ECS 152: Computer Networks

Fall 2025

Project 2

(100 points)

Due Date: 11/17/2025 at 11:59 PM

Team: The project is to be done in a team of at most 2 students. You cannot discuss your code/data with other classmates (*except* your project partner).

All submissions (including your code) will be checked for ***plagiarism*** against other submissions as well as the public Internet. Plagiarized submissions will be entitled to ***zero*** points. Generative AI code is not allowed and also entitled to ***zero*** points.

Project 2 consists of three parts:

1. DNS client from scratch
2. Web crawling and HAR file analysis

Part 1: DNS client from scratch (50 points)

In this part, you will implement a DNS client to resolve the IP address for tmz.com from scratch using socket API. You cannot use any libraries/methods that simplify DNS implementation such as gethostbyname, getaddrinfo, etc.

You will implement the following:

- You will first build the DNS request payload from scratch.
- You will then send the DNS request to a public DNS resolver from this list. If you don't get a response within 10 seconds, try another resolver from the same region in the list.
- You will then receive a response from the resolver and unpack it.
- Upon unpacking the response, you will extract the DNS records and identify the type of DNS record and IP address from each DNS record.
- Once you get the IP address of tmz.com, you will make an HTTP request to the IP address using socket API.
- Measure the RTT between your machine and the public DNS resolver.
- Measure the RTT between your machine and the tmz.com server (when you make the HTTP request to its IP).

Note that the link only contains the list of root DNS servers. You will have to extract the IP of the TLD DNS server and then the authoritative DNS server to get the IP address. You have to use the sendto() and recvfrom() methods of the socket API to send/receive UDP packets. You have to use the connect(), sendall() and recv() methods for TCP packets. You **cannot** use any packet construction APIs – we want you to construct the DNS and HTTP requests from scratch.

Report: proj2_[name1]_[student_id1]_[name2]_[student_id2].pdf

At the beginning of the page, specify the following:

1. Full name of student 1 (Student ID)
2. Full name of student 2 (Student ID)
3. Name of the Python source codes and HTML files submitted.

In your report, you will also describe how you created the DNS packet and how you parsed the response.

Part 2: Web crawling and HAR file analysis (50 points)

You will implement the following:

- You will also automatically crawl the home page of the top 100 sites from this list using Selenium. If a website is unable to be reached, move to the next site from the list until 100 sites have been successfully crawled.
- While visiting each site, you will collect the corresponding HTTP traffic (in HAR files).
  - Selenium does not download HAR files by default, so you will use it with browsermobproxy to download HAR files.
  - Java **(**version 8 will work ie version = 1.8.0_472 some newer versions might not work**)** must be installed on your system to run browsermobproxy.
  - You will need to include a binary to use browsermobproxy. You can find the binary at this link.
  - Include the harfiles in a folder for your submission named top100_harfiles.
- You will conduct the following analysis using the HAR files:
  - Track the number of requests made to third-party domains when visiting each site in a python file. A third-party domain is a domain that does not have the same second-level domain (SLD) as the site you are visiting. For example, when you are visiting google.com, ads.google.com is not a third-party since it has the same second-level domain (google) as google.com. However, doubleclick.net is considered a third party to google.com. Identify the top 10 most commonly seen third party domains across all sites and put this in your report.
  - Identify the third-party cookies on each site. Third-party cookies are those cookies that were accessed (set or read) by third-party domains. Across the top-100 sites, identify the top 10 most common third-party cookies and describe their intended functionality by referencing Cookiepedia. Put this analysis in your report

Remember to include the names of all programs in your report.

Testing Environment:

All submissions will be tested on Python 3+.

Submission Details:

Submit a zipped file with all your python files, your single report to canvas. Only one person from the group submits. Make sure both students' names are in the report.

Late Submission Policy:

No late submissions are allowed. However, if you barely miss the deadline, you can get partial points up to 24 hours. The percentage of points you will lose is given by the equation below. This will give you partial points up to 24 hours after the due date and penalize you less if you narrowly miss the deadline.

Total marks = (Actual Marks you would get if NOT late) x [1 - hours late/24]

Late Submissions (later than 24 hours from the due date) will result in zero points, *unless you have our prior permission or documented accommodation*.

Best of luck

Include this signed page along with your submission

Submission Page

I certify that all submitted work is my own work. I have completed all of the assignments on my own without assistance from others except as indicated by appropriate citation. I have read and understand the university policy on plagiarism and academic dishonesty. I further understand that official sanctions will be imposed if there is any evidence of academic dishonesty in this work. I certify that the above statements are true.

Team Member 1:

Full Name (Printed) Signature Date

Team Member 2:

Full Name (Printed) Signature Date