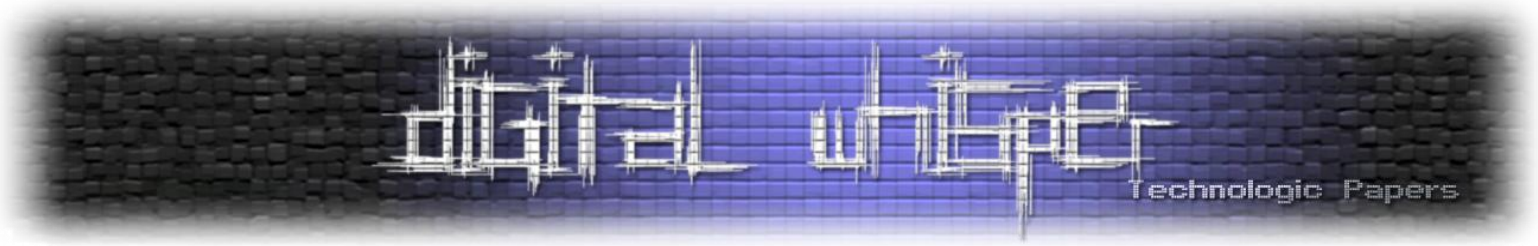


אני יודע לאן גלשת בקיץ האחרון

מאת אריק פרידמן

הקדמה או מה קורה כשמקבלים עוגיות מזרים?

האינטרנט פתח לאחרונה הזדמויות חדשות בפני מפרסמים, שלא היו קיימות לפני כן. בפנייה לקהל רחב, פרסום דרך הטלוויזיה והעיתונות אילץ את המפרסם לבסס קמפיין פרסום יחיד לקהל גדול. פרסום באינטרנט, לעומת זאת, פתח למפרסמים את ההזדמנות לפנות לקהלים מוגדרים היטב עם פרסומות מיוחדות המותאמות אליהם. תעשיית הפרסום באינטרנט מוכנה להשקיע סכומים נאים כדי להכיר טוב יותר את קהל היעד, לפלח אותו ולהתאים פרסומות לגולשים, באופן שימקסם את היענות הגולש לפרסום וכן את מכירות המוצרים. חברות רבות משתמשות ב-cookies, קבצים קטנים שאתרים שותלים במחשב המשתמש, על מנת לתעד ולעקוב אחר הרגלי הגלישה של המבקרים באתרי אינטרנט שונים, וללמוד על תחומי העניין והעדפותיהם. חלק מן החברות מאפשרות למשתמש לבחור ומכבדות את רצונם של משתמשים שאינם מעוניינים בפרסום ממוקד. לעומתן, חברות אחרות יעשו את כל שביכולתן כדי להשיג מידע רב ערך על התנהגות הגולשים.

כשאנחנו גולשים באתר באינטרנט, מה בעצם ניתן ללמוד עלינו? עד כמה התופעה נפוצה? האם יש אפשרות להגביל אותה?

בכתבה זו נסקור את השיטות הנפוצות בהן משתמשים אתרים כדי ללמוד על המבקרים בהם. בנוסף, נבחן מספר דרכים (מעשיות ותאורטיות) בהן יכולים גורמים שונים לפעול כדי לחשוף מידע נוסף על הגולשים באתרי האינטרנט.

עוגיות HTTP

עוגיית HTTP (HTTP cookie) הינה בסך הכל קובץ טקסט קטן (עד 4k) שהדפדפן יכול לשמור במחשב עבור אתר אינטרנט כאשר הוא נדרש לכך- כאשר האתר מבקש לשמור קובץ כזה. בפעם הבאה שהדפדפן ניגש לאותו אתר, אם יש קובץ כזה בסביבה, הוא ישלח אותו לאתר יחד עם הבקשה לקבלת דף אינטרנט. זה הכל. למען הסדר הטוב אזכיר שעל-פי האקדמיה ללשון העברית יש להשתמש בשם קוקית ולא בכינוי העממי עוגייה, אך אני לא מסוגל להביא את עצמי לעשות זאת, עמכם ועם האקדמיה הסליחה. להגנתי אציין שאותם אנשים טוענים שלטוקבק קוראים תגובית ושבמקום סניפר יש לומר רחרחן מנות.

העוגיות מועילות כיוון שפרוטוקול HTTP, ה"שפה" בה מדבר ה-World Wide Web, הינו חסר זכרון במהותו (stateless). כלומר, כל גישה לשרת Web נעשית ללא זכרון של גישות קודמות. על-ידי שמירת עוגיות בדפדפן, אתרים יכולים ליצור קשר בין גישות שונות שאותו משתמש עושה לאתר, למשל לקשר אותן לאותו חשבון ולאפשר זיהוי אוטומטי של המשתמש בכניסה הבאה, לשמור על תכולת עגלת קניות במעבר בין דפים באתר מסחרי, וכן הלאה.

קבצי עוגייה מאחסנים זוגות של שם וערך (למשל: ID=value). עוגייה מוגבלת לרוב לשימוש של מתחם (domain) מסויים. כאשר אתם גולשים לאתר אחד אין באפשרותו לקרוא עוגיות שנקבעו על-ידי אתרים אחרים. אורך החיים של העוגיה יכול להשתנות, סוג אחד של עוגיות הינו ארעי – עוגיות אלה נמחקות כאשר סוגרים את הדפדפן, ומטרתן רק ליצור רציפות במהלך גלישה בין דפים באתר באירוע גלישה יחיד (session). סוג שני של עוגיות הוא בעל אורך חיים ארוך יותר (עוגיות "מתמידות", persistent cookies), על-פי תאריך תפוגה שהאתר קובע.

כאשר גולשים לאתר מסויים, למשל www.amazon.com, אותו אתר יכול לשמור עוגיות לצרכיו. במקרה זה אלו עוגיות צד א' (first-party cookies). רוב מכריע של אתרים דורשים שתהיה יכולת לאחסן עוגיות צד א' בכדי ליצור חשבון משתמש ולשמור את המשתמש מחובר, כך שמניעת יצירתן תפגום בפונקציונליות של אתרים רבים וחווית הגלישה תיפגם. שימוש נפוץ נוסף שבעלי אתרים עושים בעוגיות הוא מדידת מספר הביקורים והמבקרים באתרם על מנת ללמוד כיצד אנשים משתמשים באתר. עם זאת, יש לעשות מדידות כאלה בזהירות, היות ולמשתמש יחיד הגולש ממספר מחשבים או ממספר דפדפנים באותו מחשב יהיו עוגיות שונות או לחלופין- למספר משתמשים הגולשים מאותו חשבון באותו מחשב ובאותו דפדפן תהיה עוגייה יחידה. כמו-כן, גם מחיקה תדירה של עוגיות יכולה לעוות מדידות מסוג זה.

האתר בו אנו גולשים יכול לטעון לדף נתונים מתחומים אחרים, כגון באנרים של פרסומות שמגיעים מ-ad.doubleclick.net. במקרה זה, אותם תחומים יכולים לשמור אף הם עוגיות בדפדפן. במקרה זה מדובר בעוגיות צד ג' (third-party cookies), מאחר והן נשמרות עבור תחום שונה מזה שאנו נמצאים בו כרגע. עוגיות של מפרסמים יכולות להישלח אליהם מאתרים רבים בהם הפרסומות שלהם מופיעות. דבר זה מאפשר להם לעקוב אחר האתרים שאנשים גולשים בהם, ובהתאם לכך לשלוח להם פרסום ממוקד (targeted). לאתרים השותלים עוגיות צד ג' אין אפשרות לגשת ולקרוא את עוגיות צד א' באתרים המצביעים אליהם, אך הם יכולים לראות מי הם אתרים אלה. את עוגיות צד ג' ניתן לרוב לחסום ללא פגיעה בפונקציונליות של אתרים על-ידי שינוי הגדרות בדפדפן. כיום בכל הדפדפנים הנפוצים ניתן לקבוע את העדפות לגבי שמירה על עוגיות ובפרט למחוק עוגיות או למנוע מלכתחילתה את יצירתן של עוגיות צד א' וצד ג'. כמו-כן, לא מעט כלי אבטחה למחשבים אישיים כוללים אפשרות למחוק עוגיות "עוקבות" (tracking cookies), כמו אלה שמפרסמים עושים בהן שימוש בכדי ללמוד על הגולשים. תופעות אלה הביאו למחקרים שונים של גופים המודדים רייטינג באינטרנט, בהתכתשות מתמדת על השיטה ה"נכונה" למדוד.

ארגון NAI (Network Advertising Initiative), תאגיד של עשרות חברות שיווק מקוון הכולל בין השאר גם את גוגל, מיקרוסופט, ו- Yahoo!, מספק דף Opt-out כללי המאפשר למשתמשים להכריז כי אינם מעוניינים בפרסום ממוקד מאף אחת מהחברות (המשתמש עדיין יקבל פרסומות כמובן, אך הן כבר לא יהיו מותאמות אליו אישית). במקרה זה תיווצר במחשב עוגיית Opt-out, אותה כל החברות בתאגיד מזהות ומכבדות. יצירת עוגייה כזו אינה מבטיחה בהכרח שהאתרים לא יצרו עוגיות כלל, אלא רק שהעוגיות לא ישמשו לצורך פרסום מקוון. יש לציין כי מאחר והעדפת ה-Opt out מבוטאת באמצעות עוגייה, מחיקת עוגיות בדפדפן תגרור את ביטול פעולת ה-Opt out.

לאלה המעוניינים במידע טכני יותר על עוגיות HTTP, הנכם מוזמנים לקרוא מאמר בנושא בשני חלקים ב-Infosecwriters.com.

עוגיות Flash

לכאורה, חסימת עוגיות צד ג' יכולה למנוע ממפרסמים למיניהם לעקוב אחר תבניות הגלישה שלנו וללמוד מי אנחנו. ישנן הערכות מ-2007 כי מעל ל-30% מהמשתמשים מוחקים את העוגיות שלהם ואף שמחקרים שונים מספקים הערכות שונות, כולם תמימי דעים שאחוז גבוה מהעוגיות נמחק בתדירות כך שלאור הזמן שעבר מאז והמודעות הגוברת לעוגיות, סביר להניח שאחוזים אלה גדלים. אף על פי כן, בידי המפרסמים אמצעי נוסף למעקב בו הם עושים שימוש תדיר- עוגיות Flash. בניגוד לעוגיות HTTP, עוגיות Flash הינן פחות מוכרות, אפילו לאנשי טכנולוגיה, אם כי המודעות לקיומן מתחילה לחלחל לציבור. עוגיית Flash, המכונה בשם LSO (Local Shared Object) ולעיתים גם supercookie, הינה קובץ שתוסף Adobe Flash מחזיק על מחשב המשתמש. קובץ זה פועל באופן דומה לעוגיית HTTP. זה המקום להזכיר שעל-פי חברת Adobe, נכון ליוני 2010, כ-99% מהמחשבים השולחניים בעלי הגישה לאינטרנט בשווקים המפותחים תומכים ב-Flash. עוגיית Flash יכולה להיווצר כאשר אתר טוען לדף תוכן Flash (של האתר עצמו או תוכן צד ג', כגון באנרים של פרסומות). בברירת המחדל, עוגיות Flash יכולות להיות בגודל של עד 100Kb, פי 25 מעוגיית HTTP, ואין להן תאריך תפוגה. בניגוד לעוגיות HTTP, עוגיות Flash אינן מנוהלות על-ידי הדפדפן. יש לכך מספר השלכות חשובות: דפדפנים שונים המותקנים על אותו מחשב יגשו לאותן עוגיות Flash, בניגוד לעוגיות HTTP המנוהלות בנפרד על-ידי כל דפדפן; מחיקת עוגיות HTTP, היסטוריית דפדפן או מחיקת כל מידע שהוא המאוחסן על-ידי הדפדפן לא תשפיע כלל על עוגיות Flash. אפילו גלישה במצב פרטיות (Private Browsing), המנטרלת עוגיות ותוספי דפדפן, לא מונעת את פעילותן של עוגיות Flash.

לעוגיות Flash שימושים שונים, כאשר הנפוצים שבהם הם שמירת העדפות המשתמש עבור עוצמת הקול של נגן וידאו Flash, שמירת מטמון מקומי של קובץ מוסיקה לצורך ביצועים טובים יותר מעל חיבור רשת איטי. יחד עם זאת, נעשה שימוש בעוגיות Flash גם לצורך מעקב אחר משתמשים. למעשה, כבר ב-2005 חברה בשם United Virtualities פרסמה את השימוש שהיא עושה בעוגיות Flash כגיבוי לעוגיות HTTP – אם עוגיות HTTP מסויימות היו נמחקות, ערכן היה משוחזר מתוך עוגיות Flash שהחזיקו ערכים דומים.

במחקר מ-2009 באוניברסיטת ברקלי בקליפורניה, חוקרים בחנו את 100 האתרים המובילים (על- פי דירוג QuantCast) ובדקו את השימוש שהם עושים בעוגיות Flash. מתוך 100 האתרים, 54 עשו שימוש בעוגיות Flash (נוצרו 157 קבצים עם 281 עוגיות Flash), ו-98 יצרו עוגיות HTTP (סך-הכל 3602 עוגיות HTTP), כאשר שני החריגים היו Wikipedia ו-Wikimedia.org. למרות ששם עוגיית ה-Flash הנפוץ ביותר היה Volume, נמצאה תפוצה רחבה ביותר של שמות עוגיות כמו userid, user ו-id, שעשויות לשמש

למעקב אחרי המשתמש. בנוסף, חלק מ-100 אתרים אלו משתמשים בעוגיות Flash כדי להחיות עוגיות HTTP שנמחקו על-ידי המשתמשים.

מעניין לציין כי באתר NAI נכתב שהחברות המשתתפות בתאגיד לא עושות שימוש בעוגיות Flash לצורך פרסום ממוקד. עם זאת, נכון לזמן ביצוע המחקר, התברר כי חברת QuantCast, החברה ב-NAI, עדיין עשתה שימוש בעוגיית Flash גם כשעוגיית ה-Opt out של NAI הייתה נוכחת. לאחר מחיקת עוגיות, QuantCast שיחזרה את עוגיית ה-HTTP שלה מתוך עוגיית ה-Flash (עוגיית ה-Opt out שנמחקה לא שוחזרה). בעקבות המחקר, באוגוסט 2009 הפסיקה החברה את מנהג החייאת העוגיות שלה. למרות זאת, ביולי האחרון הוגשה נגדה ונגד אתרים נוספים תביעה ייצוגית בגין חדירה בלתי חוקית למחשבי הגולשים. החודש התפרסמה תביעה נוספת נגד ClearSpring, דיסני ואתרים נוספים, בטענה כי השתמשו בעוגיות Flash על מנת לעקוב אחר דפוסי הגלישה של ילדים ברחבי האינטרנט.

למרות שהגדרות הפרטיות בדפדפן אינן משפיע על עוגיות Flash, ישנה דרך למנוע את יצירתן, דרך שרוב המשתמשים אינם מודעים אליה. חברת Adobe מספקת באתר שלה דף המאפשר לשלוט על עוגיות Flash, ובין השאר גם למנוע יצירת עוגיות צד ג'. ברוב האתרים אין בעיות פונקציונליות עקב חסימת עוגיות Flash צד ג'.

ייחודיות של דפדפנים

אם כן, המשתמש המתוחכם שאינו מעוניין שיתחקו אחר פעולותיו ברשת יכול לנקוט אמצעים כדי למנוע מעקב באמצעות עוגיות HTTP ועוגיות Flash. לרוע המזל, ככל הנראה אתרים עקשניים עדיין יוכלו לזהות את פעילותו הייחודית.

ארגון EFF (Electronic Frontier Foundation), ארגון אמריקאי ללא כוונת רווח הפועל להגנה על הצרכנים בעולם הדיגיטלי, ביצע בשנה האחרונה ניסוי מעניין שנועד לבחון עד כמה ייחודי הדפדפן שאיתו אנו גולשים באינטרנט. אתר הניסוי, http://panopticlick.eff.org, עדיין פעיל ומאפשר למשתמשים לגלוש ולבדוק עד כמה הדפדפן שלהם ייחודי. במאי 2010 הארגון הוציא דו"ח המסכם את ממצאיהם לפי הנתונים שנאספו עד לאותו זמן – דגימה הכוללת 470,161 דפדפנים שבעליהם ביקרו באתר. ממצאים אלה מעידים שדפדפנים נוטים להיות יחודיים מאוד – אם נבחר דפדפן כלשהו באקראי, ניתן לצפות שלכל

היותר לאחד מבין 286,777 דפדפנים אחרים (!) יהיו מאפיינים דומים. המצב גרוע יותר אם הדפדפן תומך גם ב-Flash או Java, ובמקרה זה 94.2% מהדפדפנים במדגם היו יחודיים.

זיהוי הדפדפן מסתמך על איסוף מאפיינים שדפדפנים מספקים לאתרים. חלק ממאפיינים אלה הם חלק סטנדרטי מבקשה שדפדפן שולח לאתר כדי לקבל דף אינטרנט. חלקם ניתנים לאיסוף על-ידי הרצת סקריפט במחשב המשתמש (ברוב המחשבים ניתן לעשות זאת ללא ידיעת המשתמש). להלן טבלה המתארת את הנתונים שנאספו (פירוט יתר לגבי הנתונים שנאספו ניתן לקרוא בדו"ח של EFF):

Variable	Source	Remarks
User Agent	Transmitted by HTTP, logged by server	Contains Browser micro-version, OS version, language, toolbars and sometimes other info
HTTP ACCEPT headers	Transmitted by HTTP, logged by server	
Cookies enabled?	Inferred in HTTP, logged by server	
Screen resolution	JavaScript AJAX Post	
Timezone	JavaScript AJAX Post	
Browser plugins, plugin versions and MIME types	JavaScript AJAX Post	Sorted before collection
System fonts	Flash applet or Java applet, collected by JavaScript/AJAX	Not sorted
Partial supercookie test	JavaScript AJAX post	

מה היא בעצם הבעיה ביכולת לזהות דפדפן באופן ייחודי? יכולת זו פותחת בפני אתרים אפשרות לעקוב אחר הגולשים גם כאשר הם נוקטים אמצעי זהירות כגון מחיקת עוגיות למיניהן. להבדיל מקבצי עוגייה, זיהוי דפדפן באמצעות ה"חתימה" שלו לא משאירה כל חותם על המחשב של המשתמש, ומסתמך על נתונים סטנדרטיים שכל דפדפן שולח לאינטרנט בעת גלישה.

חשוב לציין שהנתונים שנאספו על ידי EFF הם מדגם מוטה – סביר להניח שהמשתמשים שהתנדבו לגלוש לאתר של EFF לטובת הניסוי הם משתמשים מתוחכמים, בעלי מודעות גבוהה לפרטיות ובעלי מאפיינים שונים מאלה של האוכלוסיה הכללית (למשל, ניגשו לאתר פי 4.5 דפדפני Firefox מאשר Internet Explorer, בעוד שבכלל האוכלוסיה IE מחזיק את נתח השוק הגדול ביותר). עם זאת, יש מקום להניח שאותם משתמשים הם גם אלה שסביר יותר כי יפעילו אמצעים על מנת להמנע ממעקב על ידי עוגיות, ולכן הם יעד "מושך" יותר לזיהוי מסוג כזה.

באופן פרדוקסלי, לעיתים דווקא אמצעים לשיפור הפרטיות יכולים להפוך את מלאכת הזיהוי לקלה יותר. למשל, זיוף שדה User Agent שהדפדפן שולח יכול ליצור חתימות דפדפן יחודיות וקלות לזיהוי, כמו דפדפני אייפון התומכים ב-Flash (חיה שלא קיימת במציאות). תוספי דפדפן החוסמים Flash אף הם יחודיים. עם זאת, המחברים ציינו כי תוספי הפיירפוקס TorButton ו-NoScript התבררו כאמצעי הגנה יעילים בפני זיהוי חתימות דפדפן. חסימת JavaScript גם היא אמצעי יעיל להגבלת יכולת זיהוי הדפדפן, אך מאחר ואתרים רבים משתמשים בסקריפטים, היא כרוכה באובדן פונקציונאליות מהותי בעת גלישה באינטרנט.

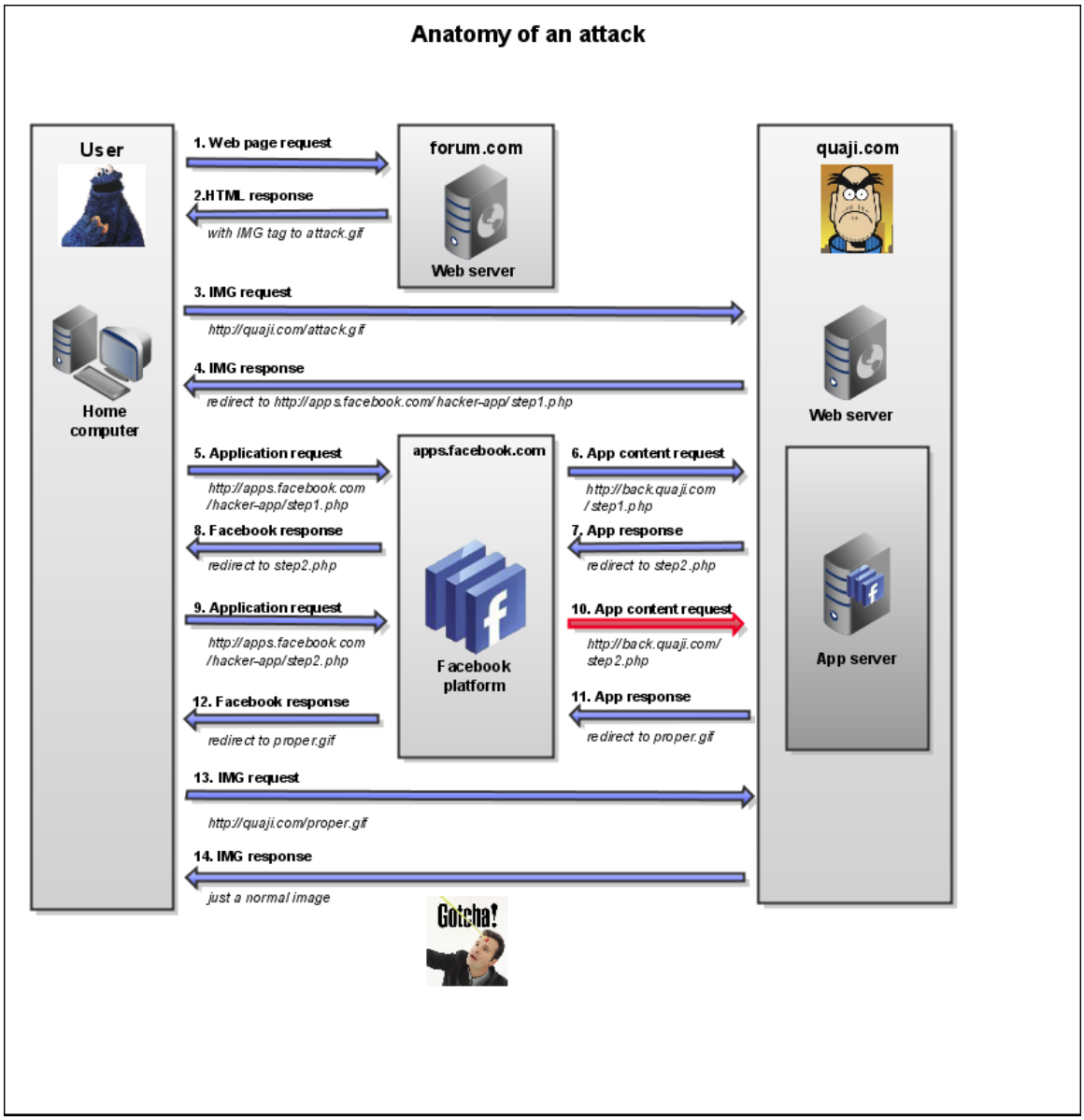
חשיפת זהות מבקר באתר באמצעות רשת חברתית

כיום אנשים רבים עושים שימוש ברשתות חברתיות, וחושפים בהן את שמם האמיתי, תמונתם, רשימת חברים, תחומי עניין וכן הלאה. לדוגמא, ברשת החברתית פייסבוק, עם השנים יותר ויותר מידע נעשה חשוף לציבור רחב יותר, בעידודה של פייסבוק עצמה. אם מפרסם או אתר כלשהו יוכל לקשר ביקור באתר למשתמש ברשת חברתית, ככל הנראה ילמד לא מעט על אותו מבקר ותחומי העניין שלו, גם אם לא עקב אחר מעשיו של אותו מבקר ברשת.

בחור ישראלי בשם רונן זילברמן הראה באוגוסט 2009 כיצד אתר כלשהו (שאינו קשור לפייסבוק) יכול לנצל נקודת תורפה בדרך בה פייסבוק מעבירה מידע לאפליקציות פייסבוק בכדי לקבל פרטים מזהים של המשתמש, ללא צורך לקבל את הסכמתו (יש לציין כי מאז פייסבוק כבר תיקנה את הבעיה). על מנת שההתקפה תעבוד, המשתמש צריך להיות מחובר לפייסבוק באחת הלשוניות של הדפדפן, או לחלופין, הדפדפן צריך לאחסן עוגייה של פייסבוק הזוכרת את פרטי ההתחברות של המשתמש (לצורך התחברות אוטומטית לאתר בעת גלישה אליו, אפשרות של "keep me logged in"). רונן העניק להתקפה את השם Cross-Site Identification (CSID), והראה כי ניתן ליישם נגזרות שלה גם על רשתות חברתיות אחרות, Bebo ו-Orkut.

אפליקציות פייסבוק אינן שונות בהרבה מאתרי אינטרנט רגילים, אך הבדל אחד מהותי הוא שגישה לאפליקציות פייסבוק נעשית לא ישירות לשרת האפליקציה, אלא תמיד דרך השרתים של פייסבוק. השרתים של פייסבוק יכולים לספק לשרתי אפליקציה מידע על המשתמש (בכפוף להרשאות שהמשתמש נתן לאפליקציה), על מנת שיוכלו להתאים את המענה למשתמש שמקבל את השירות. עד אפריל 2010, פייסבוק תמכה במנגנון של אימות אוטומטי (Automatic Authentication). משמעותו של המנגנון הייתה שאם משתמש מבקר בדף של אפליקציה, פייסבוק תעביר את פרטי המשתמש לאפליקציה גם אם המשתמש לא אישר אותה. כמנגנון הגנה, פייסבוק מעבירה את הפרטים רק במידה והמשתמש לא הקשיח את מדיניות הפרטיות שלו וכן רק כשהגישה לאפליקציה נעשית מדף באתר של פייסבוק. המגבלה הראשונה אינה תקפה לרוב המוחלט של המשתמשים. את המגבלה השנייה רונן הציע לעקוף בצורה פשוטה: גישה לדף של האפליקציה, למשל http://apps.facebook.com/hacker-app/step1.php, תבצע הפניה (redirect) לדף אחר של האפליקציה, נאמר http://apps.facebook.com/hacker-app/step2.php – גם אם לדף הראשון הגענו מחוץ לפייסבוק (ולכן לא נקבל את פרטי המשתמש), הרי שלדף השני המשתמש כבר מופנה מתוך כתובת פייסבוקית, ולכן מנגנון האימות האוטומטי ייכנס לפעולה ויעביר את פרטי המשתמש.

התרשים שלהלן, שנלקח מהבלוג של רונן, מראה תהליך מלא בו אתר יכול להשיג את פרטי המשתמש בצורה שלא מעוררת חשד, תוך העזרות באפליקציות שפייסבוק יצר. במקום לפרסם בדף האינטרנט קישור ישיר לאפליקציה, ניתן למקם קישור לקובץ תמונה. בעת גישת דפדפן לשרת המחזיק כביכול את קובץ התמונה, הדפדפן יופנה לדף הראשון של האפליקציה, ומשם לדף השני של האפליקציה, כשבשלב זה האפליקציה תקבל את פרטי המשתמש. הדף השני של האפליקציה יכול להכיל קישור לקובץ תמונה אמיתי, כך שכל התהליך יתבצע באופן שקוף למשתמש, ותוצג לו תמונה:



ניתן לראות גם סרטון המדגים את ההתקפה ביוטיוב.

חשוב לציין שההתקפה תאפשר לקבל את פרטי המשתמש רק במידה שהגדיר אותם עם הרשאת public (דבר שפייסבוק הבטיחה שיהיה נכון לגבי רוב המשתמשים). הבעייתיות היא בכך שההתקפה מאפשרת לקשור את הנתונים הפומביים הללו עם משתמש ספציפי שמבקר כרגע באתר כלשהו, ובכך פוגעת באנונימיות של המשתמש בעת הגלישה.

התקפת CSID על פייסבוק נשענה על נקודת תורפה שפייסבוק כבר תיקנה, אך עם השקעה של יותר מאמץ, אתר יכול לחשוף את פרופיל הרשת החברתית של המשתמש תוך ניצול התכונות השיתופיות הבסיסיות של הרשת וללא תלות בקיומה של נקודת תורפה זו או אחרת. ההתקפה מסתמכת על עקרונות דומים לאלו שתיארתי בכתבה קודמת בהקשר של מאגר המידע של נטפליקס. חברת נטפליקס פרסמה מאגר אנונימי המכיל דירוגי סרטים של גולשים. בפועל, כיוון שכל משתמש מדרג מספר קטן יחסית של סרטים, ומשתמשים שונים מדרגים סרטים שונים, הדירוגים שנותן כל משתמש הם יחודיים ומשמשים מעין טביעת אצבע שמאפשרת לזהות את המשתמש. תופעה דומה מתרחשת ברשתות חברתיות, המאפשרות למשתמשים להצטרף לקבוצות (למשל, Facebook groups). על-פי דף הסטטיסטיקות של פייסבוק, נכון לזמן כתיבת כתבה זו ישנם 900 מליון אובייקטים שמשתמשים יכולים לקשר אליהם (דפים, קבוצות, מאורעות ודפי קהילה), ובממוצע משתמש מקושר ל-80 מהם. קישורים אלה שונים ממשתמש למשתמש, ולכן יכולים לשמש כדי לזהות את המשתמש. עקרון זה איפשר לקבוצת מחקר לחשוף את זהותם של משתמשי רשתות חברתיות על-ידי זיהוי דפי קבוצות בהם גלשו. המחקר התמקד בעיקר ברשת החברתית Xing, המונה כ-8 מליון חברים, אך הוא בחן גם התקפות על פייסבוק ועל Linked-In. למרות שמספר המשתמשים הגדול בפייסבוק מקשה על ביצוע ההתקפה באופן שיקיף את כלל המשתמשים ברשת, התמקדות בקהל יעד מצומצם כמו, נאמר, מדינת ישראל, היא מעשית ביותר. לצורך תיאור ההתקפה אתמקד כאן בפייסבוק.

בשל הדרך בה האינטרנט והדפדפנים מעוצבים, לאתרים יש אפשרות לשאול את הדפדפן לגבי אתרי אינטרנט אחרים שביקרתם בהם. אמנם לא ניתן לקבל מהדפדפן רשימה של אתרים כאלה, אך ניתן להציג בפניו שאלות של כן/לא לגבי אתרים ספציפיים על-ידי ניצול תכונה בסיסית של דפדפנים: דפדפנים מסתמכים על היסטוריית הגלישה של המשתמש כדי לצבוע בצבע שונה לינקים בהם ביקר בעבר לעומת לינקים חדשים. אתרים יכולים לנצל זאת לחשיפת היסטוריית הגלישה על-ידי מיקום לינקים סמויים בדף אינטרנט בקוד JavaScript ובחינת הצבע שהדפדפן קובע עבורם. למשל, ניתן לראות הדגמה של התופעה באתרים http://ha.ckers.org/weird/CSS-history-hack.html (הקוד כאן) ו- http://startpanic.com (לא, אין טעם להיכנס לפאניקה). למרות שנקודת תורפה זו ידועה מזה זמן רב, לפחות מאז אוקטובר 2000, היא לא תוקנה עד כה על-ידי יצרני הדפדפנים, מאחר ופתרון הבעיה יגרור פגיעה בשימושיות הדפדפן. עם זאת, המעוניינים יכולים לבחון תוסף Firefox בשם SafeHistory המספק פתרון מסויים לבעיה.

מתברר שניתן לקחת את הטכניקה הידועה הזו צעד אחד קדימה, ולהשתמש בה בשילוב עם מידע הזמין ברשתות חברתיות, כדי לחשוף את זהותו האמיתית של משתמש המבקר באתר כלשהו. ההתקפה המלאה קצת יותר מורכבת ממה שאתאר בהמשך, אך העקרונות הבסיסיים דומים: בשלב מקדים, אוספים מידע לגבי איזה משתמשים שייכים לאיזה קבוצות. כאשר מגיע מבקר לאתר המעוניין ללמוד עליו, האתר מבצע גניבת היסטוריה כדי לגלות באיזה דפי קבוצות המשתמש היה בעבר (ולכן סביר שהוא חבר בקבוצות אלה). אז ניתן להצליב בין רשימות המשתמשים של קבוצות אלו כדי לצמצם את רשימת ה"חשודים", יתכן עד זיהוי ייחודי של המשתמש.

עבור השלב המקדים, יש לאסוף את רשימת החברים בכל קבוצה, דבר שאינו קשה כל כך לביצוע היות והנתונים נגישים מדף הקבוצה. את רשימת כלל הקבוצות בפייסבוק החוקרים השיגו באמצעות שרותי סריקה (crawling) מסחריים – איסוף המידע על 39 מליון קבוצות מהמדריך של פייסבוק עלה להם $18.47, והמידע התקבל תוך חמישה ימים. בהינתן כל מזהה של קבוצה שנאסף בסריקה זו, ניתן לקבל בקלות את רשימת החברים בקבוצה. לדוגמה, רשימת החברים בקבוצה של מונטי פייטון (שמספרה בפייסבוק 4981419559) נמצאת בקישור הזה. אמנם פייסבוק מגבילה את מספר החברים שניתן לראות באופן כזה ל-6000, אולם בקבוצות עם מעל ל-6000 חברים ניתן לעקוף את המגבלה על-ידי מעבר על שמות נפוצים וחיפוש כל החברים בקבוצה בעלי אותו שם. למשל למציאת כל הג'ון-ים, משתמשים בקישור הבא[1]. סריקה מלאה של רשימות החברים בכל הקבוצות של פייסבוק דורשת עבודה לא מעטה, אולם כהוכחת יכולת החוקרים אספו מידע על יותר מ-43.2 מליון חברי קבוצות מתוך 31,853 קבוצות תוך 23 יום באמצעות שני מחשבים בלבד. אשאיר כתרגיל לקורא את החשבון כמה עבודה תידרש כדי לאסוף מידע שיכסה את רוב המשתמשים הישראלים.

כאשר משתמש מבקר באתר כלשהו (שאינו קשור לפייסבוק), כדי לקבל את רשימת הקבוצות שהמשתמש ביקר בהן, ניתן להשתמש בגניבת היסטוריה עם קישורים לדפי הקבוצות. למשל, דף הקבוצה של מונטי פייטון בפייסבוק זמין בקישור הבא. על-ידי סריקת מספר סביר של קבוצות כאלה (ניתן לסרוק אלפי קישורים בשניות בודדות) אפשר לאתר דפי קבוצות בהן המשתמש ביקר. על-ידי שימוש במידע המוקדם שנאסף לגבי החברים בכל קבוצה, אפשר כעת להצליב בין רשימות החברים של הקבוצות הרלוונטיות במטרה לזהות את המשתמש. כדי לאתר משתמשים גם במקרים בהם ביקרו בדפי קבוצות שאינם שייכים אליהם, ניתן גם לנקוט גם בשיטות "סלחניות" יותר מהצלבה – הכל שאלה של כמה זמן מוכנים להשקיע בביצוע ההתקפה.

---

[1] דרך אגב, לא צריך לנחש שמות. ביולי האחרון חוקר אבטחת מידע בשם רון בווס (Ron Bowes) פרסם רשימה הכוללת את שמותיהם של 171 מליון ממשתמשי פייסבוק (סה"כ 100 מליון שמות יחודיים), כולל רשימה של שמות משתמש נפוצים. לא, הוא לא פרץ לפייסבוק. הוא פשוט הוריד את המידע מהמדריך הנגיש לכל שלהם.

בעקבות ההתקפה שתוארה לעיל, ארווינד נאריאנאן, אחד מהחוקרים ששברו את האנונימיות של מאגר נטפליקס, הציע בבלוג שלו את האבחנה שהאינטרנט עצמו הופך להיות יותר ויותר חברתי וכי משתמש משאיר אחריו עקבות בכל פעם שהאינטראקציה שלו עם אתר אינטרנט נרשמת באופן ציבורי, למשל טוקבק באתר, Like של פייסבוק, לינק בטוויטר, סימנייה של del.icio.us וכן הלאה. במקום להסתמך על רשימות חברים בקבוצות פייסבוק לצורך הצלבות וזיהוי משתמשים, ניתן להשתמש במקורות רבים אחרים באינטרנט כתחליף. בעקבות ניסוי שערך עם מאגר קישורים שאנשים פרסמו ב-del.icio.us, העריך כי באמצעות 4000-5000 שאילתות של גניבת היסטוריה ניתן לזהות כ-60% מהמשתמשים שפרסמו שניים או יותר קישורים ב-del.icio.us לאורך תקופה של שלושה חודשים (עם זאת בפועל התקפה כזו תהיה כנראה קשה יותר, מאחר שבמספר דפדפנים נפוצים ברירת המחדל לשמירת היסטוריה היא פחות משלושה חודשים, לכן לגניבת ההיסטוריה במקרה זה תהיה רק הצלחה חלקית).

מילות סיכום

לאורך הזמן המודעות לשיטות שמפרסמים נוקטים ללימוד המשתמשים הולכת וגוברת, עם דעות לכאן ולכאן (לדוגמה, זוג מאמרים בנושא שפורסמו החודש ב-Wall Street Journal, בעד ונגד, וכן ב- USA Today, בעד ונגד). מצד אחד, אנשים מרגישים לעיתים מנוצלים כאשר נתונים הנאספים עליהם (לעיתים ללא ידיעתם) משמשים לרווח מסחרי, וחוששים שהמידע שנאסף עליהם יכול לשמש בדרך כלשהי כנגדם. ישנה תחושה שנשחקה היכולת של אנשים לשלוט ולקבוע את האיזון הנכון עבורם בין פרטיות לבין שירותים טובים יותר, וכי לרוב האנשים לא ניתנת בחירה אמיתית. מצד שני, הפרסום באינטרנט הוא כיום המנוע העיקרי הממן שירותים חינמיים רבים באינטרנט, דבר שאנשים רבים מקבלים כמובן מאליו. הכרה טובה יותר של המשתמשים והעדפותיהם מאפשרת פרסום יעיל יותר, הכנסות גבוהות יותר, ומימון נוסף לשירותים חינמיים נוספים וטובים יותר. לדוגמה, הבלוג ars technica פרסם פוסט בו פנה לקהל הגולשים שלו בבקשה לא להשתמש בתוספים חוסמי פרסומות, והסביר את חשיבות הפרסומות למימון התוכן ממנו הגולשים נהנים.

סביר להניח שבעתיד הקרוב הפרסום ימשיך להיות מקור הכנסה מוביל לשירותים רבים באינטרנט, וכל עוד זה המצב, למפרסמים יהיה תמריץ חזק ללמוד על הגולשים ולהתאים להם תוכן אישי. הפרסום המקוון צעיר יחסית, והגבולות של מותר ואסור עדיין נתונים במשא ומתן מתמשך בין אתרי האינטרנט, גופי חקיקה והמשתמשים.

מקורות

1. Flash Cookies and Privacy, by Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas and Chris Jay Hoofnagle, August 2009, Available at SSRN, http://papers.ssrn.com/sol3/papers.cfm?abstract-id=1446862
2. Fact and Fiction: The Truth About Browser Cookies, by The How-To Geek, February 2010, http://lifehacker.com/5461114/fact-and-fiction-the-truth-about-browser-cookies
3. How Unique is Your Web Browser, by Peter Eckersley, Electronic Frontier Foundation, http://panopticlick.eff.org/browser-uniqueness.pdf
4. EPIC Flash Cookie page, http://epic.org/privacy/cookies/flash.html