SY306 Lab Five

Web Tracking

1. Introduction

Behavioral targeting is a type of online advertising where ads are displayed based on the user's 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

For this lab you will work in groups of two and turn in one common report. Work will be completed in the provided virtual machine. After logging into your workstation, open a terminal and run the following code to extract the VM:

```
unzip -d Documents/VM/ /opt/software/SEEDUbuntu12.04.zip
```

NOTE: If the file is unavailable in the above location, it can be downloaded from here **(LOCAL).**

There are two **external** mirrors from where you can also try downloading the file: mirror 1 or mirror 2. If downloaded from one of the three sites, it can extracted with the below lines:

```
unzip -d Documents/VM/ Downloads/SEEDUbuntu12.04.zip
```

After extracting the VM, click on the following link for instructions to configure and start the VM.

- Login Name: **seed**
- Password: **dees**

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 and the `Firebug` extension to examine the cookies. The pre-built Ubuntu VM image provided to you already has the Firefox web browser installed 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 might not be 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. There are 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, make sure the Apache server was 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

2.2 Clear History and cookies

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

1. Open Firefox, select History from the top menu, and click on Clear Recent History option from the menu, as shown in Figure 1.

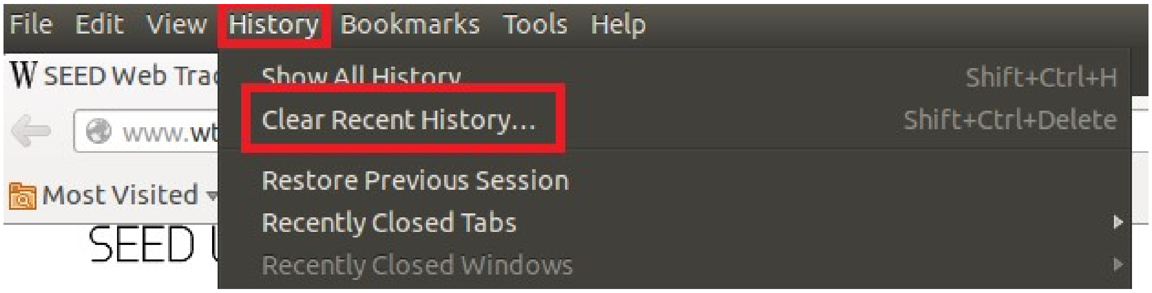
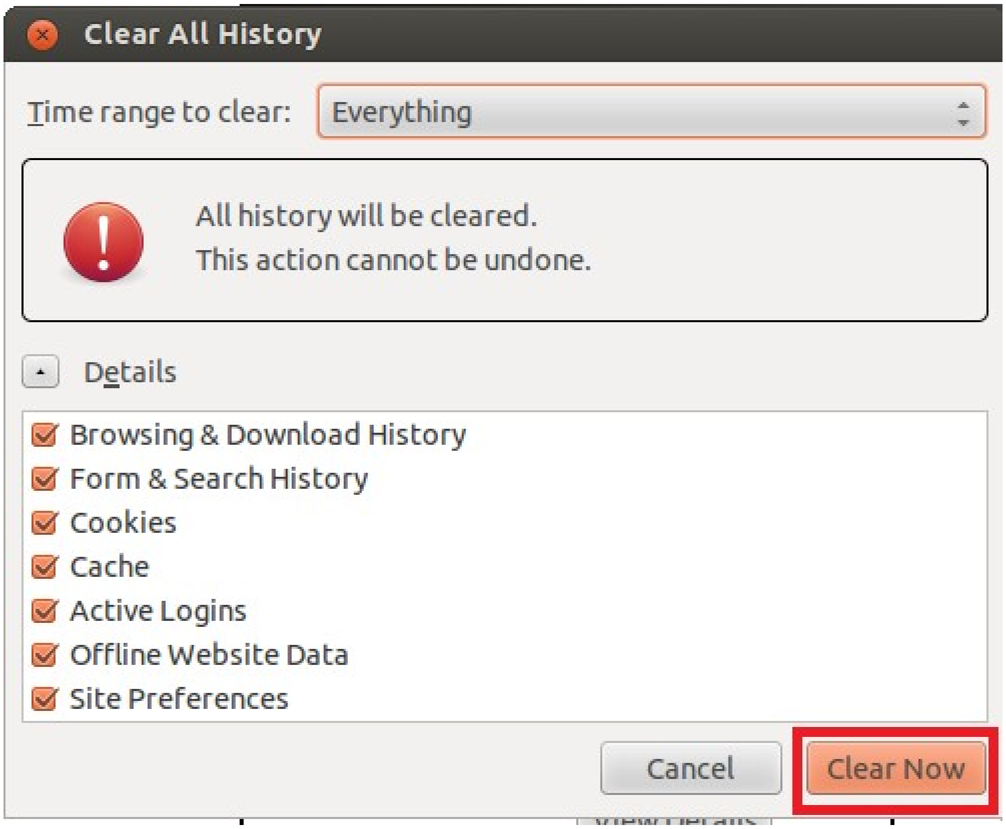


Figure 1

2. A window Clear All History pops up, as shown in Figure 2:



Time range to clear: Everything

All history will be cleared.
This action cannot be undone.

- ☑ Browsing & Download History
- ☑ Form & Search History
- ☑ Cookies
- ☑ Cache
- ☑ Active Logins
- ☑ Offline Website Data
- ☑ Site Preferences

Figure 2

3. Select all the check boxes and Click on Clear Now button in the pop up window. Close the Firefox browser, and re open.

3 Lab Tasks

Submission

You must create a folder on your public_html called "Lab05" (without the quotes) and store your work in that directory. Because you will do this lab in teams, only one submission is required for each team. The final report should be stored in one of the team members Lab05 folder. Create a file *lastname1_lastname2_lab05.docx*. For this lab you need to submit a detailed lab report (in *lastname1_lastname2_lab05.docx*) to describe what you have done and what you have observed. Please provide details using screenshots. You also need to provide explanation to the observations that are interesting or surprising. Complete this report as you go through the tasks below!!

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. Follow instructions in 2.2 to clear history and cookies. Perform the following steps using normal (not private) browsing in Firefox.
2. Open `Elgg` website without visiting any other website. Describe your observation in the lab report.
3. Open Firefox and open the CameraStore, MobileStore, ElectronicStore and ShoeStore websites.
4. Click on view details for any products in the websites.
5. Refresh the `Elgg` website in Firefox and describe your observation.
6. Close the browser, reopen it and browse the `Elgg` website. Describe your observation.

Note: If you want to repeat the observations for step 1, make sure you start at section 2.2 (clear the Browsing History and Cookies from the Firefox browser).

3.2 Task 2: Importance of cookies 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.

1. The objective of this task is to understand the importance of cookies in web tracking. In this task you need to identify the tracking cookie using the `LiveHTTPHeaders`. Please follow the steps below and give your observations.

2. Open any one of the E Commerce websites CameraStore, MobileStore, ElectronicStore and ShoeStore.

3. Open `LiveHTTPHeaders` in Firefox by going to tools (located at the top of browser screen)->Live HTTP Headers.

4. Click on view details for any product in websites and capture `LiveHTTPHeader` traffic. Make sure the Capture checkbox is checked.

5. In `LiveHTTPHeaders`, identify the HTTP request which sets the third party cookies, and take the screenshot.

6. Right click on the product Detail 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.

1. The objective of this task is to understand how third party cookies are used in web In this task you need to identify the third party cookie using `Firebug` (Firefox browser extension, which is present in right corner of the browser.) and record your observations. Please strictly follow the steps below and give your observation.

2. Open any one of the E Commerce websites CameraStore, MobileStore, ElectronicStore, ShoeStore and view details for any product.

3. Open the ad server web page http://www.wtlabadserver.com. You don't have to log in

4. Open Firefox extension `Firebug`. Observe the `Firebug` in ad server web page and product web page. 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 `Firebug` extension. 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 3.

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 users 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 the ad server database, and how is it mapped to a user. Give 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 `LiveHTTPHeader` 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 `LiveHTTPHeader` traffic in Task2).

3.5 Task 5: Tracking in a Private browser window

In Private 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 Private browsing session, this information is discarded. Once the Private 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. For this task, use private browsing. (Right click on Firefox icon and choose "Open a New Private Window")
2. Open `Elgg` website without visiting any website and describe your observation in the lab report.
3. Open Firefox and open the CameraStore, MobileStore, ElectronicStore and ShoeStore websites.
4. Click on view details for any products in the websites.
5. Refresh the `Elgg` website in Firefox and describe your observation.
6. Close the Private 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.

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 the websites given below and identify the web requests which are sent to the ad servers using the `LiveHTTPHeaders` in Firefox. The websites are:

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

Open the websites, observe the HTTP request and response in `LiveHTTPHeaders`. Capture a 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: Cyber Defense [Countermeasures]

There are certain countermeasures for web tracking but most websites will not work properly after implementing them. Many 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 link. (To get to the Options menu, go to Edit, Preferences, Privacy, Use Custom Settings For History)
2. After disabling the third party cookies, open the CameraStore, MobileStore, ElectronicStore, ShoeStore websites and `LiveHTTPHeaders`.

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

4. In `LiveHTTPHeaders`, identify the HTTP request which sets 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 `LiveHTTPHeaders`. Compare it with the HTTP request to ads server in Task 4 and explain the difference.

There are also other ways to mitigate web tracking. To opt out of targeted advertisement, add browser extensions like RequestPolicy, NoScript and Ghostery which control third party requests from the web browser. Another way to mitigage web tracking is to set a cookie policy to only keep cookies until I close my browser which will delete all the cookies after the browser window is closed.

Major web browsers provide a 'Do Not Track' option, which is a feature to let third party trackers know your preference to opt out of third party tracking, and it is done by appending a HTTP header for every web request. This Do Not Track preference may or may not be adhered to by third party trackers. Some third party trackers provide an option to Opt Out of targeted advertisement. Remember, they 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 3 below 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.

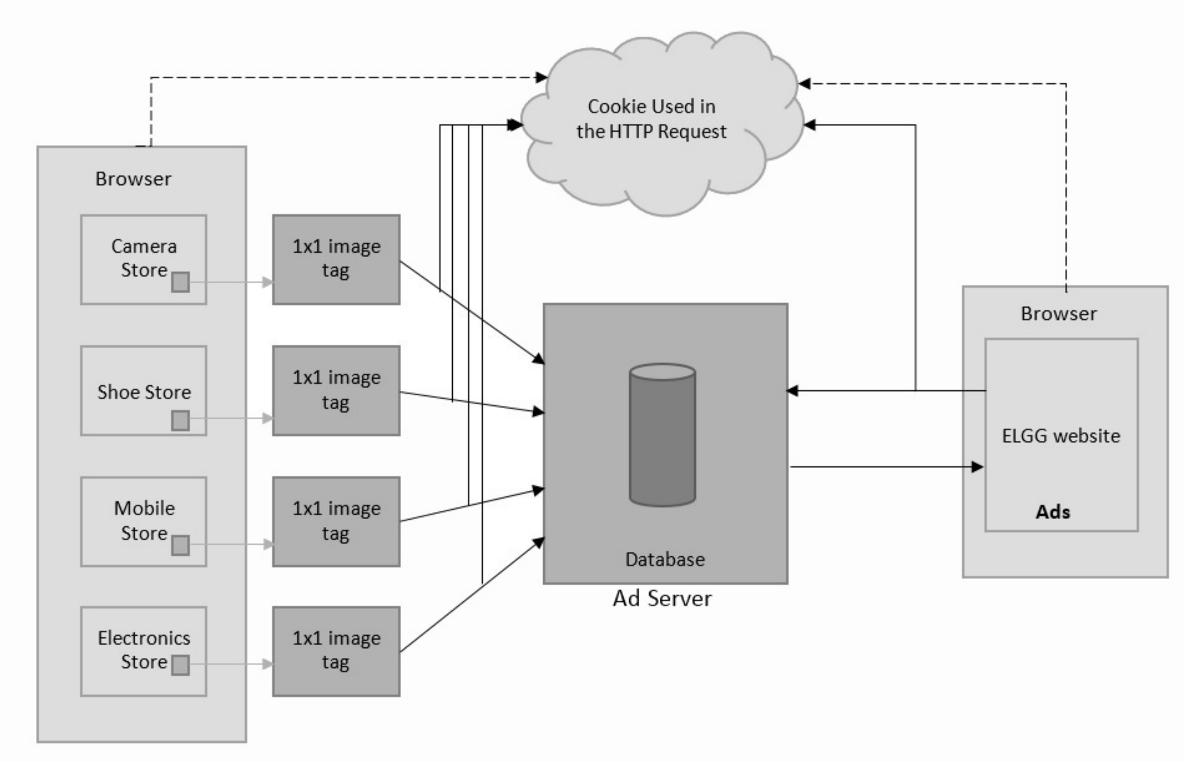


Figure 3

5 Deliverables

1. You and your teammate need to submit a report *lastname1_lastname2_Lab05.docx* describing what you have done and what you have observed. Please provide details using screen shots. You also need to provide explanation to the observations that are interesting or surprising. Only one report is required per team.

2. The file should be in a folder called Lab05 on both your public_html folders (you and your teammate). **Your instructor will assume that your web pages are viewable at http://midn.cyber.usna.edu/~m21xxxx/Lab05/lastname1_lastname2_lab05.docx**.

3. In your default.html page on your public_html, add a link to lastname1_lastname2_lab05.docx file under the heading "Lab05"

4. **Turn in (due before lab on Friday) One submission per team:**

   a. **Paper submission:** turn in the following at the beginning of Lab06:

      i. A completed assignment coversheet with both your names and comments. Your comments will help us improve the course.

      ii. A printout of *lastname1_lastname2_lab05.docx*

   b. **Electronic submission:** Submit the Lab05 report [*lastname1_lastname2_lab05.docx*] via the online system: `submit.cs.usna.edu` by the following deadlines:

      - Section **4341**: 23:59, Wednesday, February 13th
      - Sections **1151, 3351, 5551**: 23:59, Thursday, February 14th

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.

Notice/Acknowledgements:

This lab was adapted from Web Tracking lab created by Wenliang Du at Syracuse University http://www.cis.syr.edu/~wedu/seed/.