EECS 325/425: Computer Networks
Project #5
Due: December 7, 11:59 PM[1]

The last project of the semester involves conducting measurements of some network facet / behavior and reporting the results. The goal of this project is to move beyond learning about networks from the text book and towards learning about networks via observations of how they work in practice. This is a student-directed project. The broad contours of the project are outlined below, but part of the project is arriving at the theme and the details yourself. While there are a few limitations—see below—you are free to measure whatever you find most interesting about the network. You may use common tools to help with your investigation (e.g., *ping*, *traceroute*).

Project #5 submissions will take the form of a tarball that includes a (*i*) project report in PDF (called "report.pdf"), (*ii*) any tools you wrote for the project (e.g., analysis tools) and (*iii*) a pointer to the data you collected and/or used in the project (in a text file called "data.txt").

Report

Your report will consist of three parts, as follows:

1. **Introduction / Motivation:** The first section of your report should sketch what you intend to measure in high-level terms. Your project must be centered around some theme. E.g., understanding web servers. You may ultimately take different sorts of measurements, but they should all be in service of your theme. In other words, there must be some connective tissue between your measurements. The discussion in this section must be in broad terms (e.g., "delay" not "*ping* measurements"). This section should argue and ultimately motivate why this is an important aspect of the network to understand.

2. **Procedure:** This section should discuss how you go about conducting your study. This will include two parts:

   (a) **Gathering Data:** This part of the report will discuss your data collection. In the case of a set of measurements you collect yourself, this will include a discussion of any off-the-shelf tools you leveraged (e.g., *ping*), as well as tools you created to run the measurement tools. Any tools you create must be included in the project submission.

   (b) **Analyzing Data:** This part of the report will discuss *how* you analyzed your data and what tools you created to do so. These tools must also be included in the project submission.

   Note, unlike previous projects, you may write your tools in any programming/scripting language.

3. **Results:** This section of your report will discuss the results of your measurements. You must include at least N distinctive results about your chosen theme.[2] Use tables and figures as appropriate—but these must be discussed and referenced in the text (e.g., to describe what the table/plot shows and highlight particularly important aspects). Your results should not be speculation, but instead be rooted in the data you have collected.[3]

Further report guidelines:

- Writing clarity is of the utmost importance. Grades will be reduced for unclear reports. Grades will be reduced if crucial aspects of the measurement or analysis are not included. *Leave plenty of time to write a quality report.*
- There is no page requirement or limit for the report. Use as much space as is appropriate.
- The only formatting requirement is that you use at least a 10pt font.

[1] The syllabus states all projects for the course must be submitted by December 7 at 11:59pm. This will be extended until Monday December 10 at 11:59pm.

[2] For undergraduates $N = 5$. For graduate students $N = 7$.

[3] It is fine to speculate about why something may have happened, or when discussing possible measurements that could shed more light on the situation. But, you must have at least N results that are based on observations of the data you have collected.

Additional Considerations

Measurement Tools

You may liberally use existing measurement tools. E.g., *ping*, *traceroute*, *wget*. The project web page will have a non-exhaustive list of tools you can use for this project.

Data Repositories

Instead of or in addition to collecting data using some measurement tool, you may leverage data previously collected and released by researchers to the public. There are several examples on the project web page, including packet trace collections and DNS transaction logs.

Measurement Vantage Point

You may take and analyze measurements on the class servers. However, you also may also use other machines you have access to, if that is more convenient or is helpful in the measurement process. The class servers have a set of basic measurement tools installed. You can install additional tools in your home directory. If you need a tool that would benefit from being properly installed on the machine, let me know.

Limitations

Network measurements can sometimes cross into a grey area where ethical analysis is necessary to determine the appropriateness of the measurement. Further, your goal should be to be polite to the network and the operators. Full fledged measurement studies are often conducted with the blessing of network operators to ensure the measurements do not unduly impact the network and the operators have a full understanding of what is going on. In lieu of cooperating with the Case network operators we will use these conservative guidelines:

- You may not capture traffic you did not create / trigger. Your project will not involve analyzing other people's network traffic unless you are using a public dataset that an outside researcher has released.
- We will follow the "one at a time" rule. That is, you may only run one measurement at a time. Once that measurement is finished, you may conduct another measurement. E.g., if you are using *dig* to conduct 100 DNS lookups, only one invocation of dig may be running at a given time. You may not run 100 lookups in parallel.
- If you are transmitting packets at some rate into the network—e.g., using *ping*—that rate can be no more than five packets per second. If your measurement requires a quicker "burst" or "flood" of traffic, you must follow that with a silent period. E.g., you may send 100 packets into the network within one second, but then must pause for 20 seconds before your next burst, as to ensure the average sending rate is at most five packets per second.

These limitations are conservative. Please be polite and follow the spirit of these guidelines. Do not try to find "loopholes". Measurements that end up being intrusive will result in a grade reduction.

Note: Even if you're using a non-Case networks for your measurements, please adhere to these limitations.

Amount of Data

Conclusions based on only a single or a few data points are inherently weak. Such anecdotes are sometimes interesting and can be illustrative and as such, you can include them in your report to provide color. However, your N observations should be based on a decent amount of data. As a touchtone, if you can analyze the data by hand, you likely have too little. The amount of data we are interested in seeing will require a tool to analyze.

Final Bits

1. Submission specifications:
   (a) All project files must be submitted to Canvas in a gzip-ed tar file called "[CaseID]-proj5.tar.gz".
   (b) Your submission must contain a report in PDF format called "report.pdf".
   (c) Your report must include a title, your name and Case network ID.
   (d) Do not include multiple versions of your report in your submission.
   (e) Your submission must contain all code you wrote for this project.
   (f) Do not include raw data in your submission. However, you must make your raw data available when you submit your project. Place your raw data somewhere (Google Drive, Dropbox, your web server, etc.) and put a link in a "data.txt" file you include in your tarball. If this is problematic, please contact me and we will figure out a place to put your data. If you are using data collected by someone else, note this in the "data.txt" file and include a URL for that source of the data (i.e., there is no need to make a copy of the data available).
   (g) Do not include directories in your tarball.
   (h) Do not use spaces in file names.
2. If you are in doubt about your project idea, please feel free to describe it in email or come chat about it during office hours.
3. Your submission may include a "notes.txt" file for any information you wish to convey during the grading process. We will review the contents of this file, but not of arbitrary files in your tarball (e.g., "readme.txt").
4. *WHEN IN DOUBT, ASK QUESTIONS!*