# Web Tracking Lab



## 1 Lab Overview

Behavioral targeting is a type of online advertising where ads are displayed based on the users web-browsing behavior. The user leaves a trail of digital foot prints moving from one website to the other. Behavioral targeting anonymously monitors and tracks the sites visited by a user. When a user surfs internet, the pages they visit, the searches they make, location of the user browsing from, device used for browsing and many other inputs are used by the tracking sites to collect data. A user profile is created from the data and data-mined for an online behavioral pattern of the user. As a result when users return to a specific site or a network of sites, the created user profiles are helpful in reaching the targeted audience to advertise. The targeted ads will fetch more user interest, the publisher (or seller) can charge a premium for these ads over random advertising or ads based on the context of a site.

## 2 Lab Environment

This lab runs in the Labtainer framework, available at http://my.nps.edu/web/c3o/labtainers. That site includes links to a pre-built virtual machine that has Labtainers installed, however Labtainers can be run on any Linux host that supports Docker containers.

From your labtainer-student directory start the lab using:

```
labtainer webtrack
```

Links to this lab manual and to an empty lab report will be displayed. If you create your lab report on a separate system, be sure to copy it back to the specified location on your Linux system.

### 2.1 Environment Configuration

This lab includes two networked computers, one running a the browser and the other hosting each of the websites used in the lab. The computer hosting websites runs the apache server, and each site is allocated its own resources, as if each site ran on an independent web server. The Firefox browser includes the `Web Developer / Network` tools for to inspect the HTTP requests and responses.

**Starting the Apache Server.** The Apache web server will be running when the lab commences. If you need to restart the web server, use the following command:

```
% sudo systemctl restart httpd
```

**The `Elgg` Web Application.** We use an open-source web application called `Elgg` in this lab. `Elgg` is a web-based social-networking application. It is already set up in on the vuln-server. We have also created several user accounts on the `Elgg` server and the credentials are given below.

| User | UserName | Password |
|---|---|---|
| Admin | admin | seedelgg |
| Alice | alice | seedalice |
| Boby | boby | seedboby |
| Charlie | charlie | seedcharlie |
| Samy | samy | seedsamy |

**Configuring DNS.** We have configured the following URLs needed for this lab:

| URL | Description | Directory |
|---|---|---|
| `http://www.wtlabelgg.com` | Elgg web site | /var/www/webtracking/elgg |
| `http://www.wtcamerastore.com` | CameraStore | /var/www/webtracking/CameraStore |
| `http://www.wtmobilestore.com` | MobileStore | /var/www/webtracking/MobileStore |
| `http://www.wtelectronicsstore.com` | ElectronicStore | /var/www/webtracking/ElectronicStore |
| `http://www.wtshoestore.com` | ShoeStore | /var/www/webtracking/ShoeStore |
| `http://www.wtlabadserver.com` | ReviveAdserver | /var/www/webtracking/adserver |

## 2.2 Clear History and cookies

Please follow the instructions to `clear history` and `cookies` from the Firefox browser.

1. Within the Firefox browser, select `Preferences` per Figure 1 from the top-right menu button, and click on `Privacy and Security` as shown in Figure 2 and then click the `Clear Recent History` per 3. A window `Clear All History` pops up, as shown in Figure 4

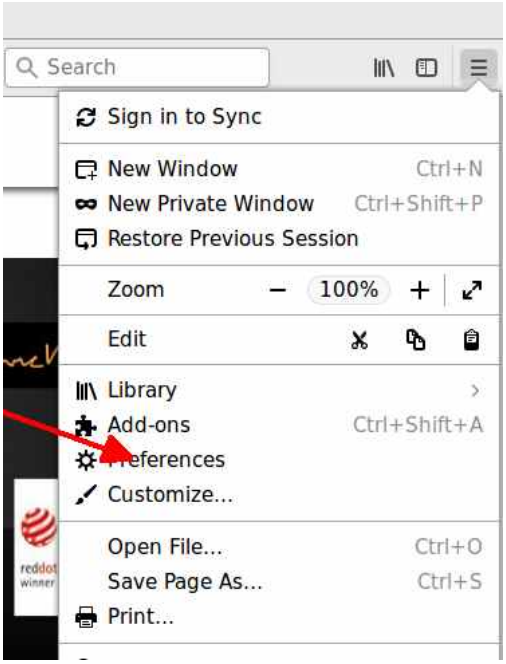


Figure 1: select Preferences.

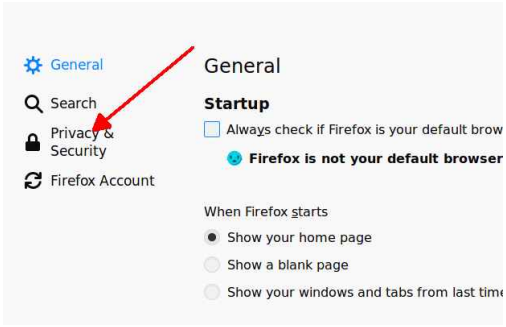


Figure 2: Privacy and Security.

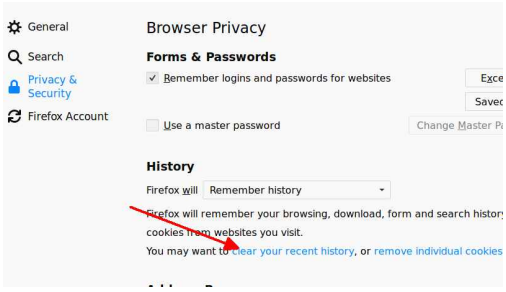


Figure 3: Clear recent history.

2. Select all the check boxes and Click on `Clear Now` button in the pop up window.

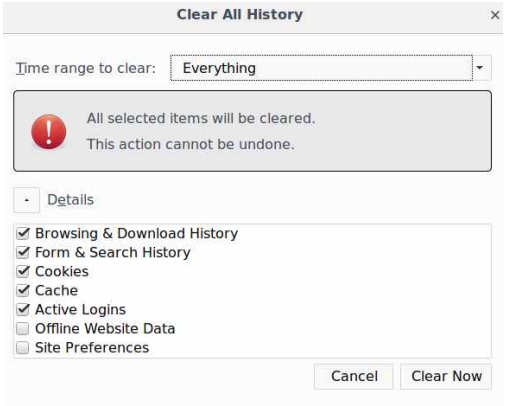


Figure 4: Clear history and cookies.

## 2.3 Open a new private window in Firefox

Please follow the instructions to open a new private window in Firefox and start a private browsing session.

1. With Firefox running, click on the menu button (upper-right) and select `Open a New Private Window` option as shown in Figure 5

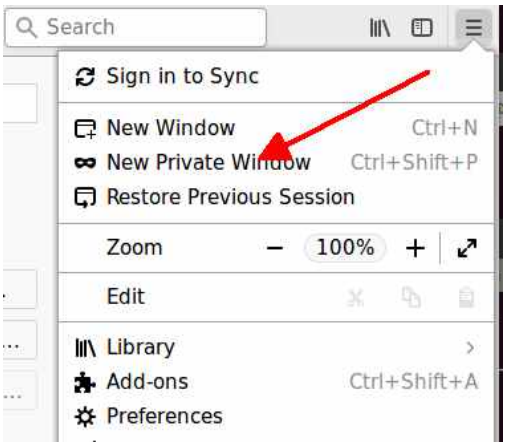


Figure 5: Open a private browser in Firefox.

2. New Private browsing Firefox window opens up, start browsing in that private browser.

## 2.4 Note for Instructors

This lab may be conducted in a supervised lab environment. The instructor may provide the following background information to students at the beginning of the lab session:

1. Information on how to use the pre-configured virtual machine.
2. How to use Firefox and the Developer Storage Inspector and the Firefox `Web Developer / Network` tools.
3. A brief overview of the tasks.

# 3 Lab Tasks

## 3.1 Task 1: Understand the basic working of the web tracking

Nowadays the online web user tracking helps in displaying ads to the targeted audience. When a user visits a website, there are certain ads, of which some of them are targeted advertisements. Say a user visits a certain product in an E-commerce website, he visits the product multiple times, checks the reviews and reads more about the product. Sometime later when the user visits another website, to his surprise he finds the previously visited product is displayed as an advertisement.

The objective of this task is to understand the basic working of the web tracking. In this task you need to open the E-commerce websites, view details of one or more products. Once you login to the `Elgg` website you should see the most visited product displayed as an advertisement.

1. Open `Elgg` website without visiting any website and describe your observation in the lab report.
2. Open Firefox and open the `CameraStore`, `MobileStore`, `ElectronicStore` and `ShoeStore` websites.
3. Click on view details for any products in the websites.
4. Refresh the `Elgg` website in Firefox and describe your observation.
5. Close the browser, reopen it and browse the `Elgg` website. Describe your observation.

Note: If you want to repeat the observations for step 1, clear the `Browsing History` and `Cookies` from the Firefox browser. Please follow the instructions to clear `history` and `cookies` in section 2.2

## 3.2 Task 2: Importance of cookie in Web tracking

`Cookies` are created when a user's browser loads a particular website. The website sends information to the browser which then creates a text file. Every time the user goes back to the same website, the browser retrieves and sends this file to the website's web server. Computer `Cookies` are created not just by the website that the user is browsing but also by other websites that run ads, widgets, or other elements on the web page which are being loaded. These `cookies` regulate the ad display and functioning of other elements on the web page.

The objective of this task is to understand the importance of `cookie` in web tracking. In this task you need to identify the tracking `cookie` using the `Web Developer / Network` tool in Firefox. Please follow the steps below and give your observation.

1. Open any one of the E Commerce websites `CameraStore`, `MobileStore`, `ElectronicStore` and `ShoeStore`.
2. Click on view details for any product in websites and capture `Web Developer / Network` traffic.
3. In `Web Developer / Network`, identify the HTTP request, which set the `third party cookies`, and take the screenshot.
4. Right click on the productDetail page and select `View Page Source`. Find out how the request for tracking `cookie` is sent from the webpage, please take a screenshot and describe your observation.

Third party cookies are cookies that are set by web site with a domain name other than the one the user is currently visiting. For example, user visits website abc.com, say the web page abc.com has an image to fetch from xyz.com. That image request can set cookie on domain xyz.com, and the cookie set on xyz.com domain is known as a third-party cookie. Some advertisers use these types of cookies to track your visits to the various websites on which they advertise.

The objective of this task is to understand how third party cookies are used in web tracking. In this task you need to identify the third party cookie using the Firefox Developer Storage Inspector, which can be viewed using the upper-right menu, select "Developer" and then "Storage Inspector". Please strictly follow the steps below and record your observation.

1. Open any one of the E Commerce websites CameraStore, MobileStore, ElectronicStore, ShoeStore and view details for any product.
2. Open the ad server web page http://www.wtlabadserver.com.
3. Open the Storage Inspector on both pages and observe its displayed Cookie values. Switch between the products webpage and ad server webpage. Describe your observation. (Please do NOT reload the products webpage).

Identify the third party cookie used for tracking in Storage Inspector. Describe your observations in the report and explain why is it called a third party cookie? Give reasons and screenshots to support your observation. A high-level architecture guideline is given in section 4, Figure 6.

Note: If you wish to redo the task from beginning, please delete history and cookies from your Firefox browser. Please follow the instructions to clear history and cookies in section 2.2

### 3.3 Task 3: Tracked user interests and data

The ad servers update their database from users browsing history. They keep track of the web pages visited, articles read, videos watched and any other footprints which user can provide. The objective of this task is to figure out the user interests and view the logged user impressions. In this task you need to understand that all the products viewed by you will be logged in the ad server database. Please follow the steps below and give your observation.

1. Open the E Commerce websites CameraStore, MobileStore, ElectronicStore and ShoeStore.
2. Click on view details for any product in the website.
3. Open www.wtlabadserver.com/preferences.php in a new tab and observe the webpage.

Explain how the user impressions are logged in ad server database, and how is it mapped to a user. Provide evidence to support your observation.

### 3.4 Task 4: How ads are displayed in website

The ad servers use the user profile (browsing history, recent product visits) to display the advertisements and now that the cookie is set to track the user, the ad servers display the targeted advertisements.

In this task you need to observe how the ad is rendered and displayed in the website. Please follow the steps below and give your observation.

1. Open the Elgg website in Firefox browser.

2. Capture and observe the `Web Developer / Network` traffic of the `Elgg` website, identify the HTTP requests which are from a different `domain` (third party).

Explain in detail how the `Elgg` website displays the targeted ads of the user. Provide evidence to support your explanation. (Hint: Use the table displayed in Task3 and `Web Developer / Network` traffic in Task2).

### 3.5 Task 5: Tracking in a Private browser window

In `InPrivate` browsing the browser stores some information such as `cookies` and temporary Internet files so the webpages you visit will work correctly. However, at the end of your `InPrivate` browsing session, this information is discarded. Once the `InPrivate` browser is closed the `cookies` are cleared, and temporary internet files are deleted for that session.

The objective of this task is to understand the working of the web tracking in a private browser window. In this task you need to open the E-commerce websites, view details of one or more products. Once you login to the `Elgg` website (in the same private browser) you should see the most visited product displayed as an advertisement.

1. Open `Elgg` website without visiting any website and describe your observation in the lab report.
2. Open Firefox and open the `CameraStore`, `MobileStore`, `ElectronicStore` and `ShoeStore` websites.
3. Click on view details for any products in the websites.
4. Refresh the `Elgg` website in Firefox and describe your observation.
5. Close the `InPrivate` browser, reopen it and browse the `Elgg` website. Describe your observation.

Compare your observations with Task1. Explain the reasons and provide evidence to support your observations.

Note: Please follow the instructions in section 2.3 to open a new private window in Firefox.

### 3.6 Task 6: Real world tracking

The web tracking in real world involves many ad servers, each ad servers have their own technique of tracking the user interests. In this task you need to visit any of the websites given below and identify the web requests which are sent to the ad servers using the `Web Developer / Network` in Firefox. The websites are:

1. `http://dictionary.reference.com`
2. `http://www.amazon.com`
3. `http://www.careerbuilder.com`

Open the websites, observe the HTTP request and response in `Web Developer / Network`. Capture screenshot of one HTTP request to the real world ad server for each web site. Also identify the `third party cookie` used for that HTTP request.

### 3.7 Task 7: Countermeasures

There are certain countermeasures for the web tracking but most of the websites wont work properly after implementing the counter measures. Most of the websites are highly dependent on `JavaScript` and `third party cookies`. You must have observed that the web tracking tasks are mostly dependent on the `third party cookies`.

The objective of this task is to understand the countermeasures. In this task you should disable the `third party cookies` in Firefox browser and figure out if your impressions are tracked. Please follow the steps below and give your observation:

1. Disable the `third party cookies` from the Firefox browser. Please follow the instructions of how to disable `third party cookies` in Firefox browser in `https://support.mozilla.org/en-US/kb/disable-third-party-cookies`.
2. After disabling the third party `cookies`, open the `CameraStore`, `MobileStore`, `ElectronicStore`, `ShoeStore` websites and `Web Developer / Network`.
3. Click on view details for any products in the websites.
4. In `Web Developer / Network`, identify the HTTP request, which set the `third party cookies`, and take the screenshot.
5. Open `Elgg` website and describe your observation. Also take the screenshot of HTTP request to ads server in `Web Developer / Network`. Compare it with the HTTP request to ads server in Task 4 and explain the difference.

Also there are other ways to mitigate the web tracking. To opt out of targeted advertisement, add browser extensions like RequestPolicy, NoScript and Ghostery which control the third party requests from the web browser. Also one can keep `cookies` for the browsing session, by setting a `cookie` policy only keep cookies until I close my browser which will delete all the `cookies` after the browser window is closed.

Major web browsers provide with an option of `Do Not Track`, which is a feature to let third party trackers know your preference to opt out third party tracking, and it is done by sending a HTTP header for every web request. This `Do Not Track` preference may or may not adhered by the third party trackers. Some third party trackers provide with an option of `Opt Out` of targeted advertisement. Some of them may interpret ”Opt Out” to mean ”do not show me targeted ads”, rather than ”do not track my behavior online”. You can check your tracked online profile created by Google in `www.google.com/settings/ads`. You can also find the Opt out option provided in the above Google URL.

## 4 Guidelines

The diagram in Figure 6 shows the high level architecture of the Web tracking. In this diagram we have three major components, the `E-Commerce` websites, `Ad server` and the `Elgg` website to display the targeted advertisements. Each of the e-commerce websites have `web bugs` or `beacons` to track user preferences. They are implanted as 1px by 1px image tags in the websites.

## 5 Submission

You need to submit a detailed lab report to describe what you have done and what you have observed. Please provide details using `Web Developer / Network`, and/or screen shots. You also need to provide

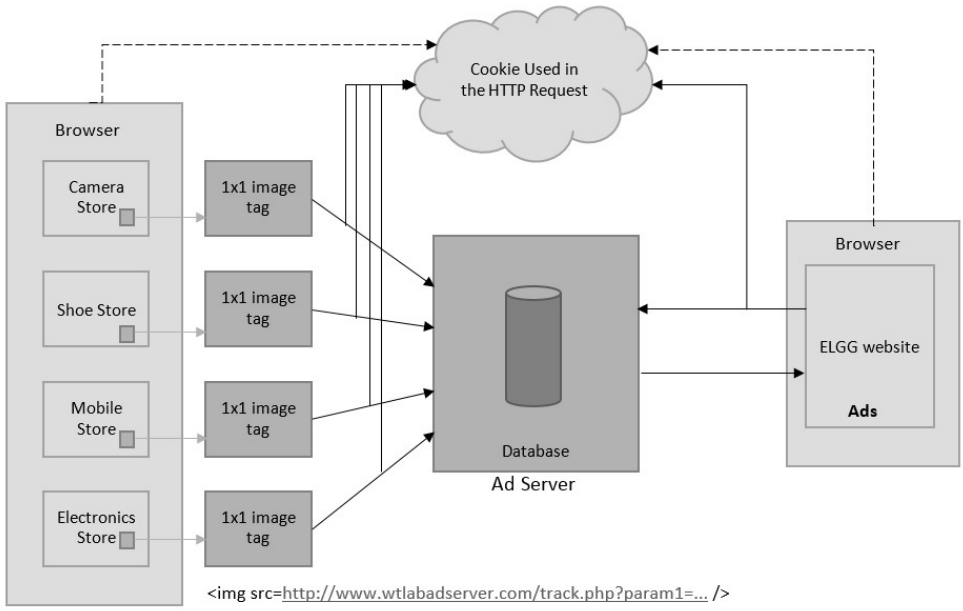


Figure 6: High level architecture diagram of web tracking

explanation to the observations that are interesting or surprising. If you edited your lab report on a separate system, copy it back to the Linux system at the location identified when you started the lab, and do this before running the stoplab command.

After finishing the lab, go to the terminal on your Linux system that was used to start the lab and type:

```
stoplab webtrack
```

When you stop the lab, the system will display a path to the zipped lab results on your Linux system. Provide that file to your instructor, e.g., via the Sakai site.

## References

[1] HTTP Cookie - Wikipedia. Available at the following URL:
`http://en.wikipedia.org/wiki/HTTP_cookie`.

[2] New Cookie Technologies : Harder to See and Remove, Widely Used to Track you
https://www.eff.org/deeplinks/2009/09/new-cookie-technologies-harder-see-and-remove-wide

[3] How Online Tracking companies know most of what you do online
`https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks`.