Data Analytics for Power Grid Planning & Operations

Anjan Bose
Washington State University
Pullman, WA, USA

Data Analytics for Smart Grid
Washington State University
Pullman, WA
August, 2017
Power Systems Data

Fixed Data (Assets)

• 7,500 generation plants
• 75,000 substations
• 300,000 miles transmission (100,000 lines and transformers)
• 2.2 million miles distribution (1 million distribution feeders)
• 300 million customers

Power Flow data for transmission system ~ 2GB
Data for all equipment in PB
Applications for System Data

• System data for assets is fixed (changes slowly)
• Each engineering application requires a (small) subset of systems data
• Are there applications that require all data?
  ▪ Asset Management
  ▪ Inventory Control
  ▪ Maintenance Records
  ▪ Automated Mapping/Facilities Management
  ▪ Etc.
Measurement Data (Variables)

- Power, Var, Voltage, Current, Frequency, etc.
- SCADA (EMS, DMS)
- PMU data
- Substation, Generating Plant, Large Load data (stored but not usually transmitted)

- Measured at various frequencies
  - EMS SCADA at 2-10 seconds
  - DMS SCADA at 10-60 seconds
  - PMU 30-120 times per second
  - Substation/Plant data stored at various rates
Average Data Flows Today

- Average Reliability Coordinator has 10 Balancing Authorities (control centers)
- Average Control Center has 100 high voltage substations
- Average substation has 100 measurement points
- Average polling rate for real time data is 5 seconds

So

- Average data rate from each substation is 20/sec
- Average data rate to a control center is 2K/sec
- Average data rate to a RC is 20K/sec
Simulated Bus Voltages by Powertech TSAT
Generated PMU Measurements
33 msecs time steps
Data Collection by PMUs

• PMU sampling rates: 30-120 per second
• Assume 100 values per second

If we assume all 100 points in a sub are PMUs

• Average data rate per sub is 10K/sec
• Average data rate for the total of 100 subs in a BA is 1M/sec
• Average data rate for the RC is then 10M/sec
Data Base Issues

• Real time data base must be distributed
  ▪ Large amounts of calculated data must be part of this data base
• Static data base must be distributed
• Historical data base will require still another design
• Substation data bases and system level data bases have to be coordinated
• All data bases in the same interconnection will have to be coordinated
• Standards will be key
Data Exchange Issues

• Within one organization
  ▪ Data movement between EMS, BMS, DMS, OMS, AMI, etc. is non-trivial

• Within one hierarchy
  ▪ Several TOs to ISO
  ▪ Several BAs to RC

• Laterally between neighbors
  ▪ Bilateral agreements too many to be manageable

• Bandwidth, volume, latency

Standardization is the key
Proposed Communications
Applications, Models and Data

• New technologies promise new applications
• Applications may be distributed
• Static data (models) and real time data (measurements) will be distributed
• Communications (hardware/middleware) needed to move data in a timely manner
• IEC/NIST has recommended standards for both data and communications

The CIM standard (IEC 61970) is being widely adopted and is evolving at the same time
Example: Power Flow Network Model

Information and Semantic Models

- **CIM UML**
  - Conforms to IEC 61970-301 CIM
  - Information Model
    - Defines all concepts needed for exchange of operational load flow models
      - Reused parts
      - New extensions

Context

- **Power System Model Profile Group**
  - Conforms to IEC 61970-452, 453, 456, others
  - Model Exchange Profile
  - Contextual layer restricts information model
    - Specifies which part of CIM is used for static/dynamic model exchange
    - Mandatory and optional
    - Restrictions
    - But cannot add to information model

Message Syntax

- **CIM/RDF Schema**
  - Conforms to IEC 61970-501 and -552 CIM XML Model Exchange Format
  - File syntax
    - Can re-label elements
    - Change associations to define single structure for message payloads
    - Mappings to various technologies can be defined
Control Centers

• The next generation of control centers will have a more flexible (decentralized) architecture
• The boundaries between various XMS functions (including protection and local controls) will fade
• The automatic coordination between entities that are interconnected will increase
• This will require large movement of data both hierarchically and laterally
• This cannot be done without wide adoption of standards across the interconnection
Large Data Applications

Historical Data (Measurement Data Only)

• Data Science has many usable tools
• Identify measurement anomalies (model-free)
• Identify trends (loads, renewable generation, outages – equipment failures, control operations)
• Identify patterns (correlations between loads, solar, wind, maintenance, external events)
Large Data Applications

Historical Data (Measurement + Model)

• Event analysis
• Identify measurement anomaly (SE bad data)
• Simulation management
• Planning scenario development
• Training scenario development
• Cross infrastructure analysis
• Controller design
Concluding Remarks

• Data science is the use of big data in NEW ways (NOT the use of existing power applications with bigger data)
• Are there uses of data that does not require physical models?
• Are there new applications that take advantage of both big data and physical models?
• Are there on-line applications that can use big data to help operators make decisions?