Big Data for Reliability and Security

Saurabh Bagchi
Purdue University

Last update: *November 23, 2019*

This is a course suitable for Masters and PhD students. Online and on-campus course to be offered in Spring 2020. This course will be available on the edX platform as well.

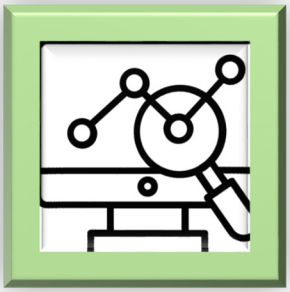
5 weeks = 15 hours of lecture and lab material.

Prerequisites: Knowledge of Python and introductory Statistics.

A. Foundational material on reliability and security [3 lectures]

1. Introduction: Motivation, System view of reliable and secure design, Terminology *[1 lecture]*
2. Security landscape for connected systems: Traditional threats, new threats due to large-scale systems, new threats due to big data *[1 lecture]*
3. Reliability landscape for connected systems: Traditional concerns, new concerns due to large-scale systems, new concerns due big data *[1 lecture]*

B. Data analytic techniques for dependability [6 lectures]



1. Supervised and unsupervised learning techniques *[2 lectures]*
2. Techniques for dealing with large scale data: regularization, feature engineering, dimensionality reduction, etc. *[2 lectures]*
3. What is our toolchest of data analytic techniques: what to use and when *[1 lecture]*
4. Data analytic techniques used for reliability and security: strengths, weaknesses, opportunities *[1 lecture]*

C. Case studies and challenge problems [6 lectures]

1. Practical case study on application of big data to reliability: Amazon Web Service *[1 lecture]*
2. Practical case study on application of big data to security: Anomaly detection *[1 lecture]*
3. Practical case study on adversarial Machine Learning: Image and video manipulation *[1 lecture]*
4. Lab exercise: Setting up Mesos and Flink cluster, TensorFlow on cluster *[1 session: hybrid of lecture and lab]*
5. Challenge problems *[2 sessions: hybrid of lecture and lab]*
    a. Challenge problem 1: Predicting computer system failure
    b. Challenge problem 2: Detecting spam and phishing emails
    c. Challenge problem 3: Detecting spurious social network content