Γεωργόπουλος Άλκης A.M.:39
Κοντογιώργης Αναστάσιος A.M.: 43

Άσκηση 1

1. Read Section 1.3.1 of Coulouris et. al or Sec 11.1 of Tanenbaum et al. (on the web). Web can be considered as an example of a distributed system. At what extend does the web satisfy the four goals of distributed systems that were set in the lecture? In particular:

   Goal 1: connecting users and resources
   What are the types of resources shared? How does web address the two problems of sharing?

   Οι πόροι που συνήθως διαμοιράζονται μέσω του Web είναι δεδομένα σε διάφορες μορφές, όπως HTML, XML, GIF, JPG, RA, MP3, MPG, AVI, PDF, PS όπως και συνδυασμοί αυτών. Επίσης παρέχεται πρόσβαση σε δεδομένα τα οποία προέρχονται από κάποια βάση, όπως για παράδειγμα τραπεζικοί λογαριασμοί, μελωδίες για κινητά τηλέφωνα, ο καιρός ανά περιοχές, το χρηματιστήριο κ.α. Η επικοινωνία με τις βάσεις δεδομένων γίνεται μέσω CGI ή μέσω server side scripts (κώδικας JScript, VBScript, Java κτλ που εκτελείται στον server), ή, στην απλούστερη περίπτωση, με HTML αρχεία τα οποία παράγονται δυναμικά από τον server.

   Άλλες υπηρεσίες (πόροι) που προσφέρονται μέσω του Web περιλαμβάνουν το webmail, τις δυνατότητες streaming (όπως είναι τα live ραδιόφωνα μέσω Web), την εκτύπωση σε απομακρυσμένο εκτυπωτή, την αποστολή Fax, multimedia μηνυμάτων σε κινητά τηλέφωνα και πολλά άλλα. Τα περισσότερα από αυτά αντιστοιχούν σε μια απλή αίτηση μέσω μιας φόρμας.

   Το πρώτο πρόβλημα της διαμοίρασης πόρων είναι η ασφάλεια. Συνήθως η ταυτοποίηση του χρήστη γίνεται μέσω ενός κωδικού πρόσβασης, ο οποίος είναι γνωστός μόνο στον χρήστη. Τις περισσότερες φορές οι servers δεν έχουν αποθηκευμένο τον κωδικό, παρά μόνο το μη αναστρέψιμο κρυπτόγραμμά του, όπως αυτό προκύπτει με την εφαρμογή κάποιας μεθόδου σαν την γνωστή crypt του UNIX. Κάθε φορά που χρειάζεται να πιστοποιηθεί ο χρήστης του ζητείται ο κωδικός, ο οποίος στη συνέχεια κρυπτογραφείται και το αποτέλεσμα συγκρίνεται με την βάση κρυπτογραμμάτων του server.

   Υπάρχουν πολλές δυνατές απόπειρες για την εξαπάτηση αυτού του μηχανισμού, με πιο γνωστές την brute force επίθεση μέσω λεξικού και την εγκατάσταση σε κάποιο σημείο ενός προγράμματος, το οποίο θα υποκλέψει τον κωδικό του χρήστη πριν αυτός κωδικοποιηθεί. Για να αποφευχθεί το δεύτερο (καθώς και πολλά άλλα προβλήματα που προκαλούνται από μη ασφαλής συνδέσεις), έχει αναπτυχθεί το πρωτόκολλο SSL, το οποίο κάθεται πάνω από το TCP έτσι ώστε τελικά να προσφέρει https. Μετεξέλιξη του SSL είναι το TLS. Η χρήση αυτών των πρωτοκόλλων, σε συνδυασμό με τη χρήση ψηφιακών υπογραφών και μεθόδων κρυπτογράφησης όπως η 3DES καθιστούν αρκετά ασφαλή την επικοινωνία.

   Το δεύτερο πρόβλημα της διαμοίρασης πόρων είναι η ακούσια επικοινωνία. Εκτός από

τα διαφημιστικά banners ή τα pop-up παράθυρα, πολλές φορές και οι βάσεις δεδομένων που χρησιμοποιούνται κατακλύζονται από ανεπιθύμητη πληροφορία. Κλασσικό παράδειγμα το spamming, το οποίο μας επηρεάζει και στο Webmail. Επίσης στα forum και στα newsgroup παρατηρείται παρόμοια συμπεριφορά.

Το Web έχει λίγα να προσφέρει κατά του Unwanted Communication. Συνήθως κάποια προγράμματα στους servers (web, mail, news, κτλ) προσπαθούν να απομακρύνουν τα ανεπιθύμητα δεδομένα, αλλά αυτό εμπεριέχει και τον κίνδυνο να απομακρυνθεί κατά λάθος και κάποιο σημαντικό δεδομένο. Από την άλλη μεριά όμως ας μην ξεχνάμε ότι στο Web όλα βασίζονται στα requests του client, δηλαδή αν δεν επισκεφτούμε κάποια ιστοσελίδα αποκλείεται να μας έρθει απ' αυτήν ανεπιθύμητο περιεχόμενο.

Goal 2: For each type of transparency, explain the transparency degree offered by the web.

access transparency: εν γένει το web δεν ασχολείται με την προσπέλαση των δεδομένων, εννοώντας ότι για παράδειγμα ένα αρχείο html στο Unix θα έχει μόνο CRs, ενώ στα Windows θα έχει CR+LFs. Το διάβασμα και η αποστολή των δεδομένων αποτελούν ευθύνη του web server, ο οποίος κρύβει τις ιδιαιτερότητες της underlying αρχιτεκτονικής. Ευθύνη όμως του web browser είναι να μπορεί να καταλάβει αυτά τα δεδομένα, και επομένως ένας μεγάλος φόρτος πέφτει σ' αυτόν. Ακόμα και τα Unicode αρχεία σε άλλα συστήματα είναι big-endian και σε άλλα little-endian. Εξάλλου η πληθώρα των τύπων αρχείων που πρέπει να υποστηρίζει ένας browser καθιστούν αναγκαία την χρήση plug-ins. Αλλά το χειρότερο είναι ότι με τον πόλεμο Netscape - Microsoft, οι HTML / DHTML / CSS / JavaScript έχουν περάσει από 40 κύματα, και πλέον είναι αδύνατο να γραφεί μία ιστοσελίδα που θα εμφανίζεται παρόμοια σε όλους τους browsers. Ακόμα και η χρήση XSL στην πολυδιαφημιζόμενη XML δεν εγγυάται ομοιογενή εμφάνιση των παραγόμενων ιστοσελίδων. Και αν σε αυτά προσθέσουμε και τις διαφορετικές δυνατότητες του υλικού, όπως βάθος χρώματος, ανάλυση οθόνης, δείγματα του wavetable της κάρτας ήχου κτλ καταλαβαίνουμε ότι μία ιστοσελίδα θα φαίνεται ίδια σε ελάχιστα μόνο μηχανήματα. Άρα αν και το web κρύβει τις λεπτομέρειες της προσπέλασης της πληροφορίας, εντούτοις μειονεκτεί στην αναπαράστασή της. Γι' αυτό και έπιασε τόσο το Flash της Macromedia.

location transparency: στο web γενικώς δεν μας ενδιαφέρει που είναι τοποθετημένα τα δεδομένα που προσπελαύνουμε. Αν και τα επιθέματα των χωρών (.gr, .uk, .it κτλ) δείχνουν κατά κάποιο τρόπο την τοποθεσία, δεν είναι δεσμευτικό ότι το ip που θα αντιστοιχούν θα είναι εντός της χώρας. Φυσικά δηλώνεται η τοποθεσία (μονοπάτι) εντός του server, γι' αυτό και λέμε ότι τα URL είναι location dependent, αλλά κατά την γνώμη μου και τα URN που υποτίθεται ότι είναι location independent δεν προσφέρουν κάτι καλύτερο. Στην γενική περίπτωση μία κατηγοριοποίηση σε καταλόγους κάνει πιο ευκολομνημόνευτες τις διευθύνεις παρά τα νούμερα των εγγράφων. Εξάλλου τα δεδομένα τα οποία ζητάμε από κάποιο μονοπάτι του server δεν είναι αναγκαίο να βρίσκονται εκεί, υπάρχει η δυνατότητα να δηλώνονται εικονικά μονοπάτια και η πληροφορία να προέρχεται ακόμα και από διαφορετικούς υπολογιστές.

migration transparency: αν καταβάλλεται η αντίστοιχη προσπάθεια, είναι δυνατόν στο web να έχουμε πολύ καλή migration transparency. Για παράδειγμα, έστω ότι κάποιος client ζητάει την διεύθυνση http://www.alkisg.tk/kairos, η οποία υποτίθεται ότι είναι στο tokylaw και ας πούμε ότι αντιστοιχεί σε ένα πρόγραμμα CGI το οποίο επικοινωνεί με κάποια βάση και επιστρέφει κάποιες πληροφορίες για τον καιρό. Αν χρειαστεί να

μετακινηθεί ο server και να έρθει στην Ελλάδα, μπορεί ακόμα να κρατήσει το επίθεμα .tk. Και αν η βάση χρειαστεί να αναβαθμιστεί σε συνδυασμό PHP/MySQL, θα πρέπει μόνο να αλλάξει το default document του φακέλου kairos, το οποίο είναι μια εσωτερική και πλήρως διαφανής δήλωση στον server.

relocation transparency: σ' αυτό το web δεν τα πάει και τόσο καλά, συνήθως μια μετακίνηση ενός server απαιτεί το κλείσιμό του για κάποιες ώρες. Χρειάζεται τέλειος συγχρονισμός με τον HostMaster, έτσι ώστε η αντιστοίχιση ενός καινούργιου IP σε ένα ήδη υπάρχον domain να πάρει λίγο μόνο χρόνο. Εξάλλου στις dialup γραμμές οι clients κάθε φορά που συνδέονται αλλάζουν IP χωρίς καν να αλλάζουν την τοποθεσία τους!

replication transparency: γι' αυτό το σκοπό υπάρχουν οι server clusters. Είτε ένα front-end καθορίζει κάποιον από μια σειρά από web servers να εξυπηρετήσει τον client, ή μέσω ενός Switch μεταβιβάζεται η κλήση σε έναν distributor κατά την πρώτη φορά και κατ' ευθείαν στον κατάλληλο web server στις επόμενες. Εξάλλου τα ίδια τα δεδομένα μπορεί να προκύπτουν από διαφορετικούς υπολογιστές με πολλές συστοιχίες δίσκων.

concurrency transparency: επειδή η φύση του web είναι mostly read, οι παράλληλες χρήσεις συνήθως δεν παρουσιάζουν πρόβλημα. Εξάλλου κατά την αποστολή δεδομένων στον server (τα οποία μπορεί να αποθηκεύονται σε κάποια βάση) οι αιτήσεις επεξεργάζονται από τον server, και άρα στον client είναι τελείως transparent. Στο πρωτόκολλο WebDAV βέβαια είναι δυνατή η απ' ευθείας εγγραφή δεδομένων, και επομένως προσφέρεται ένας μηχανισμός locking, ο οποίος όμως έχει κάποια προβληματάκια όταν χάνεται η σύνδεση με κάποιον client ο οποίος εκείνη τη στιγμή έχει το κουπόνι εγγραφής στη βάση.

failure transparency: σε κάποια σημεία είναι διαφανές (π.χ. αν χαλάσει ο πρωτεύον DNS server αντικαθίσταται αυτόματα από τον δευτερεύων), αλλά σε κάποια άλλα, όπως η μη εξυπηρέτηση μιας αίτησης, αδιαφορεί τελείως και δεν γίνεται καμία προσπάθεια recovery.

persistent transparency: είναι τελείως transparent, είναι εσωτερική δουλειά του web server η διαχείριση της μνήμης.

Goal 3: To what extent is web an open system? Briefly mention the protocols used by the web. Give an example of a web policy and a web mechanism (other than caching).

Το web είναι πολύ ανοιχτό σύστημα, και σ' αυτό βοηθάνε οι HTML, XML, τα plug-ins των browser και τα διάφορα scripts / applets. Αν και όλα αυτά προσφέρουν μεγάλες δυνατότητες ευελιξίας και επέκτασης, είναι μερικές φορές που δεν επαρκούν, όπως η εκτέλεση κώδικα στον client. Τα διάφορα scripts έχουν πολύ περιορισμένες δυνατότητες, η Java ασχολείται ελάχιστα με το interface, τα Java Beans δεν είναι ευρέως αποδεκτά, το flash δεν έχει δυνατότητες αλληλεπίδρασης με το σύστημα και μέσα σ' όλα αυτά η Microsoft λάνσαρε και την C# για να δέσει το γλυκό.
Το πρωτόκολλο μεταφοράς που χρησιμοποιείται στο web είναι το http. Επίσης υποστηρίζεται και web-based ftp. Για ασφάλεια στις επικοινωνίες χρησιμοποιούνται τα SSL και TLS, ενώ για caching υπάρχουν και pull-based και invalidation protocols.
Ένας μηχανισμός του web είναι τα cookies, τα οποία είναι μικρά αρχεία κειμένου που επιτρέπουν σε ένα script να αποθηκεύσει μια μικρή πληροφορία στον client. Η πολιτική του client θα καθορίσει αν τα cookies από τις διάφορες ιστοσελίδες είναι επιτρεπτά ή όχι.

Goal 4: Is web scalable (in each of the three dimensions)? How is scalability at each dimension achieved, that is, which are the scalability techniques used along each dimension?

size: ναι, είναι scalable. Δεν υπάρχει περιορισμός στα αρχεία που μπορεί να περιέχει ένας server, ούτε στον αριθμό των χρηστών. Οι μόνοι περιορισμοί είναι η ταχύτητα του server (το οποίο αντιμετωπίζεται εν μέρει με τους server clusters), οι γραμμές των ISPs, και τα διαθέσιμα IP, τα οποία έχουν αυξηθεί πάρα πολύ με το Internet 2.

geographical: δεν υπάρχει γεωγραφικός περιορισμός, αρκεί να υπάρχει φυσικό κανάλι επικοινωνίας.

administrative: το DNS για πιο εύκολη διαχείριση έχει χωριστεί σε μερικούς δεκάδες servers (.com, .biz, .gov, .info, .gr, .uk, .it, ...), αλλά αυτό δεν ενδιαφέρει το web. Δεν υπάρχει περιορισμός στον αριθμό των domain, ούτε άμεση συνεργασία μεταξύ τους ώστε να υπάρχουν τα προβλήματα ασφαλείας που προκύπτουν κατά την σύνδεση δύο τοπικών δικτύων. Τα δικαιώματα πρόσβασης στους web servers καθορίζονται από τις πολιτικές που εφαρμόζουν οι webmasters, καθώς και από τα δικαιώματα των χρηστών στο underlying σύστημα αρχείων (στις περιπτώσεις που γίνεται πρόσβαση με username και κωδικό).