Web Tracking Lab

Modified for COMP435, Fall 2017 by Cynthia Sturton, UNC-Chapel Hill.

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

You need to use our provided virtual machine image for this lab. The name of the VM image that supports this lab is called `SEEDUbuntu12.04.zip`, which is built in June 2014. If you happen to have an older version of our pre-built VM image, you need to download the most recent version, as the older version does not support this lab. Go to our SEED web page (`http://www.cis.syr.edu/~wedu/seed/`) to get the VM image.

2.1 Environment Configuration

In this lab, we need three things, which are already installed in the provided VM image: (1) the Firefox web browser, (2) the Apache web server, and (3) the `Elgg` web application. For the browser, we need to use the `LiveHTTPHeaders` extension for Firefox to inspect the HTTP requests and responses. The pre-built `Ubuntu` VM image provided to you has already installed the Firefox web browser with the required extensions.

Starting the Apache Server. The Apache web server is also included in the pre-built `Ubuntu` image. However, the web server is not started by default. You need to first start the web server using the following command:

```
% sudo service apache2 start
```

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 the pre-built `Ubuntu` VM image. 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. To access the URLs , the Apache server needs to be started first:

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

The above URLs are only accessible from inside of the virtual machine, because we have modified the `/etc/hosts` file to map the domain name of each URL to the virtual machine's local IP address (`127.0.0.1`). You may map any domain name to a particular IP address using `/etc/hosts`. For example you can map `http://www.example.com` to the local IP address by appending the following entry to `/etc/hosts`:

```
127.0.0.1       www.example.com
```

If your web server and browser are running on two different machines, you need to modify `/etc/hosts` on the browser's machine accordingly to map these domain names to the web server's IP address, not to `127.0.0.1`.

Configuring Apache Server. In the pre-built VM image, we use Apache server to host all the web sites used in the lab. The name-based virtual hosting feature in Apache could be used to host several web sites (or URLs) on the same machine. A configuration file named `default` in the directory `"/etc/apache2/sites-available"` contains the necessary directives for the configuration:

1. The directive `"NameVirtualHost *"` instructs the web server to use all IP addresses in the machine (some machines may have multiple IP addresses).

2. Each web site has a `VirtualHost` block that specifies the URL for the web site and directory in the file system that contains the sources for the web site. For example, to configure a web site with URL `http://www.example1.com` with sources in directory `/var/www/Example_1/`, and to configure a web site with URL `http://www.example2.com` with sources in directory `/var/www/Example_2/`, we use the following blocks:

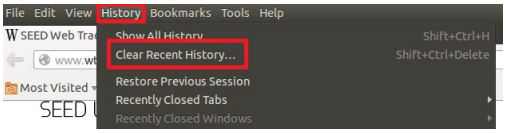

Figure 1: Open Firefox and select History.

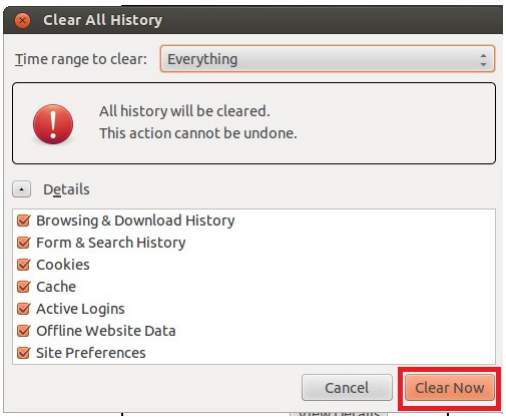

Figure 2: Clear history and cookies.

```
<VirtualHost *>
    ServerName http://www.example1.com
    DocumentRoot /var/www/Example_1/
</VirtualHost>

<VirtualHost *>
    ServerName http://www.example2.com
    DocumentRoot /var/www/Example_2/
</VirtualHost>
```

You may modify the web application by accessing the source in the mentioned directories. For example, with the above configuration, the web application `http://www.example1.com` can be changed by modifying the sources in the directory `/var/www/Example_1/`.

2.2 How to clear history and cookies

These are instructions for clearing the history and cookies from the Firefox browser. You will need to do this for some of the tasks.

1. Open Firefox browser, select `History` from the top menu, and click on `Clear Recent History` option from the menu (Figure 1). A window `Clear All History` pops up.

2. Select all the check boxes and Click the `Clear Now` button in the pop up window (Figure 2). Close the Firefox browser, reopen and start browsing.

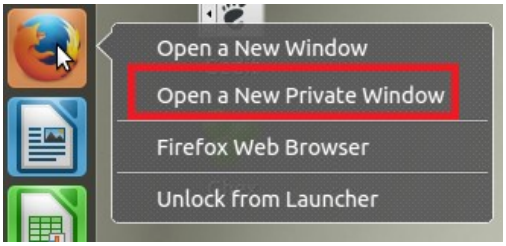

Figure 3: Open a private browser in Firefox.

2.3 How to open a new private window in Firefox

These are instructions for opening a new private window in Firefox to start a private browsing session. You will need to do this in Task 5.

1. On the left desktop menu, right click on the Firefox icon, Select `Open a New Private Window` option as shown in Figure 3.
2. New Private browsing Firefox window opens up, start browsing in that private browser.

3 Lab Tasks

For this assignment, you will submit a **pdf** file describing both your observations and requested screenshots for the following tasks. Your observations should be short - please limit yourself to 50-100 characters depending on the question (roughly one to two sentences). Your screenshots should be limited in size, specifically less than 1 MB. The expected time of completion is 3-5 hours.

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

Nowadays the online web user tracking helps in displaying ads to a targeted audience. When a user visits a website, there are certain ads, of which some of them are targeted. Say a user visits a certain product in an E-commerce website, they visit the product multiple times, check the reviews and read more about the product. Sometime later when the user visits another website, they find 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 and view details of one or more products.

1. Open Firefox and open the Elgg website (`http://www.wtlabelgg.com`) without visiting any other website.
2. Open Firefox and open the `wtCameraStore.com`, `wtMobileStore.com`, `wtElectronicStore.com` and `wtShoeStore.com` 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

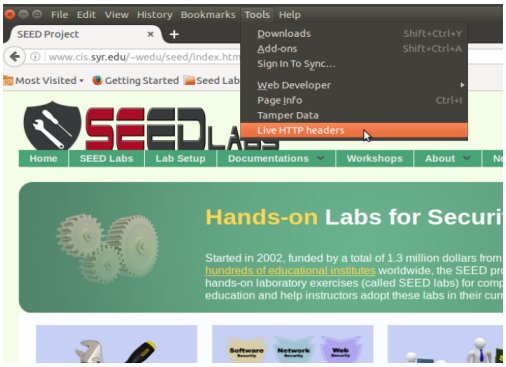

Figure 4: Open the `LiveHTTPHeader` extension in Firefox.

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. 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.

Third party cookies are cookies that are set by a 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 tracking cookie using the `LiveHTTPHeaders` Firefox extension. Please follow the steps below and give your observation.

1. Open any one of the E-Commerce websites `wtCameraStore.com`, `wtMobileStore.com`, `wtElectronicStore.com` or `wtShoeStore.com`.

2. Click on view details for any product in the website and use the `LiveHTTPHeader` Firefox extension to capture the HTTP header traffic information. You can open the `LiveHTTPHeader` Firefox extension by going to the `Tools` menu in the browser (Figure 2).

3. In `LiveHTTPHeaders` identify the HTTP request that set the third party cookies, and **take the screenshot**.

4. Right click on the Product Detail page and select `View Page Source`. Find out how the request for the tracking cookie is sent from the webpage. **Take a screenshot** with the relevant source code highlighted, and **describe your observation**.

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 the user creates. 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 `wtCameraStore.com`, `wtMobileStore.com`, `wtElectronicStore.com` and `wtShoeStore.com`.
2. Click on view details for any product on at least two different websites.
3. Open `www.wtlabadserver.com/preferences.php` in a new tab and **take a screenshot**.
4. **Answer this question**: What information is tracked about a user and how is the information mapped to a particular user?

3.4 Task 4: How ads are displayed in a 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 `LiveHTTPHeader` traffic of the Elgg website, identify the HTTP requests which are from a different domain (third party).
3. **Take a screenshot** of an HTTP request to the third party. Be sure the cookies being sent back to the server are included in the screenshot.
4. Take a moment to make sure you understand how the Elgg website displays the targeted ads to the user. Hint: take note of the cookies sent in the previous step. The value for the UserTrackID cookie should match what you saw in the table displayed in Task 3.

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 for that session are deleted. Instructions for starting an InPrivate browsing session are given in Section 2.3

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 open the Elgg website (in the same private browser) you should see the most visited product displayed as an advertisement.

1. Open the Elgg website without visiting any website.
2. Open Firefox and open the `wtCameraStore.com`, `wtMobileStore.com`, `wtElectronicStore.com` and `wtShoeStore.com` websites.
3. Click on view details for any products in the websites.
4. Refresh the Elgg website in Firefox and **describe your observations**.

5. Close the InPrivate browser, reopen it and browse the Elgg website. **Describe your observations.**

6. **Answer this question:** Why does your final observation here differ from what you saw at the end of Task 1?

3.6 Task 6: Real world tracking

Web tracking in the real world involves many ad servers, each with their own techniques for tracking user interests. In this task you need to visit one of the websites given below and identify the web requests which are sent to the ad servers using the `LiveHTTPHeaders` in Firefox. The websites are:

`http://dictionary.reference.com`

`http://www.amazon.com`

`http://www.careerbuilder.com`

Open the websites and observe the HTTP request and response in `LiveHTTPHeaders`.

1. How many HTTP requests go to third parties?

2. How many different third parties are setting cookies?

You don't have to submit your answers to this task. It's just a fun, eye-opening exercise.

3.7 Task 7: Countermeasures

There are certain countermeasures for the web tracking but many websites wont work properly after implementing the countermeasures; they are dependent on JavaScript and third party cookies for their functionality. And, as you've observed, web tracking tasks are mostly dependent on third party cookies.

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

1. Disable the third party cookies from the Firefox browser. You can find instructions on how to disable third party cookies in the Firefox browser here: `https://support.mozilla.org/en-US/kb/disable-third-party-cookies`.

2. After disabling the third party cookies, open the `wtCameraStore.com`, `wtMobileStore.com`, `wtElectronicStore.com`, `wtShoeStore.com` websites and `LiveHTTPHeaders`.

3. Click on view details for any products in the websites.

4. In `LiveHTTPHeaders`, identify the HTTP request that set the third party cookies, and **take the screenshot**.

5. Open the Elgg website and **describe your observation**.

6. **Take the screenshot** of the HTTP request in `LiveHTTPHeaders` sent to the ad server when viewing the Elgg site. Compare it with the HTTP request sent to the ads server in Task 4 and **describe the difference**.

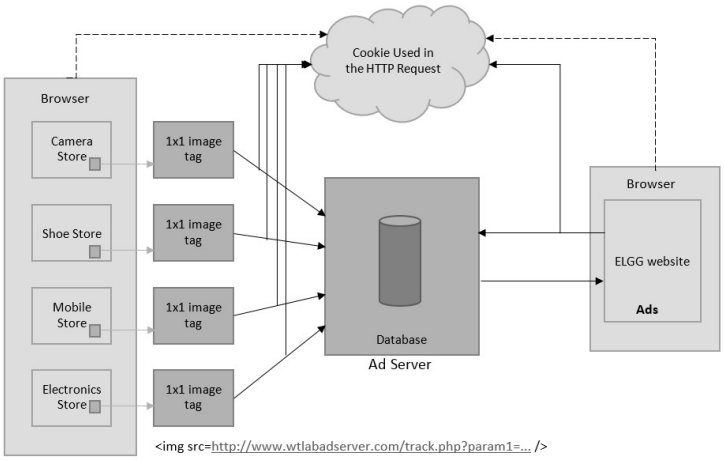

Figure 5: High level architecture diagram of web tracking

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

Major web browsers provide an option of `Do Not Track`, which is a feature to let third party trackers know your preference to opt out of third party tracking. It is done by sending a HTTP header for every web request. This Do Not Track preference may or may not be adhered to by the third party trackers. Some third party trackers provide users with an option to `Opt Out` of targeted advertisement. However, 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 5 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 will be submitting your assignment through Sakai. In your PDF submission, provide the following information:

1. Task 1: Step 4 - Briefly describe your observations.
2. Task 1: Step 5 - Briefly describe your observations after closing and repoing your web browser.
3. Task 2: Step 3 - Screenshot of HTTP request setting third party cookies
4. Task 2: Step 4 - Screenshot of product detail page source requesting the tracking cookie

5. Task 2: Step 4 - Describe how the code enables the tracking cookie and from which third party site the cookie will be sent.
6. Task 3: Step 3 - Screenshot of preferences.php output
7. Task 3: Step 4 - What information is tracked about a user and how is the information mapped to a particular user?
8. Task 4: Step 3 - Screenshot of HTTP request to a third party
9. Task 5: Step 4 - What do you see after refreshing the Elgg website?
10. Task 5: Step 5 - What happens after you close and reopen the browser in "Private/Incognito" mode?
11. Task 5: Step 6 - Does this answer differ from what you saw in task 1? Why or why not?
12. Task 7: Step 6 - Submit a screenshot of the HTTP request to the ad server.
13. Task 7: Step 6 - What is the difference between the HTTP requests sent to the ad server in this task and the ones sent in task 4?

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`.