Syllabus for CDS DS 453 / 653: Crypto for Data Science in Spring 2024

Instructor: Mayank Varia (varia@bu.edu)
TA: Shreyas Sudarsan (shreyas9@bu.edu)
Quiz dates: February 5, February 21, March 25, April 16

Lecture: Tues Thurs 9:30-10:45am in EPC 204
Discussion: Mon 9:05-9:55am / 11:15am-12:05pm in CCDS 164
Final exam: Yes, at a time chosen by BU (tentatively, 5/6 at 9-11am)

This syllabus contains information you need to know about DS 453 / 653. Please ***read this document carefully*** in the first week of class, familiarize yourself with how the course works, and maintain that familiarity throughout the semester. ***You are responsible for adhering to course policies at all times***, especially the academic code of conduct, plagiarism, AI, and collaboration policies.

Course Information

Course Description: CDS DS 453 / 653 investigates techniques for performing trustworthy data analyses without a trusted party, and for conducting data science without data sharing.

The first half of the course investigates cryptocurrencies, the blockchain technology underpinning them, and the incentives for each participant. Students will learn how to create transactions, develop smart contracts, and participate in decentralized exchanges. Then, we take a deeper dive into consensus mechanisms, historical and modern, that maintain stability if a certain fraction of the participants or computing power behaves honestly.

The second half of the course focuses on privacy and anonymity using advanced tools from cryptography. We study zero knowledge proofs and their role in preventing re-identification attacks and increasing scalability of blockchains. We also study secure multiparty computation and its role in designing private contracts and atomic swaps. The course concludes with a broader exploration into the power of conducting data science without being able to see the underlying data.

Within the undergraduate Data Sciences major, this course satisfies the DS methodology elective in the "scalable & trustworthy DS" category. *In spring 2024, this course does not satisfy any Hub units.*

Prerequisites: This course requires mathematical maturity with linear algebra and probability concepts at the level of DS 121 and 122, or equivalent, and familiarity with algorithms at the level of DS 320, CS 330, or equivalent.

Course websites: We will use two websites in this course, Piazza and Gradescope. *Please sign up for both immediately!*

- *Piazza* (https://piazza.com/bu/spring2024/ds453653/) contains the weekly course schedule, links to all textbooks, and reading and programming assignments. You should review the course schedule at least once per week. Also, all course announcements will be made using Piazza; you are responsible for knowing these announcements, so please register for Piazza using an email address that you check regularly. We also encourage you to use Piazza as a resource to ask scientific questions to the instructors and your peers, and recommend that you mark posts as public unless they contain a personal question or reveal a partial solution to an assignment. Please only email the instructor for administrative matters.
- *Gradescope* (https://www.gradescope.com/courses/710247, entry code B2Y7G5) is the website that you must use to submit completed homework assignments, view your quiz scores, and submit regrade requests for quizzes.

Course textbooks: We will use (portions of) at least three textbooks during this class, all of which are available to download for free using the links in the Piazza resources tab: https://piazza.com/bu/spring2024/ds453653/resources

Course meetings: The course instructor is Prof. Mayank Varia. I will lead lectures on Tuesdays and Thursdays at 9:30-10:45am in EPC room 204. All lectures will be livestreamed over Zoom; if you cannot come to class in person (e.g., due to illness), you can join the Zoom livestream or review the lecture notes and video on Piazza after class. Discussion labs are held Mondays at 9:05-9:55am or 11:15am-12:05pm in CDS 164, led by TA Shreyas Sudarsan. Some of the discussion sections are used for quizzes or make-up quizzes, as discussed below. We actively encourage questions and interaction during all lectures and discussion labs.

Office hours: Students are welcome and encouraged to visit office hours. We will post the date, time, and location of all office hours on Piazza at https://piazza.com/bu/spring2024/ds453653/staff, and we will also announce in class and on Piazza if the office hour times ever change. If you want to meet with me (the professor) but cannot make the scheduled times, then send me a private Piazza note with at least 3 suggestions for times that you are available to meet and we will find an alternate time.

Learning Objectives

This course has 5 parts. I list each part below, together with the objectives that students are expected to learn in each part.

1. Authenticity without interaction (weeks 1-3). By the end of this part of the course, students will be able to:

- Describe the confidentiality, integrity, and availability goals of cryptography and cryptocurrencies.
- Define the security guarantees of digital signatures and hash functions. Explain their role in authenticating people and data.

2. Currencies without centralization (weeks 3-5). By the end of this part of the course, students will be able to:

- Describe the goal of decentralization and distributed consensus, in the context of financial payments via cryptocurrencies.
- Create transactions on Bitcoin and Ethereum, determine whether a potential transaction is valid, and explain the mining algorithm and its role in incentivizing honest participation in the network.

- Analyze public blockchains using computational and data science tools, and develop new applications on a blockchain.

3. Consensus without trust (weeks 6-8). By the end of this part of the course, students will be able to:
- Describe the goal of decentralization and distributed consensus abstractly, in any scenario needing agreement and liveness.
- Explain how to achieve broadcast and agreement, with and without network synchrony or a public key infrastructure.
- Construct a blockchain as a repeated application of Byzantine broadcast.

4. Data analysis without data sharing (weeks 9-12). By the end of this part of the course, students will be able to:
- Explain how encryption protects data on the Internet, and the role that (pseudo)randomness plays within encryption.
- Describe the design of protocols that provide cryptographically secure multi-party computation. Explain how it enables a collection of data-holders to perform data science together, without sharing data with each other or anyone else.
- Define the security guarantees of zero knowledge proofs. Explain how they allow people to participate in the digital economy while preserving their privacy, and hold powerful entities to account without the need for full transparency.

5. Crypto and society (weeks 13-15). By the end of this part of the course, students will be able to:
- Engage with ethical, legal, regulatory, and policy questions about the role of crypto toward addressing social issues.
- Assess the environmental and ethical risks that blockchains introduce, and current efforts to improve their sustainability.
- Discuss the social and economic role that blockchains and secure multi-party computation can play toward democratizing access to data (especially from underserved populations) and building more transparent governance structures.

Course Assessment

At a high level, a typical week of this class contains a discussion section or quiz on a Monday, lectures on Tuesday and Thursday, and coursework due on Thursday evening at 8pm. Every assignment is graded Pass/Not passed, and you pass any assignment in which you receive at least 4 out of an available 5 points and adhere to the academic code of conduct, plagiarism, AI, and collaboration policies. The precise schedule is posted on Piazza at https://piazza.com/bu/spring2024/ds453653/info.

In more detail, the course grade in this class is based on concrete evidence of student success in four different types of tasks:
1. Engagement with the material covered in the textbooks and lectures, as demonstrated through *weekly reading assessments*.
2. Ability to apply the technical skills learned in each part of the course, as demonstrated through *programming assignments*.
3. Mastery of the fundamental technical skills within the learning objectives, as demonstrated through *quizzes*.
4. Creativity to extend the technical and social concepts in this course to new problems, as demonstrated through *projects*.

All work in this course is graded using a style known as *specifications grading*, which is a learning-focused approach that prioritizes student learning over scores or grades. The key ideas of specifications grading are for your course grade to reflect how well you have understood the learning objectives of the course in absolute terms (rather than relative to the rest of the class or the subjectivity of partial credit), and to allow you to learn from mistakes and demonstrate growth and improvement throughout the course. The upshot is: each question either receives full credit or no credit, passing a task requires you to receive at least 4 out of 5 available points, and you have multiple opportunities to complete each task.

Reading assessments: You are required to read a portion of a textbook every week. The required reading is listed in Piazza on the "Course schedule" page (post #7). To assess your understanding of the material, we will post 5 multiple-choice or short-answer questions on Gradescope each week. You must *correctly answer 4 of the 5 questions* to receive credit for the weekly reading assessment. Each assessment is posted on a Thursday, due the following Thursday at 8pm, and you can submit as many times as you wish before the due date. There are 13 reading assessments in this course – one for each week except the first and last weeks.

Programming assignments: There are a total of 9 programming assignments in this course, which require you to engage with the cryptographic concepts discussed in the course lectures, discussion sections, and textbook reading by writing code in Python. Each assignment contains questions collectively worth 5 points, and you *must receive 4 points to pass the assignment*. This is a hard cutoff; note that there is no partial credit awarded for receiving fewer than 4 points. We recommend that you work on the programming assignments using Google Colab (https://colab.research.google.com), and then you must submit on Gradescope. Just like with reading assessments: each programming assessment is also posted on a Thursday and due the following Thursday at 8pm, has auto-graded responses, and allows you to submit the assignment as many times as you wish before the due date.

Quizzes: There are 4 quizzes in this course, which test your mastery of the material in Parts 1-4 of the course. (There is no quiz for Part 5; it is evaluated instead in the assessments.) Quizzes occur either during the Monday discussion sections or Tuesday lecture periods; the exact quiz dates are stated at the top of this syllabus and posted on the Piazza course schedule.

Each quiz contains 5 questions. Your response to each question is graded on its *correctness* and *clarity*. You will receive credit for a response that meets the stated expectations of the question, sufficiently explains and demonstrates understanding of the concepts, and does not contain any significant gaps, errors, or omissions. There are four possible grades for a quiz:

- High pass (H): Receive credit for all 5 questions.
- Pass (P): Receive credit for 4 of the 5 questions.
- Conditional pass (C): Receive credit for 3 of the 5 questions.
- Not passed (N): Receive credit for 3 or fewer questions. You must redo the quiz at a later date.

You can attempt each quiz up to 3 times: on the quiz date itself, on a make-up quiz date in the following week, and during the final exam period. For each quiz, we will take the best of the 3 attempts as your grade for that quiz, except that academic misconduct on any attempt of a quiz will result in a grade of N for that quiz (as described in more detail below). Note that there is no separate final exam; instead, that time is your last opportunity to pass any remaining quizzes. Attendance at the final exam is optional; you do not need to attend if you are already satisfied with your course grade (as calculated below).

For one quiz of your choice: in lieu of taking a make-up quiz, you can "upgrade" a C to P by submitting by the make-up quiz date a written report that: provides the correct answers for the two questions that you missed, and explains why your initial quiz responses were incorrect. Otherwise, the C grade becomes an N. (Quizzes taken on the final exam day cannot be upgraded.)

Projects: There are two open-ended projects in this course, one with submission dates on March 7 and 21 on a data science investigation of a public blockchain, and one with submission dates on April 25 and May 1 on using secure computation and zero knowledge. At the start of each project, we will provide a written project description and grading rubric. As with everything else in this course, your final grade for the project is either Pass (P) or Not passed (N). You will have 3-4 weeks to work on each project, and you will have two opportunities to submit your project work. In the first submission, I will either tell you that you have passed the project, or I will provide specific revision requirements that you must complete in the second submission to pass the project.

Course Grading

Your grade in this course is based on passing a sufficient number of reading assessments, programming assignments, quizzes, and projects. For DS 453 students: to receive a course grade of A, B, C, or D, you must complete ***all*** stated requirements below.

- **A**: Pass 13 reading assessments, 9 programming assignments, 4 quizzes (with at least 2 High Pass scores), and 1 project.
- **B**: Pass 11 reading assessments, 7 programming assignments, 4 quizzes, and 1 project.
- **C**: Pass 9 reading assessments, 6 programming assignments, 3 quizzes, and 1 project.
- **D**: Pass 7 reading assessments, 4 programming assignments, and 2 quizzes.

For DS 653 students: to obtain a grade of B or higher, you must pass 2 projects rather than 1. For all students: failing to meet all of the stated requirements for a D will result in a grade of F.

A student who meets all requirements of a particular grade level and *at least one* requirement of a higher grade level will receive a + modification to their grade; for example, performing the B requirements and passing all 13 reading assessments gives a grade of B+. A student who meets all-but-one requirement of a particular grade level and one requirement at the *next lower grade level* will receive a – modification to their grade; for example passing 6 programming assignments and the other requirements of a B grade would yield a course grade of B-, but passing only 5 programming assignments would yield a grade of C- instead.

You can use the following sheet to track your progress in the course. For every passing grade that you receive, check off one box on the corresponding row, *from left to right* on each row. You must check ***all*** boxes in a specific column to receive that grade.

	D grade	**C grade**	**B grade**	**A grade**
Reading assessments	☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐	☐ ☐	☐ ☐
Programming assignments	☐ ☐ ☐ ☐	☐ ☐	☐	☐ ☐
Quizzes	☐ ☐	☐	☐	(≥ 2 High Pass scores)
Projects		☐	☐ (if in DS 653, else ignore this box)	

Course Policies

Academic code of conduct: You must read and adhere to BU's Academic Code of Conduct, which is available here: https://www.bu.edu/academics/policies/academic-conduct-code/. All work in this course is subject to BU's Academic Code, so please familiarize yourself with this code, its definitions of misconduct and plagiarism, and its sanctions.

Plagiarism: All written work in this course must be original to you. If you consult outside texts, or other forms of assistance, cite these sources in the proper format—at a minimum, include the author, title, and website link for all external sources (books, journals, lectures, web sites, AI). We are required to report all suspected cases of plagiarism to the Academic Dean for review.

Academic integrity in computing coursework has some special aspects. Please review the examples of plagiarism as provided by the BU Computer Science department: https://www.bu.edu/cs/undergraduate/undergraduate-life/academic-integrity/.

Generative AI policy: All submitted work in this course must conform to the CDS Generative AI Assistance Policy, which you can read at https://www.bu.edu/cds-faculty/culture-community/gaia-policy/. Also, keep in mind that AI tools are often wrong!

Collaboration policy: The goal of homework assignments and projects is to learn. Hence, I encourage you to use any and all resources that can help you to learn the material: computers/calculators, Piazza, lecture notes, textbooks, other websites, and your

fellow classmates. That said, you *cannot* copy solutions from anyone else, or give your solutions to a classmate to copy; this is plagiarism. You also *cannot* actively search for the solutions to the homework questions on the Internet or in any other source; this is misconduct. Finally, your submission must list all people and resources that you used, as discussed in the next section.

By contrast, the goal of the quizzes is for you to show me what you have learned. So, any form of collaboration is strictly prohibited. Also, written notes and electronic aids (e.g., computers or calculators) are all forbidden from use during quizzes. (That said, I encourage you to collaborate with classmates when studying lecture materials and preparing for the quizzes.)

Violations of the conduct code and collaboration policy will result in *receiving a score of N (not passed) on the assignment or quiz without the possibility to improve the grade*, and may be grounds for referral to BU's Academic Conduct Committee.

Documenting collaborators, sources, and AI tools: If you have any questions or concerns about the conduct code and collaboration policy, then I recommend that you ask me (in person or via private Piazza note) *before* taking an action that might be a violation. Additionally, at the bottom of each assignment and project, you *must* list (a) names of all classmates you worked with, (b) all websites you used besides the ones listed in the lecture notes or textbooks, and (c) all code that you used from other sources, including the exact prompts and responses from any Generative AI tool. This is your opportunity to document and explain any collaborators or sources used and why you believe it adheres to the code of conduct and collaboration policy.

If I discover a violation of the policy that you *have documented*, then I will not refer the case to BU's Academic Conduct Committee; the worst possible recourse is that I will ask you to redo the assignment (or create a new, similar assignment for you to do in its place). If I discover an *undocumented* violation after the fact, then this will be considered academic misconduct.

Accommodations: BU strives to be accessible, inclusive, and diverse in our facilities, programming, and academic offerings. Your experience in this course is important to the teaching staff. If you have a disability or believe that you require a reasonable accommodation, please meet with BU Disability and Access Services as soon as possible at the beginning of the semester. Their office is at 25 Buick Street, Suite 300, and they can be contacted at 617-353-3658. Requests for accommodations are sent by that office to the Academic Dean who approves and returns them. Disability and Access Services then forwards them to the instructor. If you have an accommodation letter, please send it (and only it) to me via email early in the semester, before tasks become due.

Learning environment: This course seeks to be inclusive of people of all genders, races, cultures, abilities, and sexual orientations. Please respect your fellow classmates and contribute toward a positive learning environment for everyone. While I actively encourage discussion and debate on ideas, I won't tolerate criticism of other people. Also, while you can use a computer for note-taking, you may not use a laptop or cellular phone in class for web surfing, sending messages, or anything else that can cause a distraction to your fellow classmates. Anyone who causes a disruption to the learning environment will be asked to leave.

Absence policy: This course follows BU's policy on absences for religious observance. Otherwise, students should attend the lectures and discussion labs, either in person or virtually via Zoom. Due to ongoing public health concerns: if you feel sick, please err on the side of caution and attend via Zoom or review the lecture notes and video afterward on Piazza.

Late work policy: You are responsible for submitting reading assessments, programming assignments, and projects electronically on Gradescope by the stated due date and time (typically, Thursday at 8pm). Up to *three times in the semester*, you can submit a reading assessment, programming assignment, or project up to 3 days late (e.g., by Sunday at 8pm). Any subsequent late assignments are not accepted (and any such auto-graded scores will be manually discarded after the fact), except in cases of long-term emergencies (e.g., family, medical); if this applies to you, inform me as soon as you are able.

Quiz absence policy: If you have a valid conflict with a quiz, you must send me a written note *as soon as you are aware*, and with a *minimum of 2 weeks notice* (barring extenuating circumstances, e.g., illness). If you have an excused absence from a quiz, then you retain the right to have 3 opportunities to pass that quiz; the first opportunity is during the scheduled make-up quiz date, and we can schedule another make-up opportunity afterward. If you have an unexcused absence from a quiz or a make-up quiz, then you forfeit that quiz-taking opportunity but can still attend the other opportunities to take that quiz; in other words, you would have two chances to pass that quiz rather than three.

The final exam period (i.e., the time of the final chance to take each quiz) can only be rescheduled in accordance with the university policy: https://www.bu.edu/reg/calendars/final-exams/policy/.

Regrade policy: You have the right to request a regrade of any project or quiz question. All regrade requests must be submitted via Gradescope, and must describe a factual error in our assessment. If you request a regrade for one question on a quiz or one part of a project, then we have the right to review the entire quiz or project. Beware that this may potentially result in a lower grade.

The reading assessments and programming assignments are auto-graded, so in general we do not anticipate regrades. That said, if you discover an error in our auto-grading system, please send us a Piazza note explaining the issue and we will look into it.