**ELE/COS 381 - Project Proposal**

**Princeon Network Usage**

***Angela Liu, Jean Juang, & Roberta Raileanu***

1. **Problem Statement:**

* How does network data usage at Princeton University vary with time and location?
  + Distribution of network usage in multiple areas on campus, at a given time
  + Evolution of network usage in time, for given locations
  + Find network usage patterns across locations (Frist Center, Residential College, McCosh, Engineering Building etc.) and time periods (day/night, weekday/weekend, beginning of the semester/midterm period/reading period/final period etc.)
  + Statistical analysis of network configuration with a 60 minute resolution
* Create a visualization of the campus network usage
  + Map of campus with colors or circle size to indicate amount of usage, with the ability to “play” and see it change with time
  + Graphs of usage at a given location over the year and over the day. We expect to see spikes around midterms week, for example.
* Analyze the data with respect to different ISP pricing models.
  + Which model would work the best for Princeton University as a whole, given the usage graphs and that users are willing to change? If a model was enforced upon the university, what do we predict that the traffic will end up looking like?
  + Some models to consider:
    - Time-dependent pricing
    - Usage-based pricing
  + It is difficult, given the scope, to determine whether “users” would be incentivized by the pricing structure, such as deferring their data to a later time. Maybe it’s more important to professors to download the latest paper now, regardless of price. We aim to see how well our data would fit with given models in the best case.

1. **Background and References**

Previously done work:

* Part of the data and the previous analysis done by OIT can be found at: <https://www.net.princeton.edu/statistics/>
* The OIT statistics measure traffic volume flowing through the network infrastructure. In some cases, they collect other relevant data; for example, CPU usage of key infrastructural components. They collect some statistics provided by a few hosts attached to the network.
* Campus Core and Border Switcher Traffic Statistics: comprise the core of the campus network
* Router Traffic Statistics: Traffic volume passing through OIT's core and Internet/Internet2 border routers. Also selected statistics for some departmental routers.
* Campus Aggregation Traffic volume passing through building aggregation switches and building point of entry switches.
* Remote Access Traffic Statistics: include selected remote access connections to the campus network, including OIT's VPN service and some DSL-attached sites.
* Wireless Statistics: Includes selected statistics for the campus network's wireless infrastructure.

Plan for collecting and using the available data:

* We will collect data from Statseeker, which is maintained by the OIT
* We have started to collaborate with Christopher Tengi from OIT, which will provide us the available data
* Our raw data will contain a table in a CSV file with columns corresponding to a certain time window of 60 minutes and rows corresponding to a certain network device
* We will store the received and transmitted number of bits per second for each device within each time slot
* We will then aggregate the rates for all the devices in a certain building or campus area to present the results in a more intuitive and relevant way

1. **Timeline**

* *November 24:* Submit project proposal
* *November 25 - November 30 :* Obtain the available network data from the OIT representatives and refine our goals for the project
* *December 1 - December 11:* Data mining and analysis; Raise new questions
* *December 12:* Submit Interim Report
* *December 13 - December 20:* Create Visualization Tool of the Data
* *December 21 - December 27:* Write the Report; Patterns, Conclusions, Suggestions for future work
* *December 28 - January 6:* Ask for feedback on our approach, results, report; Prepare the presentation
* *January 9:* Project Presentation
* *January 13:* Submit Final Report

1. **Division of Labor**

* Visualization Tool - Angela (and Jean)
* Data Analysis: dependence of network usage with time and location - Jean
* Compare different ISP pricing models to optimize campus network usage - Roberta

We anticipate that we will all work on the tasks collectively, so the assigned duties are flexible.