Nội dung của web tracking:

1. Giới thiệu web tracking là gì

* Định nghĩa

1. Một số kĩ thuật tracking

* Cookies
* Web beacons

- Một Web Beacon thường được ẩn đi, "tàng hình", là một object ảnh luồn lách được vào máy tính của bạn. Chúng là những object nhỏ được nhúng vào các trang web cũng như email và được kích hoạt khi bạn truy cập trang web đó cũng như là mở mail có chứa chúng.

- Để tạo ra một Web Beacon (bằng PHP), chúng ta cần kích hoạt tính năng URL rewriting module (mod\_rewrite) của Apache. Sau đó tạo một file .htaccess và viết đoạn sau vào file:

*RewriteEngine On*

*RewriteRule ^(.\*).(png|jpg|gif)$ script.php*

- Sau đó tạo file script.php:

*<?php*

*$fullpath = $\_SERVER['REQUEST\_URI'];*

*$filename = basename($fullpath);*

*$ip = $\_SERVER["REMOTE\_ADDR"];*

*$useragent = $\_SERVER["HTTP\_USER\_AGENT"];*

*echo "Path: $fullpath;<br>*

*File: $filename;<br>*

*IP address: $ip;<br>*

*User agent: $useragent";*

*?>*

Text

Description automatically generated- Kết quả sẽ in ra

- Dùng để theo dõi 1 phần hoạt động của bạn. VD: gắn beacon vào email và gửi đi. Khi bạn click vào thì beacon sẽ xác định được email này có active, và từ đó sẽ có hàng loạt email spam đến.

- Cách chống là Disable Cookies. Vì dù nó lấy được thông tin cũng k biết của ai.

* Canvas fingerprinting

- Sử dụng thẻ canvas của HTML5

- Các canvas sẽ được render khác cho khi đc sử dụng trên những máy khác nhau.

- Ngoài ra HTMML5 Canvas còn có thể theo dõi được hành động của bạn bằng cách theo dõi bàn phím, chuột, cử chỉ trên màn hình điện thoại thông minh

- Cách chống lại là Disable Cookies.

* Favicons

1. Phương thức tracking dùng cookies

- Khi vào một trang web, để định danh, website sẽ sử dụng cookies. Các trang web khác sẽ không đọc được cookies của trang web này.

- Có 2 loại cookies là cookies bên thứ nhất và cookies bên thứ 3 (tracked cookies)

- Cookies bên thứ nhất lưu trữ những thông tin của bạn khi bạn đến trang web.

- Cookies bên thứ 3 là cookies của bên thứ 3 chứ k phải của trang web hiện tại bạn đang truy cập.

- website đang truy cập sẽ đăng kí quảng cáo bên thứ 3 (google, facebook,…), bên thứ 3 sẽ trả lại 1 đoạn scripts để web sẽ gắn vào trang web và hiện quảng cáo.

- Khi bạn truy cập trang web, bạn sẽ giao tiếp với máy của của website và nhận cookies bên thứ 1. Nhưng trong trang web có những scripts sẽ điều hướng và làm bạn giao tiếp 1 cách gián tiếp với những máy chủ bên thứ 3 và sinh ra những cookies bên thứ 3.

- Và từ đây, khi bạn truy cập vào trang web thứ 2, trang web thứ 2 cũng đặt quảng cáo như web 1 thì trình duyệt của bạn sẽ gửi cookies bên thứ 3 lúc nãy cho máy chủ bên thứ 3 và họ sẽ có những hoạt động duyệt web của bạn => sở thích của bạn => quảng cáo

- Ngoài ra, còn 1 cách khác các mạng hội tracking bằng cookies. Cách này, sẽ sử dụng cookies bên thứ 1 nhưng trong ngữ cảnh bên thứ 3

- Phương thức này, thông qua những nút like share trên trang web (ví dụ như trang blog)

- Khi bạn đki 1 tài khoản mạng xã hội (vd facebook) bạn sẽ được nhận 1 cookies. Bạn truy cập vào trang web đọc blog. Bạn nhất nút like, thì trình duyệt sẽ gửi yêu cầu đến máy chủ của facebook kèm theo cookies facebook của bạn và facebook sẽ phản hồi, hiển thị trên trang web.

- Bằng cách này, facebook sẽ xây dựng được hồ sơ duyệt web

1. Phương thức tracking mới sử dụng Favicons (2021)

- Phương thức này sử dụng những biểu tượng của trang web Favicons.

- Khi truy cập 1 trang web, trình duyệt sẽ yêu cầu đến máy chủ để tra cứu 1 tệp hình ảnh Favicon (.ico) và tải nó lưu ở bộ nhớ đệm.

- Thay vì phản hồi bằng 1 favicons duy nhất, máy chủ sẽ chuyển hướng thông qua một loại tên miền phụ và mỗi miền phụ này đều có 1 favicons riêng. Những favicons này dc tạo với những ID theo dõi duy nhất. Khi bạn truy cập vào trang những trang web, từ những favicons đã được lưu trong bộ nhớ đệm, máy chủ có thể biết được bạn đã duyệt qua những trang web nào.

- Những favicons được lưu ở địa chỉ khác với bộ nhớ cache bình thường của trình duyệt nên dù bạn xóa bộ nhớ cache thường xuyên cũng không ngăn chặn được phương thức này.

- Các trang web có thể xây dựng được 1 ID theo dõi 32bit (gần 4,3 tỷ browser) trong vòng 2 giây trong khi các trang 20% trang web cần 5s để load trang.

- Thêm nữa, tracker có thể sử dụng js để ẩn các chuyển hướng, biến URL thành ảnh động hoặc biểu tưởng cảm xúc hoặc cũng có thể hiện lên bảng thông báo sử dụng cookies để che mắt bạn.

- Năm 2021, kỹ thuật này hoạt động trên các trình duyệt như Chrom, Safari, Edge, Brave.

- FireFox lại thoát được kĩ thuật này nhờ 1 lỗi (fail to load favicon from http cache) => khi fix lỗi rất có thể sẽ bị tracker sử dụng kỹ thuật này được.

Diagram

Description automatically generated with low confidenceA picture containing graphical user interface

Description automatically generated

1. Tổng kết