LE/EECS4480 A - Computer Security Project (Summer 2021-2022) Research Topic: Dark Web Cookies Research Project

Mhd Yazan Armoush
217188921
Professor : Uyen Trang Nguyen

August 16, 2022

Introduction:

- **Summary of the Project:**
  The project is centered on cookies' research and the dark web. Mainly, the research focuses on building a web crawler that passes the login page and exports the needed information into a separate text file. To be more specific, the crawler should contain the cookie of the login page which will allow it to go directly into the sign in page. After the sign in page has been passed, the crawler should be able to export any needed information such as emails, data, phone numbers, paragraphs. In other words, any information that is needed throughout the research from any site, the crawler should be capable of exporting that into a separate text file on its own. The main challenge of this research is to have access to cookies that are related to dark websites.

- **Previous Research Results:**
  You will be receiving a folder by the beginning of the project called "Dark Web Project". Within this folder, you will see all the coding files that were added by other students regarding the project. The

main file that you need to be familiar with is called “web crawler.py”.

What is the purpose of this file and what does it do?
This code contains a crawler that is responsible for copying information from any valid dark website that does not have a login requirement. In other words, any website that is valid and working on the dark web without a login page meets this crawler requirement and allows it to proceed with exporting the needed information. All you should do is copy paste the website into line 78 of the code and run the code. There will be a screen on the site called “sites” where you will be able to view all the information that is currently copied from the website you have entered.

What should you do when you receive the folder?

1. Download the folder into a new directory to keep all the information organized.
2. Download Tor browser on your machine and make sure you watch some security related videos throughout the download to avoid any viruses or bugs on your machine.
3. Open the file using any platform you wish to use (PyCharm is preferred).
4. Go to readme.md Folder and follow the given steps.
5. You might have a hard time getting the code to work after going through the previous setup, all you should do is check the permissions that you have on the WebCrawler file and check if you have read, written, or executed all on from your side. If any of the permissions are denied, I would suggest allowing them since I had an error running the code due to a permission issue where my changes to the code were not getting saved. The error was the following “HTTPConnectionPool(host=’zqktlwi4fecvo6ri.onion’, port=80): Max retries exceeded with url: /wiki/index.php/Main Page (Caused by NewConnectionError(’¡urllib3.connection.HTTPConnection object at 0x1058212b0¿: Failed to establish a new connection: Socket error: 0x01: General SOCKS server failure’))”. If you get this error, it means you do not have all the permissions fully running.

6. After the code is fully working on your machine, try to go into the dark web through the Tor browser and find a valid site that requires no login.
7. Paste the link back into the code (line 78) and run the code.
8. If everything is fully working, you should see the information getting copied on the left side of your screen (if you're using PyCharm CE).
9. Sometimes the code might take a lot of time to make the final copy, so it is totally okay to give it time while exporting.

- **What did I learn about Tor browser / Firefox/ Chrome:**

  Tor Browser

  1. Tor (short for "The Onion Router") is a completely free, open-source browser that automatically erases all browsing history.
  2. The Tor Browser is a web browser that encrypts and anonymizes web traffic completely, making it an effective method to hide identity online.
  3. The Tor Browser gives access to the hidden and dark side of the internet, also known as "the dark web".
  4. A fun fact about the Tor Browser is that it is banned in some countries around the globe for its powerful ability to grant access to all parts of the web freely.
  5. The Tor Browser was originally born for military use in order to enable anonymous online communication between military organizations.

  - **How to avoid getting any viruses or vulnerabilities while using Tor browser:**

    1. Download a VPN extension to protect your machine.
    2. Go to the Tor project website and download the right Tor browser that is associated with your machine.
    3. After installing the browser, make sure you change your VPN if you would like to hide your actual IP address while using the dark web.

Firefox

1. Tor Browser is a privacy-enhancing web browser based on Firefox, with more than a million users.
2. At the Tor Project, the Tor Browser team maintains and develops several dozen patches to Firefox that provide Tor Browser users with extra protections for their privacy and security.
3. Firefox contains some amazing extensions that might be useful for this project, and I highly recommend trying them out throughout the project.

Chrome

1. One of the best features about chrome is you can basically start checking the cookies of any chrome site throughout their developer tool option and you can view any html, css, js code of any site.
2. Chrome has a lot of great features including finding cookies and running the code on a different extension.
3. Chrome will be the best browser to start with until you fully understand the project. It helps you learn about cookies and understand what their purpose is. You can try out some live demos through some YouTube videos.

- **Selenium**

What is Selenium?

1. Selenium is a free and open-source framework utilized to test and validate web applications across various platforms and browsers.
2. Multiple programming languages like Java, C, Python etc.,can be utilized to create the 'Selenium Test Scripts'. Note that any testing done using the Selenium testing tool is referred to as "Selenium Testing".
3. Selenium Software is not a single tool, but a suite of software pieces each catering to varying Selenium QA testing that adheres to the needs of an organization. Some of the tools include: Selenium Integrated Development Environment (IDE), Selenium Remote Control (RC), WebDriver, Selenium Grid.

Selenium is a great tool that helped me a lot during the research since it can do multiple tasks at once in the background. At the beginning of the research, I used selenium to try to sign into my gym account on the chrome browser. I used python and made a code that performed the sign in on my gym membership account and that turned out great for me. There are a lot of people online who use selenium to do some amazing tasks such as creating a bot to find certain information or copy data from different sources. I personally used selenium mostly on the Chrome browser because chrome shows the entire Html code through the development tool. By having the code, I used it to guide selenium on where and how it should go on in every step. You can see everything your code is doing live while performing the task and that was cool to learn throughout the project.

Why selenium could be the solution of this project?
The reason I suggested selenium for the research is because you can run Selenium webdriver in Headless mode, which allows it to operate in the background without launching a browser window. You can achieve this by modifying the driver's set-up code to include the necessary capabilities. This way, there is no need for cookies as selenium will go manually to the code and export any available information. Since Firefox is related to the Tor browser, then I would suggest running selenium on Firefox driver and making it take certain tasks. More importantly, finding a way to let selenium export information from sites is more common online than stealing the cookies of the website.

- Cookies

  What are Cookies and what should you know about them?

  1. Cookies are the information of any visited site since it gets saved to your machine using the web browser.
  2. Cookies let sites record all the activities you make on any site. For instance, looking for a particular item on Amazon, or finding something on Facebook marketplace. The reason behind that is to keep the user more engaged later on by showing them more items that relate to their interests in different social media platforms (Facebook, Google, etc).

3. Cookies allow the user to stay signed in on any site as it stores the email address and the password of the user after the first request was made to view the site. For example, if you sign in to Facebook for the first time you will get a message at the top corner of your screen asking if you would like to save the information that you have entered. This message allows the cookies to store the given data for future use.

Usage of Cookies

1. **Session management:** cookies allow websites to recognize users and remember their individual login information and preferences. Saved preferences may include certain sports team news, politics etc.
2. **Personalization:** Customized advertising is the main way cookies are used to personalize your sessions. When viewing certain items or parts of a site, cookies use this data to help build targeted ads that users might enjoy.
3. **Tracking:** Shopping sites use cookies to track items users previously viewed, allowing the sites to suggest other goods they might like and keep items in shopping carts while they continue shopping.

What are the different types of HTTP Cookies?

1. **Authentication:** These cookies track whether a user is logged in and under what name. They also streamline login information, so users don't have to remember site passwords.
2. **Tracking:** These cookies track multiple visits to the same site over time. Some online merchants, for example, use cookies to track visits from particular users, including the pages and products viewed. The information they gain allows them to suggest other items that might interest visitors. Gradually, a profile is built based on a user's browsing history on that site.
3. **Encrypted Cookie:** Those are HTTP cookies that often come from the web server that serves to encrypt cookie values. This adds a layer of protection since the browser client can't decrypt

the data. This makes HTTP cookies meaningful only to the back-end application.

4. **Vanilla Cookie:** This is a type of cookie that never got encrypted between the user and the server. A lot of sites do not have protection on their site cookies and this could be very dangerous since it gives a big chance for errors, hacks, and vulnerabilities to occur. This could cause some serious consequences to the site.

- **Accomplishments**

  What is Cross site scripting and how did it help my research: Cross site scripting attacks are a type of injection in which malicious scripts are injected into different websites. It happens when an attacker uses a web application to send malicious code. The code mostly is a form of a browser site script to a different end user. The user on the side has no idea that this script is not trusted.Then the user will be receiving the malicious script and it will be looking the same as the normal site he was trying to access. For example, if you decide to make a purchase on amazon after placing some items in your cart, the request for the purchase will go directly to amazon server as you might imagine. But what is happening under the hood is that the request you made was sent to the hacker. Then, the hacker will be able to view the cookie you are trying to make the request to. Once the cookie is under the hacker's control, he will be sending you a fake response back that looks exactly as an amazon reply would. The response will contain a fake cookie since your original was taken by the attacker. The attacker now has full control of your cookie and he will be able to sign in into the same account that you have made the purchase from. This is why keeping your cookies safe is important. Cross site scripting has helped me by giving me a hint that the cookies I originally have on my machine can be found every time I sign in into any site.

  Ways I was able to find the cookies:

  - **First Way:**
    Let's imagine you have signed in to your Facebook account and would like to view the cookies of your Facebook site.

    1. Make sure you open your Facebook page on the Chrome site.

2. Click on the three dots at the top right → "More tools" → then "Developer tools".
3. After the page is open, you are able to see the HTML, CSS, JS to your current page.
4. Click on Console at the top, then within the console, type "document.cookie".
5. Once document.cookie has been entered, you will be able to see the cookie in a new line.
6. If the cookie is encrypted, then it is going to be long and filled with random letters and numbers.
7. If the cookie is not encrypted (vanilla cookie), then it is going to be short and readable.
8. Congratulations, that is the first way of viewing cookies.

– **Second Way:**
I found a great way to find the cookies using the Firefox browser extension.

1. Go to FireFox Addons and search "Cookie Quick manager by Ysard"
2. Download the extension to your FireFox.
3. You will be able to see it at the top right corner after the installation is done.
4. Try to play around with the extension to get an idea about what options are available. You will have the power to delete, edit, remove, and save the cookies of any visited website.
5. Congratulations, that is the second way of viewing cookies.

How do you use Vanilla cookie to pass the sign in page:
The following steps can run on any site that does not contain encrypted cookies. In other words, the cookies of the site have to be public and not secured. URL example to have a demo test:
http://testphp.vulnweb.com/userinfo.php

1. Make sure the cookie is not encrypted by following the previous videos.
2. Sign in into the website by inserting the username and password.

    3. After signing in, click on the Cookie extension that is shown on the top right corner and click on “search cookies” (Second option).
    4. You will be directed to a white screen where you will be able to view the cookie itself of the website.
    5. Click on the export/import button that is located at the bottom left, then click ‘download all’.
    6. Go back to the signed in page and save its URL on a separate text.
    7. Logout of the website and delete all the cookies that were saved from the top right corner under the “cookie extension options”. Choose to delete all cookies.
    8. Go back to the main website page then follow step 3 again.
    9. Click on the export/import tab then choose ‘restore cookie from file’. Then add the cookie file that was originally saved. A message
    10. should show after the cookie has been added, which states that the operation was successful.
    11. Open a new tab and paste back the URL that was copied in step 6.
    12. You should be directed into the sign in page.

- **Upcoming Research**
  So far I was able to sign in to any website that does not contain an encrypted cookie and was able to sign in into websites that contain vanilla cookies.

  Suggestions I feel you should focus more on are the following:

  1. Be able to sign in to encrypted cookie sites.
  2. Build a web crawler that does all the signing in steps for you.
  3. I would suggest using Selenium as a new way to pass the sign in page in case cookies were impossible to get.
  4. Selenium has the ability to run in the background instead of a web browser. Which means Selenium can run as a web crawler in

the background. My suggestion is trying to build a Selenium Web crawler to pass the sign in page.

5. Find a way to decrypt the encrypted cookies through a code or an extension.
6. Read more articles on the Dark web and how to export cookies from it.
7. Test more websites to see how many websites function with the steps that were discovered.
8. Test site scripting on multiple websites to compare the cookies of different users.