθετο διαχωρισμό.

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

Οι συγγραφείς φαίνεται να δίνουν μεγάλο βάρος στο να μπορούμε με μικρό κόστος επικοινωνιών να προσθέτουμε εύκολα κόμβους στο δίκτυο (scalability) χωρίς την χρήση centralized points και τελικά να εξασφαλίσουμε απευθείας επικοινωνία ανάμεσα στους πελάτες. Προς το τέλος (Κεφ. 6) του κειμένου επιφέρουν στις επικοινωνίες και την παράμετρο της ασφάλειας στις επικοινωνίες αλλά δεν είναι ο κύριος στόχος όπως ζητά η ερώτηση.

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

Ο πλεονασμός είναι το κύριο χαρακτηριστικό για να διατηρούμε ένα σύστημα αξιόπιστο. Έτσι για παράδειγμα σε ένα σύστημα με file sharing ακολουθούμε την τακτική της δημιουργίας πολλών αντιτύπων (replication) ενώ αντίθετα σε ένα σύστημα με ανταλλαγή μηνυμάτων στα «χαμένα» μηνύματα επαναλαμβάνουμε την διαδικασία αποστολής ή επιλέγουμε πολλά και διαφορετικά μονοπάτια για την αποστολή.

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

Στο *centralized directory* θεωρητικά έχει αρκετά scalability προβλήματα και νομικά (!) προβλήματα. Επίσης οι κεντρικοί κόμβοι μπορεί να αποτελέσουν και αιτία κατάρρευσης ή υπερβολικής κίνησης δικτύου. Αντίθετα έχουν την ικανότητα να εντοπίζουν την «ιδανική» λύση-απάντηση στην αίτηση με το μικρότερο κόστος σε επίπεδο δικτύου (λιγότερα μηνύματα ανταλλάσσονται).

Με την χρήση *flooded requests* έχουμε έναν περιορισμό στο scalability που έχουν άμεσα σχέση με την τιμή του “*flooding steps*”, επίσης έχει υπερβολικό κόστος σε πλήθος μηνυμάτων. Τελικά γίνεται χρήση αυτών με την εφαρμογή τεχνικών super peers αλλά και cache για την μείωση των μηνυμάτων.

Στα *document routing models* είναι ένας πολύ αποδοτικός αλγόριθμος για μεγάλα δίκτυα (λίγα μηνύματα ανταλλάσσονται, γρήγορη εύρεση) αλλά υπάρχουν κάποια προβλήματα που έχουν να κάνουν με την μορφή των search queries αλλά μπορεί κάποιο υποδίκτυο να αποκοπεί από το υπόλοιπο δίκτυο.

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?

Το ότι επικοινωνούν απ’ ευθείας οι δύο εμπλεκόμενοι κόμβοι είναι σημαντικό διότι δεν απασχολούν άλλους τρίτους κόμβους με αποτέλεσμα την γραμμική αύξηση του κόστους επικοινωνίας αλλά και απασχόλησης των ενδιάμεσων κόμβων. Μειονέκτημα αποτελεί ότι σε ένα τέτοιο σύστημα δεν υπάρχει περίπτωση να μειώσουμε τον φόρτο του

κόμβου που έχει το δεδομένο που θέλουμε. Δηλαδή σε μία τέτοια υλοποίηση έχουμε hot spots που δύσκολα “ανακουφίζονται” με την χρήση directory.

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

Το Gnutella με βάση το κείμενο χρησιμοποιεί flooding μοντέλο που έχει σαν μειονέκτημα κάποια δεδομένα πολλές φορές να μην μπορούν να τα εντοπίσουν ένα μεγάλο πλήθος κόμβων. Αυτό εάν το δούμε από την αντίθετη πλευρά φαίνεται σαν κάποια δεδομένα να απομονώνονται. Επίσης πρόβλημα αποτελεί ότι για να εισέλθει κάποιος κόμβος στο δίκτυο θα πρέπει να γνωρίζει τουλάχιστον έναν κόμβο. (δεν είμαι σίγουρος με το τι εννοεί closure mechanism)

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

Η scalability φαίνεται να επηρεάζεται από διάφορους παράγοντες. Ένας από αυτούς είναι επιλογή αλγορίθμου ανάμεσα στους: centralized, flooding και document routing. Ένας άλλος παράγοντας είναι η χρήση dedicated servers στο δίκτυο ή όχι όπως επίσης και η χρήση super peers στο δίκτυο ενδέχεται να επηρεάσουν σημαντικά το δίκτυο. Βέβαια όλα τα παραπάνω ενδέχεται να ανατραπούν μελλοντικά όπου αναμένουμε να αυξηθεί ακόμη περισσότερο το bandwidth αλλά και η υπολογιστική ισχύ των κόμβων όποτε να αλλάξουν και τα μέτρα σύγκρισης.

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.

Για να έχουμε αξιοπιστία στην ανταλλαγή μηνυμάτων πιο κατάλληλη θα ήταν η: “delivery-based transient synchronous communication”

13. pg 17, 1st column, last par

*“The geographical distribution of the peers helps to reduce congestion on both peers and the network.”* Explain.

Έστω ένας κόμβος A έχει κάνει μία αναζήτηση και την απάντηση την έχουν οι B και Γ τότε με την γνώση της γεωγραφικής τοπολογίας μας βοηθά έμμεσα να εκτιμήσουμε

ποιος από τους κόμβους Β,Γ είναι πιο κοντά στον Α. Έτσι είμαστε σε θέση να εκτιμήσουμε για τον κόμβο Α ποιος κόμβος θα απασχολήσει λιγότερο το δίκτυό μας από τους Β και Γ (διασχίζοντας λιγότερα hops) ώστε να παραλάβει το δεδομένα και ταυτόχρονα να τα παραλάβει όσων το δυνατόν γρηγορότερα. Επίσης γνώση της γεωγραφικής τοπολογίας μας βοηθά να εκτιμήσουμε ποιοι κόμβοι να γίνουν αποδέκτες πολλών αιτήσεων οπότε να το εκμεταλλευτούμε για την μείωση της κίνησης π.χ. χρήση αντιτύπων.

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 cache είναι μία τεχνική η οποία αποσκοπεί στο να εξυπηρετήσει παρόμοιες αναζητήσεις με άμεσο σκοπό την μείωση του χρόνου αναζήτησης – απόκρισης ενώ έμμεσα προσπαθεί να μειώσει τα hops που θα διασχίσουμε ώσπου να φτάσουμε στην πηγή. Η χρήση του caching στο return path μειώνει όπως είναι φανερό τα hops της απάντησης και φαίνεται άμεσα να πετυχαίνει ακριβώς αυτό που επιθυμούμε. Τα μειονεκτήματα είναι αρκετά: όταν τα αρχεία είναι σχετικά μεγάλα τότε έχουμε πολύ μεγάλες μεταφορές δεδομένων. Ενδεχόμενα “bursts” μπορεί να προκαλέσουν αλληπάλληλες μετακινήσεις των δεδομένων από την πηγή με αποτέλεσμα την υπερ-απασχόλησή της.

15. What does the "power-law distribution of the p2p network" (pg 17) mean?

Χωρίς να είμαι σίγουρος είναι μία μέτρηση που έκανε κάποιος για τον μέσο αριθμό hops ανάμεσα σε δύο κόμβους που ανήκουν σε ένα p2p δίκτυο. Πιο συγκεκριμένα φαίνεται να υπάρχει μία κατανομή (distribution) που κάνει εκτίμηση του δείκτη αυτού.

16. Compare/relate the definition of distributed systems in sec 5.2 (pg 21) with sec 1.4 of the textbook.

Ο ορισμός του βιβλίου κάνει εμφανή διαχωρισμό στα multiprocessor και στα multicomputer συστήματα, όπως επίσης διαχωρίζει τα ετερογενή με τα ομογενή

συστήματα. Στο κείμενο στην παράγραφο 5.2 εξετάζει από πρακτικής μεριάς τι έχει γίνει σήμερα και σε ποια αντίστοιχη κατηγορία κατανεμημένου συστήματος το εντάσσει αλλά και τι σχέση έχει το υπό εξέταση σύστημα με τα p2p δίκτυα.

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

Στα collaborative p2p systems τα δεδομένα που ανταλλάσσονται αφορούν real time εφαρμογές και ενδέχεται ο αποστολέας να περιμένει από τον παραλήπτη την απάντησή του και ου το καθεξής. Αυτά τα συστήματα είναι ευαίσθητα σε τυχόν προβληματικές επικοινωνίες και να επηρεάσουν συνολικά την λειτουργία της εφαρμογής. Κάτι τέτοιο στα p2p που έχει να κάνει με file-sharing δεν είναι τόσο επικίνδυνο.