Submitted for publication. Author copy - do not redistribute.

Investigating Synchrophasor Data Quality Issues

Kenta Kirihara, *Student Member, IEEE*, Karl E. Reinhard, *Student Member, IEEE*,
Andy K. Yoon, *Student Member, IEEE*, Peter W. Sauer, *Fellow, IEEE*
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

***Abstract*—Phasor Measurement Units (PMUs) installed across North America, compute and store synchrophasor data at select locations. However, the efforts to process the synchrophasor data to meet very high operator trust expectations has been challenging. This paper describes Synchrophasor Data Quality (SDQ) Activity, Trustworthy Cyber Infrastructure for the Power Grid (TCIPG) Project efforts to investigate synchrophasor data archived from American Transmission Company's system including both observations from the field to characterize synchrophasor data quality.**

I. Introduction

Since April 2012, the SDQ Activity has been working with American Transmission Company (ATC) to investigate its archived synchrophasor data. Fig. 1 shows an example nominal synchrophasor data network from the point of measurement to the regional ISO. Data archives are created at the third level before the data is sent to MISO. Each level has an identifiable role in computing, processing, and forwarding data to meet power system requirements and as such becomes a possible point of failure. Attributing defective synchrophasor data sources and corresponding error rates is key to prioritizing efforts to improve data quality. Fig. 2 shows the possible sources of error. These act as a guideline to diagnose and investigate issues.

During summer 2013, Kenta Kirihara worked as an Energy Management System intern working for ATC's synchrophasor system project engineer, James Kleitsch. His experience working with field synchrophasor data in the field highlighted fundamental challenges to providing very reliable, high integrity, and very accurate synchrophasor data to the system control room.

The first challenge was analyzing synchrophasor data flow to identify and classify data error types. There were occurrences in which a single PMU data failed to report to the PDC, or when a group of PMUs that are associated with a PDC failed to report to the data archives. To aid understanding the issues, he developed and implemented an application showing real-time data status at each PMU location on a service area map. This tool enabled the control center to have continuous, real-time situational awareness of data availability.

Second challenge was productively using synchrophasor data. The volume of the PMU data collected, which is over 1GB of system-wide relevant measurement information per hour: unlike conventional data, such as video streams, the data is difficult to visualize and overlay. This data stream contains a plethora of information about the system dynamics.

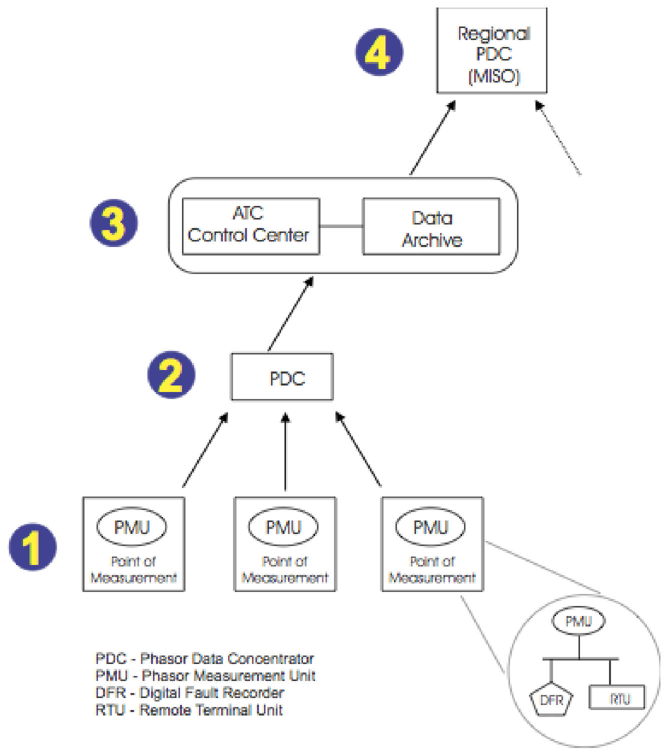

Fig. 1. Example PMU Data Flow

Identified Error Sources and Proposed Error Type Classifications[1]		
Error Source	**Level(s)**	**Error Type**
Status code errors	1,2,3	data processing
Data streams disordered / shifted in processing	1,2,3	data processing
Loss of PDC configuration	2,3	data processing
Improperly configured PMUs (window length/windowing method)	1	digital signal processing
Frequency calculation discrepancies (C37.118.2005)	1	digital signal processing
Quality of metering	1	equipment specification
Accuracy issues (CT/PTs not properly rated for application)	1	equipment specification
Calculation uncertainty – vendor equipment operating differences	1	equipment specification
Metering locations separated by breakers	1	installation
Meters not installed at recorded locations	1	installation
PMU data streams not named as per system policies	1	installation
Asynchronous local behaviors (e.g., DC bias injections during solar storm)	1	measurement
Malformed network packets	2,3	network failure
Network data loss	2,3	network failure
Mislabeled phasor data streams	1,2,3	PMU configuration
Differences between PMU manufacturer calculation approaches	1	PMU standards

Fig. 2. Possible Sources of Error [1]

For instance, during an unusual event in a system, the voltage phase may present signature behavior. In order to investigate this, a tool to visualize the phase angle difference through

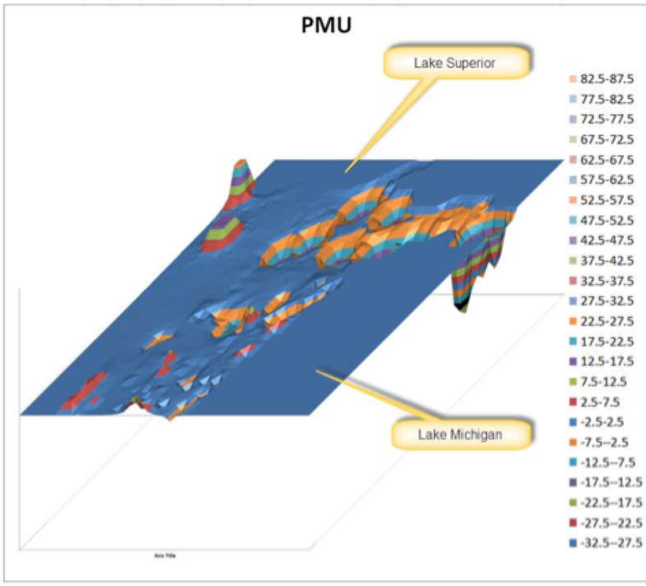

Fig. 3. PMU Angle Data Visualization [3]

system in both real-time and post-event was created as shown in Fig. 3. The tool was also used to compare the PMU angle measurements to the State Estimator derived from the SCADA measurements. The tool allowed to compare the proximity of the State Estimator to real-time measurements, which could not have been done in SCADA measurements. Also, visualization of the phase angle measurements, highlights the system stability and dynamics. Having a tool which allows the large amount of information to be presented immediately adds value to the data.

Although both of these solutions were significant steps in developing a utilization of PMUs, the full potential of synchrophasor measurements is yet to be discovered fully. Through the summer internship, ATC and TCIPG researchers gained valuable insights into ATC, which subsequently proved to be invaluable to setting up practical, secure data sharing arrangements between ATC and TCIPG that enable further investigation of synchrophasor data quality.

II. Research Approach

To effectively investigate synchrophasor data quality and the signatures of defective data, robust partnership between industry and researchers is very desirable. This partnership facilitates synchrophasor data discovery research which allows both investigation of synchrophasor such tool data quality issues and data analysis with detailed context such as system topology and operating state. The fundamental basis to this project is in the building of a robust 3-way collaboration including American Transmission Company (ATC), Pacific Northwest National Laboratory (PNNL), and University of Illinois (UIUC).

With 2 years of ATC synchrophasor data, knowledge of the system topology, and system events, R-project statistical analysis program (R) is used to screen synchrophasor data for signatures of both known events and synchrophasor data issues.

A. *Investigating Data with R*

Along with a R code written for this project to analyze the data, Situational Awareness and Alerting Report (SitAAR), a Big Data analyzing tool developed by PNNL, is also utilized. SitAAR is designed to detect any atypicalities using advanced statistical analysis tools. Taking into account the data point multivariate distance from the center of all the points, data point is considered atypical if it lies outside of the multi-dimensional cloud of all data points (i.e. an outlier) or if the data point is clustered by itself or a few other points (i.e. an inlier) [2]. Thus, to build a more precise cloud of system information, or different categories of normal system behavior, a long-term data is necessary.

R is an open source program which is optimized for statistical application. R is a direct successor to the S-language, which is a C-based programming language developed for statistical usage. Both elementary and advanced statistical methods can be accessed as functions. In the case in which certain functions are not available on the native package of R, additional packages can be downloaded or scripted with C,C++, or Fortran to meet the specific task. For these reasons, R was the desired resource for this project.

B. *ATC Proposed Preliminary Event Analysis*

The data used for this project is located in the data archives at ATC. From this archive, relevant information can be extracted, and transferred to UIUC for analysis. From a commercial computational capacity, this amount of data to analyze is considered ”Big Data” which proposes difficulty in computing all of the signatures.To reduce the computational intensity of this potential ”Big Data”, certain sections of the ATC territory for a relatively short amount of time are analyzed to provide a foundation for the project.

ATC has provided preliminary set of data for analysis. Two weeks of seven different substations throughout the system has been extracted out of the ATC system in CSV format. In order to look for characteristics in certain events, ATC has proposed preliminary event analysis categories:

1) unit trips and/or loss of significant loads,
2) system fault
3) capacitor bank failures
4) predecessor events to trips
5) imbalanced line phase angle differences

By identifying these signatures in the PMU data, alarming events in the grid and data quality issues are distinguished. Signatures for these events are used to detect similar behavior; thus, the remnant abnormalities in the data are due to synchrophasor data quality: as a result, synchrophasor data quality problems are extracted.

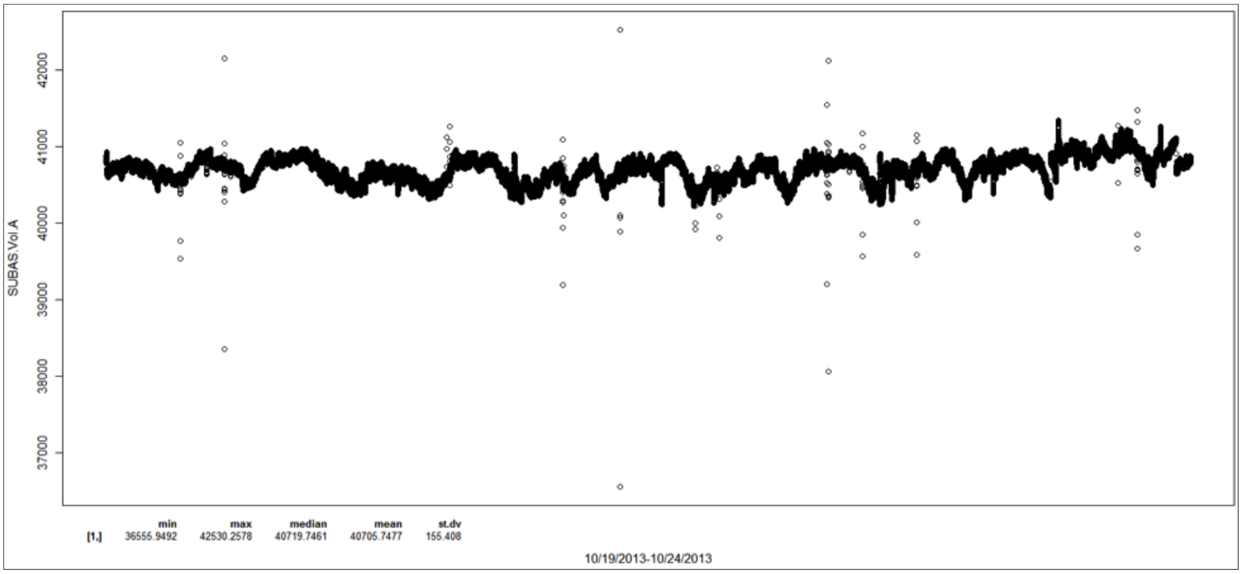

Fig. 4. Primary Analysis on Voltage Magnitude for Substation AS

III. Research Results

A. Primary Analysis with Principal Statistical Values for Stable Operations Extraction

Analysis are completed in week intervals to reduce stress on the computation. Since computation time is non linear in practice, shorter computation period is preferable for preliminary understanding. Fig. 4 shows the results for primary analysis. The data points are vectorized to depict a general trend in operation during this time frame. The specific time each data point is associated with has been removed since the focus is to calculate the statistical quantities associated over a period of time. Basic statistical values; min, mean, median, mean, and standard deviation; are calculated for the voltage magnitude for this substation. The mean and the standard deviation implies the acceptable range for normal and stable operation. From the exclusion-inclusion principle, these points that lie outside of these expected normal operation are unusual behaviors. However, as seen from the graph, many points exist outside of this standard deviation, but the cause is unknown; more advanced data mining techniques are needed.

B. Secondary Analysis with SitAAR

To find specific cause for these abnormality, SitAAR is utilized. SitAAR divides the data set into 1 minute intervals and calculates a quadratic regression to group the behaviors into clusters. The group of these clusters create the cloud.The 1 minute interval equates to 1800 PMU data points, and this allows for general robustness in the analysis. As of October 2013, ATC system has 92% of the PMUs delivering 99.9% reliable data with a 99.9% success rate [3]. Using these data and averaging all of the PMUs, the expected value of reliable data per PMU is 1600 data points per minute: an accurate quadratic regression can be made with this many points. Having a quadratic regression over the micro-intervals, allows a more granular detail on the behavior by having the change in slope: this allows a differentiation between gradual change in value, sudden permanent change in value, and impulse change. For instance, when focused on voltage magnitude only, these three factors signify gradual increase in voltage, capacitor bank switching, and a line trip. Again, similar behavior can be identified from the current magnitude as well.

Fig.5 shows an analysis for current magnitude abnormality for the span of 2 weeks. Discrete voltage magnitude change is identified and denoted by the dotted line. Possible explanation may be a change in load at such point due to line trip elsewhere.Also, Fig.6 shows the atypicality score corresponding to the same SitARR report. This section shows that this cluster was less populated with similar behavior, implying that this behavior was atypical. Further inspection

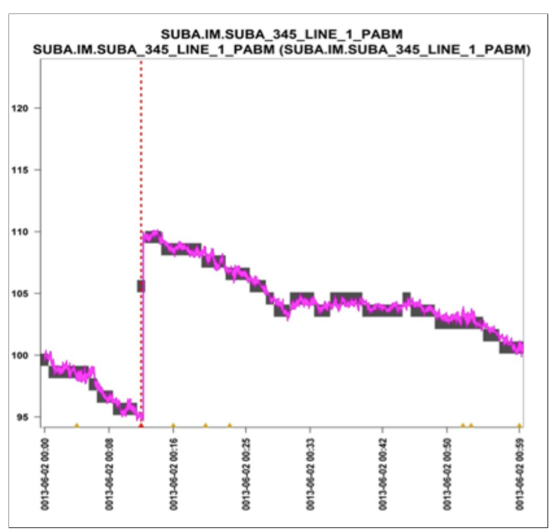

Fig. 5. Sample SitAAR report showing line current data

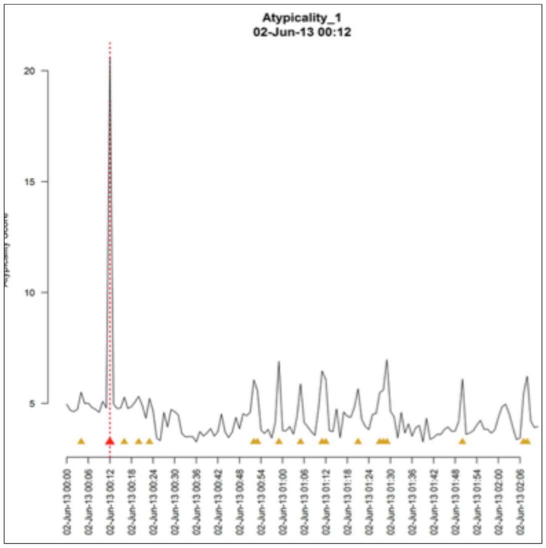

Fig. 6. Atypicality score corresponding to SitAAR report

of the SitAAR report revealed similar alerts associated to similar behavior.

Having access to the possible events that has occurred on the system, the signatures of these events and corresponding system events can be linked. However, if there exists signatures that are abnormal, but no system event is evident at the time, the signature is likely to be a synchrophasor data quality issue or a system event that has yet to identified. If certain signatures are identified predating the system event, that is not caused by sudden weather conditions, the events can be used to forecast the events.

IV. Broader Impact

Progression of the project brings about a benefit to ATC both in real-time system awareness and system forecasting. Present SCADA only allows system event to be identified by the result of the action; by relays or capacitor banks switching; but does not link the cause to the result. The strength of having a stream line of data, unlike a Digital Fault Recorder, allows precursor events to be identified. In other words, studying PMU data may link a system behavior to specific causes of faults. Advancement in algorithm will allow ATC to have a better system situational analysis: thus, rapid fault-condition location and identification is possible while possibly identifying the cause in real-time. Replacement of assets can be done quicker as location and cause is known.

Implementation of this information may also be used in parallel with State Estimators to forecast system events. By having the precursor signatures to known events, operations can be done to avoid such events from occurring.

Both of these will support the business cases for ATC. A more optimized and smoother control of the transmission, as well as an identification to specific points of improvements on the system, will allow stake holders to support business improvement initiatives.

PNNL, which provided the tools for the analysis, also has benefits. Close interaction with power system data allows for improvement recommendations of SitAAR. Improvements in SitAAR will allow a more detailed analysis for system events outside of electrical grids, and can advance analysis on other power systems such as automobiles and aircrafts.

Lastly, for consumers of the power systems, which includes majority of the U.S Population, will benefit as wellt. Improved electrical power grid operations through the Midwestern state, which is essential is connecting the two coasts, will result in less outages: there will be less losses to corporations and locally, less food supply spoilage in households which is often a problem in areas that are dependent on electricity.

V. Conclusions

Creating a profile for normal operation and known signatures aid in the development of a real time alarming system. Unlike SCADA with cloud capabilities, real-time analysis on the PMU network will allow faster, granular, and more secure wide area awareness of the system. Due to the PMU network being built in parallel to the SCADA system, data can be retrieved from the system without compromising the control aspect. A line of effort exist in a development of such tool to aid operations, identify new synchrophasor data issues, and possibly identify a phenomenon which was never identified.

References

[1] Research Activity Fact Sheet, Trustworthy Cyber Infrastructure for the Power Grid, Champaign, IL, 2012, pp. 57-59

[2] Amidan, Brett. Baselining Studies and Analyses., presented at 2013 Transmission Reliability Program Peer Review, Washington DC, 27 June 2013 [Online]. Available: http://energy.gov/sites/prod/files/2013/07/f2/2013TRR-2Amidan.pdf

[3] Kleitsch, James. Smart Grid Investment Grant Update, presented at October 2013, North American Synchro-Phasor Initiative (NASPI) Work Group Meeting, Rosemont, IL, 22-24 Oct 2013. [Online]. Available: https://www.naspi.org/File.aspx?fileID=1177