Εργαστήριο Wireshark: HTTP

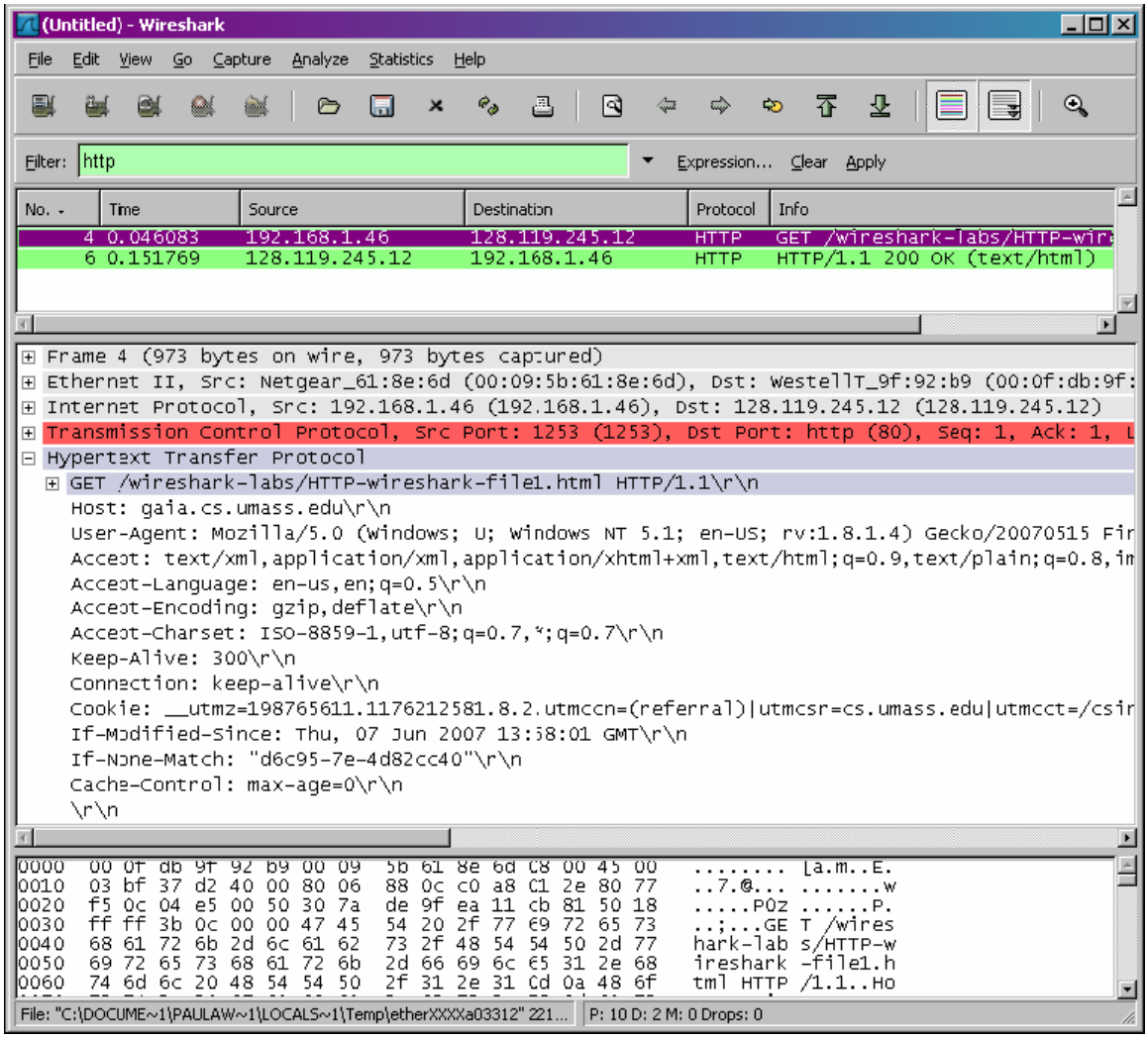
Μετά την πρώτη γεύση του packet sniffer Wireshark στο εισαγωγικό εργαστήριο, είμαστε έτοιμοι να χρησιμοποιήσουμε το Wireshark για να εξετάσουμε τα πρωτόκολλα σε λειτουργία. Στο εργαστήριο αυτό θα διερευνήσουμε μερικές πλευρές του πρωτοκόλλου HTTP: τη βασική αλληλεπίδραση GET/απόκριση, τις μορφές των μηνυμάτων HTTP, την ανάκτηση μεγάλων αρχείων HTML, την ανάκτηση αρχείων HTML με ενσωματωμένα αντικείμενα και την εξουσιοδότηση και ασφάλεια στο HTTP. Ενδεχομένως να θέλετε να κάνετε μία ανασκόπηση της Ενότητας 2.2 του βιβλίου πριν ξεκινήσετε αυτό το εργαστήριο.

1. Η Βασική Αλληλεπίδραση GET/Απόκριση στο HTTP

Ας ξεκινήσουμε την διερεύνηση του HTTP φορτώνοντας ένα πολύ απλό αρχείο HTML, ένα αρχείο το οποίο είναι πολύ μικρό και δεν περιέχει ενσωματωμένα αντικείμενα. Ακολουθήστε τα παρακάτω βήματα:

1. Ξεκινήστε τον web browser σας.
2. Ξεκινήστε τον packet sniffer Wireshark όπως περιγράφεται στο εισαγωγικό εργαστήριο (αλλά όμως μην ξεκινήσετε τη σύλληψη πακέτων ακόμη). Εισάγετε "http" (χωρίς τα εισαγωγικά) στο παράθυρο προδιαγραφών του φίλτρου παρουσίασης ώστε στο παράθυρο καταλόγου πακέτων να παρουσιάζονται αργότερα μόνο τα συλλαμβανόμενα μηνύματα HTTP. (Στο σημείο αυτό ενδιαφερόμαστε μόνο για το πρωτόκολλο HTTP και δεν θέλουμε να δούμε όλα τα πακέτα που συλλαμβάνονται.)
3. Περιμένετε λίγο περισσότερο από ένα λεπτό (σύντομα θα διαπιστώσετε για ποιο λόγο) και μετά αρχίστε τη σύλληψη πακέτων από το Wireshark.
4. Εισάγετε το ακόλουθο URL στον browser
   http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html.
   Στον browser θα πρέπει να εμφανισθεί ένα πολύ απλό αρχείο HTML με μία γραμμή μόνο.
5. Διακόψτε τη σύλληψη πακέτων από το Wireshark.

Το παράθυρο Wireshark θα πρέπει να μοιάζει με το παράθυρο που φαίνεται στο Σχήμα 1. Εάν δεν είστε σε θέση να τρέξετε το Wireshark σε μία ζωντανή σύνδεση δικτύου, μπορείτε να φορτώσετε ένα trace πακέτων το οποίο δημιουργήθηκε ακολουθώντας τα παραπάνω βήματα.



Σχήμα 1: Το παράθυρο Wireshark μετά την ανάκτηση από τον browser σας του trace http://gaia.cs.umass.edu/wireshark-labs/ HTTP-ethereal-file1.html

Στο παράδειγμα του Σχήματος 1 φαίνεται, από το παράθυρο καταλόγου πακέτων, ότι

συνελήφθησαν δύο μηνύματα HTTP: το μήνυμα GET (από τον browser στον web server gaia.cs.umass.edu) και το μήνυμα απόκρισης από τον server στον browser. Στο παράθυρο περιεχομένων πακέτου φαίνονται λεπτομέρειες του επιλεγμένου μηνύματος (στην περίπτωση αυτή, του μηνύματος HTTP GET, το οποίο έχει επισημειωθεί στο παράθυρο καταλόγου πακέτων). Υπενθυμίζεται ότι, αφού το μήνυμα HTTP μεταφέρθηκε μέσα σε ένα TCP segment, το οποίο μεταφέρθηκε μέσα σε ένα IP datagram, το οποίο μεταφέρθηκε μέσα σε ένα πλαίσιο Ethernet, το Wireshark παρουσιάζει πληροφορίες και για τα πακέτα Frame, Ethernet, IP και TCP. Επειδή θέλουμε να ελαχιστοποιήσουμε το ποσό των πληροφοριών που παρουσιάζονται για δεδομένα που δε σχετίζονται με το HTTP (στο εργαστήριο αυτό ενδιαφερόμαστε για το HTTP, ενώ θα εξετάσουμε αυτά τα άλλα πρωτόκολλα σε επόμενα εργαστήρια), βεβαιωθείτε ότι τα κουτάκια που φαίνονται στα αριστερά των πληροφοριών για τα πακέτα Frame, Ethernet, IP και TCP έχουν το σύμβολο (+) (που σημαίνει ότι υπάρχει κρυμμένη πληροφορία η οποία δεν εμφανίζεται) και ότι η γραμμή HTTP έχει το σύμβολο (-) (που σημαίνει ότι εμφανίζονται όλες οι πληροφορίες σχετικά με το μήνυμα HTTP).

> (*Σημείωση:* Αγνοείστε οποιαδήποτε μηνύματα HTTP GET και αποκρίσεων για το favicon.ico. Εάν δείτε μία αναφορά σε αυτό το αρχείο, πρόκειται για τον δικό σας browser που ρωτά αυτόματα τον server εάν έχει το αρχείο ενός εικονιδίου το οποίο θα έπρεπε να απεικονισθεί δίπλα στο URL που εμφανίζεται στο browser σας. Στο εργαστήριο αυτό θα αγνοήσουμε αναφορές σε αυτό το αρχείο.)

Εξετάζοντας την πληροφορία των μηνυμάτων HTTP GET και απόκριση, απαντήστε στις ακόλουθες ερωτήσεις. Εκτυπώσετε πρώτα το μήνυμα GET και το μήνυμα απόκρισης (ο τρόπος με τον οποίο μπορείτε να πάρετε μία εκτύπωση εξηγείται στο εισαγωγικό εργαστήριο Wireshark). Σε κάθε απάντησή σας να υποδεικνύετε το σημείο του μηνύματος που περιέχει την πληροφορία που την αιτιολογεί.

1. Ποια έκδοση του HTTP τρέχει στον browser σας; Ποια έκδοση του HTTP τρέχει στον server;
2. Ποιες γλώσσες υποδεικνύει ο browser στον server ότι μπορεί να αποδεχθεί;
3. Ποια είναι η διεύθυνση IP του υπολογιστή σας; Ποια είναι η διεύθυνση IP του server gaia.cs.umass.edu;
4. Ποιος είναι ο κώδικας κατάστασης (status code) που επιστρέφει ο server στον browser σας;
5. Πότε τροποποιήθηκε για τελευταία φορά στον server το αρχείο HTML το οποίο ανακτήσατε;
6. Πόσα bytes περιεχομένου επιστρέφονται στον browser σας;
7. Εξετάζοντας τα ανεπεξέργαστα δεδομένα στο παράθυρο περιεχομένων πακέτου, διαπιστώνετε ότι μέσα στα δεδομένα περιλαμβάνονται επικεφαλίδες οι οποίες δεν εμφανίζονται στο παράθυρο καταλόγου πακέτων; Εάν υπάρχουν τέτοιες επικεφαλίδες, κατονομάστε μία.

Στην απάντησή σας στην ερώτηση 5 παραπάνω, ενδεχομένως να σας έχει προκαλέσει έκπληξη η διαπίστωση ότι το έγγραφο (document) που έχετε μόλις ανακτήσει τροποποιήθηκε για τελευταία φορά μέσα στο τελευταίο λεπτό πριν το φορτώσετε. Αυτό

οφείλεται στο γεγονός ότι, για το συγκεκριμένο αρχείο, ο server gaia.cs.umass.edu θέτει το χρόνο τελευταίας τροποποίησης του αρχείου στον τρέχοντα χρόνο μία φορά ανά λεπτό. Έτσι, εάν περιμένετε για ένα λεπτό ανάμεσα σε διαδοχικές προσπελάσεις του αρχείου, το αρχείο θα εμφανίζεται ως προσφάτως τροποποιημένο και επομένως ο browser σας θα φορτώνει ένα "νέο" αντίγραφο του εγγράφου.

2. Η Αλληλεπίδραση Υπό συνθήκη GET/Απόκριση στο HTTP

Υπενθυμίζεται, από την Ενότητα 2.2.5 του βιβλίου, ότι οι περισσότεροι web browsers εφαρμόζουν την προσωρινή αποθήκευση (caching) των αντικειμένων και επομένως εκτελούν ένα υπό συνθήκη GET (conditional GET) κατά την ανάκτηση ενός αντικειμένου HTTP. Πριν εκτελέσετε τα παρακάτω βήματα, βεβαιωθείτε ότι η cache του browser σας είναι άδεια. (Για να αδειάσετε την cache στο Netscape 7.0, επιλέξτε *Edit* → *Preferences* → *Advanced* → *Cache* και αδειάστε την cache στη μνήμη και το δίσκο. Για το Firefox, επιλέξτε *Tools* → *Clear Private Data* ή για τον Internet Explorer, επιλέξτε *Tools* → *Internet Options* → *Delete File*. Οι ενέργειες αυτές θα απομακρύνουν τα αποθηκευμένα αρχεία από την cache του browser σας.) Ακολουθήστε τώρα τα παρακάτω βήματα:

- Ξεκινήστε τον web browser σας και βεβαιωθείτε ότι η cache του είναι άδεια όπως συζητήθηκε παραπάνω.
- Ξεκινήστε τον packet sniffer Wireshark.
- Εισάγετε το ακόλουθο URL στον browser
  http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html .
  Στον browser θα πρέπει να εμφανισθεί ένα πολύ απλό αρχείο HTML με πέντε γραμμές.
- Γρήγορα εισάγετε για άλλη μία φορά το ίδιο URL στον browser σας (ή απλά επιλέξτε το κουμπί refresh του browser).
- Σταματήστε τη σύλληψη πακέτων από το Wireshark και εισάγετε "http" στο παράθυρο προδιαγραφών του φίλτρου παρουσίασης ώστε στο παράθυρο καταλόγου πακέτων να παρουσιάζονται αργότερα μόνο τα συλλαμβανόμενα μηνύματα HTTP.

Απαντήστε στις παρακάτω ερωτήσεις:

8. Ελέγξτε τα περιεχόμενα της πρώτης αίτησης HTTP GET από τον browser σας στον server. Υπάρχει η γραμμή "**`If-modified-since`**" στην αίτηση HTTP GET;
9. Ελέγξτε τα περιεχόμενα της απόκρισης του server. Επέστρεψε ο server τα περιεχόμενα του αρχείου; Που βασίζεται το συμπέρασμά σας;
10. Ελέγξτε τώρα τα περιεχόμενα της δεύτερης αίτησης HTTP GET από τον browser σας στον server. Υπάρχει η γραμμή "**`If-modified-since`**" στην αίτηση HTTP GET; Εάν υπάρχει η γραμμή αυτή, τι είδους πληροφορία ακολουθεί την επικεφαλίδα "**`If-modified-since`**";

11. Τι κώδικα και φράση κατάστασης HTTP επιστρέφει ο server ως απόκριση στην δεύτερη αίτηση HTTP GET; Επέστρεψε ο server τα περιεχόμενα του αρχείου; Εξηγείστε.

3. Ανάκτηση Μεγάλων Εγγράφων

Τα έγγραφα που ανακτήθηκαν στα προηγούμενα παραδείγματα ήταν απλά και μικρά αρχεία HTML. Στη συνέχεια θα εξετάσουμε τι συμβαίνει όταν φορτώνουμε ένα μεγάλο αρχείο HTML. Ακολουθήστε τα παρακάτω βήματα:

- Ξεκινήστε τον web browser σας και βεβαιωθείτε ότι η cache του είναι άδεια όπως συζητήθηκε παραπάνω.
- Ξεκινήστε τον packet sniffer Wireshark.
- Εισάγετε το ακόλουθο URL στον browser
  http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html .
  Στον browser θα πρέπει να εμφανισθεί το Καταστατικό των Ανθρώπινων Δικαιωμάτων των ΗΠΑ.
- Σταματήστε τη σύλληψη πακέτων από το Wireshark και εισάγετε "http" στο παράθυρο προδιαγραφών του φίλτρου παρουσίασης ώστε να παρουσιαστούν μόνο τα συλλαμβανόμενα μηνύματα HTTP.

Στο παράθυρο καταλόγου πακέτων θα πρέπει να δείτε το μήνυμα HTTP GET, ακολουθούμενο από μία απόκριση πολλαπλών πακέτων στην αίτηση HTTP GET. Αυτή η απόκριση πολλαπλών πακέτων χρειάζεται λίγη εξήγηση. Υπενθυμίζεται, από την Ενότητα 2.2 (βλ. Σχήμα 2.9 του βιβλίου), ότι το μήνυμα απόκρισης HTTP αποτελείται από μία γραμμή κατάστασης (status line), ακολουθούμενη από γραμμές επικεφαλίδας (header lines), ακολουθούμενες από μία κενή γραμμή, ακολουθούμενη από το σώμα οντότητας (entity body). Στην περίπτωση της δικής μας αίτησης HTTP GET, το σώμα οντότητας της απόκρισης είναι *ολόκληρο* το αιτούμενο αρχείο HTML. Σε αυτήν την περίπτωση, το αρχείο HTML είναι σχετικά μεγάλο και με 4500 bytes είναι πολύ μεγάλο για να χωρέσει σε ένα TCP segment. Έτσι, το ένα μήνυμα απόκρισης HTTP τεμαχίζεται σε αρκετά κομμάτια από το TCP και κάθε κομμάτι περιέχεται σε ξεχωριστό TCP segment (βλ. Σχήμα 1.22 του βιβλίου). Κάθε TCP segment καταγράφεται ως ξεχωριστό πακέτο από το Wireshark και το γεγονός ότι η μία απόκριση HTTP κατατεμαχίσθηκε σε πολλαπλά πακέτα TCP υποδεικνύεται με τη φράση "Continuation" που εμφανίζει το Wireshark. Τονίζουμε στο σημείο αυτό ότι δεν υπάρχει μήνυμα "Continuation" στο HTTP!

Απαντήστε στις παρακάτω ερωτήσεις:

12. Πόσα μηνύματα αιτήσεων HTTP GET στάλθηκαν από τον browser σας;
13. Πόσα TCP segments που περιείχαν δεδομένα χρειάσθηκαν για τη μεταφορά της μίας απόκρισης HTTP;
14. Τι κώδικας και φράση κατάστασης σχετίζονται με την απόκριση στην αίτηση HTTP GET;

15. Υπάρχουν γραμμές κατάστασης HTTP στα μεταδιδόμενα δεδομένα που να σχετίζονται με τον τεμαχισμό του σώματος οντότητας από το TCP;

4. Έγγραφα HTML με Ενσωματωμένα Αντικείμενα

Αφού είδαμε τον τρόπο με τον οποίο το Wireshark παρουσιάζει τη συλλαμβανόμενη κίνηση πακέτων για μεγάλα αρχεία HTML, μπορούμε να εξετάσουμε τι συμβαίνει όταν ο browser σας φορτώνει ένα αρχείο με ενσωματωμένα αντικείμενα, δηλαδή ένα αρχείο που περιλαμβάνει άλλα αντικείμενα (για παράδειγμα, αρχεία εικόνων) τα οποία είναι αποθηκευμένα σε έναν ή περισσοτέρους διαφορετικούς servers.

Ακολουθήστε τα παρακάτω βήματα:

- Ξεκινήστε τον web browser σας και βεβαιωθείτε ότι η cache του είναι άδεια όπως συζητήθηκε παραπάνω.
- Ξεκινήστε τον packet sniffer Wireshark.
- Εισάγετε το ακόλουθο URL στον browser
  http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file4.html .
  Στον browser θα πρέπει να εμφανισθεί ένα μικρό αρχείο HTML με δύο εικόνες. Το αρχείο HTML βάσης περιέχει αναφορές στις δύο αυτές εικόνες. Αυτό σημαίνει ότι το αρχείο HTML δεν περιέχει τις ίδιες τις εικόνες παρά περιέχει τα URL για τις εικόνες. Όπως συζητήθηκε στο βιβλίο, ο browser θα πρέπει να ανακτήσει αυτά τα λογότυπα από τους υποδεικνυόμενους ιστοτόπους. Το λογότυπο του εκδότη του βιβλίου θα ανακτηθεί από τον ιστότοπο www.aw-bc.com. Η εικόνα του εξώφυλλου του βιβλίου είναι αποθηκευμένη στον server manic.cs.umass.edu.
- Σταματήστε τη σύλληψη πακέτων από το Wireshark και εισάγετε "http" στο παράθυρο προδιαγραφών του φίλτρου παρουσίασης ώστε να παρουσιαστούν μόνο τα συλλαμβανόμενα μηνύματα HTTP.

Απαντήστε στις παρακάτω ερωτήσεις:

16. Πόσα μηνύματα αιτήσεων HTTP GET στάλθηκαν από τον browser σας; Σε ποιες διευθύνσεις IP στάλθηκαν αυτές οι αιτήσεις GET;
17. Μπορείτε να διακρίνετε εάν ο browser σας φόρτωσε τις δύο εικόνες σειριακά ή αν οι εικόνες φορτώθηκαν παράλληλα από τους δύο ιστοτόπους; Εξηγείστε.

5. Εξουσιοδότηση στο HTTP

Τέλος, ας προσπαθήσουμε να επισκεφθούμε έναν ιστότοπο που προστατεύεται με κωδικό πρόσβασης (password-protected) και ας εξετάσουμε την ακολουθία μηνυμάτων HTTP που ανταλλάσσονται για έναν ιστότοπο αυτού του είδους. Το URL
http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html

προστατεύεται με κωδικό πρόσβασης. Το όνομα χρήστη (username) είναι “wireshark-students” (χωρίς τα εισαγωγικά) και ο κωδικός πρόσβασης (password) είναι “network” (επίσης χωρίς τα εισαγωγικά). Ας προσπελάσουμε λοιπόν αυτόν τον “ασφαλή”, προστατευόμενο με κωδικό πρόσβασης ιστότοπο. Ακολουθήστε τα παρακάτω βήματα:

- Βεβαιωθείτε ότι η cache του web browser σας είναι άδεια, όπως συζητήθηκε παραπάνω, και τερματίστε τον browser. Κατόπιν, ξεκινήστε τον web browser σας.
- Ξεκινήστε τον packet sniffer Wireshark.
- Εισάγετε το ακόλουθο URL στον browser
  http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html
  Πληκτρολογήστε το όνομα χρήστη και τον κωδικό πρόσβασης που σας ζητούνται.
- Σταματήστε τη σύλληψη πακέτων από το Wireshark και εισάγετε “http” στο παράθυρο προδιαγραφών του φίλτρου παρουσίασης ώστε αργότερα να παρουσιαστούν μόνο τα συλλαμβανόμενα μηνύματα HTTP στο παράθυρο καταλόγου πακέτων.

Ας εξετάσουμε τώρα την έξοδο του Wireshark. Ενδεχομένως να θέλετε να διαβάσετε πρώτα για την εξουσιοδότηση στο HTTP ανατρέχοντας στο ευανάγνωστο υλικό σχετικά με το “Πλαίσιο Εξουσιοδότησης Πρόσβασης στο HTTP” (“HTTP Access Authentication Framework”) στην ιστοσελίδα http://frontier.userland.com/stories/storyReader$2159 .

Απαντήστε στις παρακάτω ερωτήσεις:

18. Ποια η απόκριση του server (κωδικός κατάστασης και φράση) στο αρχικό μήνυμα HTTP GET από τον browser σας;
19. Όταν ο browser σας στέλνει το μήνυμα HTTP GET για δεύτερη φορά, ποιο νέο πεδίο περιλαμβάνεται στο μήνυμα HTTP GET;