CSCI E-11 (15525):
Introduction to the Challenges and Opportunities of Big Data, the Internet of Things, and Cybersecurity

Prof. Brian Subirana ©

Draft Syllabus as of July 25, 2018.

COURSE OVERVIEW

In this course, we review use cases and challenges of three interrelated areas in computer science: big data, Internet of Things (IoT), and cybersecurity. It includes an overview of the possibilities and challenges of building complex information systems that take advantage of recent advances in these fields. The course is divided into three units, each presented by leading MIT experts in their respective fields.

Unit 1 surveys state-of-the-art topics in **big data**: data collection (smartphones, sensors, the Web), data storage and processing (scalable relational databases, Hadoop, Spark), extracting structured data from unstructured data, systems issues (exploiting multicore processors, security), analytics (machine learning, data compression, efficient algorithms), visualization, and a range of applications. In this first part, you will

- learn to distinguish big data (volume, velocity, variety), learn where it comes from, and the key challenges in gathering and using it
- determine how and where big data challenges arise in a number of domains, including social media, transportation, finance, and medicine
- investigate multicore challenges and how to engineer around them
- explore the relational model, SQL, and capabilities of new relational systems in terms of scalability and performance
- understand the capabilities of NoSQL systems, their capabilities and pitfalls, and how the NewSQL movement addresses these issues, and

- maximize the MapReduce programming model: its benefits, how it compares to relational systems, and new developments that improve its performance and robustness.

Unit 2 looks at the **Internet of Things (IoT)**. While the promise of the Internet of things brings many new business prospects, it also presents significant challenges ranging from technology architectural choices to security concerns. This part of the course offers important insights on how to overcome these challenges and thrive in this exciting space. The concept of IoT has begun to make an impact in industries ranging from industrial systems to home automation to healthcare. Researchers continue to conduct ground-breaking research on topics that are presented ranging from RFID to cloud technologies, from sensors to the authentication technologies.

Unit 3 covers **cybersecurity** issues related to hardware, software, cryptography, and policy in order to make better and safer decisions. Topics include:

- systems (secure architectures, network security, secure programming languages, system verification)
- algorithmic solutions (public key cryptography, multi-party computation, secret sharing, distributing trust, computing on encrypted data)
- public policy issues in cybersecurity, and
- case studies (BitLocker, web security, mobile phone security).

About this Course Offering: This course is a collaboration between MITProfessionalX and Harvard Extension School. The video lectures are of MIT Professors recorded for three MITProfessionalX courses. The student work and implementation in Canvas was co-developed with Harvard Extension School faculty and staff. This document provides a summary of the course. Students will be expected to review the Canvas site for updated and complementary instructions.

About the Instructor: Dr. Brian Subirana is the Director of the MIT Auto-ID Laboratory.

Prerequisites: An introductory computer science course (e.g., CSCI E-3, E-7 or E-10a) plus familiarity with precalculus mathematics (Math E-10).

Who is this Course For? This course was designed to expose participants to the core concepts of and current trends in Big Data, Internet of Things, and Cybersecurity, and to give suggestions for software, relevant articles, and research material. The assignments do not require any form of coding or working with any specific software tools, since these are not topics covered in this course. The material instead provides a general overview and it does not get into many technical details. In other words, the objective of this course is to provide a broad overview, not depth in a particular dimension.

Course Schedule In Brief: A detailed schedule of lecture topics is available at the end of this Syllabus.

Week	Start Date	Lecture Topic	Lecture Video Length	Student Submissions
Week 1	09/03/18	Course and Big Data Intro	1:28:55	Weekly Assessment
Week 2	09/10/18	Big Data Collection	2:04:58	Weekly Assessment

				Term Paper Proposal
Week 3	09/17/18	Big Data Storage	3:46:26	Weekly Assessment
Week 4	09/24/18	Big Data Systems	3:36:11	Weekly Assessment
Week 5	10/1/18	Big Data Analytics	4:13:21	Weekly Assessment Term Paper: Big Data Analysis Proctorio Practice Quiz (Recommended)
Week 6	10/08/18	Internet of Things Intro IoT Architectures	2:42:10	Weekly Assessment Unit 1 Quiz
Week 7	10/15/18	IoT Technologies Part 1	3:03:17	Weekly Assessment
Week 8	10/22/18	IoT Technologies Part 2	3:03:55	Weekly Assessment Term Paper: Peer Review #1
Week 9	10/29/18	IoT Applications IoT Conclusion	3:19:22	Weekly Assessment Term Paper: IoT Analysis
Week 10	11/05/18	Cybersecurity Intro	1:00:25	Weekly Assessment Unit 2 Quiz (end of week)
Week 11	11/12/18	Systems Security	3:33:46	Weekly Assessment Term Paper: Peer Review #2
Break	11/19/18			
Week 12	11/26/18	Cryptography and Network	4:06:34	Weekly Assessment Optional Assignment Due (unless you pick a lecture in Week's 12, 13 or 14)
Week 13	12/03/18	Cybersecurity Case Studies	4:18:21	Weekly Assessment
Week 14	12/10/18	Cybersecurity Policy	1:27:30	Weekly Assessment Final Term Paper Unit 3 Quiz
Week 15	12/17/18	Course Closeout	None	Optional Assignment Due Dec 15th if you pick a lecture in Week's 12, 13 or 14. Submit earlier if at all possible.

Weekly Cycle of Work: Each week of the course runs from 9:00 AM U.S Eastern Time on Monday of the week listed in the "Course Schedule In Brief" until 9:00 AM Monday of the following week. If you like to work ahead, new material will be released sometime on the Friday before each Monday listed. Once course material is released, it will be visible for the rest of the course. Students will have until the end of the week (9:00 AM Monday) to complete the work of the week:

- watching the lecture video
- completing the Weekly Assessment
- participating in the optional Discussions, and
- making any Term Paper submissions

Each Unit Quiz will be available for 48 hours, starting at 9:00 AM U.S. Eastern Time on the Friday of the week in which it is listed. You can start the quiz anytime during that period, but once you start, you will have only two hours to complete it. It is very important to be aware that the Quiz will not be available after Sunday at 9:00 AM which means that it should be started no later than Sunday at 6:00 AM (preferably earlier in case there are issues with Proctorio). Starting the Quiz at 7:00 AM may be cutting it too short, for example, if there are issues with setting up.

Discussion Participation: It is required to make at least one meaningful contribution to discussions per Unit or you will get a penalty of 5 points over 100. For clarification: this is all or nothing, either you make all the three contributions and are not penalized or you do not make all and you lose 5 points out of 100 of your overall grade for the class. If you make three meaningful contributions in one unit and none in the other two you will still be penalized. Contributions must be made within the weeks of the unit - i.e. you can not make the three contributions on week 10. A meaningful contribution is defined around 4-5 lines of thoughtful comment or idea, and should definitely not consist of suggesting corrections that do not fundamentally change the meaning of another post ("Actually it was 36% adoption rate, not 38%"), compliments that do not add new information ("+1", "Great Point!"), Criticisms which do not add anything constructive ("This is not true.") . Instead try to add value, criticise when constructive and with reasoning, and complement and add further weight to discussions.

Late Policy: There will be a short grace period if you are unable to complete the Weekly Assessments on time in a particular week, but please do not do this more than twice during the semester. There is no grace period for Term Paper submissions or Unit Quizzes. Any exceptions for extreme circumstances must be negotiated individually, and in advance if possible.

Grading Summary:

Weekly Assessments:	30%
Three Unit Quizzes	30% (10% each)
Term Paper:	40% (See the Term Paper assignment for further details.)
Discussion Participation	- 5% (details above)
Extra-Credit Task	+ 5% (details below)

COURSE COMPONENTS

Lecture Video: The lectures are provided by two dozen of the most prominent thinkers and researchers in the respective fields, to ensure you get a diversity of views. Some weeks have much more video time than others, so plan accordingly. The video is hosted on an EdX platform outside of Canvas, which is how it was originally presented. All video is fully closed-captioned. PDFs of the presentations used are also available on the same EdX platform.

Weekly Assessments: After each week's lectures, you will take a Weekly Assessment, in the form of a short quiz. You can take each quiz twice. After you submit the quiz the first time, you will see which answers were correct and incorrect. If you didn't do as well as you'd like, you can then review any relevant lecture content, and take the quiz again. Canvas will keep the average of your two quiz scores. Each of these 14 assessments are about 2% of your final grade, for a total of 30%.

Discussions: There will be opportunities every week to discuss course materials, ask questions of the course staff, and share ideas. These are optional and will not be graded.

Unit Quizzes: At the end of each of the three units, you will take a Unit Quiz that may include any topic in that unit. The questions will be similar to questions asked in the Weekly Assessments, but there will be less emphasis on basic facts and skills, and more emphasis on critical thinking and applying what you have learned. Each quiz will be worth 10% of your final grade in the course, for a total of 30%.

Term Paper: You will write a paper during the course of the semester, comprising material from all three unit topics. You will be presented with a hypothetical future scenario (described in a separate document), and choose which of several aspects of that scenario to explore. You will consider hypothetical future technologies and imagine how current technologies might be used or adapted to make that future a reality. While the paper will require basic qualitative understanding of the material in the course, no coding or software development will be required. You will be submitting drafts of parts of the paper throughout the semester, as well as reviewing other students' work and giving feedback. Students taking the course for graduate credit will be required to conduct additional research and chose between submitting an outline of a Research Paper or submitting a Business Plan. Detailed instructions will be provided on an expanded document.

Extra Credit Task: Deepen your understanding of one of the course lectures by creating flash cards you can use to review the material for increased content retention, mastery and versatility. First choose one of the long video lectures (the ones that are about 30-60 minutes) and then develop a set of flashcards that summarize the contents of the lecture. The flash cards in the set should cover part, or the whole class, and be between 5 and 20 flashcards in number (depending on the amount of detail each card covers. More detail -> less cards). Then create a question similar to the weekly assessment questions to test the knowledge acquired. Flashcards that are of good quality may be circulated to this class or others (let us know if you don't want your cards circulated).

Textbook and Supplemental Reading: There is no textbook for the course. The weekly assessments and three unit quizzes will be based solely on the material in the video lectures. Each unit and week link to recommended supplemental reading and video, but these resources are optional. They may useful to you if you wish to deepen your knowledge in a particular area. You may find these resources useful in completing your Term Paper.

Getting Help: You will have access to course staff (the Instructor and Teaching Assistants) in a variety of ways:

- Weekly question-and-answer (Q&A) discussion forums.
- Periodic all-class or by-section web conferences
- Review sessions before each Unit Quiz
- Email and individual meeting support

All of these activities are optional. Participate in them if you think they will be useful to you.

HARVARD EXTENSION SCHOOL POLICIES

Special accommodations:
The Extension School is committed to providing an accessible academic community. The Accessibility Office offers a variety of accommodations and services to students with documented disabilities. Please visit www.extension.harvard.edu/resources-policies/resources/disability-services-accessibility for more information.

Academic Honesty

The work you submit must be your own work. You may build your code on samples from class or examples from texts, and we encourage students to discuss problems and techniques. Your homework should be all your own work or a combination of your own work and your synthesis and extension of examples. Please state the sources of any piece of code you use.

You are responsible for understanding Harvard Extension School policies on academic integrity (www.extension.harvard.edu/resources-policies/student-conduct/academic-integrity) and how to use sources responsibly. Not knowing the rules, misunderstanding the rules, running out of time, submitting the wrong draft, or being overwhelmed with multiple demands are not acceptable excuses.

There are no excuses for failure to uphold academic integrity. To support your learning about academic citation rules, please visit the Harvard Extension School Tips to Avoid Plagiarism (www.extension.harvard.edu/resources-policies/resources/tips-avoid-plagiarism) , where you'll find links to the Harvard Guide to Using Sources and two free online 15-minute tutorials to test your knowledge of academic citation policy. The tutorials are anonymous open-learning tools.

DETAILED SCHEDULE OF LECTURE TOPICS

UNIT 1: BIG DATA

Week 1 - Big Data Introduction and Use Cases

Introduction: Big Data Challenges (Sam Madden)

- Identify and understand the application of existing tools and new technologies needed to solve next generation data challenges
- Challenges posed by the ability to scale and the constraints of today's computing platforms and algorithms
- Addressing the universal issue of Big Data and how to use the data to align with a company's mission and goals

Case Study: Transportation (Daniela Rus)

- Data driven models for transportation
- Coresets for Global Positioning System (GPS) data streams
- Congestion aware planning

Case Study: Visualizing Twitter (Sam Madden)

- Understand the power of geocoded Twitter data
- Learn how Graphic Processing Units (GPUs) can be used for extremely high throughput data processing
- Utilize MapD, a new GPU based database system for visualizing Twitter in action

Week 2 - Big Data Collection

Data Cleaning and Integration (Michael Stonebraker)

- Available tools and protocols for performing data integration
- Curation issues (cleaning, transforming, and consolidating data)

Hosted Data Platforms and the Cloud (Matei Zaharia)

- How performance, scalability, and cost models are impacted by hosted data platforms in the cloud
- Internal and external platforms to store data

Week 3 - Big Data Storage

Modern Databases (Michael Stonebraker)

- Survey data management solutions in today's market place, including traditional RDBMS, NoSQL, NewSQL, and Hadoop
- Strategic aspects of database management

Distributed Computing Platforms (Matei Zaharia)

- Parallel computing systems that enable distributed data processing on clusters, including MapReduce, Dryad, Spark
- Programming models for batch, interactive, and streaming applications
- Tradeoffs between programming models

NoSQL, NewSQL (Sam Madden)

- Survey of new emerging database and storage systems for Big Data
- Tradeoffs between reduced consistency, performance, and availability
- Understanding how to rethink the design of database systems can lead to order of magnitude performance improvements

Week 4 - Big Data Systems

Security (Nickolai Zeldovich)

- Protecting confidential data in a large database using encryption
- Techniques for executing database queries over encrypted data without decryption

Multicore Scalability (Nickolai Zeldovich)

- Understanding what affects the scalability of concurrent programs on multicore systems
- Lock-free synchronization for data structures in cache-coherent shared memory

User Interfaces for Data (David Karger)

- Principles of and tools for data visualization and exploratory data analysis
- Research in data-oriented user interfaces

Week 5 - Big Data Analytics

Fast Algorithms I (Ronitt Rubinfeld): Efficiency in data analysis

Fast Algorithms II (Piotr Indyk): Advanced applications of efficient algorithms Scale-up properties

Data Compression (Daniela Rus)

- Reducing the size of the Big Data file and its impact on storage and transmission capacity
- Design of data compression schemes such as coresets to apply to Big Data set

Machine Learning Tools (Tommi Jaakkola)

- Computational capabilities of the latest advances in machine learning
- Advanced machine learning algorithms and techniques for application to large data sets

Case Study: Information Summarization (Regina Barzilay)

Applications: Medicine (John Guttag)

- Utilize data to improve operational efficiency and reduce costs
- Analytics and tools to improve patient care and control risks
- Using Big Data to improve hospital performance and equipment management

Applications: Finance (Andrew Lo)

- Learn how big data and machine learning can be applied to financial forecasting and risk management
- Analyze the dynamics of the consumer credit card business of a major commercial bank

- Recognize and acquire intuition for business cases where big data is useful and where it isn't

UNIT 2: THE INTERNET OF THINGS

Week 6 - Internet of Things Introduction & Architectures

IoT and the Connected World *(Sanjay Sarma)*

- What is IoT?
- How to think about IoT?

The Architecture of IoT *(Sanjay Sarma)*

- RFID Story
- Opportunities for IoT
- Some interesting IoT projects
- Architecture of IoT

The Web of Things *(Tim Berners-Lee)*

- Linked data- value is greatest when linked
- Enterprise data – shared v. public v. private
- Importance of security, privacy and authenticity
- Standards
- Web of Things layer – driver for IOT systems

Lessons from the Internet *(David Clark)*

- Is the Internet the right technology to hook together a network of things?
- The key lessons that our experience with the Internet teaches us about a future of things.
- A focus on network management, security, mobility and longevity.
- The desirable features of a distributed architecture for a system of things.

RFID Deep Dive *(Sanjay Sarma):* Case Study - RFID

Week 7 - IoT Technologies Part 1

Network Connectivity for IoT *(Hari Balakrishnan)*

- A simplified IoT network architecture
- Room/body-area networks: Bluetooth Low Energy
- Extending communication range

Data Processing and Storage *(Sam Madden)*

- Managing high rate sensor data
- Processing data streams
- Data consistency in an intermittently connected or disconnected environment
- Identifying outliers and anomalies

Localization *(Daniela Rus)*

- Localization algorithms
- Indoor localization
- Localization for mobile systems
- Applications

Week 8 - IoT Technologies Part 2

Security in IoT *(Srini Devadas)*

- Why is security for IoT so hard?

- Threat models
- Defensive strategies and examples

HCI in an IoT World *(Jim Glass)*

- Theory and applications of spoken dialogue for human-computer interaction
- Combining speech with other modalities for natural interaction
- Considerations for multilingual interactions
- Paralinguistic information from speech for enhanced HCI
- Future challenges for ubiquitous speech interfaces

Robotics and Autonomous Vehicles *(John Leonard)*

- Potential benefits of self-driving vehicles and service robots
- Sensing and data processing
- Simultaneous mapping and localization
- Levels of autonomy
- Future research challenges

Week 9 - IoT Applications and Roadmap

Beyond IoT - Ubiquitous Sensing and Human Experience *(Joe Paradiso)*

- Emerging Descriptive data standards for IoT and sensors
- Immersive visualization of diverse sensor data using game engines (part of IoT's 'control panel')
- Wearable sensing for IoT (future user interfaces for IoT - new ways to control and interact with your environment)
- Sensors and paradigms for seamless Interaction with the Built Environment (lighting, heating, etc.)
- Smart Tools for IoT
- Smart, sensate materials

Wireless Technologies for Indoor Localization, Smart Homes, and Smart Health *(Dina Kitabi)*

- Smart health
- Home automation
- Location tracking

Smart Cities *(Carlo Ratti)*

- The city as a cyber physical system
- Principles of cybernetics: sensing and actuating
- Collection of information: opportunistic sensing (a)
- Collection of information: crowd sensing (b)
- Collection of information: ad hoc sensing (c)
- Response of the system: analytics and optimization
- Response of the system: distributed action, people as intelligent actuators
- Price of anarchy
- Hacking the city: the risk for cyber attacks in centralized and distributed systems
- Smart city equals Smart Citizens

Roadmap of IoT *(Sanjay Sarma)*

UNIT 3: CYBERSECURITY

Week 10: Cybersecurity Introduction

Introduction: Cybersecurity (Howard Shrobe)

- Learning from the past: Multics
- Examples of what can go wrong
- Capability architectures
- Tagged architectures, including Memory safety, Type safety, Information flow, and “Zero Kernel”

Security Overview (Srini Devedas)

- Why security is a hard goal to achieve
- Broad strategies that one can employ to create secure systems

Week 11: Systems Security

Hardware Architectures for Security (Howard Shrobe)

- How novel hardware architectures can help to enforce the security properties that Operating Systems and Programming Languages expect, including memory safety, type safety, information flow, and access control
- How to enforce properties in hardware can be much more systematic and dramatically more efficient than enforcement by software alone

Operating Systems Security (Frans Kaashoek)

- Taking a global, systems-wide view of security. Viewing security as a "negative goal," considering all possible paths to security breaches- permissions, access, trojans, bugs, and many others
- Discussion of various design approaches to securing systems, including complete mediation, separation privilege, and minimizing the trusted computer base

Verifying Systems (Adam Chlipala)

- How to formulate requirements on secure behavior of C-like programs as rigorous logical formulas
- How to argue that programs meet such requirements

Secure Programming Languages (Armando Solar Lezama)

- Languages and low-level security properties
- Languages and high-level security properties
- Type safety

Week 12: Cryptography and Network

Public Key Cryptography (Ron Rivest)

- A quick overview of basic public-key cryptography, including the RSA and El-Gamal public-key cryptosystems, relevant number theory, security definitions, and complexity assumptions
- A brief intro to more advanced notions, such as digital signatures, certificates, homomorphic encryption, and elliptic curves

Multi-party Computation, Secret Sharing, Distributed Trust (Shafi Goldwasser)

- A cryptographic paradigm shift: from communication of private data to computation over private data
- How to define secure multiparty computation and its applications
- How to achieve multi-party secure computation through polynomial secret sharing and computation on shares
- Using the principles of multiparty computation to distribute trust and power

Homomorphic and functional encryption (Vinod Vaikuntanathan)

- Solving the all or nothing paradigm of encryption - how do we keep data secure, while allowing computations on underlying data: how to “have your cake and eat it too”
- Discussion on the developments in homomorphic encryption from the ‘80s through today, and constructing partially and fully homomorphic crypto-systems

- How functional encryption systems can grant third-party access to only the precise information it needs to perform a task

Network Security and Protocol Design (Dave Clark)

- What the different aspects of network security are and what mechanisms are used to address them
- Why the current state of network security is so poor

Week 13: Cybersecurity Case Studies

Bitlocker (Nickolai Zeldovich)

- Discussion of the challenges of building a disk encryption system using passwords, removable devices and trusted hardware methods
- Overview of the Trusted platform method and case study of how BitLocker uses this method
- How BitLocker actually encrypts data using the 'poor man's authentication' method

Resilient Software (Martin Rinard)

- An explanation of how errors can cause serious security vulnerabilities using examples of buffer and integer overflow errors
- Overview of methods to automatically find and patch these errors, including transferring correct code from one application to another, and generating patches to identify errors and produce the correct output
- Discussion of how automating these difficult and expensive manual tasks can produce software that is more reliable and secure, with enhanced functionality

Web security (Daniel Jackson)

- What makes web applications especially vulnerable to attack
- How attacks against web applications work: two broad classes
- How to defend against attacks, at the design and implementation levels

Mobile Phone Security (Nickolai Zeldovich)

- Discussion of the unique characteristics of a mobile phone that necessitate different applications from existing desktop or web applications
- Case study of how computer system security works in mobile phones, using the Android's application model, and mechanisms used by Android to secure data while allowing applications to share information
- A look back on Android's development: What worked? What didn't work? What changes have been made to overcome challenges?

Week 14: Cybersecurity Policy

This systems module discusses policy aspects of cybersecurity.

Management, Strategy and Organizational Issues (Michael Siegel)

- Security of conventional information systems is recognized as important, but is still not fully effective.
- The number and magnitude of recent cyber-attacks (Target, Home Depot, SONY, etc.) is growing weekly. Also there is growing concerns for the security of our Cyber-Physical Infrastructure and increase exposure from the IoT (Internet of Things).
- At the same time it has become more apparent that people represent the weakest link in the security stack. In this module we examines managerial, strategic and organizational issues that can help improve performance and reduce the growing cyber threat.

The Landscape of Cyber Policy (Danny Weitzner)

- Inquiring into the role public policy plays in sustaining and securing the Internet

- Reviewing six key policy goals embodied in today's digital communications environment, how they arose, and what keeps them on track
- Exploring the way progress requires policymakers to understand the background against which new rules are made