Fall 2019 ELEN6774 Topics in Networking:

Internet Measurement

Basic Information
Time and place: Thursday 1:10pm to 3:40pm, 303 Hamilton Hall
Instructor: Ethan Katz-Bassett
Prerequisites: CSEE 4119 or equivalent, or permission of the instructor.
Appropriate for graduate students or advanced undergraduates with previous classwork in networking. Students from non-systems/networking areas are welcome.
Format: Reading, writing about, and discussing research papers. Short presentations. Research projects in small groups.
Course homepage: CourseWorks

For course page and draft syllabus, see:

Course Description
In this course, we will investigate important problems, techniques, results, and challenges that arise in measuring the Internet. We will explore both what measurements tell us about the Internet and how we can leverage what they tell us to improve the Internet and services that run on it. We will focus on why certain questions are hard to answer, how we might start to answer them, and why different measurements might reveal what seem to be conflicting answers. We will learn to measure various aspects of the Internet, including topology, routing and routing policies, performance, failures, traffic, and applications. Researchers often talk about Internet measurement as being analogous to astronomy, in that we take observations from afar in order to understand how a system works. We will learn to leverage and integrate the various sources of information that leak out from networks and services about their internal operations.

Course Motivation
The Internet now plays a central role in many aspects of our lives. Internet measurement as a field seeks to understand the Internet by assessing its operation. The field is interesting for a range of reasons:

- We depend on the Internet, so we need to understand its operation, and we need to discern its problems in order to improve it.
- Despite the myriad ways we have come to depend on it, many aspects of it can be
opaque even to network operators. Engineers at Google and Facebook cite the difficulty of Internet measurement as one of the impediments they face.

- The Internet is a loose federation of networks that must cooperate to provide global connectivity, even as they compete for business. Some of the opacity arises from this tension, as networks may lack incentive to expose their inner workings and may lack the visibility necessary to optimize their performance.
- The Internet protocols were designed in an era when networks looked very different than they do today. The protocols do not naturally expose some of the information that we want to learn. The standards defining them can be underspecified and evolve over time, resulting in differences in implementations. Some of the fun of Internet measurement comes from finding novel ways to manipulate the protocols to reveal information.
- The Internet is one of the largest systems humankind has ever engineered, and it has emergent properties. Measurement provides a basis for determining these properties.

**Textbooks and Course Materials**

We will read 2-4 research papers a week. Recommended supplemental textbook: Crovella and Krishnamurthy’s *Internet Measurement: Infrastructure, Traffic & Applications*. Recommended background textbooks: Kurose and Ross’s *Computer Networking : A Top-Down Approach*. Tanenbaum and Wetherall’s *Computer Networks*.

**Primary Course Work**

The work for the course will primarily consist of:

- Paper reading and written paper responses: Students are expected to read and write responses to 2-4 papers a week, giving short answers to questions and writing a novel response of a few paragraphs.
- Class discussions: Students are expected to participate in in-class discussions weekly, including on topics from their written responses.
- A research project, including ~6 pg writeup and ~20 minute presentation: The semester-long project is an open-ended Internet measurement project. The instructor will provide some possible project topics, or (with instructor approval) you can work on a project of your own devising. Most projects will likely involve either making new measurements or performing new analysis of existing data. Projects can be done in groups of two.
Grading

There are no exams in this class. However, the final exam period may be used for final project presentations.

The course grade will be determined based on:

- Class participation (50%)
  - Written paper responses
  - Discussions online and in class
- Research project (50%)
  - Research novelty, quality, writing/presentation
  - Few page proposal and midterm report
  - ~6 page final report
  - ~20 minute presentation
  - Written review of one other group’s project/report
  - When done in a group, all students in the group will receive the same grade.

Schedule by Week (subject to change)

- Week 1: Course overview.
  - How to read papers:
    - These two 2-page papers describe similar (but not identical) 3 pass approaches to reading research papers. Please read at least one of them.
      - Efficient Reading of Papers in Science and Technology. Michael J. Hanson.
      - How to Read a Paper. Michael J. Hanson.
    - Why measure the Internet? How to conduct measurements?
      - Optional:
- Weeks 2-4: Internet topology
  - What is the Internet’s topology?
    - On power-law relationships of the Internet topology. Faloutsos, Faloutsos,
and Faltoutsos. SIGCOMM 1999.


■ **What are our standards for validation of measurement-based networking research?** Balachander Krishnamurthy and Walter Willinger. HotMetrics 2008.

■ Optional
  


  ○ AS Relationships and Missing links


    ■ **Bigfoot, Sasquatch, the Yeti and Other Missing Links: What We Don’t Know About the AS Graph.** Matthew Roughan, Jonathan Tuke, and Olaf Maenenn. IMC 2008.

  ■ Optional


■ **Week 5-9: Internet routing and performance**

  ○ Traceroute and Its Limitations


- Optional

- Understanding Routes

- Path changes

- Outages and anomalies

- Tomography

**Week 10-12: Internet traffic**
- **Backbone traffic**
  - Internet inter-domain traffic. Craig Labovitz (Arbor Networks), Scott Ilekel-Johnson (Arbor Networks), Danny McPherson (Arbor Networks), Jon Oberheide (University of Michigan), Farnam Jahanian (University of Michigan). SIGCOMM 2010.
- **Edge traffic, censorship, and traffic differentiation**
  - **A Large-Scale Analysis of Deployed Traffic Differentiation Practices.** Fangfan Li, Arian Niaki, David Choffnes, Phillipa Gill, Alan Mislove. SIGCOMM 2019.
- **Anomaly detection**

**Week 13-14: Internet applications**
○ P2P file sharing

○ Spam, botnets, and the underground economy

● Finals period: Project presentations