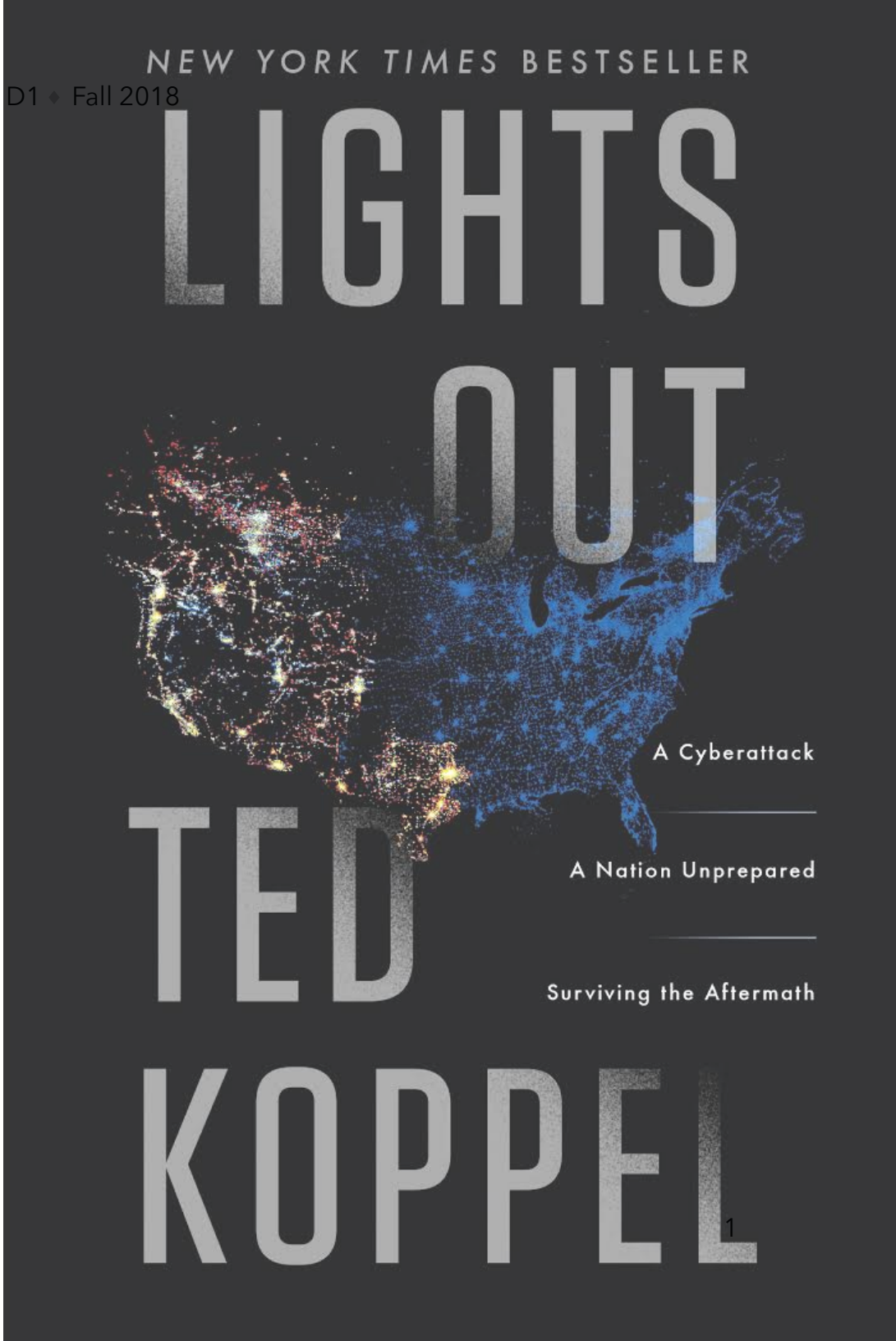
Course Project
Critical Infrastructure Protection

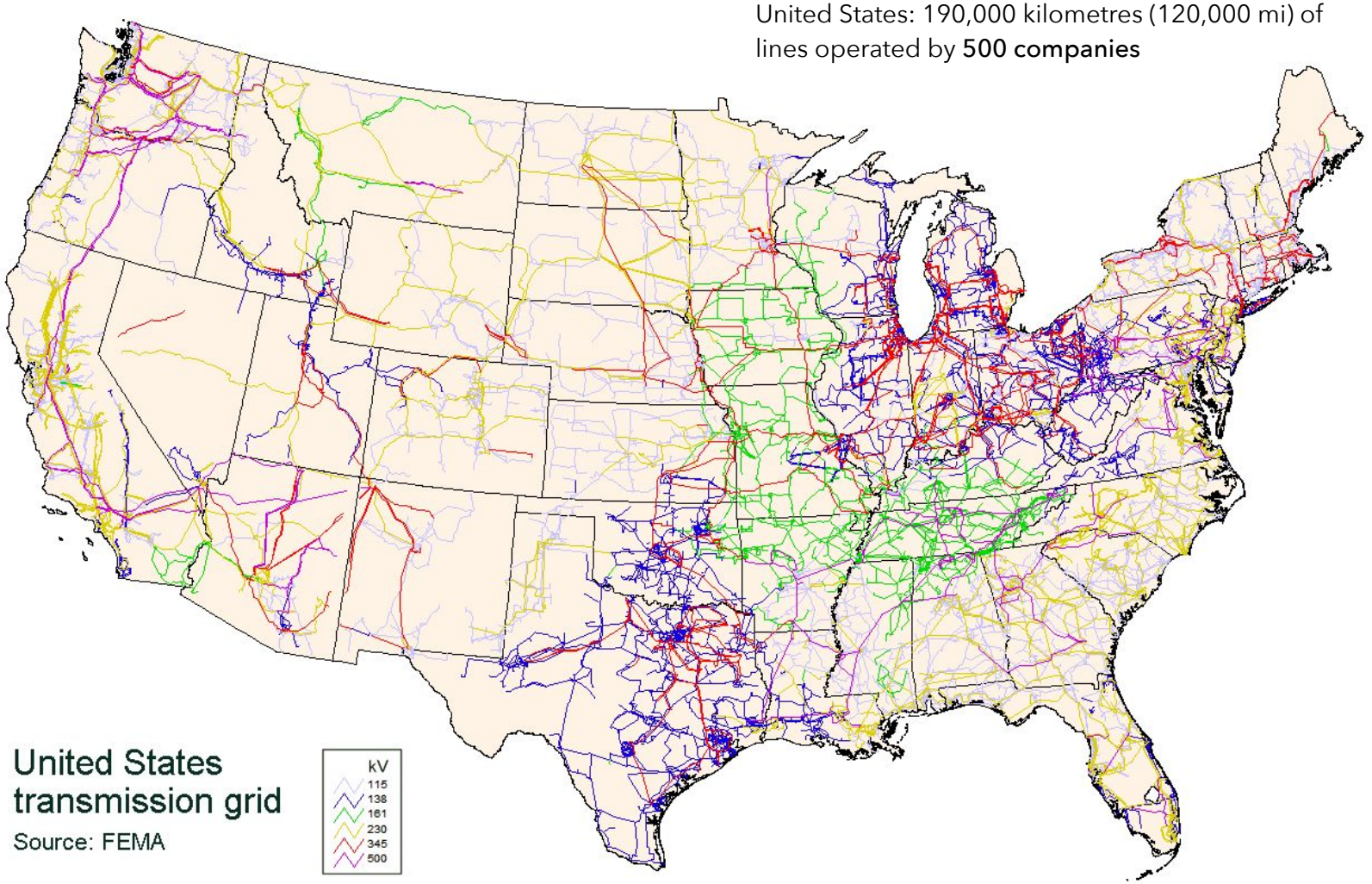
"So many of our transactions are now conducted in cyberspace that we have developed dependencies we could not even have imagined a generation ago. **To be dependant is to be vulnerable.** We have grown cheerfully dependent on the benefits of our online transactions, even as we observe the growth of cyber crime. We remain largely oblivious to the potential catastrophe of a well-targeted cyberattack."

– Ted Koppel, 2015



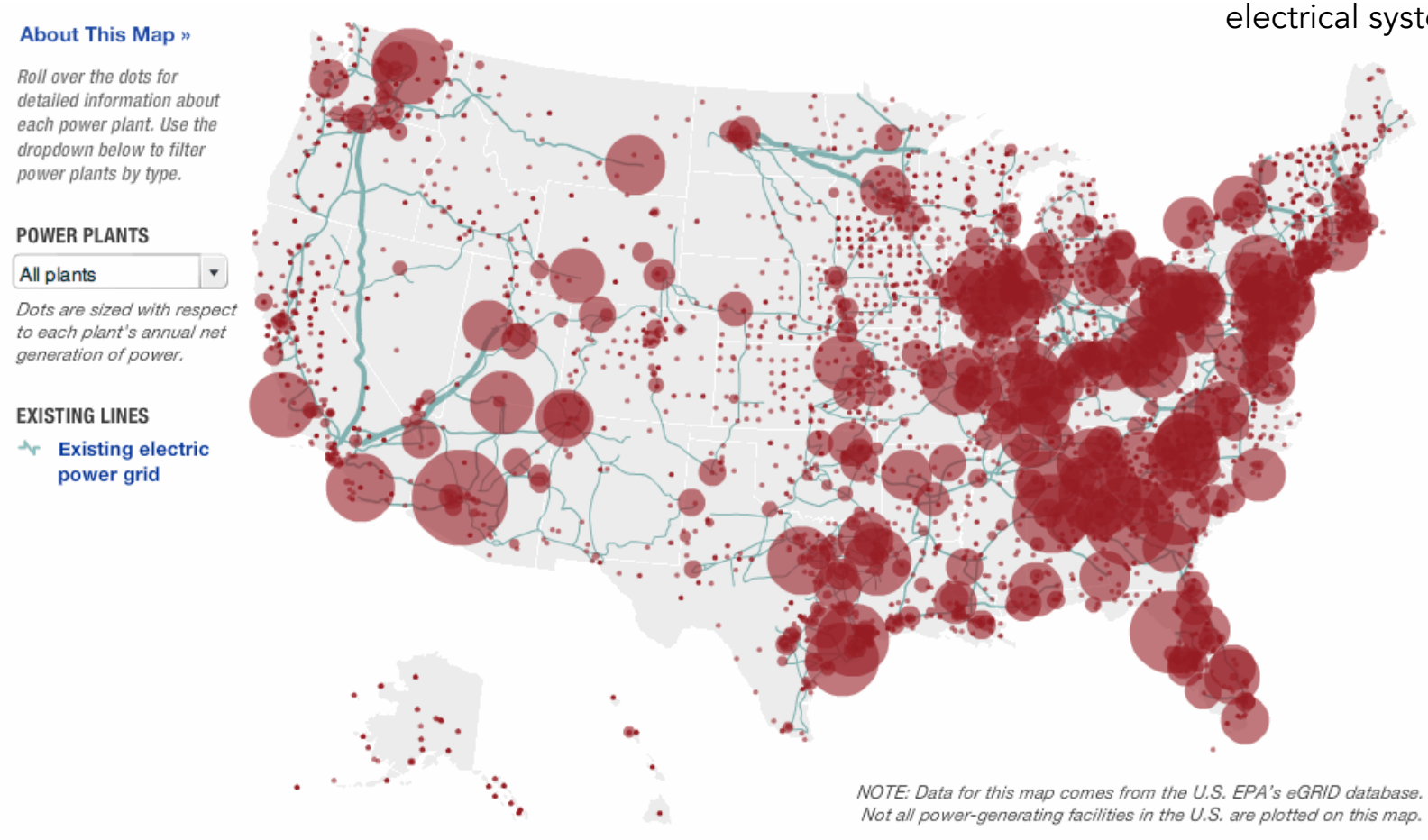
Electric Power Grid

Electric power transmission grid of the contiguous United States: 190,000 kilometres (120,000 mi) of lines operated by **500 companies**



Security Problems

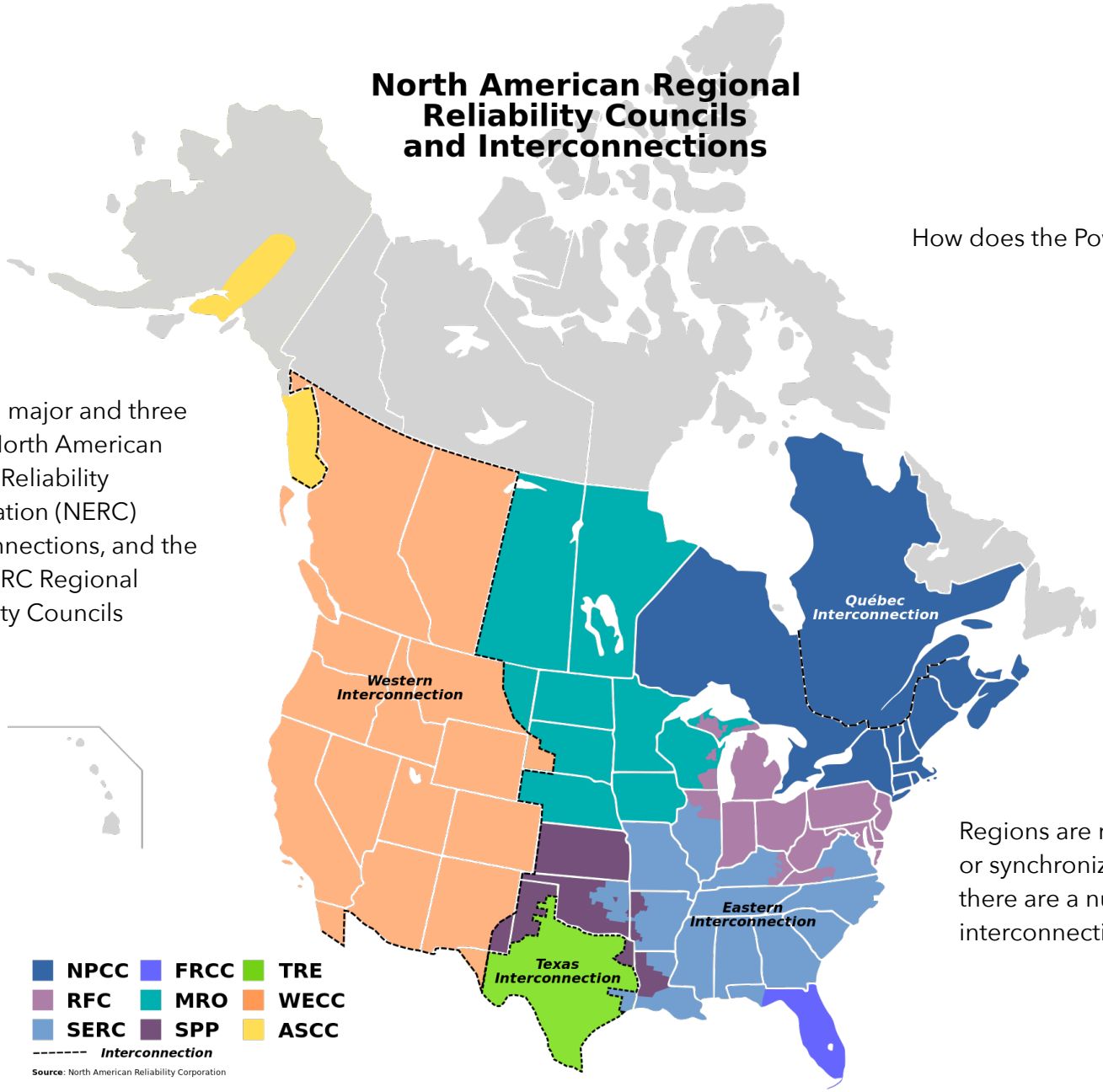
"The U.S. electric grid is a complex network of independently owned and operated power plants and transmission lines. Aging infrastructure, combined with a rise in domestic electricity consumption, has forced experts to critically examine the status and health of the nation's electrical systems."



Source: https://www.infrastructureusa.org/interactive-map-visualizing-the-us-electric-grid/

How does the Power Grid work? (video)

The two major and three minor North American Electric Reliability Corporation (NERC) interconnections, and the nine NERC Regional Reliability Councils



Regions are not directly connected or synchronized to each other, but there are a number of HVDC interconnections.

The U.S. power grid consists of three loosely connected parts, referred to as interconnections: eastern, western and Texas.

Within each, high-voltage power lines transmit electricity from generating sources such as coal or hydroelectric plants to local utilities that distribute power to homes and businesses, where lights, refrigerators, computers and myriad other "loads" tap that energy.

Because electricity in power lines cannot be stored, generation and load have to match up at all times or the grid enters blackout territory. The interconnectedness of the grid makes it easier to compensate for local variations in load and generation but it also gives blackouts a wider channel over which to spread.

Transmission system operators scattered across some 300 control centers nationwide monitor voltage and current data from SCADA (supervisory control and data acquisition) systems placed at transformers, generators and other critical points.

Source: Scientific American
https://www.scientificamerican.com/article/2003-blackout-five-years-later/

The 2003 Northeast Blackout

On **August 14, 2003**, shortly after 2 P.M. Eastern Daylight Time, a high-voltage power line in northern Ohio brushed against some overgrown trees and shut down—a fault, as it's known in the power industry. The line had softened under the heat of the high current coursing through it.

Normally, the problem would have tripped an alarm in the control room of FirstEnergy Corporation, an Ohio-based utility company, but the alarm system failed.

Over the next hour and a half, as system operators tried to understand what was happening, three other lines sagged into trees and switched off, forcing other power lines to shoulder an extra burden. Overtaxed, they cut out by 4:05 P.M., tripping a cascade of failures throughout southeastern Canada and eight northeastern states.

All told, 50 million people lost power for up to two days in the biggest blackout in North American history. The event contributed to at least 11 deaths and cost an estimated $6 billion.

Source: Scientific American
https://www.scientificamerican.com/article/2003-blackout-five-years-later/

SCADA Systems

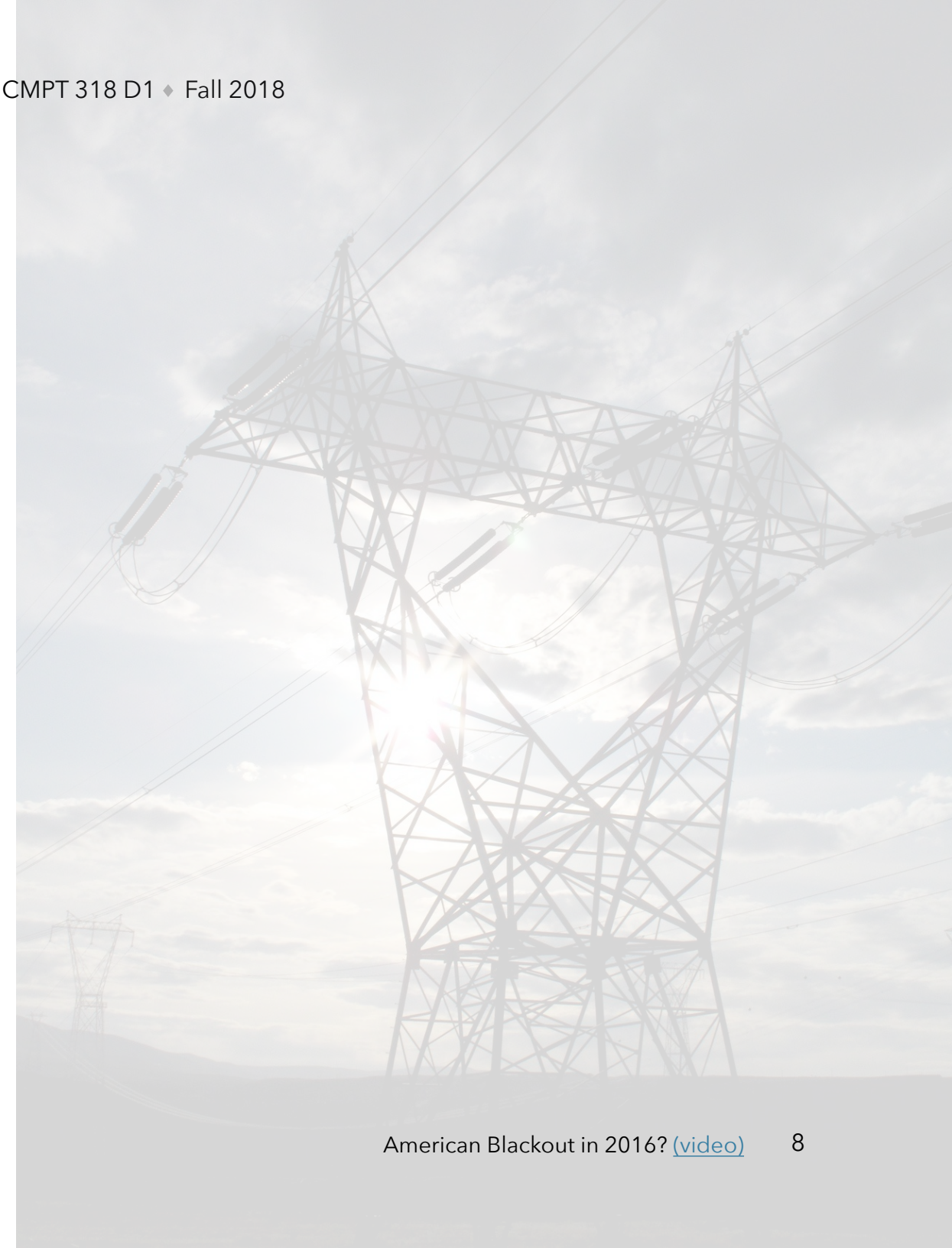
Supervisory control and data acquisition (SCADA) is a **control system architecture**

Functional manufacturing levels using computerized control

Source: Wikipedia

Electric Grid Security (video)

"These days, within any one of the three U.S. grids, almost all operational phases of thousands of power companies are interconnected. Coordinating operations are run using the same supervisory control and data acquisition (SCADA) systems. Most of the systems are manufactured by a relative handful of companies, and while they are not quite interchangeable, there are similarities in programming and structure. This presents a **web of pathways** connecting the thousands of power companies and enabling transactions … The overall system has been designed for **maximum efficiency**, eliminating waste while establishing a precise balance between the power needed and the power generated."



Course Project Description

Our course project is a group project comprising **three separate parts** as detailed below. Each group has five members working together as a project team.

Important Dates

- A project report is to be completed and **submitted** by November 21, 2018.
- Final presentations of project outcomes follow on November 26, 28 and 30. Please note that in addition to the regular class hour on November 28, we need an extra time slot to accommodate presentations.[1]
- Finally, there will be a **final test session** on December 3 (last day of classes) to assess the quality of the technical solution and contribution of all individual team members.

[1] Extra presentation session (selected groups only): Wednesday, **November 28, 12-14 PM,** TASC 1 - Room 9204 West.

Project Scope

In light of increasing cyber threats, especially advanced persistent threats, and existing vulnerabilities that expose critical infrastructure to a variety of adversarial scenarios, the project explores behaviour-based intrusion detection methods used for cyber situational analysis of automated control processes. Electric power grids, intelligent transportation systems, public water utilities, oil and gas pipelines, and other critical infrastructure routinely **rely on automated control** for their continuous operation. Automation is essential for operating equipment and monitoring conditions of machinery, production processes and plants; it enhances efficiency and quality of service delivery, safe operation of critical assets and their protection in case of internal or external disruptions.

Supervisory Control and Data Acquisition (SCADA) is a control system architecture that uses computers, networked data communications and peripheral control devices for high-level process supervisory management—widely used for operating critical infrastructure all over North America and beyond. Situational awareness requires continuous situation analysis performed on real-time control data from the operation of critical systems. This is to facilitate early detection and warning about suspicious and potentially harmful anomalies in order to mitigate the impact of attacks by launching countermeasures and facilitate cyber forensics.

Methodology

Intelligent monitoring and control of critical infrastructure produces massive volumes of *time series data* from heterogeneous sensor networks. Whenever comprehensive historic data from past operation of critical infrastructure is available, Big Data analytics is a sensible approach to advanced anomaly detection and has been studied in the scientific literature. Various types of probabilistic models such as Hidden Markov Models have been proposed as a formal basis for predictive analytics to represent normal operation in a compact form allowing to effectively differentiate between **normal behaviour** and **anomalies**.

Challenges

A number of inescapable 'external factors' can make anomaly detection in time series data challenging whenever data originates from the operation of a real-world system. Typical examples include: imperfections in the data, such as missing or corrupted values; lack of *ground truth* in historic data, unavailability of labels to differentiate normal data points from outliers; various types of anomalies depending on the application context; striking a good balance between ***precision and recall***, specifically also reducing the false alarm rate to make anomaly detection practical in any real application context with resource constraints.

Data Source

In this project we use variations of a real data set compiled from monitoring **household power consumption** for some part of the electrical power grid. The data is available in compressed form from the course page and comes in several related versions: a training data set and several test data sets (with more test sets coming further down the road). You may use any tool or language of your choice. We do recommend though using R to those who are looking for an easy and powerful starting point. R has many analytic packages which make your analysis easier. For example: seqHMM or MHSMM are some popular HMM packages in R; Python also has HMM packages like np-HMM and others.

Logical Organization

The project is logically organized in three separate parts, each of which comprises a number of different tasks. The marking of your course project will ultimately take into account all three parts: the breadth and depth as well as the quality of the technical solution, the project report and the presentation.

Part 1 - Data Analytics

This part is about exploring and understanding the specific characteristics of the data and developing an analytic approach to anomaly detection for different types of simple and complex anomalies. Specifically, you need to find suitable probabilistic models for the purpose of representing 'normal' system behaviour to the extend possible with no ground truth available. Think about common usage patterns one can identify and possible anomalies one may observe while continuously monitoring system operation in a control centre. Generally, there may not be the one best solution but a number of reasonable alternatives how to address the problem. Highlight and explain your major design choices, providing a **proper rational**.

Part 2 - Project Report

Each project team is supposed to describe their methodical approach, their experimental analysis, the key findings from their experiments, challenges encountered and lessons learned in the form of a **technical report**. Details about what is expected from a technical report in terms of overall structure, logical organization and writing style will be provided in a separate document (to come). Generally, a technical report is a clearly written, well structured document to communicate technical ideas and insights to an audience with a technical background.

Part 3 - Presentation

Each project team will present the outcome of their work in a 10-12 minutes technical presentation in class during the second last week of classes. This means to properly summarize the **essence of your project report** in a formal presentation using slides with intuitive textual and graphical illustrations. Normally, only two members of each team will actively present while the other team members only have to answer questions.

Getting Started

With the information provided here you should be able to get your team organized and your project under way. Meet early and meet often!

You will receive **additional input** verbally and in writing as you proceed through your project. A clear breakdown of tasks and responsibilities shared by all team members helps in developing a clear roadmap allowing the team to work more productively. Please take into account though that the team as a whole is responsible for their project and team members are expected to help each other in managing project tasks.

We hope you will find this project an interesting and rewarding experience.

Thank you in advance for your cooperation!