

Εργαστηριακή Άσκηση 9 Πλοήγηση στον παγκόσμιο ιστό

Ο σκοπός αυτού του εργαστηρίου είναι η εξέταση του πρωτοκόλλου εφαρμογής HTTP που χρησιμοποιείται ευρύτατα στο διαδίκτυο για την πλοήγηση στον ιστό, με τη βοήθεια του αναλυτή πρωτοκόλλων Wireshark.

Εισαγωγή

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, 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, ένα **Ομοιόμορφος Εντοπιστής Πόρων URL (Uniform Resource Locator)** που στην ουσία παίζει το 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 της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ είναι: <http://www.ece.ntua.gr/index.php>. Το URL αυτό αποτελείται από τρία μέρη: το πρωτόκολλο (*http*), το όνομα DNS της μηχανής στην οποία εντοπίστηκε η σελίδα (*www.ece.ntua.gr*) και το όνομα του αρχείου που περιέχει τη σελίδα (*index.php*), με κάποια σημεία στίξης που διαχωρίζουν τα κομμάτια.

Ο σκοπός αυτού του εργαστηρίου 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!).

1. Ανάκτηση HTML σελίδας

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

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

transfer protocol). Το HTTP μεταφέρει πόρους (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. PHP script), είτε να είναι έγγραφο διαθέσιμο σε πολλές γλώσσες 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 request) και ο εξυπηρετητής στέλνει μία απάντηση (HTTP response) που περιέχει διάφορα αντικείμενα. Μετά η σύνδεση απελευθερώνεται, δηλαδή, το HTTP δε διατηρεί πληροφορία κατάστασης (stateless) μεταξύ διαδοχικών δοσοληφιών (transactions).

Όπως έχει ήδη αναφερθεί, το HTTP είναι πρωτόκολλο επιπέδου εφαρμογής και ακολουθεί το μοντέλο πελάτη/εξυπηρετητή (client/server model). Οι μορφές της αίτησης και απόκρισης HTTP είναι ταυτόσημες και χρησιμοποιούν Αγγλικές λέξεις. 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), 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 και μια Αγγλική λέξη που περιγράφει τον κωδικό κατάστασης. Οι γραμμές επικεφαλίδων παρέχουν πληροφορίες για την αίτηση ή την απόκριση ή για το αντικείμενο που περιέχει το κυρίως σώμα. Εμφανίζονται όπως το συνήθες 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).

Μέχρι σήμερα έχουν αναπτυχθεί 2 εκδόσεις του πρωτοκόλλου, η HTTP 1.0 (περιγράφεται στο RFC 1945) και η HTTP 1.1 (περιγράφεται στο RFC 2068). Οι συνδέσεις HTTP διακρίνονται σε μη επίμονες (Non-persistent HTTP) και σε επίμονες (Persistent HTTP). Στις μη επίμονες συνδέσεις HTTP, πάνω από μία σύνδεση TCP μπορεί να αποσταλεί το πολύ ένα αντικείμενο κάθε φορά. Το πρωτόκολλο HTTP/1.0 χρησιμοποιεί μη επίμονες συνδέσεις. Στην περίπτωση των επίμονων συνδέσεων HTTP, πάνω από την ίδια σύνδεση TCP μπορούν να αποσταλούν πολλαπλά αντικείμενα μεταξύ πελάτη-εξυπηρετητή. Οι επίμονες συνδέσεις χρησιμοποιούνται από το πρωτόκολλο HTTP/1.1. Περισσότερες πληροφορίες για το HTTP μπορείτε να βρείτε στην ιστοθέση http://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol καθώς και <http://www.w3.org/Protocols/>.

Σε αυτή την άσκηση θα καταγραφούν τα μηνύματα HTTP που παράγονται κατά την επίσκεψη μιας ιστοσελίδας με χρήση του Internet Explorer. Προτού αρχίσετε την καταγραφή φροντίστε να αδειάσετε την προσωρινή μνήμη (cache) του πλοηγού. Επιλέξτε *Tools* → *Internet Options*, στην πινακίδα (tab) *General* πιέστε το κουμπί *Delete*, στην οθόνη που θα εμφανισθεί επιλέξτε το *Delete Files*, επιβεβαιώστε την πρόθεσή σας, περιμένετε να ολοκληρωθεί η διαγραφή και κλείστε τα παράθυρα διαλόγου.

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

Επειδή τα μηνύματα HTTP μεταφέρονται σε περισσότερα από ένα πακέτα, αφού ξεκινήσετε το Wireshark, ακολουθήστε από το μενού του κεντρικού παραθύρου τη διαδρομή *Edit* → *Preferences...*, κάνετε κλικ στο σύμβολο δίπλα στο Protocols στην αριστερή πλευρά του παραθύρου, κατόπιν εντοπίστε και κάνετε κλικ στο πρωτόκολλο HTTP και βεβαιωθείτε ότι όλες οι επιλογές περί ανασύνθεσης και αποσυμπίεσης είναι επιλεγμένες. Στη συνέχεια, στο πρωτόκολλο TCP βεβαιωθείτε ότι το *Allow subdissector to reassemble TCP streams* είναι επιλεγμένο και φροντίστε το *Validate the TCP checksum if possible*³ να **μην** είναι επιλεγμένο. Τέλος, πιάστε OK για να κλείσει το παράθυρο και εφαρμοστούν οι αλλαγές σας. Με τις επιλογές αυτές μηνύματα που μεταφέρονται σε περισσότερα από ένα πακέτα θα αποκωδικοποιηθούν από το Wireshark ως πλήρη μηνύματα HTTP και όχι αποσπασματικά.

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 μήνυμα HTTP από τον υπολογιστή σας προς τον εξυπηρετητή ιστού; [Υπόδειξη: Συμβουλευθείτε την ιστοσελίδα <http://www.networksorcery.com/enp/protocol/http.htm>].
- 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;
- 1.7 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;
- 1.8 Πόσες συνδέσεις TCP έγιναν και ποιες οι αντίστοιχες θύρες πηγής;
- 1.9 Αφού απενεργοποιήσετε το φίλτρο απεικόνισης της ερώτησης 1.7, εφαρμόστε νέο φίλτρο ώστε να παραμείνουν μόνο οι εντολές HTTP προς τον εξυπηρετητή ιστού (HTTP requests). Ποια είναι η σύνταξή του;
- 1.10 Πόσες εντολές HTTP απέστειλε ο υπολογιστής σας προς τον εξυπηρετητή ιστού;
- 1.11 Εκτός από την ίδια την ιστοσελίδα, ο πλοηγός ιστού ζήτησε και κάποιες εικόνες. Πόσες εικόνες κατέβασε ο πλοηγός ιστού;

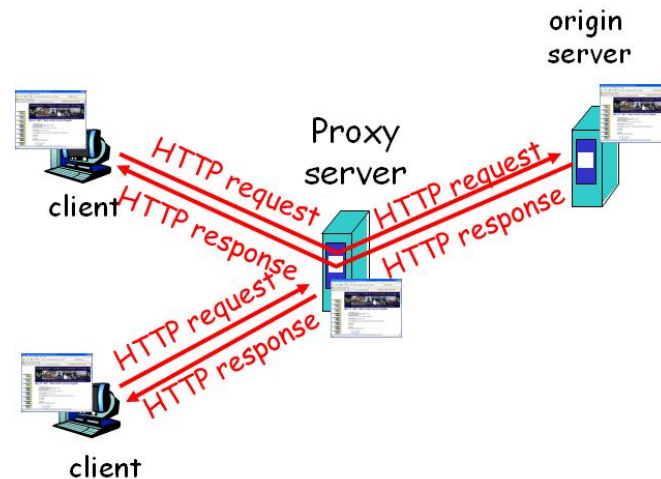
Εφαρμόστε και πάλι το φίλτρο της ερώτησης 1.1 ώστε να παραμείνουν τα μηνύματα του πρωτοκόλλου HTTP.

- 1.12 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.
- 1.13 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.
- 1.14 Μια από τις εικόνες που κατέβασε ο πλοηγός ιστού σας εμφανίζεται δίπλα από το URL <http://my.ntua.gr> στο πεδίο διευθύνσεων (address bar). Ποιο είναι το όνομά της; [Υπόδειξη: Αναζητείστε τις λέξεις *favorites icon* στο Google.]
- 1.15 Καταγράψτε την εντολή HTTP με την οποία ο πλοηγός ιστού σας κατέβασε αυτήν την εικόνα.

³ Η επιβεβαίωση του TCP checksum (επειδή αυτή γίνεται στην κάρτα δικτύου) παρενοχλεί τη διαδικασία επανασύνθεσης.

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 ήταν 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 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 περισσότεροι πλοηγοί σήμερα 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, 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.

Σε αυτή την άσκηση θα καταγραφούν τα μηνύματα HTTP που παράγονται κατά την επίσκεψη μιας ιστοσελίδας με χρήση του Firefox. Προτού ξεκινήσετε την καταγραφή φροντίστε να αδειάσετε την προσωρινή μνήμη (cache) του πλοηγού. Ακολουθείστε τη διαδρομή *Tools* → *Clear Private Data*, επιλέξτε το cache στον πίνακα που θα εμφανισθεί, επιβεβαιώστε την πρόθεσή σας, περιμένετε να ολοκληρωθεί η διαγραφή και κλείστε το παράθυρο διαλόγου.

Αφού ξεκινήσετε μια νέα καταγραφή με το Wireshark, επισκεφθείτε με το Firefox 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 ης σελίδας get2.html;
- 2.6 Ποια σύνολα χαρακτήρων (character sets) δηλώνει ο πλοηγός ιστού ότι μπορεί δεχτεί; [Υπόδειξη: Παρατηρείστε το πρώτο μήνυμα τύπου GET που έστειλε ο πλοηγός ιστού σας.]
- 2.7 Σε ποιες γλώσσες είναι γραμμένο το κείμενο της ιστοσελίδας get2.html;

- 2.8 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.]
- 2.9 Τι παρατηρείτε σχετικά με την κωδικοποίηση των ελληνικών χαρακτήρων της σελίδας get2.html; [Υπόδειξη: Παρατηρείστε στο παράθυρο του Wireshark την αναπαράσταση του περιεχομένου του πεδίου δεδομένων της απάντησης σε δεκαεξαδική και ASCII μορφή.]
- 2.10 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;
- 2.11 Εάν ναι, ποια ημερομηνία και ώρα αναφέρεται;
- 2.12 Ποιος είναι ο κωδικός κατάστασης (status code) που επιστρέφει ο εξυπηρετητής ως απόκριση στο μήνυμα αυτό;
- 2.13 Ο εξυπηρετητής επέστρεψε ρητά τα περιεχόμενα του αρχείου get2.html; Αιτιολογήστε την απάντησή σας.
- 2.14 Στην καταγραφή των μηνυμάτων HTTP φαίνεται ότι ο πλοηγός ζήτησε το κατέβασμα του εικονιδίου favicon.ico. Ποιος είναι ο κωδικός κατάστασης (status code) που επιστρέφει ο εξυπηρετητής ως απόκριση στο μήνυμα αυτό;

3. Επανάκτηση τροποποιημένης HTML σελίδας

Όπως είδαμε παραπάνω το ζητούμενο αντικείμενο δεν είχε τροποποιηθεί και γι' αυτό το 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 στη μνήμη cache και παράλληλα θα το απεικόνιζε στο χρήστη.

Ξεκινήστε μια νέα καταγραφή με το Wireshark. Με τον Firefox επισκεφθείτε τη σελίδα <http://edu-dy.cn.ntua.gr/get.html>. 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 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;
- 3.6 Πότε τροποποιήθηκε τελευταία φορά το περιεχόμενο σύμφωνα με την απάντηση στο δεύτερο μήνυμα HTTP τύπου GET;
- 3.7 Ποιο είναι το μέγεθος σε byte του περιεχομένου που επιστρέφεται στον πλοηγό ιστού;
- 3.8 Γιατί σας ζητήθηκε να περιμένετε περισσότερο από ένα λεπτό προτού ανανεώσετε τη σελίδα; (Σημείωση: κάποιιοι εξυπηρετητές θέτουν την ώρα τελευταίας τροποποίησης των αρχείων ίση με την τρέχουσα ώρα κάθε λεπτό)

4. Ανάκτηση HTML σελίδας με ενσωματωμένα αντικείμενα

Μια σελίδα HTML είναι μια συλλογή αντικειμένων (π.χ. ενός αρχείου HTML, διαφόρων εικόνων JPEG, αρχείων ήχου, Java applets κλπ) και μπορεί να περιέχει παραπομπές προς άλλες ιστοσελίδες. Στην άσκηση αυτή θα μελετήσουμε την κίνηση που παρατηρείται 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.

- 4.1 Εφαρμόστε κατάλληλο φίλτρο ώστε να παραμείνουν μόνο τα **πρώτα** τεμάχια TCP των τριμερών χειρασιών που διεξήχθησαν. Πόσες συνδέσεις TCP έγιναν;
- 4.2 Με ποιους υπολογιστές έγιναν οι συνδέσεις TCP;
- 4.3 Εφαρμόστε κατάλληλο φίλτρο ώστε να παραμείνουν μόνο οι εντολές HTTP προς εξυπηρετητές ιστού. Πόσες εντολές HTTP τύπου GET έχει καταγράψει το Wireshark;
- 4.4 Ποια είναι η διεύθυνση IP προορισμού κάθε εντολής HTTP τύπου GET; Εξηγήστε.
- 4.5 Καταγράψτε τα ονόματα των αρχείων εικόνων που κατέβασε ο πλοηγός από κάθε εξυπηρετητή.

5. Ανάκτηση εκτενούς σελίδας HTML

Στα προηγούμενα ερωτήματα τα αρχεία που ζητήσατε να κατέβουν ήταν απλά και σύντομα HTML αρχεία. Ωστόσο 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.

- 5.1 Πόσες συνδέσεις TCP έγιναν;
- 5.2 Ποιο είναι το μέγεθος της MSS που συμφωνήθηκε;

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

- 5.3 Πόσες αιτήσεις HTTP τύπου GET στάλθηκαν από τον πλοηγό ιστού;
- 5.4 Να καταγραφεί η γραμμή κατάστασης (status line) της απόκρισης που επιστρέφει ο εξυπηρετητής.
- 5.5 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.

- 5.6 Ποια είναι η σύνταξη του παραπάνω φίλτρου;
- 5.7 Πόσα τεμάχια TCP, που περιείχαν δεδομένα, χρειάστηκαν για τη μεταφορά της απόκρισης στο μήνυμα HTTP τύπου GET; [Υπόδειξη: Μαζί με το τελευταίο τεμάχιο της σειράς το Wireshark εμφανίζει στο παράθυρο καταγεγραμμένων πακέτων την ανασύνθεση της απόκρισης HTTP.]
- 5.8 Πιο τεμάχιο TCP του προηγούμενου ερωτήματος περιλαμβάνει τη γραμμή κατάστασης του πρωτοκόλλου HTTP; [Υπόδειξη: Αναζητήστε το περιεχόμενο της γραμμής κατάστασης στο παράθυρο με τα περιεχόμενα του επιλεγμένου πλαισίου.]
- 5.9 Ποιο είναι το μέγεθος του περιεχομένου HTTP που μεταφέρει κάθε ένα από τα τεμάχια αυτά; [Υπόδειξη: Αναπτύξτε το περιεχόμενο της επικεφαλίδας TCP στο παράθυρο με τις λεπτομέρειες επικεφαλίδας.]
- 5.10 Γιατί το μέγεθος των πλαισίων Ethernet που μεταφέρουν τα προηγούμενα τεμάχια TCP (πλην του τελευταίου) είναι σταθερό;
- 5.11 Πώς προκύπτει το μέγεθος του τελευταίου εξ αυτών;

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

Εργαστηριακή Άσκηση 9 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

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
-
-
- 2.14

- 3**
- 3.1
- 3.2
- 3.3
- 3.4
- 3.5
- 3.6
- 3.7
- 3.8
-
-

- 4**
- 4.1
- 4.2
-
- 4.3
- 4.4
-
- 4.5
-

- 5**
- 5.1
- 5.2
- 5.3
- 5.4

5.5

5.6

5.7

5.8

5.9

5.10

.....

5.11

.....

.....