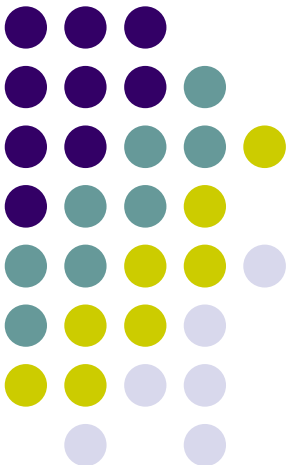
Προγραμματισμός Ιστοσελίδων (Web Design)

Πανεπιστήμιο Πελοποννήσου
Τμήμα Επιστήμης & Τεχνολογίας Τηλεπικοινωνιών

Ευάγγελος Α. Κοσμάτος

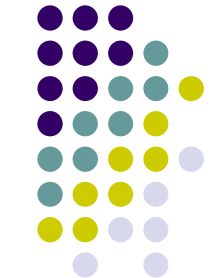


Ορισμός Web

- Ο Παγκόσμιος Ιστός (World Wide Web) ή απλά Web αποτελεί τον κόσμο των πληροφοριών που είναι προσβάσιμες από διασυνδεσεμένους υπολογιστές

- Το Διαδίκτυο (Internet) διαφέρει από το Web
  - είναι ένα δίκτυο από υπολογιστές
    - κάθε υπολογιστής δεν είναι απαραίτητο να ενεργεί ως πελάτης ή εξυπηρετητής.

Στοιχεία του Web



- Το Web αποτελείται από τρία βασικά συστατικά στοιχεία:
  - Υποδομή ονοματολογίας (Uniform Resource Identifier - URI)
  - Γλώσσα δημιουργίας εγγράφων (Hypertext Markup Language - HTML)
  - Πρωτόκολλο ανταλλαγής μηνυμάτων (Hypertext Transfer Protocol - HTTP)

Uniform Resource Identifier - URI

- Η πρόσβαση και η διαχείρηση πληροφοριών και πόρων κατανεμημένων στο Web απαιτεί έναν τρόπο προσδιορισμού της ταυτότητάς τους.
- Το URI αποτελεί ένα παγκόσμιο μηχανισμό ονοματοδοσίας για τον προσδιορισμό της ταυτότητας πόρων στο Web
  - ανεξαρτησία από τοποθεσία
- Το URI είναι ένας δείκτης προς ένα μαύρο κουτί στο οποίο αιτήσεις μπορούν να προκαλέσουν διαφορετικές αποκρίσεις σε διαφορετικές χρονικές στιγμές
  - Αίτηση: απλές διαδικασίες φόρτωσης, αλλαγής ή διαγραφής πληροφοριών.
- Παράδειγμα URI:
  - http:// www.uop.gr/coolpic.gif

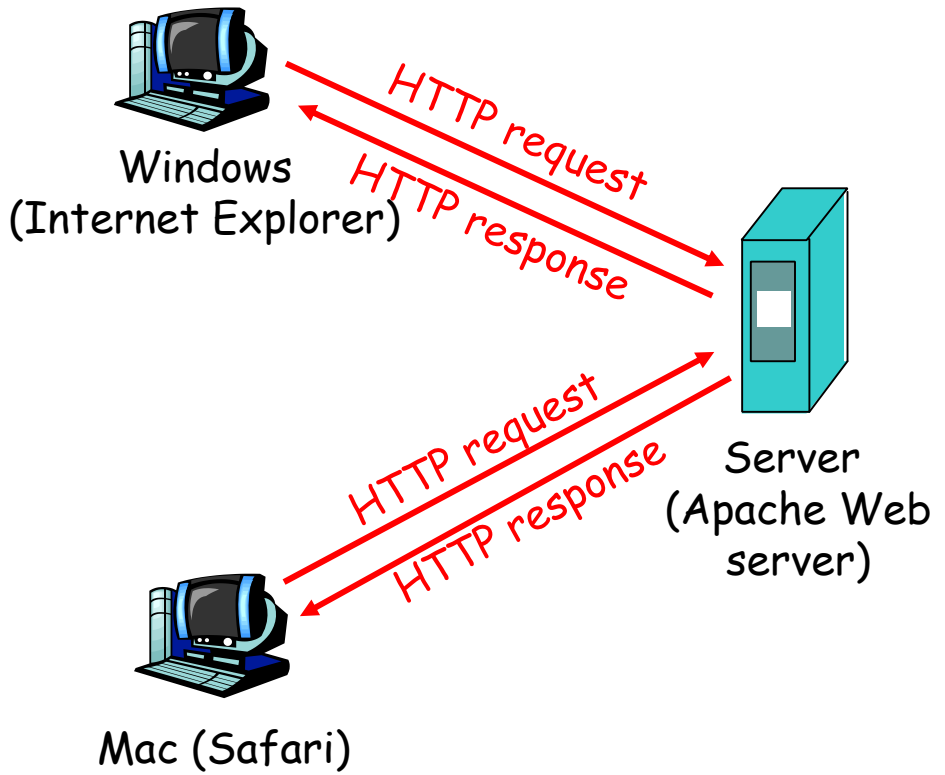
Hypertext Markup Language - HTML

- Η γλώσσα HTML παρέχει έναν τρόπο παρουσίασης υπερκειμένων (hypertexts) σε μορφή ASCII
  - Ένα HTML αρχείο είναι ένα αρχείο κειμένου που περιέχει **markup tags**
  - Τα tags λένε στον Web Browser πώς να παρουσιάσει το περιεχόμενο της σελίδας

Hypertext Transfer Protocol - HTTP

HTTP: Hyper-Text Transfer Protocol

- Πρωτόκολλο επιπέδου εφαρμογής
- μοντέλο πελάτη/εξυπηρετητή (client/server)
  - *πελάτης (client):* **browser** που αιτεί, λαμβάνει και παρουσιάζει αντικείμενα από το Web
  - *εξυπηρετητής (server):* **Web server** που στέλνει αντικείμενα σε απάντηση των αιτήσεων



- HTTP 1.0: RFC 1945
- HTTP 1.1: RFC 2068

HTTP

Χρήση TCP:

- Ο client εκκινεί ένα TCP connection (δημιουργεί και συνδέεται δηλαδή με socket) στον server, στο port 80
- Ο server δέχεται το TCP connection του client
- Μηνύματα HTTP (application-layer protocol μηνύματα) ανταλλάζονται μεταξύ του browser (HTTP client) και του Web server (HTTP server)
- Τέλος, κλείνει το TCP connection

Το HTTP είναι “stateless”

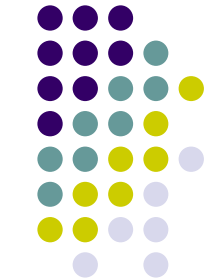
- Ο server δεν συγκρατεί πληροφορία κατάστασης για τα παλαιότερα requests του client (απλή υλοποίηση)

αντιθέτως

Τα πρωτόκολλα εφαρμογής που διατηρούν πληροφορία κατάστασης είναι σύνθετα!

- Πρέπει να διατηρείται η παλιά κατάσταση
- Αν ένα από τα δύο μέρη (server ή client) αποκοπεί, η κατάσταση των δύο μερών πρέπει να ανακτηθεί απ΄ την αρχή σε τέτοιου είδους πρωτόκολλα

HTTP συνδέσεις



Non-persistent HTTP (μη επίμονες) συνδέσεις

- Ένα το πολύ αντικείμενο μπορεί να σταλεί πάνω από μία TCP σύνδεση μεταξύ ενός client και ενός server
- Το HTTP/v.1.0 χρησιμοποιεί non-persistent HTTP

Persistent HTTP (επίμονες) συνδέσεις

- Πολλαπλά αντικείμενα μπορούν να σταλούν πάνω από μία TCP σύνδεση μεταξύ ενός client και ενός server
- Το HTTP/v.1.1 χρησιμοποιεί persistent connections σε default mode

Non-persistent HTTP

Έστω ότι ο χρήστης επιθυμεί το URL
`www.uop.gr/index.html`

(έστω επίσης ότι το URL περιέχει text και references σε 5 jpeg εικόνες)

1a. Ο HTTP client εκκινεί ένα TCP connection στον HTTP server (1 process στο socket) www.uop.gr στο port 80

1b. Ο HTTP server www.uop.gr βρίσκεται σε κατάσταση αναμονής περιμένοντας ένα TCP connection στο port 80. "Δέχεται" την αίτηση για σύνδεση (connection) και ειδοποιεί τον client

2. Ο HTTP client στέλνει ένα HTTP *request message* (που περιέχει το URL) στο socket του TCP connection. Το message υποδεικνύει ότι ο client επιθυμεί το object index.html

3. Ο HTTP server λαμβάνει το request message, συνθέτει το *response message* που περιέχει το ζητούμενο αντικείμενο (object) και στέλνει το message αυτό στο socket

time

Non-persistent HTTP

4b. Ο HTTP client λαμβάνει το μήνυμα απόκρισης (response message) που περιέχει το html αρχείο και το εμφανίζει.

Κατά την επεξεργασία του html file, βρίσκει 5 σχετιζόμενα αντικείμενα τύπου jpeg

4a. Ο HTTP server κλείνει το TCP connection

time

5. Τα βήματα 1-4 επαναλαμβάνονται για κάθε ένα από τα 5 jpeg objects

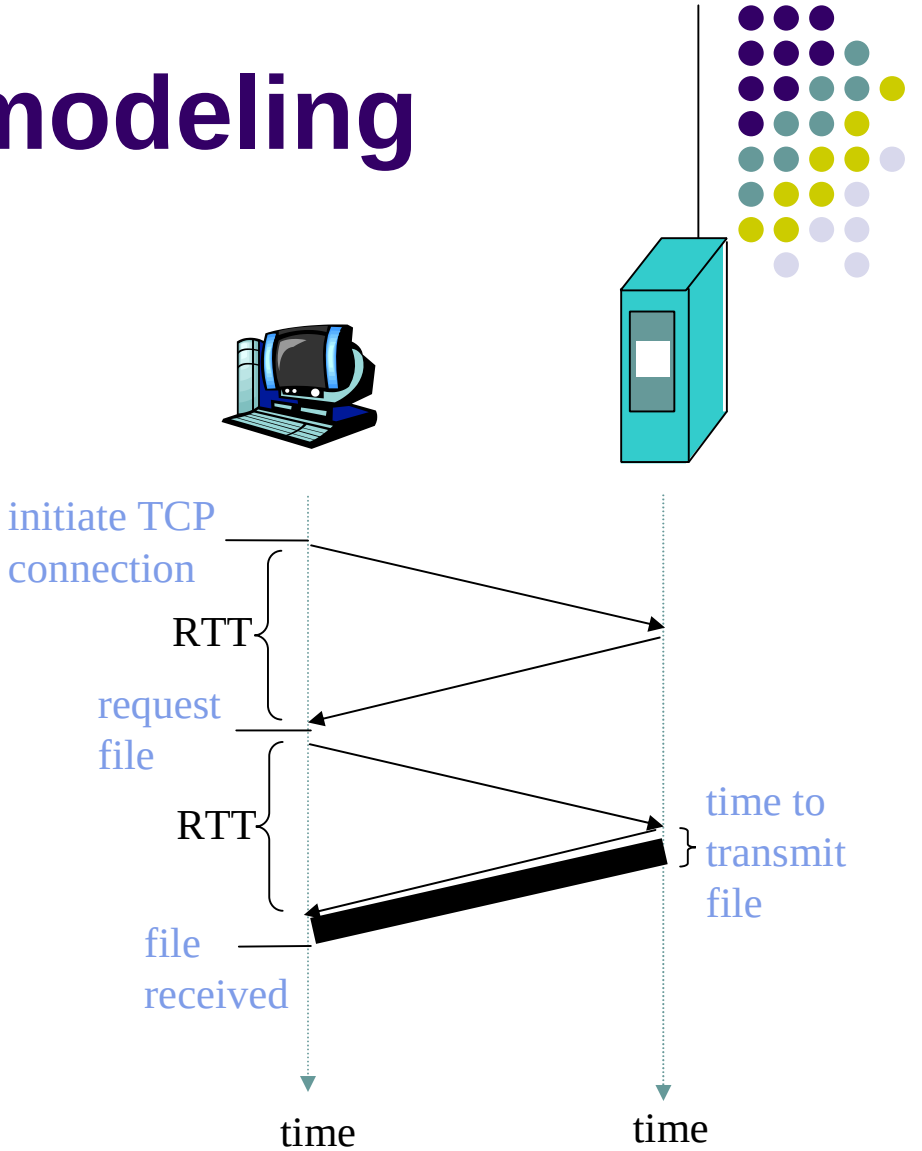
Response time modeling

RTT (Round Trip Time): ο χρόνος που απαιτείται για την αποστολή ενός μικρού πακέτου για να διανύσει την απόσταση από τον client στον server και πίσω

Response time:

- Ένα RTT για την έναρξη του TCP connection
- Ένα RTT για την αποστολή του HTTP request έως και να ληφθούν τα πρώτα bytes του HTTP response
- file transmission time

Total = 2RTT+Transmit Time



Persistent HTTP

Προβλήματα Non-persistent HTTP :

- Απαιτεί 2 RTTs ανά ένα αντικείμενο
- Το λειτουργικό σύστημα πρέπει να απασχολείται συνεχώς για τον εντοπισμό resources για κάθε TCP σύνδεση (αφού κλείνει η σύνδεση)
- Όμως οι browsers συχνά ανοίγουν παράλληλα TCP connections προς τα διάφορα αντικείμενα

Persistent HTTP

- Ο server αφήνει ανοιχτή τη σύνδεση αφού στείλει το response message
- Έτσι τα επακόλουθα HTTP messages μεταξύ των ίδιων client/server ανταλλάσσονται από την ίδια σύνδεση

Persistent χωρίς pipelining:

- Ο client μπορεί να κάνει νέο request αν και μόνο αν έχει λάβει το προηγούμενο response
- Synchronous
- Χρονικά απαιτείται ένα RTT για κάθε αντικείμενο

Persistent με pipelining:

- Default στο HTTP/v.1.1
- Ο client μπορεί να στείλει νέα requests για νέα objects, χωρίς να έχει λάβει ακόμα απάντηση για προηγούμενα requests
- Asynchronous
- Θεωρητικά μπορεί χρονικά σε ένα RTT να αιτηθεί (να στείλει requests) για όλα τα αντικείμενα

HTTP request message

- Διακρίνουμε HTTP messages δύο τύπων:

request & response

- HTTP request message:
  - ASCII (human-readable format)

request line (GET, POST, HEAD commands)

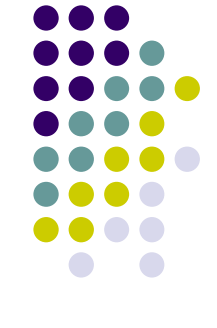
header lines

Carriage return, line feed indicates end of message

```
GET /mylab/index.html HTTP/1.1
Host: www.uop.gr
User-agent: Mozilla/1.1
Connection: close
Accept: text/html,image/jpg,image/gif
Accept-language:fr
```

(extra carriage return, line feed)

Μηνύματα Αίτησης HTTP



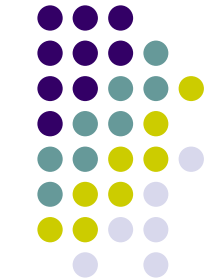
HTTP/1.0

- GET
  - Αίτηση μεταφοράς αρχείου
- POST
  - Για αποστολή πληροφορίας από την πλευρά του client (π.χ. συμπλήρωση στοιχείων ηλεκτρονικής φόρμας)
- HEAD
  - Παραπλήσιο με το μήνυμα GET, με τη σημαντική όμως διαφορά ότι άπαξ και ο client στείλει μήνυμα HEAD, ο server ΔΕΝ ΠΡΕΠΕΙ να περιέχει ΤΙΠΟΤΑ μέσα στο message body του response μηνύματος (χρησιμοποιείται για testing ή από τους proxies για τον έλεγχο ανάγκης ανανέωσης τοπικά αποθηκευμένων σελίδων)

HTTP/1.1

- GET, POST, HEAD
- PUT
  - Κάνει upload το αρχείο που μεταφέρει στο entity body του μηνύματος, στο path που αναφέρει το URL πεδίο του μηνύματος
- DELETE
  - Διαγράφει το αρχείο που περιγράφεται στο URL πεδίο του μηνύματος

GET request



- Η πιο συχνά χρησιμοποιούμενη αίτηση
- Χρησιμοποιείται για την αίτηση μεταφοράς αρχείου από τον Server ή για την εκτέλεση προγράμματος σε αυτόν
- Το URI περιέχει τον πόρο που ζητήθηκε
- Δεν περιέχει κάποιο μήνυμα αλλά το URI μπορεί να περιέχει παραμέτρους προς τον συγκεκριμένο πόρο του Web Server

http://www.uop.gr/index.php?option=cources_content&task=view&id=36&Itemid=67

Οι παράμετροι ξεκινούν μετά το ?

HTTP response message

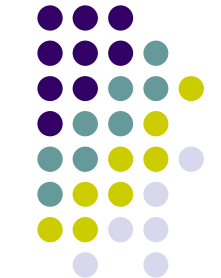
status line (protocol, status code, status phrase)

header lines

data, e.g., requested HTML file

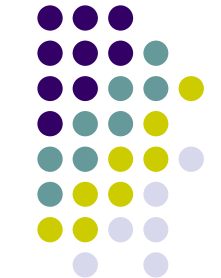
```
HTTP/1.1 200 OK
Connection close
Date: Thu, 22 Dec 2008 13:02:15 GMT
Server: Apache/1.3.0 (Unix)
Last-Modified: Mon, 22 Jun 2004 ...
Content-Length: 6821
Content-Type: text/html

data data data data data ...
```

Περιεχόμενο στο Web



- Κάθε πόρος στο Web μπορεί να είναι διαθέσιμος σε διαφορετική μορφή:
  - HTML
  - XML
- Κάθε πόρος μπορεί να είναι:
  - ένα στατικό αρχείο σε κάποιον εξυπηρετητή
  - να δημιουργείται δυναμικά κατά την στιγμή του request

Δομικά στοιχεία του Web

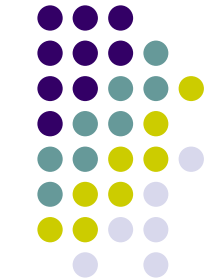


- Πελάτης/Αντιπρόσωπος (Web Client/User agent)
  - Ο πελάτης είναι ένας browser που δημιουργεί αιτήσεις για λογαριασμό ενός χρήστη
  - Παρουσιάζει μια Web σελίδα
  - Αποθηκεύει τα bookmarks του χρήστη
- Πληρεξούσιος (Proxy)
  - Ο proxy αποτελεί έναν ενδιάμεσο μεταξύ πελάτη και εξυπηρετητή και εκτελεί διάφορες λειτουργίες
    - φιλτράρισμα των αιτήσεων
    - παροχή ανωνυμίας
    - προσωρινή αποθήκευση (caching) δημοφιλών πόρων

Δομικά στοιχεία του Web

- Εξυπηρετητής (Server)
  - Ο server δέχεται τις αιτήσεις του πελάτη και δημιουργεί αποκρίσεις
  - Δίνει εντολές στον πελάτη να διατηρήσει την σύνδεση του με τον Server για μια σειρά από αιτήσεις και αποκρίσεις
    - αποθήκευση cookie

Πελάτης (Web Client)

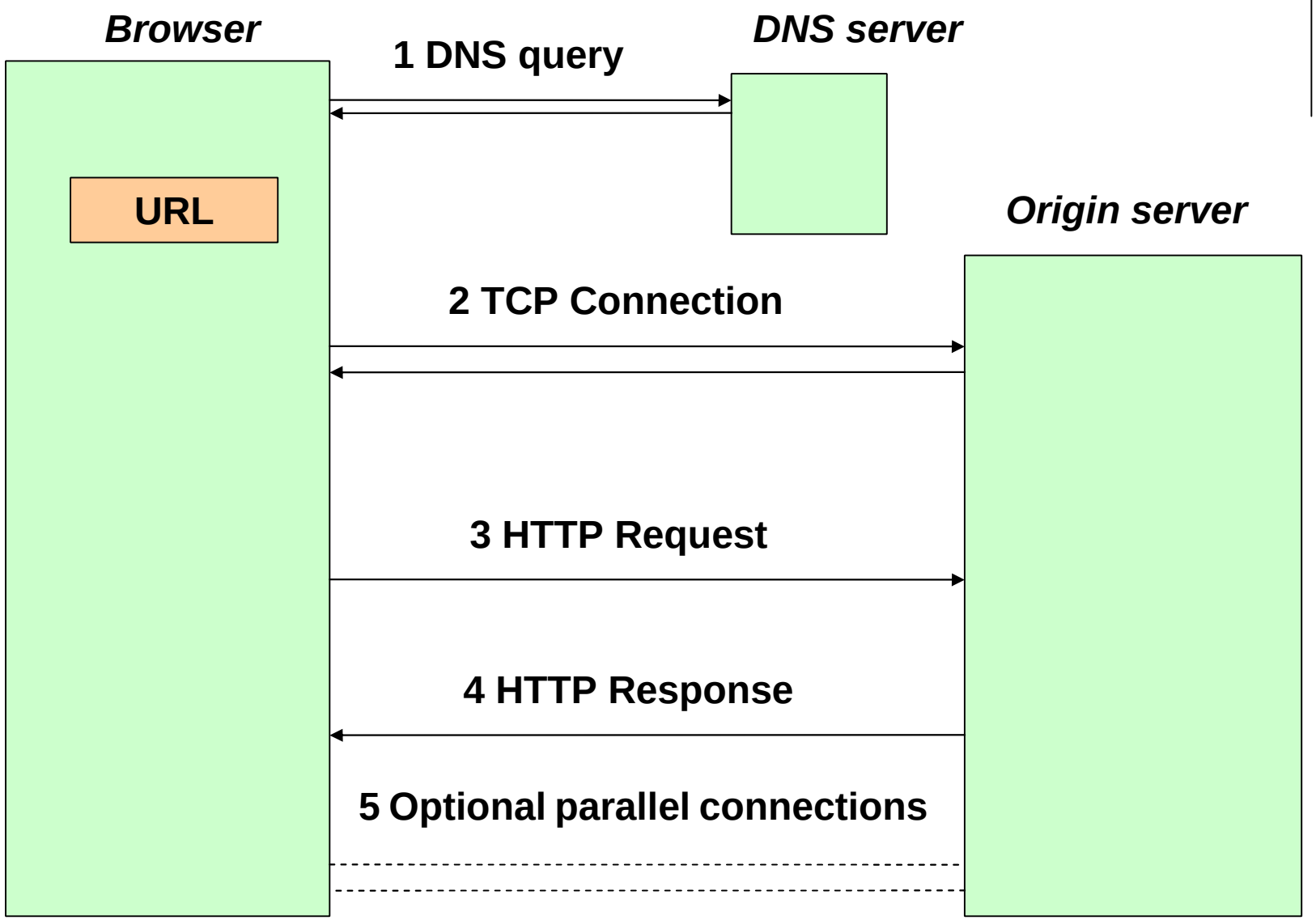


- Τρεις διαφορετικοί τύποι Web clients:
  - Browser
    - δημιουργεί μια Web request
  - Spider
    - Ο spider είναι ένα πρόγραμμα που δεν εκκινείτε από κάποιον χρήστη
    - Τα Spiders είναι απαραίτητα για την αναζήτηση και ευρετηριοποίηση των πόρων
  - Agent
    - Ο agent είναι ένα πρόγραμμα που τρέχει για λογαριασμό των χρηστών σε συγκεκριμένες εφαρμογές

Browser

- Είναι ο ποιος κοινός τύπος Web client.
- Μια σύνοδος σε ένα browser αποτελείται από μια σειρά από requests που στέλνονται από το χρήστη σε απάντηση των responses που λαμβάνονται σε κάθε χρονική στιγμή.
- Το βασικό πλεονέκτημα των μοντέρνων browser είναι η ικανότητά τους να μεταπηδούν ταχύτατα από πόρο σε πόρο

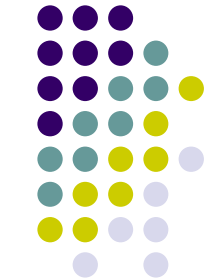
Browser


Browser
DNS server
1 DNS query
URL
Origin server
2 TCP Connection
3 HTTP Request
4 HTTP Response
5 Optional parallel connections

Browser Caching

- Το browser caching χρησιμοποιεί δύο περιοχές του υπολογιστή:
  - Ένα τμήμα της μνήμης (της τρέχουσας διαδικασίας)
  - Ένα τμήμα του σκληρού δίσκου
- Επικαιροποίηση της Cache
  - Ελέγχεται το αντίγραφο στην cache με το τρέχον αντίγραφο στο Server.
    - Αν το αντίγραφο στην cache είναι παλαιότερο τότε ανανεώνεται από τη νέα έκδοση

Browser Caching



- Συνοχή στην Cache
  - Έλεγχος για το αν το αντίγραφο στην cache είναι πρόσφατο
    - Ισχυρός
      - Ανανέωση του αντιγράφου της cache κάθε φορά που δημιουργείται αίτηση για το συγκεκριμένο πόρο.
    - Ασθενής
      - Περιοδική ανανέωση του αντιγράφου της cache
        - Σταθερά διαστήματα
        - Ανάλογα με τα χαρακτηριστικά των πόρων (μέγεθος, τύπος περιεχομένου, κτλ.)

Παραμετροποίηση Browser

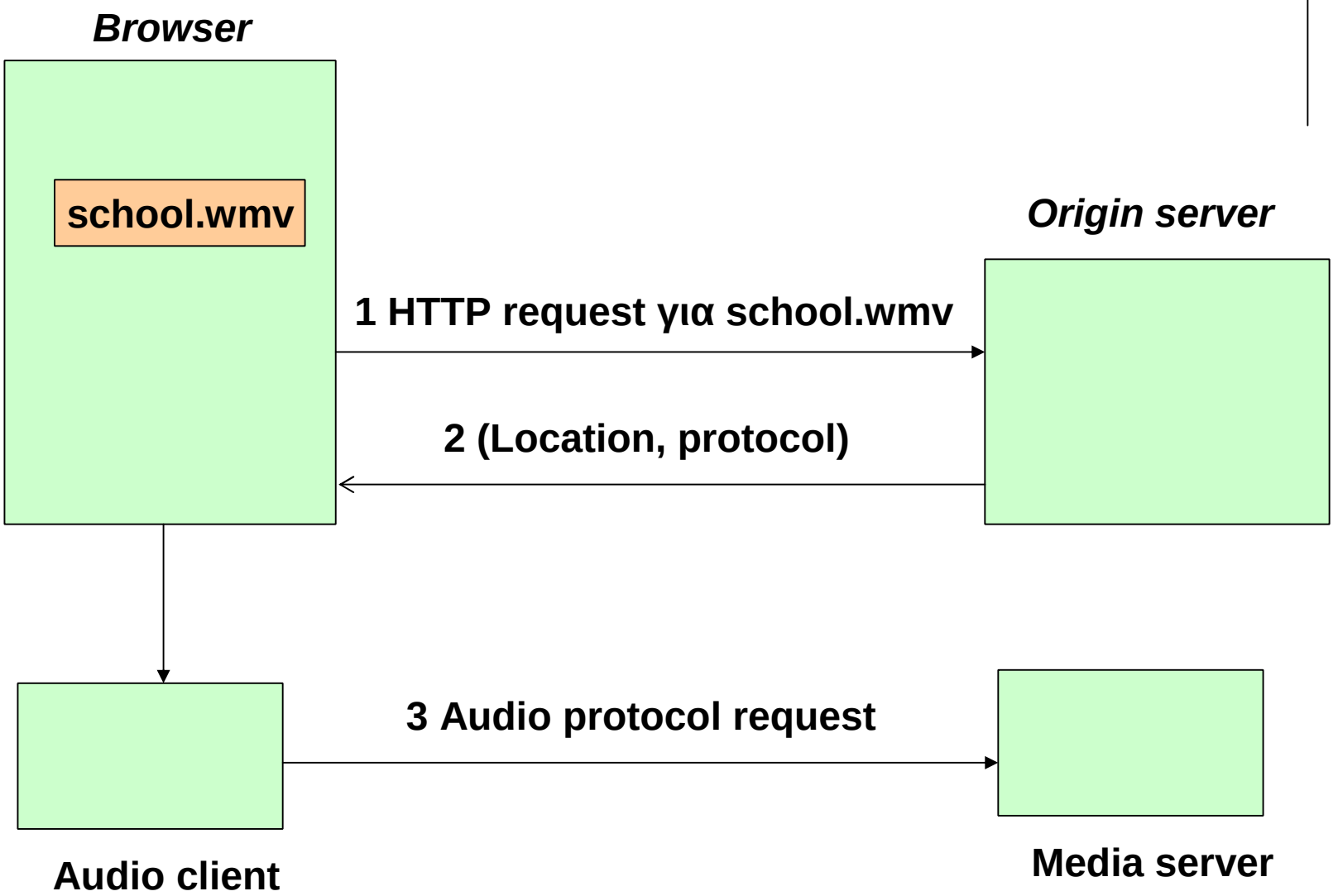
- Ο browser μπορεί να παραμετροποιηθεί από τον χρήστη
  - Καθορίσει πως θα δρομολογούνται οι αιτήσεις
    - Μέσω πληρεξούσιου (proxy)
    - Μέσω κάποιου άλλου ενδιάμεσου
    - Κατευθείαν στον Server
  - Καθορίσει πως θα λαμβάνονται οι αποκρίσεις
    - Εμφάνιση των στοιχείων
    - Υποστήριξη γλωσσών
- Ο browser μπορεί να χρησιμοποιήσει διάφορα βοηθητικά προγράμματα για να χειριστεί τις αποκρίσεις
  - παραμετροποίηση της επιλογής και τις διαχείρησης των βοηθητικών εφαρμογών

Παραμετροποίηση Browser

Content type	MIME type	Helper application
Zip compressed data	application/x-zip-compressed	gunzip/WINZIP32
PostScript document	application/postscript	ghostview/GSVIEW32
Word document	application/msword	catdoc/WINWORD
PDF document	application/pdf	acroread/ACRORD32
audio/video	video/x-mpeg-2	raplayer/MPLAYER2

- Τύποι περιεχομένου και βοηθητικά προγράμματα που καλούνται για την παρουσίασή τους

Παραμετροποίηση Browser



Παραμετροποίηση Browser

- Plug-in
  - Κομμάτι κώδικα που χρησιμοποιείται για να ερμηνεύσει δεδομένα κάποιου τύπου
    - Shockwave: χρήση του plug-in για την εμφάνιση animations
- Η ικανότητα του browser να υποστηρίζει διαφορετικά πρωτόκολλα
  - Ενιαίο front end για ένα σύνολο από εφαρμογές.

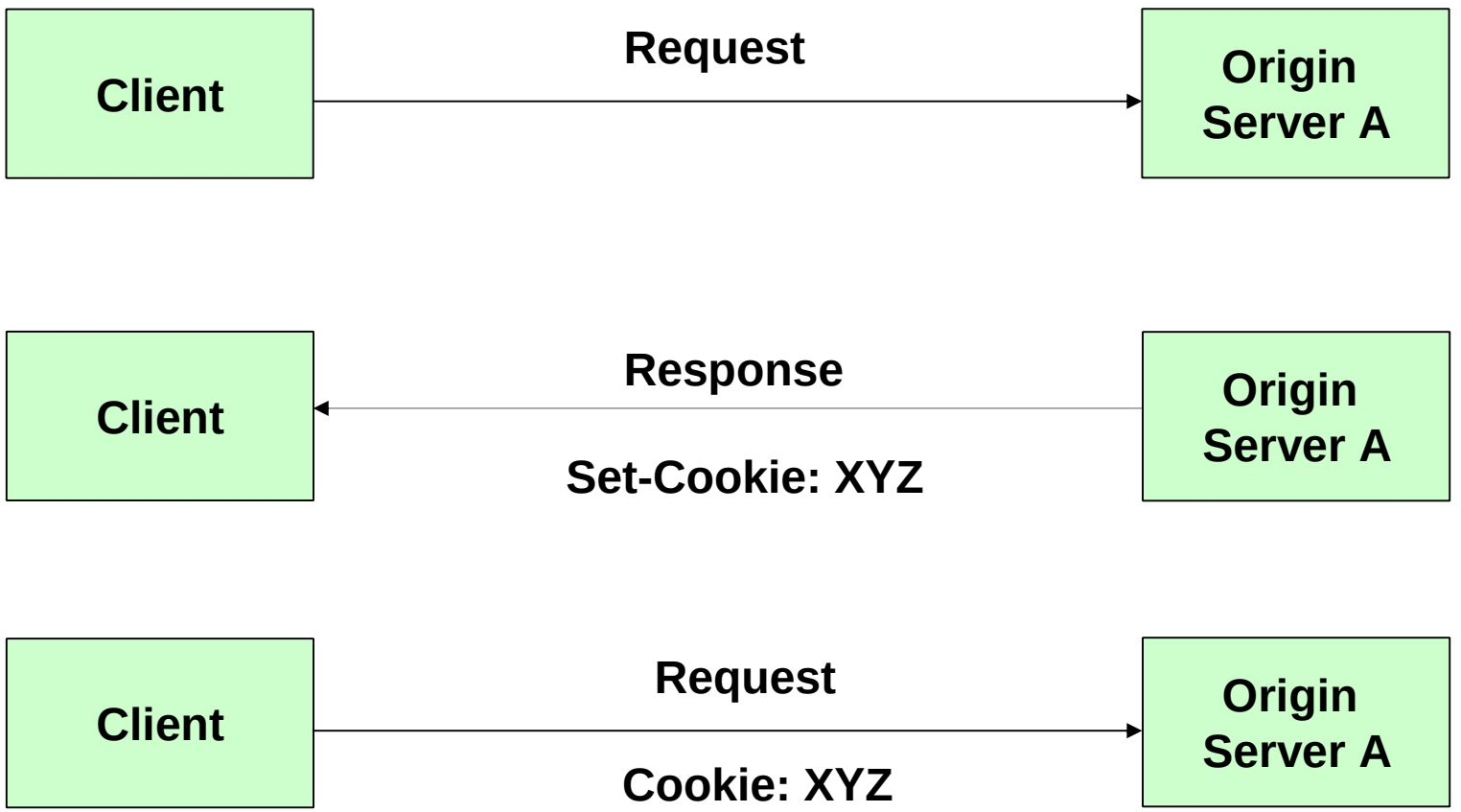
Cookies

- Ένας τρόπος να διαχειριστούμε state στο πρωτόκολλο HTTP.
- Ένα μικρό κομμάτι δηλώσεων
  - Στέλνεται από τον Server μιας ιστοσελίδας
  - Αποθηκεύεται στο μηχάνημα του χρήστη για λογαριασμό του Server
- π.χ. «κάρτα αγορών».
  - τι έχουμε ήδη επιλέξει για να αγοράσουμε
  - τι αγοράσαμε κατά την τελευταία μας επίσκεψη

Cookies

- Την επόμενη φορά που ο χρήστης θα επισκεφθεί την ιστοσελίδα, ο browser θα περιλάβει στην επικεφαλίδα του request την πληροφορία στο cookie.
- Χρησιμοποιούνται σαν δείκτης στην βάση δεδομένων του server για να αναγνωρίσουν την κατάσταση του χρήστη
- Μπορεί να έχουν μέγεθος μέχρι 4KB.
- Κάθε browser μπορεί να υποστηρίξει μέχρι 20 cookies ανά server.

Cookies



Ανταλλαγή πληροφοριών μεταξύ πελάτη εξυπηρετητή με τη βοήθεια των cookies

Web Server

- Ο Web server είναι ένα πρόγραμμα που δημιουργεί και μεταδίδει responses σε απόκριση requests από πελάτες.
- Ο χειρισμός των requests αποτελείται από τα βήματα:
  - ανάλυση του request μηνύματος
  - ελέγχει αν το request είναι εξουσιοδοτημένο να ζητήσει το συγκεκριμένο πόρο
  - συνδέει το URL που περιέχεται στο request με το όνομα του ζητούμενου αρχείου
  - κατασκευάζει το μήνυμα response
  - μεταδίδει το μήνυμα response στον πελάτη

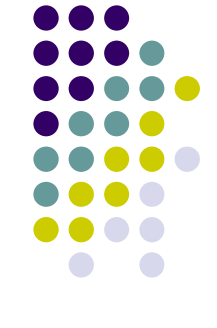
Web Server

- Ο server μπορεί να δημιουργήσει το μήνυμα response με διαφορετικούς τρόπους:
  - Ο server απλά προσπελαύνει το αρχείο που είναι συσχετισμένο με το URL και επιστρέφει το περιεχόμενό του στον πελάτη
  - Ο server μπορεί να καλέσει ένα script που επικοινωνεί με άλλους servers ή βάσεις δεδομένων για να κατασκευάσει το μήνυμα response.

Έλεγχος πρόσβασης

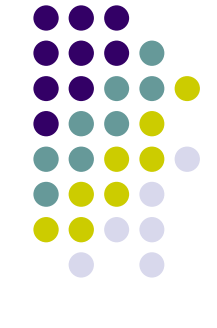
- Ο Web server μπορεί να καθορίσει ποιοι χρήστες έχουν πρόσβαση σε ποιους πόρους.
- Ο έλεγχος πρόσβασης απαιτεί τον συνδυασμό των διαδικασιών πιστοποίησης και εξουσιοδότησης
  - Η διαδικασία Πιστοποίησης (Authentication) αναγνωρίσει ποιος χρήστης έχει δημιουργήσει το request.
  - Η διαδικασία Εξουσιοδότησης (Authorization) αποφασίζει ποιοι χρήστες έχουν πρόσβαση σε ένα συγκεκριμένο πόρο.

Πιστοποίηση



- Τα περισσότερα συστήματα client-server υλοποιούν την πιστοποίηση μέσω ενός ζεύγους username και password.
- Ο Web server πρέπει να εκτελέσει την διαδικασία πιστοποίησης σε κάθε request για έναν πόρο που έχει περιορισμούς πρόσβασης
- Ο server επιστρέφει ένα HTTP response που δηλώνει ότι η συγκεκριμένη αίτηση απαιτεί πιστοποίηση
  - Το μήνυμα response δηλώνει τι είδους πιστοποίηση απαιτείται.
  - Το μήνυμα response αναγνωρίζει τον χώρο επιρροής της πιστοποίησης (realm)
    - οι πόροι στους οποίους μπορεί να έχει πρόσβαση ο συγκεκριμένος χρήστης

Εξουσιοδότηση



- Για να ελέγχει την πρόσβαση στους πόρους, ο server πρέπει να υιοθετήσει μια πολιτική εξουσιοδότησης.
- Μια πολιτική εξουσιοδότησης αποτελείται από μια λίστα ελέγχου πρόσβασης
  - αναφέρει τους χρήστης που έχουν ή δεν έχουν πρόσβαση στον συγκεκριμένο πόρο
- Ο server μπορεί να επιτρέπει ή να απαγορεύει την πρόσβαση χρησιμοποιώντας άλλες πληροφορίες εκτός από το username
  - την διεύθυνση IP του πελάτη
  - το υποδίκτυο στο οποίο ανήκει ο πελάτης.
- Η πιστοποίηση HTTP requests μπορεί να δημιουργήσει τεράστιο φόρτο στον Web server.

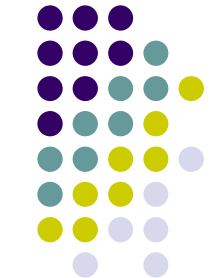
Δημιουργία Cookies

- Τα Cookies δημιουργούνται, χρησιμοποιούνται και τροποποιούνται από scripts που εκτελούνται για να δημιουργήσουν δυναμικά responses και όχι από τον Web server.
- Ο Server δίνει εντολή στον browser να περιλαμβάνει κάποιο cookie σε κάθε HTTP request.
  - Αν ένα request δεν περιλαμβάνει το cookie, τότε το script δημιουργεί ένα νέο cookie και περιλαμβάνει το cookie στην επικεφαλίδα του μηνύματος response.

    Set-Cookie: Customer="user17"; Version="1"; Path="/book"
  - Το επόμενο μήνυμα request από το χρήστη θα περιλαμβάνει το cookie

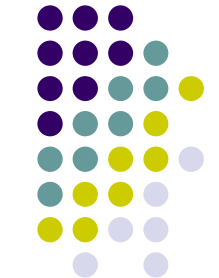
    Cookie: Customer="user17"; Version="1"; Path="/book"

Χρήση Cookies



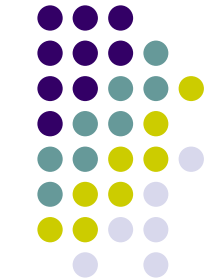
- Ένα script μπορεί να χρησιμοποιήσει το cookie ως αναγνωριστικό για τον χρήστη
  - αλληλεπιδράσει με μια βάση δεδομένων
- Αποθηκεύοντας προσωπικές πληροφορίες σε ένα cookie αποφεύγουμε να τις διατηρήσουμε σε μια βάση δεδομένων.

Πληρεξούσιος (Proxy)



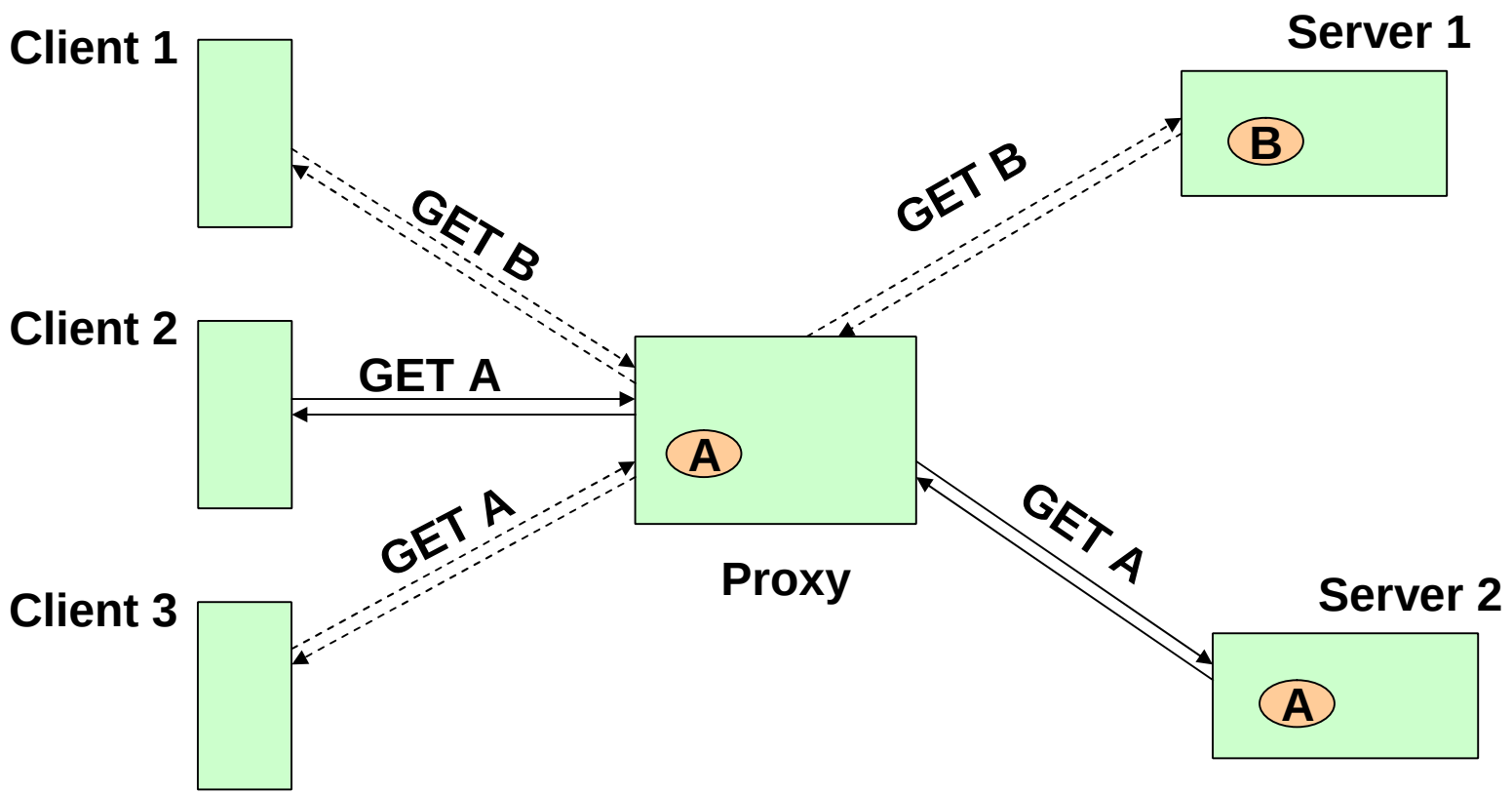
- Ο Proxy αποτελεί ένα ενδιάμεσο πρόγραμμα μεταξύ πελάτη και εξυπηρετητή
  - μπορεί να βρίσκεται στον ίδιο ή σε διαφορετικό υπολογιστή από αυτόν που στέλνει το request.
- Ο Proxy ενεργεί για λογαριασμό μιας ομάδας χρηστών κατά την επικοινωνία με τον server.
- Ο Proxy ενεργεί
  - ως ένας server για τους πελάτες
  - ως πελάτης για άλλα proxies ή servers

Πληρεξούσιος (Proxy)



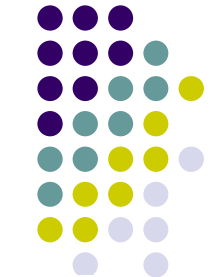
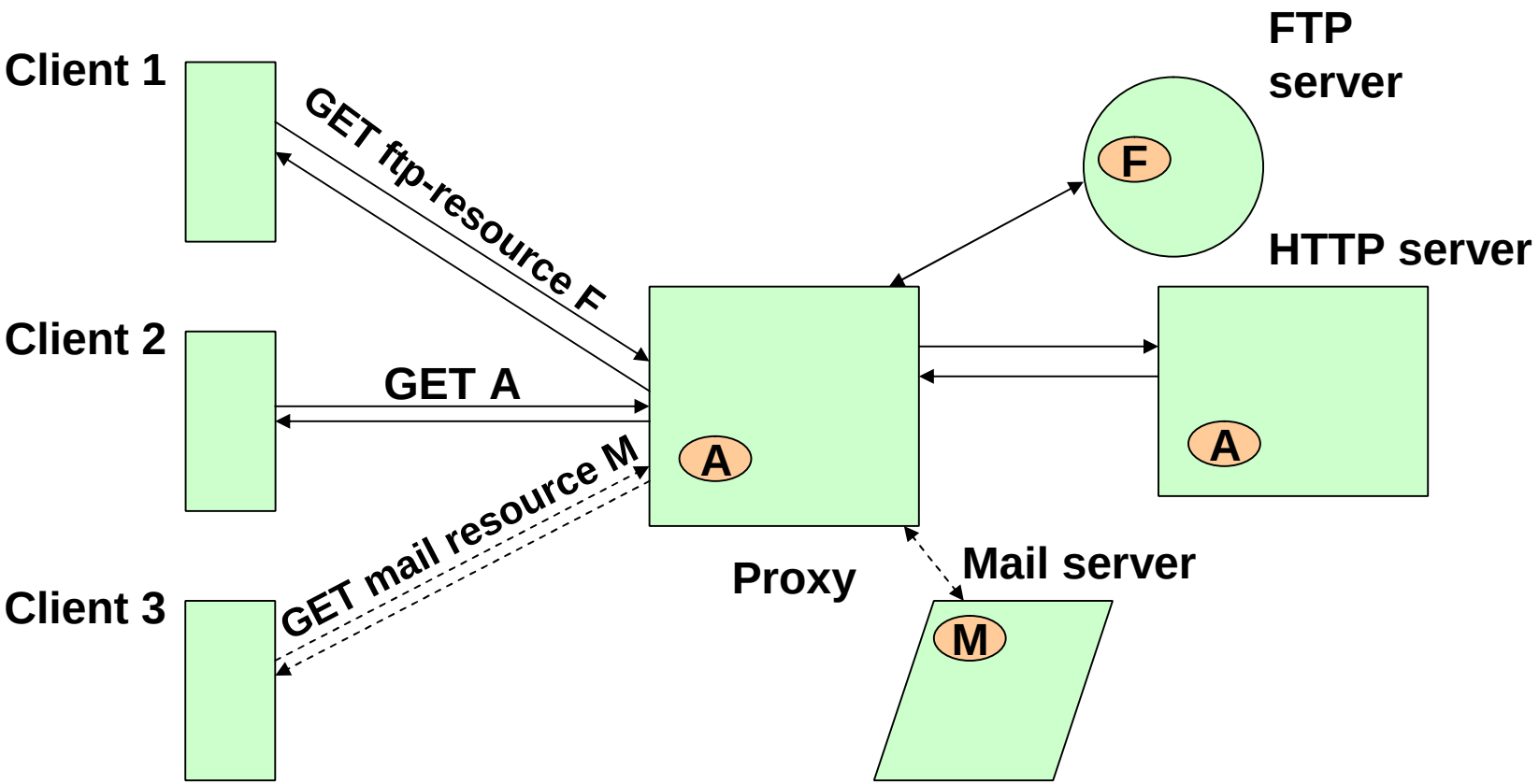
- Ο Proxy μπορεί να χρησιμοποιηθεί
  - διαμοιρασμό της πρόσβασης σε πόρους
  - αποθήκευση (caching) των responses
  - υποστήριξη ανωνυμίας
  - μετατροπή requests/responses
  - φιλτράρισμα requests/responses

Πληρεξούσιος (Proxy)



- Ο Proxy λειτουργεί ως ενδιάμεσος μεταξύ πελάτη και εξυπηρετητή
  - αποθήκευση (caching) των responses

Gateway προς non-HTTP συστήματα

- Ένας Proxy μπορεί να λειτουργήσει ως μεσάζοντας (όπως μια πύλη) για την επικοινωνία με δίκτυα που δεν χρησιμοποιούν το πρωτόκολλο HTTP.
- π.χ. κατά την επικοινωνία μεταξύ ενός πελάτη και ενός FTP server, ο proxy ενεργεί ως πύλη
  - ο πελάτης και ο εξυπηρετητής δεν υποστηρίζουν το ίδιο πρωτόκολλο.

Gateway προς non-HTTP συστήματα