Experiment: Networking Data Analytics

Prepared by: Babak Alipour, Ph.D. student, CISE, UFL
For: CIS4930/6930 Mobile Networking (Special Topics)
Instructor: Ahmed Helmy
(Due Feb 19 beginning of class)

In this experiment, students get hands-on experience in dealing with network measurements wireless LAN, WLAN, and DHCP traces in this case. The traces can be used for analysis of network activity throughout the network (campus in this case), for activity and mobility modeling purposes, or to deploy mobile services to serve the student and faculty population.

Students are provided with data samples of WLAN DHCP logs, in addition to external data (prefixes of buildings around campus, categories of buildings, etc.).

The goal is to use simple data analytics tools and methods to read the data, analyze it, and produce representations of various statistics and distributions of activity and load over time, space, and building categories.

Following is a description of the data files (Input.a and Input.b) then four parts to the experiment to analyze the data:

- Input.a

  • **2 days** of derived-DHCP logs from campus WLAN in CSV format
  - ***outputwireless-logs-20120407.DHCP_ANON.csv***
  - ***outputwireless-logs-20120409.DHCP_ANON.csv***

  • 5 columns are:
  - userIP: User's assigned IP address
  - userMAC: MAC address of user device (anonymized to integers, each number uniquely identifies a user **in a single day**)
  - APNAME: Name of AP the session belongs to
  - APMAC: MAC address of AP
  - startTime: Beginning timestamp of session
  - endTime: Finish timestamp of session

- Input.b

  •Mapping from AP prefixes to location
  - **prefix_lat_lon_name_category.csv**

  •5 columns are:
  - Prefix: Prefix of APNAME (see input.a)
  - Latitude
  - Longitude
  - Building name
  - Building Category

- Part I.

  1. Produce the time series plot of events in "***outputwireless-logs-20120409.DHCP_ANON"***

•x-axis: Time (15-minute bins)
•y-axis: number of events in the bin

2. What time of the day is the most active (in terms of number of events)?
3. What time of the day is the least active (in terms of number of events)?
4. Considering the class periods (start/end times), please explain your observations in the time series around beginning or end of classes.
5. Try different bins (1m, 5m, 10m, 30m, 60m) and discuss how it affects your analyses.

- Part II.

1. Repeat the same tasks of Task 1, on "***outputwireless-logs-20120407.DHCP_ANON"***
2. In addition, do you notice any similarities or differences?
    - •Please explain how you observed them and why you think these similarities and differences exist.
    - •Include any supporting graphs and plots

- Part III.

• Using "***outputwireless-logs-20120409.DHCP_ANON"***
• Can you identify the user devices that have many sessions in early morning hours (the early-birds)? How about during lunch time (the munchers)? Or the evening (the stompers)?
• Use **Input.b** to map these devices (and corresponding sessions) to locations. Which buildings are the most popular among 'early-birds'? Or 'munchers? Or 'stompers'?
• Sample output of this task: 50% of munchers munch at the hub.
• Please include any necessary graphs or plots to support your output.
• Does the most popular building of different users change as the day progresses? What are the most popular ***building categories*** in the morning? How about in the evenings?

- Part IV. (bonus fieldwork)

• Can you find the mysterious 'b' APs?
• A few sample of their names:
• b420r100-win-lap1231-1, b411r115hall-win-lap1242-1, b430mail-win-lap1231-1
• These APs have names that look like: b(\d{3,4}).*
• You need to find their names and MAC addresses from **input.a**, then scan WiFi APs around campus to find a matching AP MAC address.
• Output of this part is a mapping from a 'b'-class AP to its location (latitude, longitude, building name) with high accuracy
• If you do find any, repeat parts 1-3 for those APs and describe the similarities and differences with your previous observations.

• The experiment relies on a few files available (zipped) here (UserMAC anonymized):
    • https://www.dropbox.com/s/tfdlrh4zj0cfrsh/Experiments4ForClass.zip?dl=0