CS8803-O01 Modern Internet Research Methods Course Syllabus for Fall 2024

Instructor information
Course instructor: Dr. Maria Konte, mkonte@gatech.edu
Head TA: Erick Herring, erickherring@gatech.edu

About this course
Welcome! This is a research oriented course that covers new developments in Internet measurement techniques, with an emphasis on topics related to reliability, freedom and security of modern Internet platforms.

The goals of this course are to:
  a. Explore new research topics in the modern Internet interdisciplinary research areas.
  b. Familiarize and experiment with techniques, tools, platforms and datasets.
  c. Develop new research ideas and deliver an academic research paper. Use the course material as a starting point to brainstorm on new research ideas, and select a topic of interest. Perform the entire cycle from selecting a research topic, focusing to a specific research question, following through (e.g. data collection and analysis, system design and evaluation, etc.) and finally delivering the results through an academic paper.

The topics the course discusses span across three areas:
  1. Core Internet measurements and security: techniques to study and map the Internet host population along with the services it offers.
  2. Freedom: techniques currently used to block access to specific content enforcing censorship, censorship observatories, techniques that are used to overtake the control of the Internet infrastructure, security of voting systems.
  3. Social platforms: identify abusive entities on social platforms and map out the current landscape of abuse (misinformation, tracking narratives, etc.). Techniques that leverage social platforms for early warning such as identifying emerging threats as soon as possible.
Finally, the course covers topics related to ethics guidelines when performing large scale Internet measurements. Last, the course discusses elements of sustainable research such as transparency and reproducibility.

Lectures format.
The instructor presents weekly lectures, each lecture covering a topic along with associated techniques, platforms or datasets (see course calendar below). The lectures are delivered across 16 modules. Through each module the instructor presents 2-4 research papers delivered over 2-4 mini videos. In parallel, the students are assigned milestones that guide them towards the completion of their research project and paper (see assignments section below).

Course Learning Objectives

The learning objectives of this course are to:

1. Describe the current state of research in the intersection of Internet measurements and cybersecurity. Specifically, you will learn about modern topics of broad and current interest related to the risks that the Internet infrastructure faces (e.g. Internet infrastructure hijacking), Internet censorship, abuse and entities on social platforms, web trust management, the ecosystem of false information on the web. Finally, you will learn about how to perform Internet measurements using ethical guidelines, and principles of sustainable research (e.g. replicability).
2. Describe a plethora of passive and active measurement techniques, data collection and analysis approaches for each of the above topics.
3. Demonstrate the ability to apply the learned techniques to different or new research questions.
4. Demonstrate the ability to put together a research project; from identifying a broad idea, to specifying a well-defined research question, outlining and executing a research approach to address it.
5. Demonstrate the ability to transfer a research project into an academic paper, and deliver a presentation of the paper.

Course Assignments and Grading

As a research oriented course, the main component and focus of the course is a semester long research project. Each student, either by themselves or in a group, will work on a research project that will run through the entire semester. More specifically the assignments, along with their weight on the final grade, are the following:

- o **Project idea brainstorming – 15%.** Each group starts by working towards a project proposal, that eventually becomes the final project. Towards this goal and as intermediate steps, each group will submit three mini write-ups that summarize the groups progress towards the proposal. The mini write-ups (5% each) that reflect the progress as the group refines the project idea and getting ready for the proposal.
- o **Project proposal – 10%.** Each group will put together a 1-2 page project proposal describing a clear research problem they will address, including the related work section, and an outline of their suggested approach.
- o **Project weekly check-ins – 5%.** After the proposal submission, each group will be asked to keep in touch and check-in with the TA team, on a weekly basis through the end of the semester. The goal of these check-ins are to discuss problem solves any challenges the group might be facing through any stage of their project.
- o **Project initial and interim submissions – 20%.** Towards the final submission of the deliverables and as intermediate steps, each group will submit two updates: the initial (10%) and the interim (10%). These are essentially draft versions of the final submissions for the three deliverables (code, paper, presentation).

- **Final Project code – 20%**. Each group submits the code of the project. Each group is required to be using Georgia Tech's GitHub repository to host the project code throughout the course.
- **Final Paper presentation – 5%.** Each group will prepare and record a 15-20 minute presentation on their project.
- **Final Paper – 25%.** Each group will write a 10-12 page final report written in academic style-format (*Latex template*). Each group is required to use Georgia Tech's Overleaf to host the paper throughout the course.

Forming a team, team size and proposed work

Team size. You can work on a project individually or in a group. You are highly encouraged to form a team of 3 students to 5 students (max) to work on a project of your choice. Also, you are highly encouraged to form teams that are consisted of students with <u>diverse backgrounds, expertise and skills set</u>. Therefore, the overall work of each group will benefit and each student will be provided the opportunity to bring in their unique skills and perspective to their team.

Proposed work and team size. The scope and the complexity of the project is expected to be consistent with the size of the team. A single student planning to work on a project individually is expected to propose a 100-hour project. Whereas a team consisted of five students is expected to propose a 500-hour project.

Grading Scale. The final grade will be assigned as a letter grade according to the following scale:

- A 90-100%
- B 80-89%
- C 70-79%
- D 60-69%
- F 0-59%

Course Prerequisites

The course is geared towards students who have completed the vast majority of the courses in the program, and they are interested in leveraging their knowledge to pursue a research project and write an academic paper.

This course requires a strong background in core topics in computer science such as data structures, algorithms, computer architecture. This is an advanced research oriented course that intersects topics in Internet protocols, computer networks, cybersecurity and data analysis. Having taken courses in topics related to systems, ML/AI, data visualizations is a plus, since the student will be able to leverage their background in these areas to pursue a research project and write an academic paper. Therefore, the course will not go over undergraduate material typically covered in undergraduate networking, cybersecurity or data analysis courses. The students are expected to code in Python (or a language of their choice) at an intermediate level (e.g. comfortably using object-oriented programming, data structures, control structures, etc. as well as testing and debugging tools/strategies).

In lieu of a readiness questionnaire, prospective students are expected to be comfortable with and/or passionate about:
- Reading and understanding the paper "*An Open Platform to Teach How the Internet Practically Works*".
- Defining their own research questions/ideas, and therefore working with open ended projects rather than predefined assignments.
- Student-led projects that require more autonomy and taking ownership of the work and the progress/pace.
- Working with projects that require coding skills, as well as technical writing and presentation skills.
- Receiving peer-to-peer feedback.
- Working in a group of students with multidisciplinary backgrounds.

Course Calendar

Week	Lecture Topics & Tasks	Deliverables & Due Date
1	Lecture: Course Intro & Crash Review "How the Internet Works" Lecture quiz (ungraded)	[Aug 19, First day of classes] Course survey (i)
2	Lecture: Surveying the Internet Address Space (I) Lecture quiz (ungraded)	Project idea brainstorming (i) [Due Aug 25]
3	Lecture: Surveying the Internet Address Space (II) Lecture quiz (ungraded)	Project idea brainstorming (ii) [Due Sept 1]
4	Lecture: Overtaking the Internet Infrastructure Control Lecture quiz (ungraded)	Project idea brainstorming (iii) [Due Sept 8]
5	Lecture: Internet Censorship (I) Lecture quiz (ungraded)	
6	Lecture: Internet Censorship (II) Lecture quiz (ungraded)	Project proposal [Due Sept 22]
7	Lecture: Web and Trust Management	Check-in with the TA team

	Lecture quiz (ungraded)	[Due Sept 29]
8	Lecture: Measuring the Security of Voting Systems Lecture quiz (ungraded)	Check-in with the TA team [Due Oct 6]
9	Lecture: The Landscape of Abuse and Threats on Social Platforms Lecture quiz (ungraded)	Check-in with the TA team [Due Oct 13]
10	Lecture: Entities on Social Platforms Lecture quiz (ungraded)	Project initial submission Check-in with the TA team Course survey (ii) [Due Oct 20]
11	Lecture: False Information on the Web and Social Platforms Lecture quiz (ungraded)	Check-in with the TA team [Due Oct 27]
12	Lecture: The False Information Ecosystem Lecture quiz (ungraded)	Check-in with the TA team [Due Nov 3]
13	Lecture: Online Social Platforms as a Vantage Point to Identify Emerging Threats Lecture quiz (ungraded)	Project interim submission Check-in with the TA team [Due Nov 10]
14	Lecture: Ethics in Internet measurements Lecture quiz (ungraded)	Check-in with the TA team [Due Nov 17]
15	Lecture: Sustainable Research: The Importance of Transparency, Reproducibility & Replicability Lecture quiz (ungraded)	Check-in with the TA team [Due Nov 24]

16 (Final Instructional Days, Dec 1-3)	Lecture: Reading, Writing and Presenting Papers	Final submission: Code Presentation Paper CIOS survey Course survey (iii) [Due Dec 3]

Course Policies

Late submissions & extensions. The students are expected to complete the work on time by the due dates. In case of an emergency, please reach out to TA team through a private Ed Stem post, so we can come up with a plan to make up for the work or alternative solutions, depending on the type of the emergency and the impact it has.

Plagiarism & academic integrity. Students are expected to follow the Georgia Tech Honor Code (https://policylibrary.gatech.edu/student-life/academic-honor-code), including the Graduate Addendum. All incidents of suspected dishonesty will be reported to and handled by the Office of Student Integrity. In addition, the following specific policies apply to this course. If in doubt as to whether an action is allowed in this course, please ask the Instructor/TAs.

EdStem code of conduct. The students are expected to be respectful with others when interacting on Ed Stem. Please review the students' code of conduct. https://policylibrary.gatech.edu/student-life/student-code-conduct

Georgia Tech student resources & student accommodation. The Disability Services team collaborates with the students to find creative solutions and reasonable accommodation. Please contact the Office of Disability Services at (404)894-2563 or http://disabilityservices.gatech.edu/, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodation letter. Please also send a private message to "Instructors" on Ed Stem as soon as possible. Please note that the TA team is not able to provide any accommodation or extensions without an accommodation letter, nor the accommodations can be provided retroactively.

Final note. Please note that the current syllabus is subject to change at any time.