

Εργαστηριακή Άσκηση 11

Πλοήγηση στον παγκόσμιο ιστό

Ο παγκόσμιος ιστός (World Wide Web) είναι μια αχανής συλλογή εγγράφων που εμφανίζεται με τη μορφή ιστοσελίδων. Η ιστοσελίδα είναι μια συλλογή αντικειμένων, όπως αρχείο HTML, εικόνες JPEG, αρχεία ήχου, βίντεο, Javascript, Java applet, κλπ, και μπορεί να περιέχει παραπομπές προς άλλες ιστοσελίδες¹. Η ανάγνωση των ιστοσελίδων μπορεί να γίνει με τη βοήθεια ενός προγράμματος-πελάτη (client) γνωστού ως πλοηγού ιστού ή φυλλομετρητή (web browser). Οι χρήστες μπορούν να ακολουθήσουν τις υπερζεύξεις με ένα κλικ του ποντικιού, διαδικασία γνωστή, ως πλοήγηση (navigation). Ο πλοηγός ιστού προσκομίζει την ιστοσελίδα και την απεικονίζει στην οθόνη του υπολογιστή, μεταφράζοντας το κείμενο και τις εντολές μορφοποίησης. Οι υπερζεύξεις απεικονίζονται έντονα, είτε με υπογράμμιση είτε με τη χρήση ενός συγκεκριμένου χρώματος, είτε και με τα δύο.

Σε κάθε ιστοσελίδα έχει προσδεθεί ένα μονοσήμαντο όνομα, ώστε να μην υπάρχει καμία αμφιβολία όσον αφορά την αναγνώρισή της. Συγκεκριμένα, σε κάθε σελίδα εκχωρείται μια ταυτότητα, ένας *Ομοιόμορφος Εντοπιστής Πόρων URL (Uniform Resource Locator)* που στην ουσία παίζει τον ρόλο του παγκόσμιου ονόματος της σελίδας. Τα URL αποτελούνται από τρία μέρη: το πρωτόκολλο, το όνομα DNS ή τη διεύθυνση IP της μηχανής στην οποία βρίσκεται η σελίδα και ένα τοπικό όνομα που προσδιορίζει μονοσήμαντα τη συγκεκριμένη σελίδα (συνήθως απλώς ένα όνομα καταλόγου ή αρχείου που υπάρχει στη μηχανή όπου βρίσκεται η σελίδα). Για παράδειγμα, το URL <http://www.telecom.ntua.gr/index.html> αποτελείται από τρία μέρη: το πρωτόκολλο (*http*), το όνομα DNS της μηχανής στην οποία εντοπίστηκε η σελίδα (*www.telecom.ntua.gr*) και το όνομα του αρχείου που περιέχει τη σελίδα (*index.html*), με κάποια σημεία στίξης που διαχωρίζουν τα κομμάτια. Για να επισκεφθεί, ο χρήστης μια ιστοσελίδα αρκεί να συμπληρώσει σωστά τη διεύθυνση της ιστοσελίδας στη γραμμή διευθύνσεων του πλοηγού ιστού.

Για τη μεταφορά των πληροφοριών που περιέχει η ιστοσελίδα χρησιμοποιείται το πρωτόκολλο στρώματος εφαρμογής [HTTP](#) (hypertext transfer protocol). Η πρώτη έκδοση του πρωτοκόλλου ήταν η HTTP/1.0 που περιγράφεται στο [RFC 1945](#). Η επόμενη και πλέον διαδεδομένη έκδοση HTTP/1.1 ορίστηκε το 1997 στο [RFC 2068](#), που στη συνέχεια αντικαταστάθηκε από το [RFC 2616](#). Σε συντομία, η βασική λειτουργία του HTTP είναι η μεταφορά πόρων (resources). Πόρος είναι ένα κομμάτι πληροφορίας που προσδιορίζεται μοναδικά από ένα URL (το R στο ακρωνύμιο URL). Η πιο κοινή περίπτωση πόρου είναι το αρχείο. Όμως, ένας πόρος μπορεί να δημιουργείται δυναμικά ως αποτέλεσμα εντολών (π.χ. από κώδικα PHP), είτε να είναι έγγραφο διαθέσιμο σε πολλές γλώσσες, είτε κώδικας Javascript που θα εκτελεσθεί δυναμικά ή οτιδήποτε άλλο.

Σε κάθε ιστοθέση (Web site) υπάρχει μία διεργασία εξυπηρετητή (web server), η οποία παρακολουθεί την TCP θύρα (port) 80 περιμένοντας εισερχόμενες συνδέσεις από πελάτες πλοηγούς (web clients). Μόλις εδραιωθεί η σύνδεση, ο πελάτης στέλνει μία αίτηση HTTP (HTTP request) και ο εξυπηρετητής στέλνει μία απάντηση HTTP (HTTP response) που περιέχει διάφορα αντικείμενα. Μετά η σύνδεση απελευθερώνεται, δηλαδή, το HTTP δεν διατηρεί πληροφορία κατάστασης (stateless) μεταξύ διαδοχικών δοσοληψιών (transactions). Για την πρόσβαση σε εξυπηρετητές ιστού έχει όμως πλέον επικρατήσει η χρήση του πρωτοκόλλου [HTTPS](#) (HTTP Secure). Ο εξυπηρετητής τυπικά ακούει στη θύρα 443 και η επικοινωνία με αυτόν είναι κρυπτογραφημένη ακολουθώντας το πρωτόκολλο TLS ή παλιότερα το SSL (που θα δείτε στην επόμενη εργαστηριακή άσκηση).

Οι συνδέσεις HTTP διακρίνονται σε μη επίμονες (Non-persistent HTTP) και σε επίμονες (Persistent HTTP). Το πρωτόκολλο HTTP/1.0 χρησιμοποιεί μη επίμονες συνδέσεις HTTP. Σε αυτές, πάνω από μία σύνδεση TCP μπορεί να αποσταλεί το πολύ ένα αντικείμενο κάθε φορά. Το

¹ Κείμενο που παραπέμπει σε άλλο κείμενο αναφέρεται ως υπερκείμενο (hypertext) και οι παραπομπές είναι γνωστές ως υπερζεύξεις (hypertext links).

πρωτόκολλο HTTP/1.1 χρησιμοποιεί επίμονες συνδέσεις HTTP. Σε αυτές, πάνω από την ίδια σύνδεση TCP μπορούν να ζητηθούν και αποσταλούν πολλαπλά αντικείμενα μεταξύ πελάτη-εξυπηρετητή. Έτσι μειώνονται οι επιπτώσεις της καθυστέρησης κατά την τριπλή χειραψία TCP (μία για κάθε αντικείμενο στο HTTP/1.0). Ως εκ τούτου, η μεταφορά των αντικειμένων και η εμφάνιση της ιστοσελίδας γίνεται γρηγορότερα. Με την πάροδο των χρόνων όμως, το μέγεθος των ιστοσελίδων έχει μεγαλώσει, τυπικά ξεπερνούν τα 2 Mb, και ο αριθμός των αντικειμένων που περιέχει η αρχική σελίδα μιας ιστοθέσης άνετα ξεπερνά τα 100. Το HTTP/1.1 δεν αποδίδει καλά και για αυτό αναπτύχθηκαν διάφορα πολύπλοκα τεχνάσματα.

Προκειμένου να παρακαμφτούν αυτά, το 2007 ξεκίνησαν οι εργασίες της ομάδας HTTPbis Working Group που κατέληξαν το 2014 στην αναθεώρηση του HTTP/1.1 από μια σειρά έξι νέων προτύπων [RFC 7230](#) έως [RFC 7235](#) και το 2015 στη δημοσίευση της τελευταίας έκδοσής του, το HTTP/2, στο [RFC 7540](#). Το [HTTP/2](#) διατηρεί τη συμβατότητα με το HTTP/1.1 χρησιμοποιώντας τις ίδιες μεθόδους (δείτε παρακάτω) με αυτό και προσθέτει νέα λειτουργικότητα ώστε το φόρτωμα των ιστοσελίδων να γίνεται γρηγορότερα. Στο HTTP/2 χρησιμοποιούνται λιγότερες συγκριτικά με το HTTP/1.1 συνδέσεις και η πληροφορία μεταδίδεται σε δυαδική μορφή (όχι ως κείμενο όπως στο HTTP/1.1) σε παράλληλες ροές εντός μιας σύνδεσης TCP, με αποτέλεσμα το ταχύτερο κατέβασμα των ιστοσελίδων. Πλέον όλοι οι πλοηγοί ιστού έχουν ενσωματώσει το HTTP/2 και σύμφωνα με τα πιο πρόσφατα στοιχεία (Αύγουστος 2020) το 49% των πιο δημοφιλών ιστοθέσεων το υποστηρίζει.

Τέλος, από το 2018, ξεκίνησαν επίσημα οι εργασίες για την προτυποποίηση της επόμενης έκδοσής του, το [HTTP/3](#). Στο HTTP/3, για τις παράλληλες συνδέσεις χρησιμοποιείται ένα άλλο πρωτόκολλο μεταφοράς, το [QUIC](#), που βασίζεται στο UDP. Σκοπός του QUIC είναι να ελαχιστοποιήσει το εγγενές στο HTTP/2 φαινόμενο καθυστέρησης στην κορυφή της ουράς (head-of-line blocking) λόγω της πολυπλεξίας παράλληλων ροών σε μια σύνδεση TCP. Προς το παρόν (2020), παρότι έχει υλοποιηθεί στους δημοφιλείς πλοηγούς ιστού είναι απενεργοποιημένο και πολύ λίγοι εξυπηρετητές ιστού το υποστηρίζουν.

Σκοπός αυτής της εργαστηριακής άσκησης είναι η εξέταση του πρωτοκόλλου εφαρμογής HTTP που χρησιμοποιείται ευρύτατα στο διαδίκτυο για την πλοήγηση στον ιστό, με τη βοήθεια του αναλυτή πρωτοκόλλων Wireshark. Εδώ θα χρησιμοποιήσετε τη λειτουργία *Capture* με φίλτρο, ώστε τα πλαίσια που καταγράφονται να περιέχουν κάποια συγκεκριμένα χαρακτηριστικά. Υπενθυμίζεται ότι το φίλτρο απεικόνισης (*Display*), που επιλέγετε από το μενού *Analyze*, μπορεί να (απ)ενεργοποιηθεί οποιαδήποτε στιγμή κατά τη διάρκεια της καταγραφής, καθώς επίσης και μετά την ολοκλήρωση αυτής, προκειμένου να αποκρύπτει (αποκαλύπτει) κάποια από τα συλληφθέντα πλαίσια, ενώ το φίλτρο σύλληψης, που επιλέγετε από το μενού *Capture*, ενεργοποιείται πάντοτε πριν ξεκινήσει η διαδικασία καταγραφής, με αποτέλεσμα να καταγράφεται μόνο ένα μέρος των διερχόμενων πλαισίων. Επιπλέον, ας σημειωθεί ότι η απενεργοποίηση του φίλτρου απεικόνισης γίνεται πιέζοντας το κουμπί *Clear* (η διαγραφή του φίλτρου στο πεδίο εισαγωγής δεν το ακυρώνει!).

Για τις παρακάτω ασκήσεις απαντήστε στο συνοδευτικό φυλλάδιο, το οποίο θα υποβάλλετε ως αρχείο pdf.

1. Ανάκτηση HTML σελίδας

Όπως έχει ήδη αναφερθεί, το HTTP είναι πρωτόκολλο επιπέδου εφαρμογής και ακολουθεί το μοντέλο πελάτη/εξυπηρετητή (client/server model). Το HTTP ορίζει ένα σύνολο μεθόδων αίτησης HTTP (HTTP request methods) ως λέξεις της Αγγλικής γλώσσας (συνήθως αναφέρονται ως HTTP verbs) που δηλώνουν την επιθυμητή δράση σε ένα συγκεκριμένο πόρο. Στο HTTP/1.1 έχουν ορισθεί οι μέθοδοι: GET, HEAD, POST, PUT, DELETE, CONNECT, OPTIONS, TRACE και PATCH. Εξ αυτών, η μέθοδος GET χρησιμοποιείται μόνο για τη λήψη δεδομένων και ζητά την αναπαράσταση ενός συγκεκριμένου πόρου. Η μέθοδος POST χρησιμοποιείται για την αποστολή δεδομένων προς τον καθοριζόμενο πόρο (π.χ. ένα συνθηματικό) προκαλώντας συνήθως κάποια

αλλαγή στην πλευρά του εξυπηρετητή. Οι αποκρίσεις HTTP (HTTP responses) είναι κωδικοί κατάστασης (status codes) που δείχνουν το κατά πόσο μια συγκεκριμένη αίτηση HTTP ικανοποιήθηκε. Ομαδοποιούνται σε πέντε κατηγορίες: informational, successful, redirects, client errors και servers errors. Για παράδειγμα, το 200 OK δηλώνει επιτυχία και το 404 Not Found δείχνει ότι ο εξυπηρετητής δεν βρήκε τον αιτούμενο πόρο. Για την πλήρη λίστα των κωδικών κατάστασης δείτε https://en.wikipedia.org/wiki/List_of_HTTP_status_codes.

Οι μορφές των μηνυμάτων της αίτησης και της απόκρισης HTTP είναι ταυτόσημες. Και στις δύο περιπτώσεις τα μηνύματα περιλαμβάνουν:

- Μια αρχική γραμμή,
- Καμία ή πολλές γραμμές επικεφαλίδων,
- Μια κενή γραμμή (blank line), δηλαδή, τους χαρακτήρες <CR><LF>², και
- Προαιρετικά το κυρίως σώμα του μηνύματος HTTP (π.χ. ένα αρχείο, τα δεδομένα μιας ερώτησης, κλπ).

Η αρχική γραμμή είναι διαφορετική για τις αιτήσεις και αποκρίσεις. Η αρχική γραμμή των ερωτήσεων περιέχει τρία μέρη που χωρίζονται με κενά: το όνομα της μεθόδου (πάντα με κεφαλαία γράμματα, π.χ. GET), την τοπική διαδρομή (local path) του αιτούμενου πόρου και τέλος την έκδοση του πρωτοκόλλου HTTP που χρησιμοποιείται. Η αρχική γραμμή των αποκρίσεων αποκαλείται γραμμή κατάστασης. Περιέχει και αυτή τρία μέρη που χωρίζονται με κενά: την έκδοση του HTTP, ένα κωδικό κατάστασης που υποδεικνύει το αποτέλεσμα της αίτησης (π.χ. 200) και μια Αγγλική λέξη που περιγράφει τον κωδικό κατάστασης (π.χ. OK). Οι γραμμές επικεφαλίδων παρέχουν πληροφορίες για την αίτηση ή την απόκριση ή για το αντικείμενο που περιέχει το κυρίως σώμα. Εμφανίζονται όπως το σύνηθες κείμενο: μια επικεφαλίδα ανά γραμμή της μορφής "Header-Name: value" και καταλήγουν σε <CR><LF>. Το κυρίως σώμα του μηνύματος HTTP ακολουθεί τις επικεφαλίδες. Στις αποκρίσεις, το κυρίως σώμα επιστρέφει τον αιτούμενο πόρο στον πελάτη ή κάποιο επεξηγηματικό κείμενο σε περίπτωση λάθους. Στις αιτήσεις, είναι το μέρος όπου αρχεία ή δεδομένα που εισάγει ο χρήστης αποστέλλονται στον εξυπηρετητή. Όταν το μήνυμα HTTP περιέχει σώμα, συνήθως, οι επικεφαλίδες περιγράφουν το σώμα (τον τύπο και το μήκος του).

Σε αυτή την άσκηση θα καταγραφούν τα μηνύματα HTTP που παράγονται κατά την επίσκεψη μιας ιστοσελίδας. Για μια λεπτομερή περιγραφή της ροής μηνυμάτων όλων των εμπλεκόμενων πρωτοκόλλων κατά το κατέβασμα μιας ιστοσελίδας δείτε το παράδειγμα στην ιστοθέση http://www.eventhelix.com/RealtimeMantra/Networking/http_sequence_diagram.pdf.

Προτού αρχίσετε την καταγραφή θα χρειαστεί να αδειάσετε την προσωρινή μνήμη (cache) του πλοηγού σας. Ξεκινήστε τον πλοηγό ιστού και πιέστε τα πλήκτρα Shift+Ctrl+Del ή επιλέξτε στο μενού τη σχετική εντολή, π.χ. *History* → *Clear Recent History* στον Firefox. Στο παράθυρο που θα εμφανισθεί προσδιορίστε το είδος (cache) της προς διαγραφή πληροφορίας, δηλώστε ως διάστημα τουλάχιστον μία ημέρα, επιβεβαιώστε την πρόθεσή σας, περιμένετε να ολοκληρωθεί η διαγραφή και κλείστε τα παράθυρα διαλόγου. Στη συνέχεια ξεκινήστε μια καταγραφή, επισκεφτείτε την ιστοσελίδα <http://my.ntua.gr> και σταματήστε την καταγραφή όταν φορτωθεί πλήρως η σελίδα. Εφαρμόστε κατάλληλο φίλτρο απεικόνισης ώστε να παραμείνουν μόνο μηνύματα του πρωτοκόλλου HTTP.

- 1.1 Ποιο είναι το φίλτρο απεικόνισης που εφαρμόσατε;
- 1.2 Ποια είναι η έκδοση του πρωτοκόλλου HTTP που χρησιμοποιεί ο πλοηγός ιστού σας;
- 1.3 Ποια είναι η έκδοση του πρωτοκόλλου HTTP που χρησιμοποιεί ο εξυπηρετητής ιστού;
- 1.4 Ποια είναι η ονομασία της μεθόδου στο πρώτο αίτημα HTTP από τον υπολογιστή σας προς τον εξυπηρετητή ιστού; [Υπόδειξη: Για τα ονόματα των διαθέσιμων μεθόδων συμβουλευθείτε την ιστοσελίδα <http://www.networksorcery.com/enp/protocol/http.htm>].

² Η αρχική γραμμή και οι επικεφαλίδες οφείλουν και αυτές να λήγουν με τους χαρακτήρες <CR><LF>.

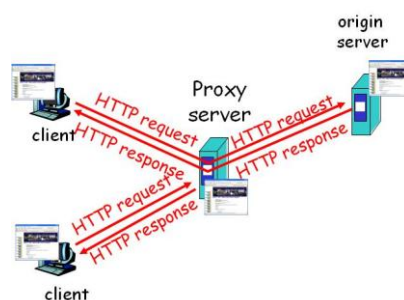
- 1.5 Ποιες γλώσσες εκτός της Αγγλικής δηλώνει ενδεχομένως ο πλοηγός ιστού σας ότι μπορεί να δεχτεί από τον εξυπηρετητή;
- 1.6 Ποιος είναι ο κωδικός κατάστασης και ποια η επεξηγηματική λέξη που επιστρέφει ο εξυπηρετητής ιστού στην απόκριση HTTP;
- 1.7 Τι μήκος έχει το σώμα της απόκρισης και τι είδους περιεχόμενο μεταφέρει;
- 1.8 Στο περιεχόμενο του σώματος υπάρχει ο τίτλος της ιστοσελίδας. Ποιος είναι αυτός και σε ποιο σημείο εμφανίζεται στον πλοηγό σας;
- 1.9 Εφαρμόστε κατάλληλο φίλτρο ώστε να παραμείνουν μόνο τα πρώτα τεμάχια TCP των τριμερών χειρασιών που διεξήχθησαν με τον εξυπηρετητή my.ntua.gr. Ποια είναι η σύνταξή του;
- 1.10 Πόσες συνδέσεις TCP έγιναν και ποιες οι αντίστοιχες θύρες πηγής;
- 1.11 Εφαρμόστε νέο φίλτρο απεικόνισης ώστε να παραμείνουν μόνο τα αιτήματα HTTP (HTTP requests) προς τον εξυπηρετητή ιστού. Ποια είναι η σύνταξή του;
- 1.12 Πόσες εντολές HTTP απέστειλε ο υπολογιστής σας προς τον εξυπηρετητή ιστού;
- 1.13 Εκτός από την ίδια την ιστοσελίδα, ο πλοηγός ιστού ζήτησε και κάποιες εικόνες. Εφαρμόστε νέο φίλτρο ώστε να παραμείνουν μόνο οι αποκρίσεις HTTP (HTTP responses). Ποια είναι η σύνταξή του;
- 1.14 Πόσες εικόνες κατέβασε ο πλοηγός ιστού;

Εφαρμόστε και πάλι το φίλτρο της ερώτησης 1.1 ώστε να παραμείνουν τα μηνύματα του πρωτοκόλλου HTTP.

- 1.15 Ζητήθηκε από τον πλοηγό ιστού σας να κατέβουν *παράλληλα* κάποιες από τις παραπάνω εικόνες; Αιτιολογήστε την απάντησή σας παρατηρώντας το κατά πόσον οι αιτήσεις γίνονται με χρήση διαφορετικών συνδέσεων TCP.
- 1.16 Ζητήθηκε από τον πλοηγό ιστού σας να κατέβουν *ακολουθιακά* κάποιες από τις παραπάνω εικόνες; Αιτιολογήστε την απάντησή σας παρατηρώντας το κατά πόσο χρησιμοποιείται η ίδια σύνδεση TCP για διαφορετικές αιτήσεις.
- 1.17 Μια από τις εικόνες που κατέβασε ο πλοηγός σας εμφανίζεται στην καρτέλα πριν από το όνομα της σελίδας. Ποιο είναι το όνομα του αρχείου που την περιέχει; [Υπόδειξη: Αναζητήστε τις λέξεις *favorite icon browser* στο Google.]
- 1.18 Καταγράψτε την εντολή HTTP με την οποία ο πλοηγός ιστού σας κατέβασε αυτήν την εικόνα.

2. Επανάκτηση μη-τροποποιημένης HTML σελίδας

Η ανάγκη για μείωση του χρόνου εξυπηρέτησης των αιτήσεων των πελατών και η ανάγκη για μείωση του φορτίου στις ζεύξεις πρόσβασης των τοπικών δικτύων οδήγησαν στην επινόηση της τεχνικής της ενδιάμεσης αποθήκευσης (Web caching). Στόχος είναι η ταχύτερη ικανοποίηση του αιτήματος του πελάτη ακόμη και χωρίς την ανάμιξη του εξυπηρετητή πηγής. Ο χρήστης, λοιπόν, ρυθμίζει τον πλοηγό του ώστε η περιήγηση στον παγκόσμιο ιστό να γίνεται μέσω ενδιάμεσου εξυπηρετητή (proxy server). Εάν τα ζητούμενα αντικείμενα περιέχονται εκεί (δηλαδή, τα έχει ήδη αναζητήσει κάποιος προηγουμένως) επιστρέφονται από αυτόν. Διαφορετικά, ο ενδιάμεσος εξυπηρετητής τα ζητά από τον εξυπηρετητή πηγής, κρατά ένα αντίγραφο και τα προωθεί στον πελάτη.



Μια συμπληρωματική λύση στο ίδιο πρόβλημα δόθηκε μέσω των πλοηγών ιστού. Συγκεκριμένα, οι πλοηγοί κάνουν τοπική αποθήκευση, δηλαδή, κρατούν στο δίσκο του υπολογιστή διάφορα αντικείμενα από τις ιστοσελίδες που επισκέπτονται. Όταν λοιπόν ο χρήστης επισκεφτεί ξανά την ίδια σελίδα ή ζητήσει ανανέωση (refresh) αυτής που έχει επισκεφτεί, τότε αντί για την HTTP μέθοδο GET εκτελείται η conditional GET. Σύμφωνα με την τελευταία, το ζητούμενο αντικείμενο μεταφέρεται μόνο εάν ισχύουν οι συνθήκες που περιγράφονται στην επικεφαλίδα της εντολής. Για μια πιο λεπτομερή περιγραφή δείτε το παράδειγμα ανταλλαγής μηνυμάτων μεταξύ ενός πλοηγού ιστού και ενός εξυπηρετητή κατά την επανεπίσκεψη ή την ανανέωση μιας ιστοσελίδας του αρχείου https://www.eventhelix.com/RealtimeMantra/Networking/web_browser_caching_tutorial.pdf.

Αφού ξεκινήσετε μια νέα καταγραφή με το Wireshark, επισκεφθείτε με τον πλοηγό ιστού τη σελίδα: <http://edu-dy.cn.ntua.gr/get2.html>. **Αμέσως** ανανεώστε τη σελίδα πατώντας το κουμπί Reload. Κατόπιν σταματήστε την καταγραφή και εφαρμόστε κατάλληλο φίλτρο απεικόνισης ώστε να παραμείνουν μόνο μηνύματα του πρωτοκόλλου HTTP.

- 2.1 Παρατηρήστε τις επικεφαλίδες πρωτοκόλλου HTTP του πρώτου μηνύματος τύπου GET με το οποίο ζητείται το κατέβασμα της σελίδας get2.html. Υπάρχει γραμμή IF-MODIFIED-SINCE;
- 2.2 Ποιος είναι ο κωδικός κατάστασης (status code) που επιστρέφει ο εξυπηρετητής ως απόκριση στο μήνυμα αυτό;
- 2.3 Παρατηρήστε τις επικεφαλίδες πρωτοκόλλου HTTP που περιέχονται στην απάντηση. Πότε τροποποιήθηκε για τελευταία φορά το αρχείο που κατεβάσατε;
- 2.4 Ποιο είναι το μέγεθος σε byte του περιεχομένου που επιστρέφεται στον πλοηγό ιστού;
- 2.5 Τι είδους αρχείο είναι το περιεχόμενο της σελίδας get2.html;
- 2.6 Σε ποιες γλώσσες είναι γραμμένο το κείμενο της ιστοσελίδας get2.html;
- 2.7 Ποιο σύνολο χαρακτήρων χρησιμοποιείται για το κείμενο της ιστοσελίδας get2.html; [Υπόδειξη: Αναπτύξτε το περιεχόμενο του πεδίου δεδομένων της απάντησης στην οθόνη με τις λεπτομέρειες επικεφαλίδας.]
- 2.8 Τι παρατηρείτε σχετικά με την κωδικοποίηση των ελληνικών χαρακτήρων της σελίδας get2.html; [Υπόδειξη: Παρατηρήστε στο παράθυρο του Wireshark την αναπαράσταση του περιεχομένου του πεδίου δεδομένων της απάντησης σε δεκαεξαδική και ASCII μορφή.]
- 2.9 Παρατηρήστε τις επικεφαλίδες πρωτοκόλλου HTTP του δεύτερου μηνύματος τύπου GET με το οποίο ζητείται το κατέβασμα της σελίδας get2.html. Υπάρχει γραμμή IF-MODIFIED-SINCE;
- 2.10 Εάν ναι, ποια ημερομηνία και ώρα αναφέρεται;
- 2.11 Ποιος είναι ο κωδικός κατάστασης (status code) που επιστρέφει ο εξυπηρετητής ως απόκριση στο μήνυμα αυτό;
- 2.12 Ο εξυπηρετητής επέστρεψε ρητά τα περιεχόμενα του αρχείου get2.html; Αιτιολογήστε την απάντησή σας.
- 2.13 Πλην των παραπάνω αιτημάτων HTTP υπήρξε άλλο προς τον εξυπηρετητή ιστού; Εάν ναι, ποιος ήταν ο σκοπός του;

3. Επανάκτηση τροποποιημένης HTML σελίδας

Όπως είδατε παραπάνω το ζητούμενο αντικείμενο δεν είχε τροποποιηθεί και γι' αυτό τον λόγο ο πλοηγός ιστού απεικόνισε στο χρήστη το ήδη αποθηκευμένο αντικείμενο. Στην αντίθετη περίπτωση, δηλαδή, εάν το ζητούμενο αντικείμενο είχε τροποποιηθεί, τότε ο πλοηγός ιστού θα κατέβαζε το τροποποιημένο αντικείμενο από τον εξυπηρετητή, θα το αποθήκευε στην προσωρινή μνήμη cache και παράλληλα θα το απεικόνιζε στο χρήστη.

Ξεκινήστε μια νέα καταγραφή με το Wireshark. Με τον πλοηγό ιστού επισκεφθείτε τη σελίδα <http://edu-dy.cn.ntua.gr/get.html>. **Αμέσως** ανανεώστε τη σελίδα πατώντας το κουμπί Reload. Περιμένετε **λίγο περισσότερο** από ένα λεπτό και στη συνέχεια ανανεώστε τη σελίδα πατώντας το κουμπί Reload. Κατόπιν σταματήστε την καταγραφή και εφαρμόστε κατάλληλο φίλτρο απεικόνισης ώστε να παραμείνουν μόνο μηνύματα του πρωτοκόλλου HTTP.

- 3.1 Πόσες και τι τύπου εντολές πρωτοκόλλου HTTP έδωσε ο πλοηγός ιστού σας;
- 3.2 Ποιος είναι ο κωδικός κατάστασης (status code) που επιστρέφει ο εξυπηρετητής ως απόκριση στο πρώτο μήνυμα HTTP τύπου GET του πλοηγού ιστού για το κατέβασμα της σελίδας get.html;
- 3.3 Πότε τροποποιήθηκε για τελευταία φορά το περιεχόμενο αυτό, σύμφωνα με την απάντηση στο πρώτο μήνυμα HTTP τύπου GET;
- 3.4 Ποιο είναι το μέγεθος σε byte του περιεχομένου που επιστρέφεται στον πλοηγό ιστού;
- 3.5 Παρατηρήστε τις επικεφαλίδες πρωτοκόλλου HTTP του δεύτερου μηνύματος τύπου GET με το οποίο ζητείται το κατέβασμα της σελίδας get.html. Υπάρχει γραμμή IF-MODIFIED-SINCE;
- 3.6 Ποιος είναι ο κωδικός κατάστασης που επιστρέφει ο εξυπηρετητής ως απόκριση στο δεύτερο μήνυμα HTTP τύπου GET του πλοηγού ιστού για το κατέβασμα της σελίδας get.html;
- 3.7 Παρατηρήστε τις επικεφαλίδες πρωτοκόλλου HTTP του τρίτου μηνύματος τύπου GET με το οποίο ζητείται το κατέβασμα της σελίδας get.html. Υπάρχει γραμμή IF-MODIFIED-SINCE;
- 3.8 Ποιος είναι ο κωδικός κατάστασης που επιστρέφει ο εξυπηρετητής ως απόκριση στο τρίτο μήνυμα HTTP τύπου GET του πλοηγού ιστού για το κατέβασμα της σελίδας get.html;
- 3.9 Πότε τροποποιήθηκε τελευταία φορά το περιεχόμενο σύμφωνα με την απάντηση στο τρίτο μήνυμα HTTP τύπου GET;
- 3.10 Ποιο είναι το μέγεθος σε byte του περιεχομένου που επιστρέφεται στον πλοηγό ιστού;
- 3.11 Γιατί σας ζητήθηκε να περιμένετε περισσότερο από ένα λεπτό προτού ανανεώσετε τη σελίδα; (Σημείωση: κάποιοι εξυπηρετητές θέτουν την ώρα τελευταίας τροποποίησης των αρχείων ίση με την τρέχουσα ώρα κάθε λεπτό)

4. Ανάκτηση εκτενούς σελίδας HTML

Υπάρχουν περιπτώσεις όπου το ζητούμενο αρχείο είναι αρκετά μεγάλο και δεν χωράει σε ένα τεμάχιο TCP. Στην περίπτωση αυτή το μήνυμα HTTP τεμαχίζεται στο στρώμα μεταφοράς. Ξεκινήστε μια νέα καταγραφή με το Wireshark. Με τον πλοηγό ιστού επισκεφθείτε τη σελίδα τη σελίδα <http://edu-dy.cn.ntua.gr/long.html>. Σταματήστε την καταγραφή μόλις ολοκληρωθεί το φόρτωμα της σελίδας.

- 4.1 Πόσες συνδέσεις TCP έγιναν;
- 4.2 Ποιο είναι το μέγεθος του MSS που ανακοινώνει η κάθε πλευρά;

Εφαρμόστε κατάλληλο φίλτρο απεικόνισης ώστε να παραμείνουν μόνο μηνύματα του πρωτοκόλλου HTTP.

- 4.3 Πόσα τεμάχια TCP απαιτήθηκαν για να μεταδοθεί το αίτημα HTTP;
- 4.4 Να καταγραφεί η γραμμή κατάστασης (status line) της απόκρισης του εξυπηρετητή.
- 4.5 Πόσα τεμάχια TCP απαιτήθηκαν για να μεταδοθεί η απόκριση HTTP;
- 4.6 Να καταγραφεί το μήκος του περιεχομένου του αρχείου long.html όπως αναφέρεται στην προηγούμενη απόκριση HTTP.

Εφαρμόστε νέο φίλτρο απεικόνισης στο Wireshark ώστε να παραμείνει μόνο η κίνηση IP που προέρχεται από τον εξυπηρετητή ιστού.

- 4.7 Ποια είναι η σύνταξη του παραπάνω φίλτρου;
- 4.8 Πιο τεμάχιο TCP από αυτά του ερωτήματος 4.5 περιλαμβάνει τη γραμμή κατάστασης του πρωτοκόλλου HTTP που καταγράψατε προηγουμένως; [Υπόδειξη: Αναζητήστε το περιεχόμενο της γραμμής κατάστασης στο παράθυρο με τα περιεχόμενα του επιλεγμένου πλαισίου.]
- 4.9 Ποιο είναι το μέγεθος του περιεχομένου HTTP που μεταφέρει κάθε ένα από τα τεμάχια αυτά πλην του τελευταίου; [Υπόδειξη: Αναπτύξτε το περιεχόμενο της επικεφαλίδας TCP στο παράθυρο με τις λεπτομέρειες επικεφαλίδας.]
- 4.10 Γιατί το μέγεθος των πλαισίων Ethernet που μεταφέρουν τα προηγούμενα τεμάχια TCP πλην του τελευταίου είναι σταθερό;
- 4.11 Πώς προκύπτει το μέγεθος του τελευταίου εξ αυτών;

5. Ανάκτηση HTML σελίδας με ενσωματωμένα αντικείμενα

Μια σελίδα HTML είναι μια συλλογή αντικειμένων (π.χ. ενός αρχείου HTML, διαφόρων εικόνων JPEG, αρχείων ήχου, βίντεο, κλπ) και μπορεί να περιέχει παραπομπές προς άλλες ιστοσελίδες. Στην άσκηση αυτή θα μελετήσετε την κίνηση που παράγεται στην περίπτωση όπου ο πλοηγός ιστού κατεβάζει μια σελίδα που περιέχει ενσωματωμένα αντικείμενα (embedded objects).

Ξεκινήστε μια νέα καταγραφή με το Wireshark. Με τον πλοηγό ιστού επισκεφθείτε τη σελίδα <http://edu-dy.cn.ntua.gr/links.html>. Η σελίδα αυτή περιλαμβάνει διευθύνσεις URL που αναφέρονται σε δύο εικόνες που βρίσκονται σε διαφορετικούς από τον edu-dy.cn.ntua.gr εξυπηρετητές ιστού. Μόλις η σελίδα φορτωθεί πλήρως, σταματήστε την καταγραφή.

- 5.1 Εφαρμόστε κατάλληλο φίλτρο ώστε να παραμείνουν μόνο τα αιτήματα HTTP προς εξυπηρετητές ιστού. Πόσες εντολές HTTP τύπου GET έχει καταγράψει το Wireshark;
- 5.2 Πόσες συνδέσεις TCP έγιναν για το κατέβασμα της σελίδας long.html;
- 5.3 Καταγράψτε τα ονόματα των αρχείων εικόνων που ζήτησε ο πλοηγός.
- 5.4 Ποια είναι τα ονόματα των εξυπηρετητών ιστού με τους οποίους έγιναν συνδέσεις TCP; *[Υπόδειξη: Κάντε κλικ στους υπερσυνδέσμους.]*
- 5.5 Ποια είναι η διεύθυνση IPv4 ή IPv6 προορισμού κάθε εντολής HTTP τύπου GET; Εξηγήστε.

Όνοματεπώνυμο:	Ομάδα:
Όνομα PC/ΛΣ:	Ημερομηνία: / /
Διεύθυνση IP: . . .	Διεύθυνση MAC: - - - - -

Εργαστηριακή Άσκηση 11

Πλοήγηση στον παγκόσμιο ιστό

Απαντήστε στα ερωτήματα στον χώρο που σας δίνεται παρακάτω και στην πίσω σελίδα εάν δεν επαρκεί. Το φυλλάδιο αυτό θα παραδοθεί στον επιβλέποντα.

1

- 1.1
- 1.2
- 1.3
- 1.4
- 1.5
- 1.6
- 1.7
- 1.8
-
- 1.9
- 1.10
-
-
-
- 1.11
- 1.12
- 1.13
- 1.14
- 1.15
-
-
-
- 1.16
-
-
-
- 1.17
- 1.18

2

- 2.1

2.2
2.3
2.4
2.5
2.6
2.7
2.8

2.9
2.10
2.11
2.12

2.13

3	
3.1
3.2
3.3
3.4
3.5
3.6
3.7
3.8
3.9
3.10
3.11

4	
4.1
4.2
4.3
4.4
4.5
4.6

4.7

4.8

4.9

.....

4.10

.....

4.11

.....

.....

5

5.1

5.2

5.3

.....

5.4

.....

.....

5.5

.....

.....