 **Wireshark -HTTP**

**הקדמה**

אנחנו מוכנים כעת להשתמש ב-Wireshark כדי לחקור פרוטוקולים בפעולה. במעבדה זו נחקור כמה היבטים של פרוטוקול HTTP: אינטראקציה בסיסית של בקשת GET ותגובה, פורמטים של הודעות HTTP, אחזור קובצי HTML גדולים, אחזור קובצי HTML עם אובייקטים משובצים ואימות HTTP ואבטחה. לפני שתתחילו את העבודה, מומלץ לחזור על פרוטוקול HTTP.

**1. אינטראקציה בסיסית של HTTP: בקשת GET ותגובה**

נתחיל להתנסות בעבודה עם פרוטוקול HTTP באמצעות הורדת קובץ HTML פשוט מאוד - קובץ קצר שאינו מכיל אובייקטים משובצים. בצעו את השלבים הבאים:

1. הפעילו את דפדפן האינטרנט שלכם.
2. הפעילו את ה-packet sniffer של Wireshark, כפי שעשיתם במעבדת ההקדמה (אך אל תתחילו ללכוד את החבילות). הזינו "http" (רק את האותיות, לא את המרכאות) בחלון ה-display-filter-specification, כך שרק הודעות HTTP שיילכדו יוצגו בהמשך בחלון ה-packet listing. שימו לב שבתרגיל זה אנחנו מעוניינים לבחון רק את פרוטוקול HTTP, ואיננו רוצים לראות את כל החבילות הרבות שנלכדו.
3. המתינו קצת יותר מדקה (בקרוב נראה מדוע) ואז פקדו על Wireshark להתחיל בלכידת החבילות.
4. הזינו את הכתובת הבאה לדפדפן שלכם: <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html>. על הדפדפן שלכם להציג את קובץ ה-HTML. זהו קובץ פשוט ביותר, המציג שורה אחת.
5. עצרו את לכידת החבילות של Wireshark.  
     
   חלון ה-Wireshark שתראו אמור להיראות דומה לחלון המוצג באיור 1. אם אינכם מצליחים להריץ את Wireshark בחיבור ישיר לרשת, תוכלו להוריד קובץ trace, שנוצר לאחר ביצוע השלבים שתוארו למעלה[[1]](#footnote-1).



איור 1: תצוגת Wireshark לאחר שהדפדפן הביא את המשאב בכתובת: http://gaia.cs.umass.edu/wireshark-labs/ HTTP-wireshark-file1.html

הדוגמה באיור 1 מראה בחלון ה-packet-listing את שתי הודעות ה-HTTP שנלכדו: הודעת ה-GET (מהדפדפן שלכם לשרת האינטרנט gaia.cs.umass.edu) והודעת התגובה מהשרת לדפדפן שלכם. חלון ה-packet-contents מציג את פרטי ההודעה שנבחרה (במקרה זה הודעת ה- HTTP OK, המודגשת בחלון ה-packet-listing). נזכיר כי מאחר שהודעת ה- HTTP הועברה בתוך סגמנט TCP - אשר הועבר בתוך חבילת IP, שהועברה בתוך Ethernet Frame, Wireshark - מציג גם את המידע על כל ההודעות הללו. אנחנו רוצים למזער את כמות הנתונים שאינם HTTP המוצגים (במעבדה זו אנחנו מעוניינים לחקור את פרוטוקול HTTP. את הפרוטוקולים האחרים נחקור במעבדות מאוחרות יותר), אז ודאו שבתיבות בקצה השמאלי של ה-Frame, של ה-Ethernet, של ה-IP ושל ה- TCP יש סימן פלוס או משולש המצביע ימינה (מה שאומר שיש מידע מוסתר ולא מוצג), ולשורת ה-HTTP יש סימן מינוס או משולש המצביע כלפי מטה (מה שאומר שכל המידע על הודעת ה-HTTP מוצג) .

שימו לב: התעלמו בבקשה מכל הודעת GET ותגובה של HTTP המתייחסת לקובץ favicon.ico. אם אתם רואים הפניה לקובץ זה, מה שקורה בפועל הוא שהדפדפן שלכם שואל את השרת אוטומטית אם יש לו (לשרת) קובץ אייקון קטן שצריך להיות מוצג לצד כתובת האתר המוצגת בדפדפן שלכם. אנחנו נתעלם מהפניות לקובץ הזה במעבדה על פרוטוקול HTTP.

השיבו על השאלות שלפניכם תוך בחינת המידע הנמצא בהודעות ה-HTTP GET ובהודעות התגובה. כשאתם עונים על השאלות, עליכם להדפיס את הודעות ה-GET והתגובות (עיינו במעבדת ההקדמה של Wireshark לקבלת הסבר כיצד לעשות זאת) ולציין היכן בהודעה מצאתם את המידע העונה על השאלות. בדו"ח המסכם על המעבדה, עליכם להראות בפלט שקיבלתם מהיכן אתם שואבים את המידע שעל בסיסו כתבתם את התשובה שלכם (למשל, באמצעות הדגשת הטקסט הרלוונטי). השאלות שעליכם לענות עליהן הן:

1. האם הדפדפן שלכם מריץ את פרוטוקול HTTP בגרסה 1.0 או בגרסה 1.1? איזו גרסה של HTTP מפעיל השרת?
2. אילו שפות (אם בכלל) הדפדפן שלכם מציין שהוא יכול לקבל מהשרת?
3. מהי כתובת ה-IP של המחשב שלכם? מהי כתובת ה-IP של השרת gaia.cs.umass.edu?
4. מהו קוד הסטטוס המוחזר מהשרת לדפדפן שלכם?
5. מתי שונה לאחרונה קובץ ה-HTML שקיבלתם מהשרת?
6. כמה בתים של תוכן מוחזרים לדפדפן שלכם?
7. בדקו בעזרת הנתונים הגולמיים בחלון ה-packet content אם אתם רואים כותרות (headers) כלשהן בנתונים שאינן מוצגות בחלון ה-packet-listing. אם כן, ציינו אחת.

בתשובתכם לשאלה 5 לעיל, ייתכן שהופתעתם לגלות שהמסמך שזה עתה קיבלתם שונה לאחרונה רק דקה לפני שהורדתם את המסמך. הסיבה לכך היא שעבור קובץ מסוים זה, שרת gaia.cs.umass.edu מגדיר את הזמן הנוכחי כזמן האחרון שבו שונה הקובץ, והוא עושה זאת פעם בדקה. כך, אם תחכו דקה בין שתי גישות עוקבות, ייראה הקובץ ככזה שהשתנה לאחרונה, ולכן הדפדפן שלכם יוריד עותק "חדש" של המסמך.

**2. אינטראקציה של HTTP מותנה: בקשת GET ותגובה**

למדנו שמרבית הדפדפנים משתמשים בזיכרון מטמון, ולכן כשהם רוצים להביא אליהם אובייקט באמצעות פרוטוקול HTTP, הם משתמשים ב-conditional GET. לפני ביצוע חלק זה של המעבדה, ודאו שזיכרון המטמון של הדפדפן שלכם ריק. בדפדפן Firefox לחצו על Tools->Clear Recent History ואז בחרו בתיבת ה-Cache. בדפדפן אינטרנט אקספלורר לחצו על Tools->Internet Options->Delete File. פעולות אלה תמחקנה את הקבצים, שנמצאים בזיכרון המטמון של הדפדפן שלכם. כעת בצעו את הפעולות האלה:

* הפעילו את דפדפן האינטרנט שלכם וּוַדאו שזיכרון המטמון של הדפדפן שלכם נקי.
* הפעילו את ה-packet sniffer של Wireshark.
* הזינו בדפדפן שלכם את כתובת האתר הזאת: <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html>.

על הדפדפן שלכם להציג קובץ HTML פשוט מאוד בן חמש שורות.

* הזינו במהירות את אותה כתובת אתר לדפדפן שלכם (או פשוט בחרו בלחצן הרענון, ה-refresh, בדפדפן שלכם).
* עצרו את לכידת החבילות של Wireshark והזינו "http" בחלון ה-display-filter-specification, כך שרק הודעות HTTP שנלכדו יוצגו בחלון ה-packet-listing.
* הערה: אם אינכם מצליחים להריץ את Wireshark בחיבור ישיר לרשת, באפשרותכם להשתמש בקובץ ה-trace: http-ethereal-trace-2 כדי לענות על השאלות שלהלן; ראו הערת שוליים 1. קובץ trace זה נוצר תוך ביצוע הפעולות שתוארו לעיל.

ענו על השאלות שלפניכם:

1. בדקו את תוכן בקשת ה-HTTP GET הראשונה מהדפדפן שלכם לשרת. האם אתם רואים שורת "IF-MODIFIED-SINCE" ב-HTTP GET?
2. בדקו את תוכן תגובת השרת. האם השרת החזיר במפורש את תוכן הקובץ? איך אתם יודעים?
3. עיינו כעת בתוכן של בקשת ה-HTTP GET השנייה ששלח הדפדפן שלכם לשרת. האם אתם רואים שורת: "IF-MODIFIED-SINCE: " ב-HTTP GET? אם כן, איזה מידע מופיע לאחר הכותרת ""IF-MODIFIED-SINCE:?
4. מהו קוד הסטטוס של HTTP והביטוי הנלווה אליו, המוחזרים מהשרת בתגובה לבקשת ה-HTTP GET השנייה? האם השרת החזיר במפורש את תוכן הקובץ? הסבירו.

**3. אחזור, קבלה של, קבצים גדולים**

בדוגמאות שלנו עד כה, המסמכים שאוחזרו היו קובצי HTML פשוטים וקצרים. כעת נראה מה קורה כשאנחנו מורידים קובץ HTML גדול. בצעו את הפעולות הבאות:

* הפעילו את דפדפן האינטרנט שלכם וּוַדאו שזיכרון המטמון של הדפדפן שלכם נקי.
* הפעילו את ה-packet sniffer של Wireshark.
* הזינו בדפדפן שלכם את כתובת האתר הזאת: <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html>.

הדפדפן שלכם אמור להציג את המסמך הארוך של מגילת הזכויות של ארצות הברית של אמריקה.

* עצרו את לכידת החבילות של Wireshark והזינו "http" בחלון ה-display-filter-specification, כך שרק הודעות HTTP שנלכדו יוצגו בחלון ה-packet-listing.
* הערה: אם אינכם מצליחים להריץ את Wireshark בחיבור ישיר לרשת, באפשרותכם להשתמש בקובץ ה-trace: http-ethereal-trace-3 כדי לענות על השאלות שלהלן; ראו הערת שוליים 1. קובץ trace זה נוצר תוך ביצוע הפעולות שתוארו לעיל.

בחלון ה-packet-listing, אתם אמורים לראות את הודעת ה-HTTP GET שלכם, ולאחריה את התגובה מרובת החבילות ששלח פרוטוקול TCP לבקשת ה-HTTP GET שלכם. לתגובה מרובה המנות הזו מגיע קצת הסבר. נזכור כי הודעת התגובה של HTTP מורכבת משורת מצב, אחריה שורות כותרת, ואחריה שורה ריקה, ואחריה תוכן הדף. במקרה של בקשת ה-HTTP GET שלנו, תוכן הדף בתגובה הוא כל קובץ ה-HTML המבוקש. במקרה שלנו, קובץ ה- HTML ארוך למדיי, ו-4500 הבתים שלו אינם יכולים להתאים לחבילת TCP אחת. לכן התשובה לבקשת ה-GET תחולק ע"י פרוטוקול TCP לכמה חלקים, כל חלק כלול בתוך סגמנט TCP נפרד. בגרסאות האחרונות של Wireshark מציין Wireshark כל סגמנט TCP כחבילה נפרדת, והעובדה שתשובת HTTP בודדת חולקה על פני חבילות TCP מרובות מצוינת על ידי הכיתוב: "TCP segment of a reassembled PDU" בעמודת ה-info של Wireshark. גרסאות קודמות של Wireshark השתמשו בביטוי "Continuation" כדי לציין כי התוכן של הודעת ה-HTTP חולקה לכמה מקטעי TCP. חשוב להדגיש שאין הודעת "Continuation" ב-HTTP!

ענו על השאלות שלפניכם:

1. כמה הודעות של בקשת HTTP GET שלח הדפדפן שלכם? מהו מספר החבילה המכיל את הודעת ה-GET עבור הדף של מגילת הזכויות של ארה"ב?
2. מהו מספר החבילה המכיל את קוד הסטטוס ואת הביטוי המשויכים לתגובה לבקשת ה-HTTP GET?
3. מה קוד הסטטוס והביטוי בתגובה?
4. כמה סגמנטים של TCP המכילים נתונים נדרשים כדי לשלוח את תגובת ה-HTTP הבודדת ואת הטקסט של מגילת הזכויות של ארה"ב?

**4. מסמכי HTML עם אובייקטים משובצים**

כעת, לאחר שראינו כיצד Wireshark מציג את תעבורת החבילות שנלכדה עבור קובצי HTML גדולים, נוכל לבדוק מה קורה כאשר הדפדפן שלכם מוריד קובץ עם אובייקטים משובצים, כלומר קובץ הכולל אובייקטים אחרים (בדוגמה למטה, קובצי תמונה) המאוחסנים בשרתים אחרים.

בצעו את הפעולות האלה:

* הפעילו את דפדפן האינטרנט שלכם וּוַדאו שזיכרון המטמון של הדפדפן שלכם נקי.
* הפעילו את ה-packet sniffer של Wireshark.
* הזינו בדפדפן שלכם את כתובת האתר הזאת: <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file4.html>.

הדפדפן שלכם אמור להציג קובץ HTML קצר עם שתי תמונות. קובץ ה-HTML מפנה אל שתי תמונות אלה. במילים אחרות, התמונות עצמן אינן כלולות בקובץ ה-HTML; במקום זאת, הקובץ כולל את הכתובות לתמונות אלה. הדפדפן שלכם יצטרך לפנות לשרתים, המכילים את התמונות תוך שימוש בכתובות, הנמצאות בקובץ ה-HTML. שתי התמונות מגיעות משני שרתים שונים. האחת מגיעה מ-gaia.cs.umass.edu, ואילו השנייה מגיעה מ-caite.cs.umass.edu. אלה הם שני שרתי אינטרנט שונים בתוך cs.umass.edu.

* עצרו את לכידת החבילות של Wireshark והזינו "http" בחלון ה-display-filter-specification, כך שרק הודעות HTTP שנלכדו יוצגו בחלון ה-packet-listing.
* הערה: אם אינכם מצליחים להריץ את Wireshark בחיבור ישיר לרשת, באפשרותכם להשתמש בקובץ ה-trace: http-ethereal-trace-4 כדי לענות על השאלות שלהלן; ראו הערת שוליים 1. קובץ trace זה נוצר תוך ביצוע הפעולות שתוארו לעיל.

ענו על השאלות שלפניכם:

1. כמה בקשות HTTP GET שלח הדפדפן שלכם? לאילו כתובות אינטרנט נשלחו בקשות GET אלה?
2. האם אתם יכולים לדעת אם הדפדפן שלכם הוריד את שתי התמונות בצורה סדרתית, או האם הן הורדו משני אתרי האינטרנט במקביל? הסבירו.

**5. אימות HTTP**

לבסוף, נבקר באתר אינטרנט המוגן ע"י סיסמה ונבחן את רצף הודעות ה-HTTP שמוחלפות בין השרת והלקוח באתר מסוג זה. כתובת האתר היא:

<http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html>.

האתר מוגן באמצעות סיסמה. שם המשתמש הוא "wireshark-students" (ללא מרכאות) והסיסמה היא "network" (שוב, ללא מרכאות). בצעו את הפעולות האלה:

* הפעילו את דפדפן האינטרנט שלכם וּוַדאו שזיכרון המטמון של הדפדפן שלכם נקי.
* הפעילו את ה-packet sniffer של Wireshark.
* הזינו את כתובת האתר הבאה בדפדפן שלכם: <http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html> והקלידו את שם המשתמש והסיסמה לתוך החלון הקופץ.
* עצרו את לכידת החבילות של Wireshark והזינו "http" בחלון ה-display-filter-specification, כך שרק הודעות HTTP שנלכדו יוצגו בחלון ה-packet-listing.
* הערה: אם אינכם מצליחים להריץ את Wireshark בחיבור ישיר לרשת, באפשרותכם להשתמש בקובץ ה-trace: http-ethereal-trace-5 כדי לענות על השאלות שלהלן; ראו הערת שוליים 1. קובץ trace זה נוצר תוך ביצוע הפעולות שתוארו לעיל.

כעת נבחן את הפלט של Wireshark, אולם לפני כן ייתכן שתרצו לקרוא על אימות ב-HTTP. תוכלו למצוא הסבר קל לקריאה במסמך " HTTP AccessAuthentication Framework". תוכלו למצוא אותו בכתובת: <http://frontier.userland.com/stories/storyReader$2159>.

ענו על השאלות שלפניכם:

1. מהי תגובת השרת (קוד סטטוס והביטוי הנלווה) בתגובה לבקשת ה-HTTP GET הראשונית מהדפדפן שלכם?
2. כאשר הדפדפן שולח את בקשת ה-HTTP GET בפעם השנייה, איזה שדה חדש כלול בהודעת ה-HTTP GET?

שם המשתמש "wireshark-students" והסיסמה "network" מקודדים במחרוזת התווים: (d2lyZXNoYXJrLXN0dWRlbnRzOm5ldHdvcms=) המופיעה מייד לאחר הכותרת: “Authorization: Basic” בבקשת ה-HTTP GET של הלקוח. אמנם זה נראה כאילו שם המשתמש והסיסמה שלכם מוצפנים, אך למעשה הם מקודדים בפורמט המכונה Base64. שם המשתמש והסיסמה אינם מוצפנים! כדי לראות זאת, עברו אל http://www.motobit.com/util/base64-decoder-encoder.asp והזינו את המחרוזת המקודדת base64 d2lyZXNoYXJrLXN0dWRlbnRz ופענחו. בכך תרגמתם את המחרוזת מקידוד Base64 לקידוד ASCII, ולכן אתם אמורים לראות את שם המשתמש שלכם! לצפייה בסיסמה, הזינו את שארית המחרוזת Om5ldHdvcms = ולחצו על פענוח. מכיוון שכל אחד יכול להוריד כלי כמו Wireshark ולעקוב אחר חבילות שעוברות ליד מתאם הרשת שלו, ומאחר שכל אחד יכול לתרגם מקידוד Base64 לקידוד ASCII (כפי שעשיתם זה עתה), צריך להיות ברור לכם שסיסמאות פשוטות באתרי אינטרנט אינן מאובטחות אלא אם ננקטים אמצעים נוספים.

1. הורידו את קובץ ה- zip הנמצא בכתובת: http://gaia.cs.umass.edu/wireshark-labs/wireshark-traces.zip וחלצו ממנו את הקובץ http-ethereal-trace-1. הקבצים בקובץ ה-zip כוללים traces שנאספו על ידי תוכנת Wireshark תוך כדי ביצוע השלבים המתוארים במעבדת Wireshark. לאחר שאיתרתם את קובץ ה-trace המתאים, תוכלו לטעון אותו ל-Wireshark ולהציג את ה-traces באמצעות תפריט ה-File: בחרו באפשרות Open ואז בחרו את קובץ ה-trace שנקרא: http-ethereal-trace-1. התצוגה המתקבלת אמורה להיות דומה לזו המתוארת באיור 1. שימו לב שהתצוגה של ממשק המשתמש של Wireshark עשויה להיות שונה מעט במערכות הפעלה שונות ובגרסאות שונות של Wireshark. [↑](#footnote-ref-1)