SYNCHROPHASOR STARTER KIT DATA QUALITY

North American Synchrophasor Initiative
Chicago, IL

October 14, 2015

Ken Martin

Electric Power Group

What is Data Quality?

- Means different things to different people
- Generally can be defined:

Data Quality is any aspect of data that bears on its ability to satisfy a given purpose*

- Quality requirements depend on the application
- Multi-application use must address all data aspects

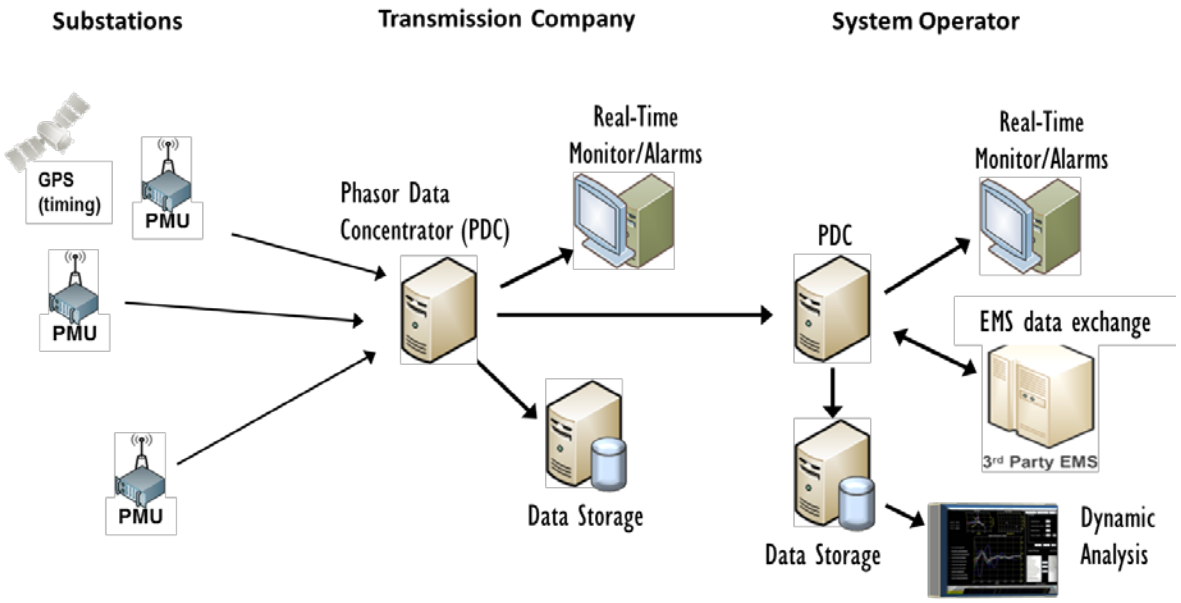
* from Wikipedia

Basic categories of DQ

- Data Loss
- Data Corruption
- Inaccurate representation of engineering quantity
- Lack of precision
- Incorrect measurement identification
- Excessive or inconsistent latency

Data Loss

- Multilevel
- Many steps
- Multiple handoffs

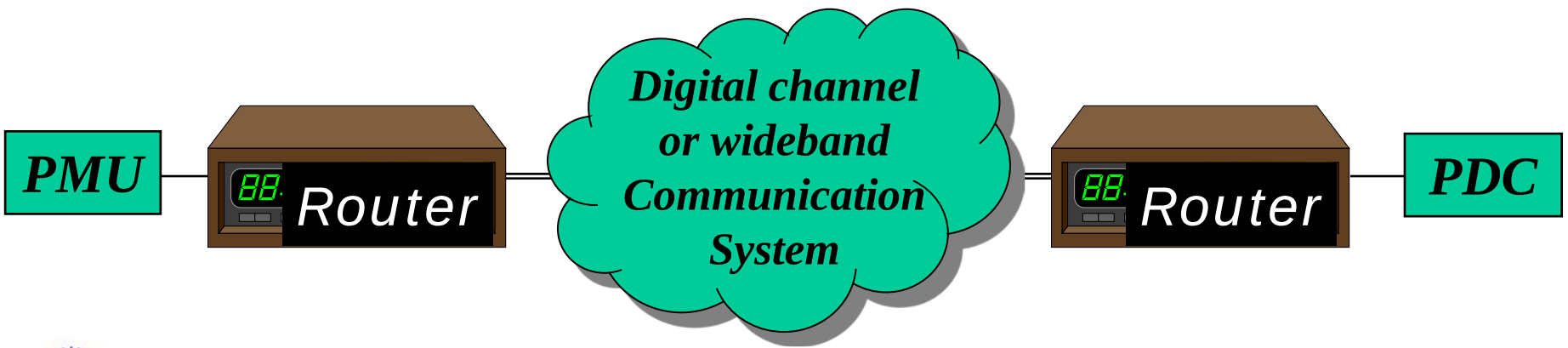


Typical phasor measurement system

- Communications – insufficient capacity (links, buffers, etc.), routing errors, priority errors
- Processing – overloaded PDCs, data exchange mismatch
- Equipment or program failure

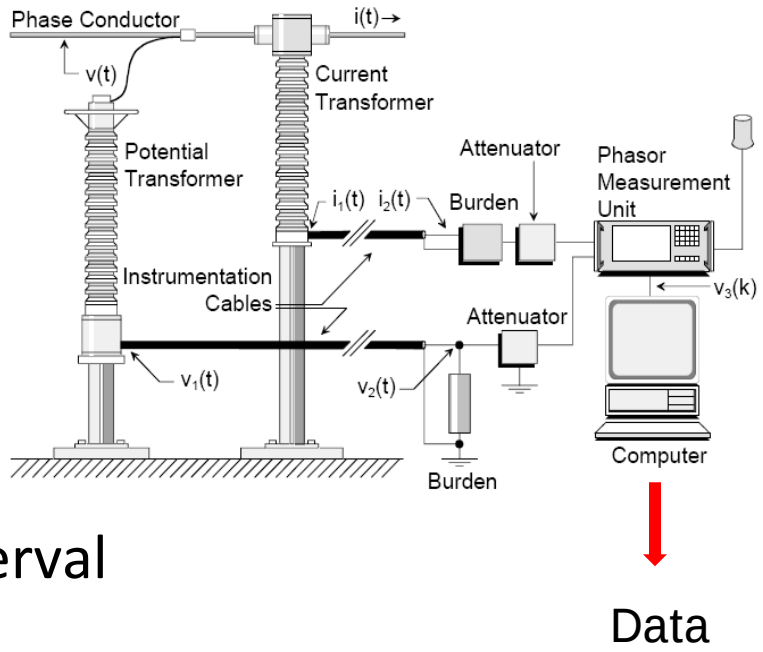
Data corruption

- Scaling and corrections misapplied
- Data type misread
  - Integer/floating point, int15/int32, etc.
- Communication problem
  - Clocking error, overruns, etc.
- Message fragments lost
- Computer and program mishandling



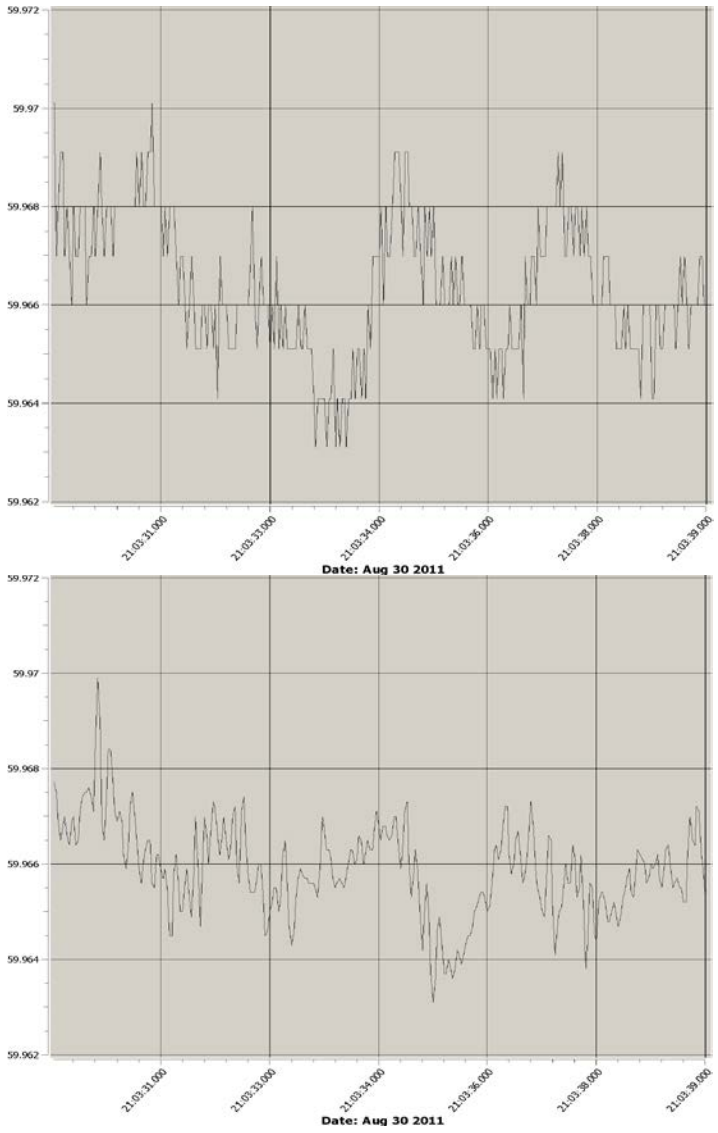
Inaccurate Representation

- Primary transducer scaling – CT & PT
  - Imprecise calibration, aging, temperature
  - Ratio selection
  - Mag & phase errors
- Timing errors
  - Primary sync & local clock
  - Phase angle & time errors
- Phasor/frequency estimates
  - Noise in signal
  - Dynamic changes in estimate interval
- Processing errors (PDC & apps)
  - Scaling incorrectly applied
  - Time alignment errors



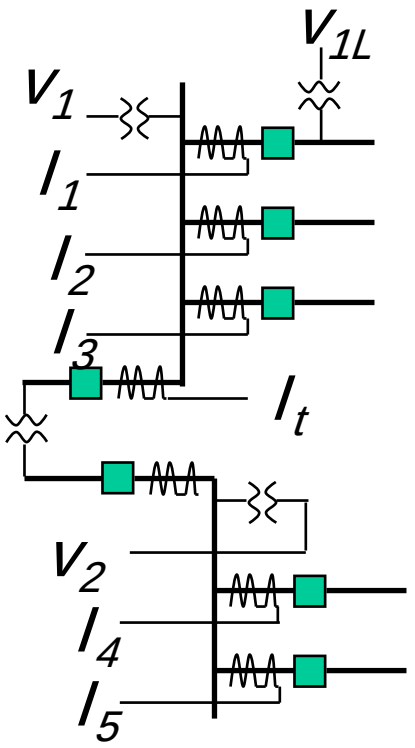
Lack of Precision

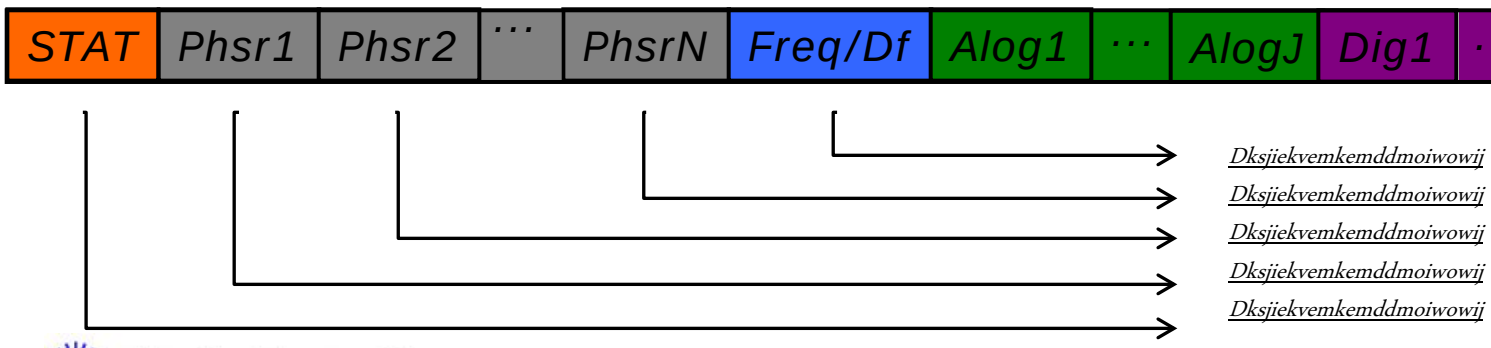
- Input waveform scaling
  - Waveform “steppy” or clipped
  - Inaccurate or noisy measurements
- Output data scaling
  - Overrun with high values
  - Loss of precision in low values
  - Insufficient bits in format
    - Floating point vs integer
- Misapplied compression
  - Loss of measurement detail



Incorrect Identification of Data

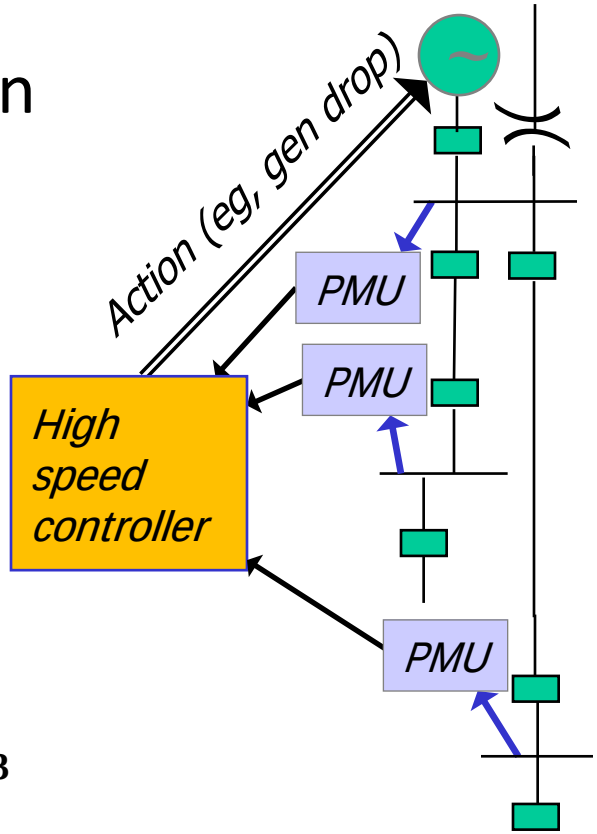
- Line PT vs bus PT
- Wrong bus or line
- Measurement mismatch
  - Voltage, current, or frequency
- Configuration error
- Naming error

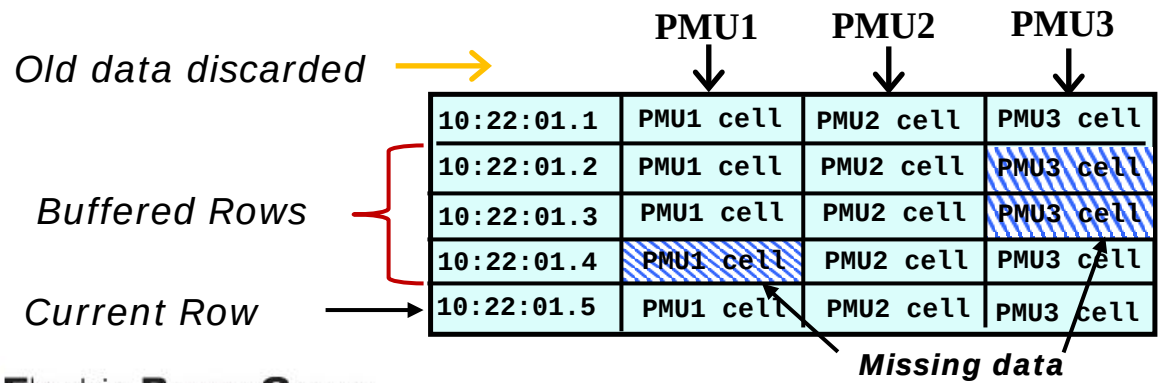




Excessive or Inconsistent Latency

- Excessive latency
  - Data too late for use by application
    - Real-time monitor
    - Automatic control
- Inconsistent latency
  - Application processing errors
  - Exceeds PDC wait





DQ Practice

- **How should data quality be approached?**
- Planning and design
  - Adequate communication – sufficient capacity and high reliability (probably the biggest problem area)
  - Timing system with alarming (2nd biggest problem)
  - PMU coverage serving application needs
- Validate installation
  - Calibrate PMU & components
  - Compare measurements with substation measurements
  - Compare at control centers with other systems (SCADA)
- Ongoing monitoring
  - Continuous or periodic problem alarms & cross check
  - Organized maintenance plan for prompt servicing

Summary

- **Data quality:**
  - Depends on the actual intended use of the data
  - Best practice examines all aspects of data
  - Can be broken down into 6 categories: loss, corruption, accuracy, precision, identification, and delivery delays
    - Other categorization forthcoming
- **Address issues by:**
  - System planning and design
  - Installation and validation
  - Monitoring and maintenance program

Thank you

Questions?