



Θέματα Προγραμματισμού Διαδικτύου Εισαγωγή - Πρωτόκολλα

Στελιος Σφακιανάκης
Εαρινό 2019



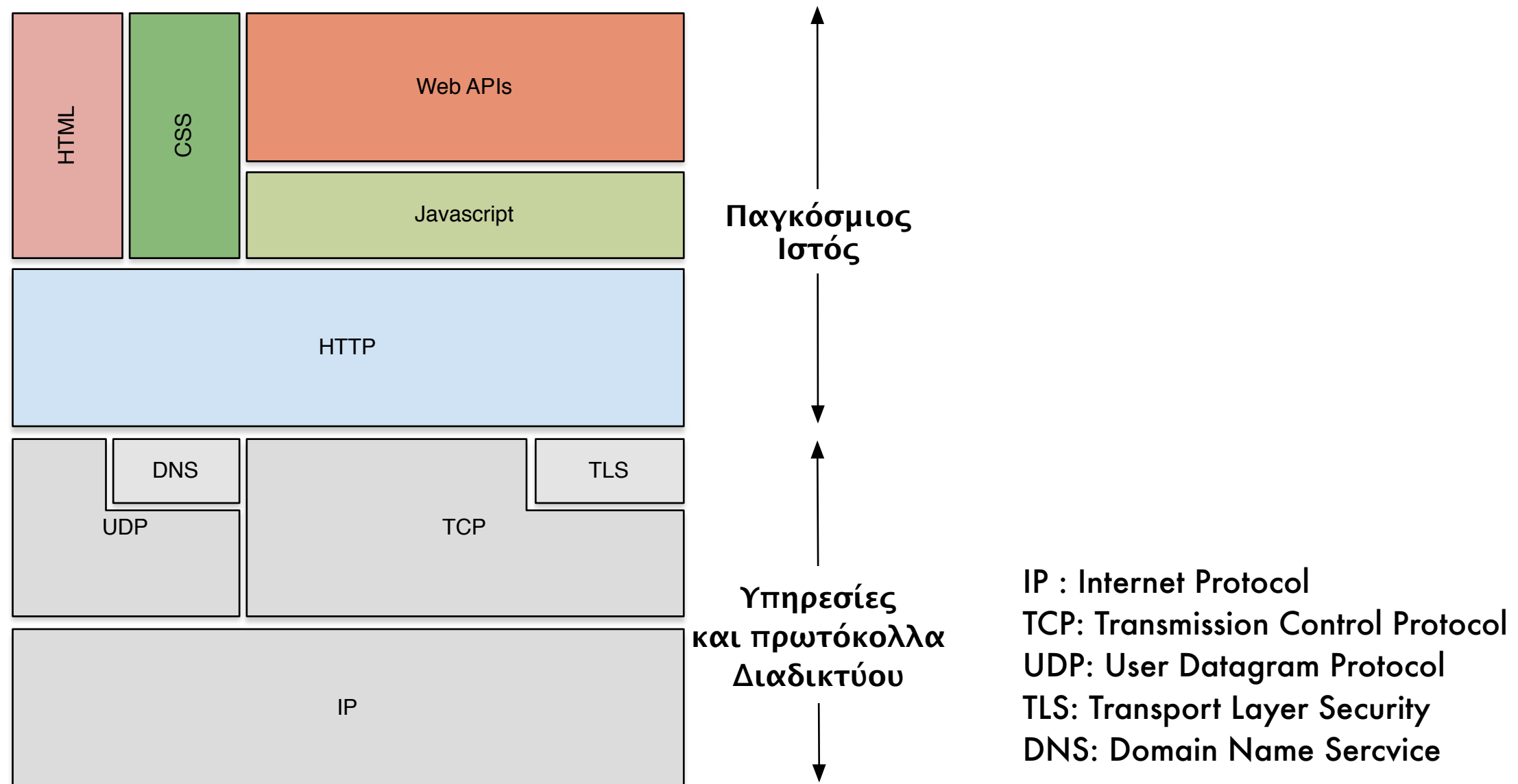


Περιεχόμενα

- Εισαγωγή στο Διαδίκτυο
 - Διευθύνσεις
 - Πρωτόκολλα
- Παγκόσμιο Ιστός - HTTP

Διαδίκτυο - Πρωτόκολλα - Υπηρεσίες

Αρχιτεκτονική του Ιστού ως υπηρεσία διαδικτύου





Internet Protocol (IP) και οι διευθύνσεις του

- Καθε υπολογιστής προσβάσιμος στο Ιντερνετ έχει μια διεύθυνση IP (Internet Protocol address)
 - Στην πιο διαδεδομένη έκδοση version 4 του IP, οι διευθύνσεις είναι 4 bytes (δηλ. 32 bit) που συνήθως τις γράφουμε ως 4 αριθμοί χωρισμένοι με τελεία π.χ. **147.95.40.60**
- Οι διευθύνσεις (όπως και στην καθημερινότητα μας) επιτρέπουν την εύρεση μιας "τοποθεσίας" και τη "δρομολόγηση" σε αυτήν
- Το πρωτόκολλο IP ορίζει πώς γίνεται η μεταφορά *πακέτων δεδομένων* στο Διαδίκτυο, μέσω ενδιάμεσων δικτύων και υπολογιστών-κόμβων. Καθε πακέτο IP περιέχει την IP διεύθυνση του "αποστολέα" και την διεύθυνση του προορισμού.



Transmission Control Protocol (TCP)

- Τα πακέτα δεδομένων του IP μπορεί να χαθούν ή να φτάσουν με διαφορετική σειρά στον προορισμό τους . Το TCP προσφέρει *αξιοπιστία*, χειρισμό λαθών, και αλγορίθμους για την βελτιστοποίηση της μεταφοράς των δεδομένων.
- Χρησιμοποιεί την έννοια της **σύνδεσης** δηλ. ορίζει μηνύματα για την αρχικοποίηση του "καναλιού" επικοινωνίας πριν τη μεταδοση των δεδομένων των εφαρμογών, καθώς και μηνύματα για το "κλείσιμο" της σύνδεσης
- Τα πακέτα του TCP περιέχονται μέσα σε πακέτα του IP και με τη σειρά τους περιέχουν της **θύρες** (ports) του αποστολέα και του παραλήπτη.
 - Ενώ οι IP διευθύνσεις προσδιορίζουν τον "υπολογιστή" από τον οποίο ξεκινάνε ή καταλήγουν τα πακέτα, οι θύρες προσδιορίζουν τις εφαρμογές (π.χ. web server) που επικοινωνούν.
 - Έτσι π.χ. το λειτουργικό σύστημα του υπολογιστή προορισμού μπορεί να παραδώσει το πακέτο πληροφορίας στο κατάλληλο πρόγραμμα που το περιμένει.



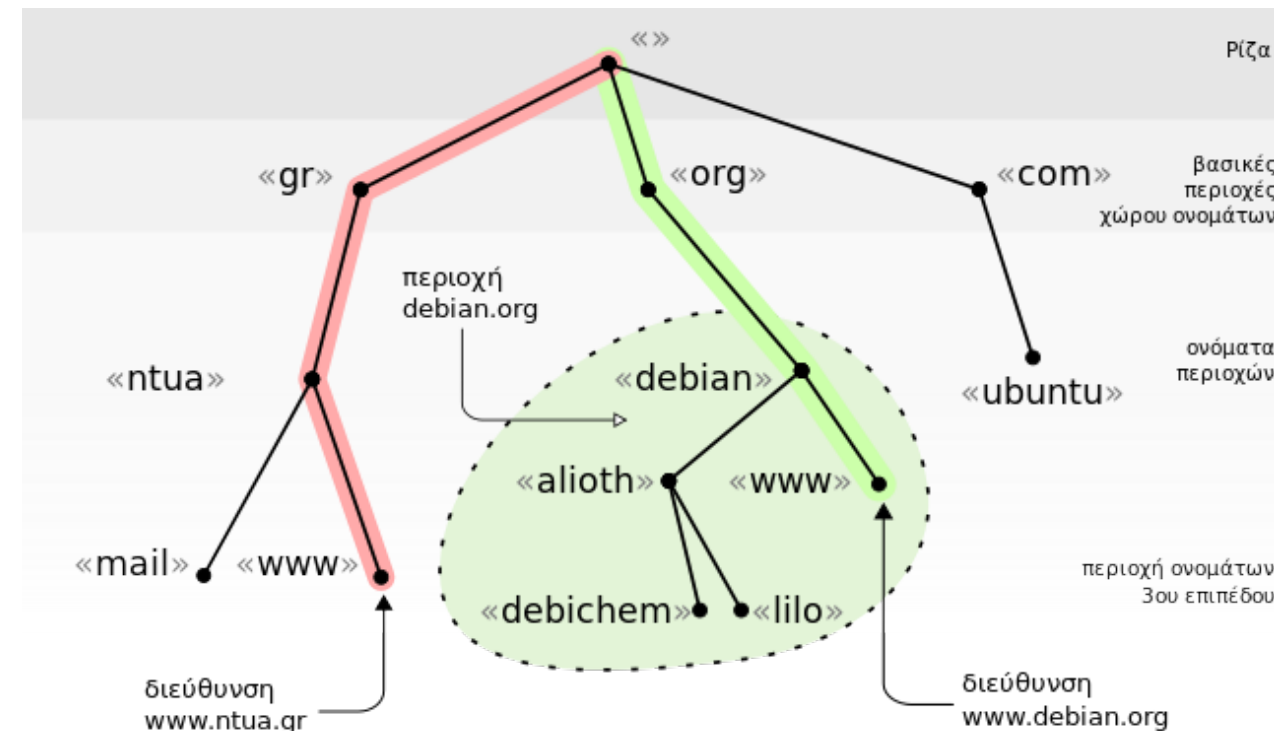
UDP και DNS

- Η αξιόπιστη λειτουργία του TCP όμως εισάγει πολυπλοκότητα και καθυστερήσεις.
- Σε περιπτώσεις που η γρήγορη παράδοση της πληροφορίας (π.χ. για βίντεο και ήχο) είναι πιο σημαντική από την αξιόπιστη μπορεί να χρησιμοποιηθεί το User Datagram Protocol (UDP)
 - Τα UDP πακέτα που ανταλλάσσονται μεταξύ δυο εφαρμογών στο διαδίκτυο μπορεί να φτάσουν με λαθος σειρά και δεν υπάρχουν αναφορές παράδοσης οπότε ο αποστολέας δεν γνωρίζει αν ένα μήνυμα έφτασε στον προορισμό.
- Όπως για το TCP έτσι και τα πακέτα του UDP μεταδίδονται μέσα σε πακέτα IP και περιέχουν ομοίως αριθμούς *θυρών*
- Ενώ η πλειονότητα των εφαρμογών χρησιμοποιεί TCP, υπάρχει μια πολύ κρίσιμη υπηρεσία που το χρησιμοποιεί: το DNS!



Domain Name Service (DNS)

- Η Υπηρεσία Ονομάτων (DNS) του Διαδικτύου επιτρέπει την αντιστοίχιση των αριθμητικών διευθύνσεων IP σε φιλικά για τον χρήστη **ονόματα**
- Τα ονόματα είναι δομημένα σε μια ιεραρχική δομή (δένδρο) που επιτρέπει την κατανεμημένη κατασκευή και διαχείριση τους
- Η σχέση μεταξύ ονομάτων και διευθύνσεων IP δεν είναι 1 προς 1. Δηλαδή σε ένα όνομα μπορεί να αντιστοιχούν πολλές διευθύνσεις IP και σε μια διεύθυνση πολλά ονόματα.



By DNS_Tree.svg: Sylvain Leroux (talk) derivative work: Ggia (talk) - DNS_Tree.svg, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=15525424>

- Π.χ. Το google.com έχει τις 108.177.127.102, 108.177.127.100, 108.177.127.138, 108.177.127.101, 108.177.127.113, 108.177.127.139, ...



TLS

- Το Transport Layer Security (TLS) εξασφαλίζει την ασφάλεια της επικοινωνίας στο Διαδίκτυο
 - ✓ Κρυπτογράφηση, που εξασφαλίζει την ιδιωτικότητα (privacy)
 - ✓ Ακεραιότητα δεδομένων (δηλ. δεν μπορεί ένας ενδιάμεσος κόμβος να αλλάξει την πληροφορία που ανταλλάσσεται)
 - ✓ Αυθεντικοποίηση, δηλ. ταυτοποίηση των επικοινωνούντων μερών
- Χρησιμοποιείται ευρέως για να ασφαλίσει πρωτόκολλα και υπηρεσίες εφαρμογής, π.χ. HTTP**S**, SMTP**S** (email), ...

HTTP και ο Παγκόσμιος Ιστός



Ο Παγκόσμιος Ιστός

- Ο Παγκόσμιος Ιστός (**World Wide Web**) είναι ένα παγκόσμιο σύστημα διασυνδεδεμένων εγγράφων υπερκειμένου που είναι διαθέσιμα στο *Ιντερνετ*

★ Τι είναι το υπερκείμενο?

- ➔ Είναι το *μη γραμμικό* κείμενο, δηλαδή αν μπορεί ο αναγνώστης να μεταβεί από κάποιο σημείο του εγγράφου σε κάποιο άλλο (ή σε άλλο έγγραφο) χωρίς να χρειαστεί να “προσπελάσει” σειριακά ό,τι βρίσκεται “ενδιάμεσα”
- ➔ Επομένως, το χαρακτηριστικό στοιχείο του υπερκειμένου είναι η ύπαρξη **(υπερ)συνδέσμων** (hyperlinks)



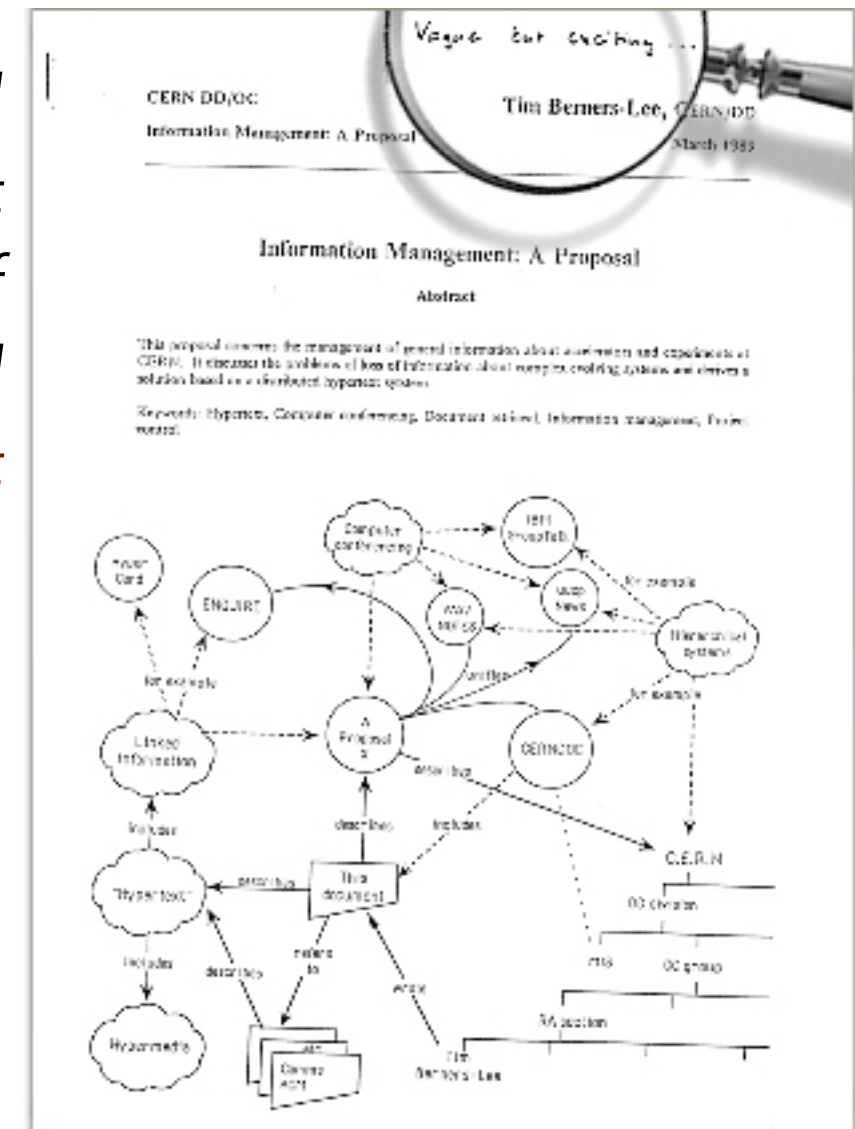
Ο Παγκόσμιος Ιστός: λίγη ιστορία

- Τη δεκαετία του 1960 ο πρόγονος του Ιντερνετ, το λεγόμενο **ARPANET**, δημιουργείται από το Υπουργείο Άμυνας των ΗΠΑ
 - Το 1974 δημοσιεύεται το TCP (Transmission Control Protocol) από τους Cerf και Kahn και το 1983 το TCP/IP γίνεται ο θεμέλιος λίθος του ARPANET
 - Το 1984 υλοποιείται το πρώτο Domain Name System (DNS), η υποδομή “ονοματοδοσίας” του Internet
 - Το Ηλεκτρονικό Ταχυδρομείο (e-mail) και η ανταλλαγή εγγράφων (File Transfer Protocol - FTP) οι κύριες υπηρεσίες του
- Το 1989 ο Sir Tim Berners-Lee προτείνει ένα “κατανεμημένο σύστημα υπερκειμένου” στο CERN (δες επόμενο slide) και πριν τα Χριστούγεννα του 1990 υλοποιεί τον πρώτο εξυπηρετητή (*server*) και τον πρώτο πλοηγητή (*web browser*)
 - ..μαζί (επανα)εισαγει τον όρο Hypertext (υπερκείμενο) και υλοποιεί το **Πρωτόκολλο Υπερκειμένου (HTTP)** και την **Γλώσσα Σήμανσης Υπερκειμένου (HTML)**
- Η πρώτη τεκμηριωμένη του HTTP είναι η 0.9 το 1991, το 1996 η έκδοση 1.0 (RFC 1945), ενώ το 1997 η έκδοση 1.1 (RFC 2068) με βελτιώσεις το 1999 (RFC 2616)
- Ο πρώτος ευρέως διαθέσιμος browser ήταν ο Mosaic (1993) στον οποίο βασίστηκε ο δημοφιλής Netscape Navigator (1994)
- Ξεκινάει ο (1ος) “πόλεμος των πλοηγητών” (browser war) με τη διάθεση από την Microsoft του Internet Explorer (1995). Η Netscape χάνει και δημιουργεί το Mozilla Foundation που το 2004 διαθέτει τον (ανοικτού λογισμικού) Mozilla Firefox
- Το 1998 δημιουργείται η Google και σε λίγα χρόνια “εξαφανίζει” τις άλλες μηχανές αναζήτησης (Lycos, Altavista, Yahoo!)
- Το 2008 παρουσιάζεται ο Chrome browser από την Google
- Το HTTP 2 δημοσιεύεται ως RFC 7540 τον Μαίο 2015

“Information Management: A Proposal”

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Tim Berners-Lee, CERN
March 1989



<https://www.w3.org/History/1989/proposal.html>

Το πρωτόκολλο HTTP

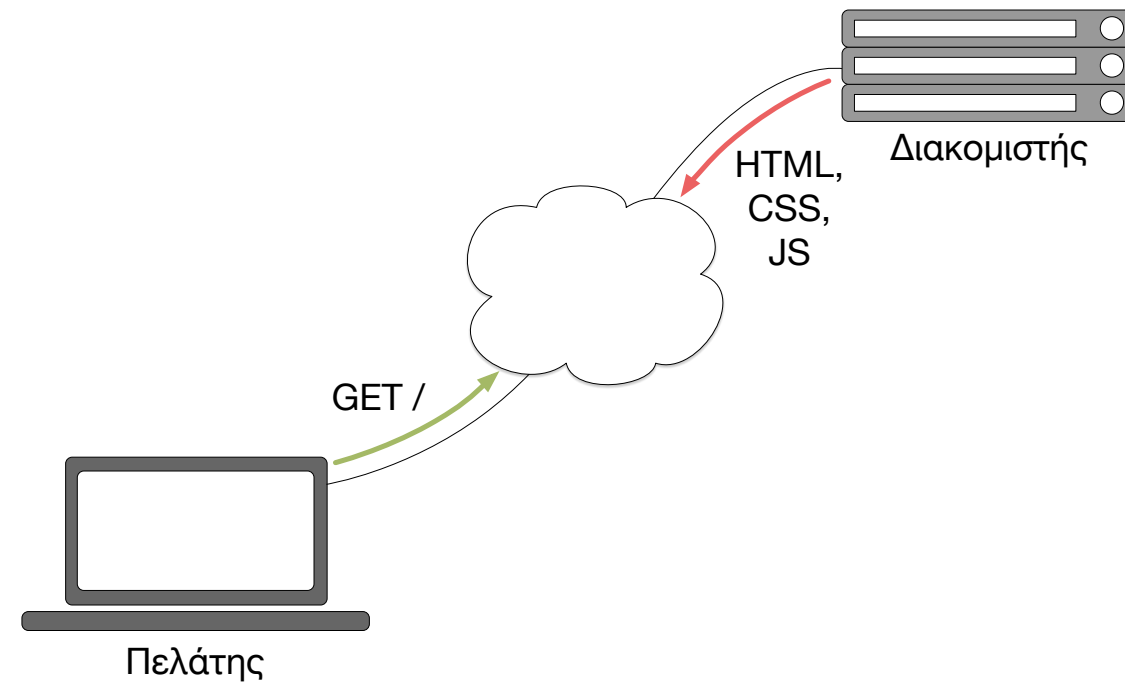


HTTP

- Είναι ένα πρωτόκολλο εφαρμογής "πάνω" από τα TCP/IP
- Η προτυποποιημένη θύρα επικοινωνίας είναι το **80** και στην περίπτωση που χρησιμοποιείται το TLS (δηλ. για το **HTTPS**) είναι το **443**
- Τα μηνύματα που ανταλλάσσονται είναι σε κείμενο που κάνει εύκολο το να δούμε τί πληροφορία ανταλλάσσεται (π.χ. στο Chrome DevTools)



Διακομιστές και Πελάτες

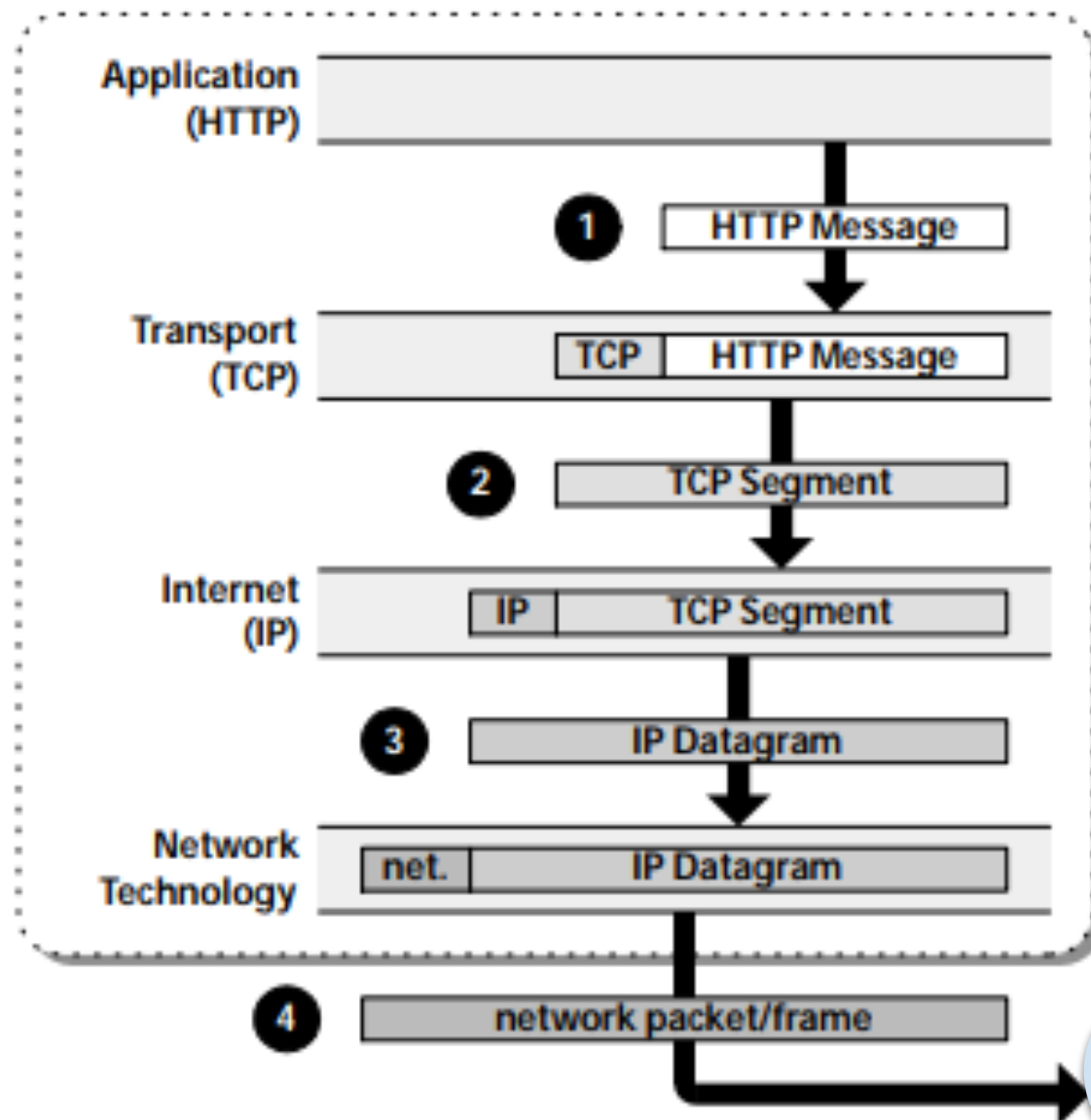


- Το HTTP ακολουθεί μια λογική *αίτησης - απάντησης* (request-response) μεταξύ του πελάτη (π.χ. browser σε ένα PC) και του διακομιστή (server)
- Το πρόγραμμα Πελάτης ξεκινάει την επικοινωνία στέλνοντας μια αίτηση στον Διακομιστή
- Ο Διακομιστής απαντάει στέλνοντας την αιτούμενη πληροφορία

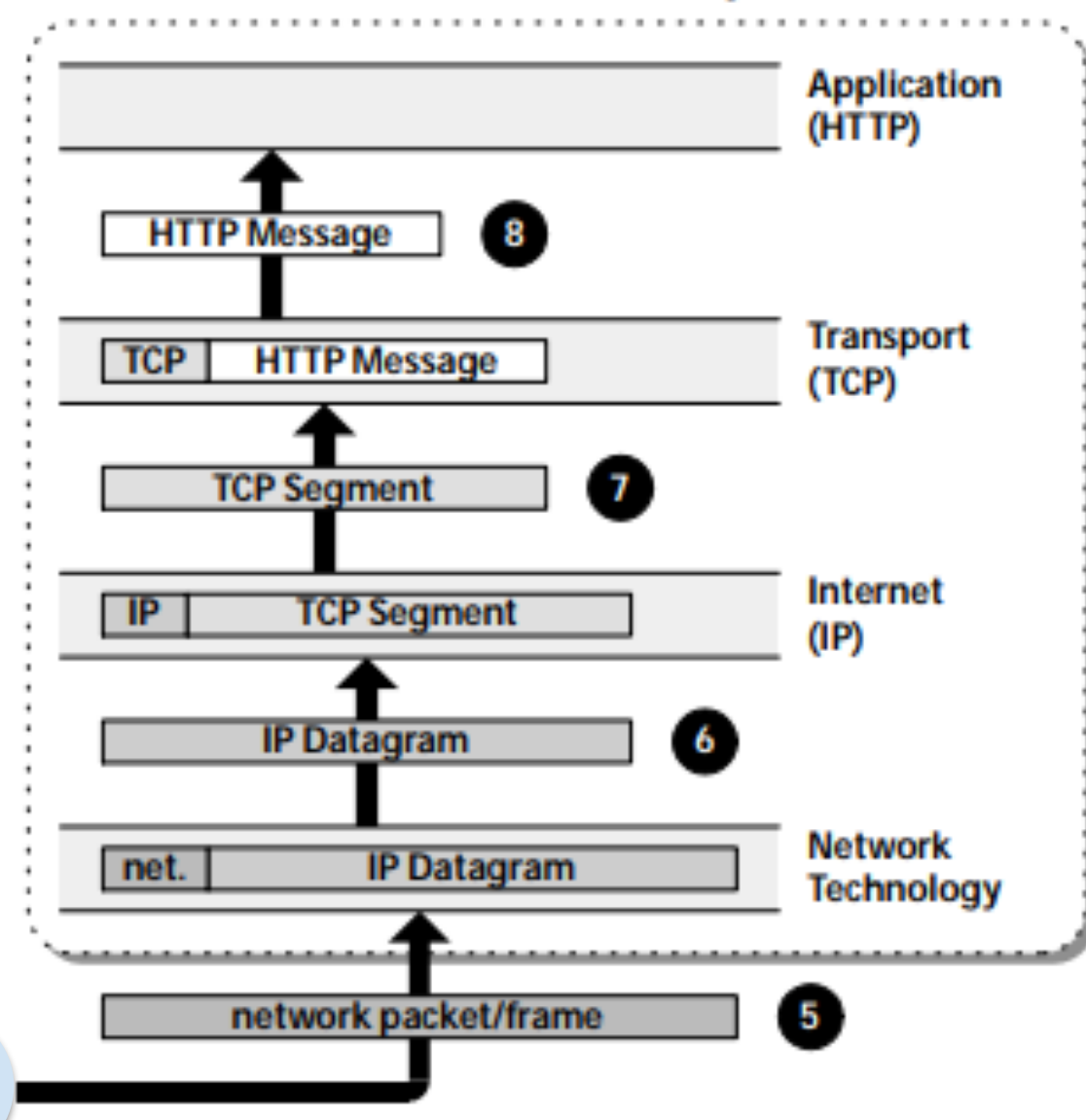


"Under the hood"

Πελάτης (Browser)



Διακομιστής (Web Server)

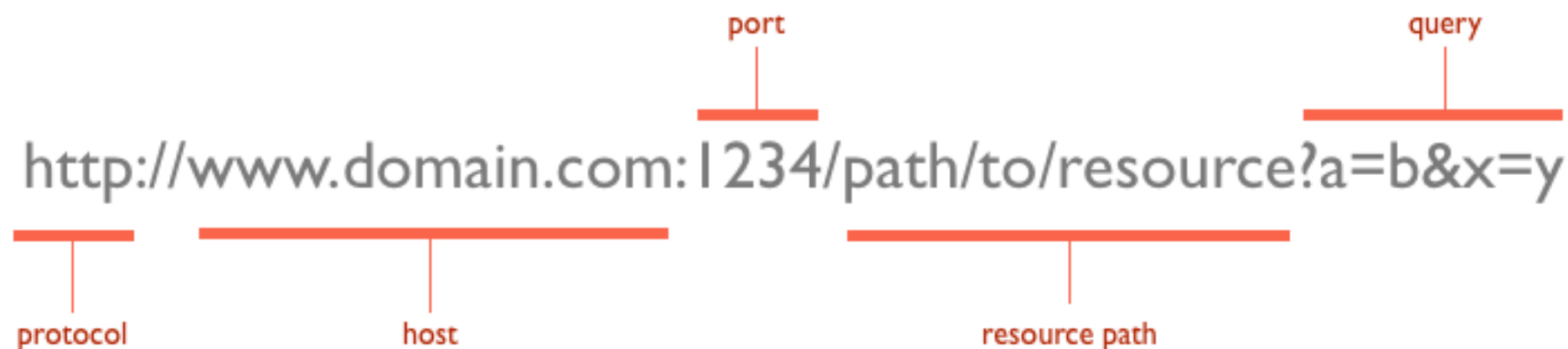


Internet



URLs

- Τα **Uniform Resource Locators** (URLs) είναι διευθύνσεις για σελίδες στον Ιστό (web pages)
- Ενσωματώνουν πληροφορία για το πρωτόκολλο επικοινωνίας (http/https), την διεύθυνση/όνομα του διακομιστή, την θύρα επικοινωνίας, το "μονοπάτι" στο συγκεκριμένο αρχείο/έγγραφο, και άλλες πληροφορίες






HTTP μηνύματα

- Όπως συνηθίζεται σε όλα τα πρωτόκολλα τα μηνύματα HTTP περιέχουν:
 - Μια κεφαλίδα (headers), και
 - (προαιρετικά) ένα "σώμα" (body/payload)
- Τα headers μηνυμάτων που στέλνουν οι πελάτες περιέχουν την "μεθοδο" (π.χ. GET) και τη διύθυνση της σελίδας, ενώ τα μηνύματα που επιστρέφει ο διακομιστής περιέχουν κωδικούς κατάστασης και άλλα μεταδεδομένα



TEI Κρήτης | x Stelios


← → ↻ 🏠 <https://www.teicrete.gr> ☆ 📧 📧 📧 📧 📧

 **TEI Κρήτης**
Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

Αρχική A-Ω MyTEI Webmail Σύνδεση English

Αναζήτηση


🏠 TEI Κρήτης Σπουδές Ζωή στο TEI Σχολές Διεθνείς Σχέσεις Έρευνα Νέα



Elements Console Sources **Network** Performance Memory Application Security Audits

View: ☐ Group by frame ☐ Preserve log ☒ Disable cache ☐ Offline Online

Filter ☐ Hide data URLs All XHR JS CSS Img Media Font **Doc** WS Manifest Other

Name	× Headers Preview Response Cookies Timing
 www.teicrete.gr	<p>Vary: Cookie,Accept-Encoding Content-Encoding: gzip Vary: Accept-Encoding</p> <p>▼ Request Headers view parsed</p> <p>GET / HTTP/1.1 Host: www.teicrete.gr Connection: keep-alive Pragma: no-cache Cache-Control: no-cache Upgrade-Insecure-Requests: 1 User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.67 Safari/537.36 DNT: 1 Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8 Accept-Encoding: gzip, deflate, br Accept-Language: el,en-US;q=0.9,en;q=0.8 Cookie: _ga=GA1.2.1707404579.1536050637; has_js=1; _gid=GA1.2.2067371965.1539805314; _gat=1</p>

1 / 58 requests | 18.3 KB / 870 KB transferred | Finish:...



Request

```
GET / HTTP/1.1
Host: www.teicrete.gr
Connection: keep-alive
Pragma: no-cache
Cache-Control: no-cache
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Macintosh...
DNT: 1
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8
Accept-Encoding: gzip, deflate, br
Accept-Language: el,en-US;q=0.9,en;q=0.8
```

Response

```
HTTP/1.1 200 OK
Connection: Keep-Alive
Transfer-Encoding: chunked
Expires: Sun, 19 Nov 1978 05:00:00 GMT
Date: Thu, 18 Oct 2018 07:00:20 GMT
Content-Type: text/html; charset=utf-8
Etag: "1539844751-0"
Server: Apache/2.4.7 (Ubuntu) PHP/5.5.9-1ubuntu4.26 OpenSSL/1.0.1f
X-Powered-By: PHP/5.5.9-1ubuntu4.26
X-Drupal-Cache: HIT
Content-Language: el
X-Frame-Options: SAMEORIGIN
X-Generator: Drupal 7 (https://www.drupal.org)
Cache-Control: public, max-age=600
Last-Modified: Thu, 18 Oct 2018 06:39:11 GMT
Vary: Cookie,Accept-Encoding
Content-Encoding: gzip
```



(Βασικές) Μέθοδοι HTTP

- **GET**: ανάκτηση σελίδας από διακομιστή
- **HEAD**: επιστροφή μόνο της κεφαλίδας (header) του μηνύματος δηλ μόνο μεταδεδομένα (π.χ. μέγεθος αρχείου) χωρίς το περιεχόμενο της σελίδας/αρχείου
- **POST**: υποβολή φόρμας ή άλλων δεδομένων στον διακομιστή για δημιουργία νέας "σελίδας"
- **PUT**: υποβολή δεδομένων για την ενημέρωση σελίδας
- **DELETE**: αίτηση για κατάργηση (διαγραφή) της σελίδας στον διακομιστή



HTTP κώδικες κατάστασης

- Τα HTTP Status codes είναι αριθμοί 3 ψηφίων, από τα οποία το πρώτο προσδιορίζει την γενική κατηγορία της απάντησης (response):
 - 1xx : πληροφοριακά μηνύματα
 - 2xx : Επιτυχία!!
 - 3xx : ανακατεύθυνση σε άλλο URL
 - 4xx: Αποτυχία, οφείλεται στον πελάτη
 - 400: Bad request
 - 401: Unauthorized (π.χ. λαθος κωδικός)
 - 403: Forbidden
 - 404: Not Found
 - 5xx : Αποτυχία, οφείλεται στον διακομιστή
 - 500: Internal Server Error (π.χ. ο διακομιστής δεν μπορεί να επικοινωνήσει με τη βάση δεδομένων)

