A Wide-Area, Wide-Spectrum Big Data System

Gregary C. Zweigle
Schweitzer Engineering Laboratories, Inc.

This paper was presented at the 2015 IEEE Power & Energy Society General Meeting and can be accessed at: http://dx.doi.org/10.1109/PESGM.2015.7285873.

For the complete history of this paper, refer to the next page.

Published in
Wide-Area Protection and Control Systems: A Collection of Technical Papers Representing Modern Solutions, 2017

Originally presented at the
2015 IEEE Power & Energy Society General Meeting, July 2015

A Wide-Area, Wide-Spectrum Big Data System

Gregary C. Zweigle
Schweitzer Engineering Laboratories, Inc.
Pullman, WA, USA

***Abstract*—Power system data undergo significant signal filtering and downsampling at measuring devices before being communicated to wide-area visualization and analysis applications. For these applications, such processing acts in part to reduce network bandwidth requirements while maintaining essential characteristics of the original signal. However, modern communications systems are advancing in their capacity, and the price of data storage continues to drop. The constraints under which existing systems were built are being mitigated. Meanwhile, the need for higher-resolution data across larger areas continues to increase. This paper discusses the design and implementation of a system that delivers wide-spectral bandwidth data with precise time stamps over a wide area. Processing steps, such as converting to phasor representation, are moved to receivers. The system architecture allows a broader utilization of power system data. Several applications are described. Construction of a test system, operating between several cities, is described, and performance results are shown.**

***Index Terms*—big data, event collection, SCADA, synchrophasors, wide-area visualization.**

I. Introduction

Supervisory control and data acquisition (SCADA) is the primary method for communicating power system data for visualization, analysis, and human-interacting operational control. Data are scanned asynchronously from measurement devices, preprocessed, and transmitted at intervals in the range of 1 to 5 seconds. Recently, synchrophasors have become a complementary approach. These data are measured, filtered, and transmitted in a streaming manner by phasor measurement units (PMUs). Rates are as fast as 60 times per second. With the use of a global timing signal, the time-stamp accuracy of synchrophasor data is better than 1 microsecond.

SCADA data include the calculation of the magnitude component of phasor values for measured voltages and currents but not the angle component. Magnitudes are useful because in many cases their dynamics change slower than the time stamp and sampling uncertainty. With better timing accuracy, synchrophasor systems allow the calculation of angles for voltages and currents.

The developed and deployed system described in this paper communicates wide-area power system data without first converting the data to a phasor representation. Unlike modern PMUs, the measuring devices are kept simple, and all sophisticated signal processing is performed at the receivers. Signals are sampled at a fixed rate, attached to a precise time stamp, and transmitted as the original wide-spectrum signal. The receivers are responsible for storage and applications. Signal processing is performed in the applications. The volume and rate of data to manage are more than an order of magnitude higher than what is generated by present synchrophasor systems. Applications are those related to data analysis, asset management, modeling, and power system operations.

II. Solution Overview

A. *Quasi-Stationary Phasor Representation*

Although SCADA and synchrophasor-based systems differ in data rates and timing mechanisms, they both use the quasi-stationary approximation for representing power system values. Consider a measurement x(t), either of a voltage or current signal. In an ideal steady-state operating mode, the signal x(t) is a consistent sinusoidal signal scaled by A, frequency f, and phase offset δ.

$$x(t) = A\cos(2\pi f t + \delta) \tag{1}$$

For this ideal case, x(t) can be represented by a phasor quantity that removes time dependence.

$$X = Ae^{j\delta} \tag{2}$$

A more realistic representation of the signal x(t) includes the fact that the peak value and nominal frequency referenced phase change with time. Under the quasi-stationary assumption, changes are approximated as happening relatively slowly [1]. In this case, the phasor includes a time-varying component.

$$X(t) = A(t)e^{j\delta(t)} \tag{3}$$

Both SCADA and synchrophasor systems use representation, as shown in (3). In the SCADA case, the magnitude $A(T_s)$ is calculated at the measuring device and communicated, with measurement time T_s having a relatively large uncertainty. In the synchrophasor case, the magnitude $A(T_s)$ and phase angle $\delta(T_s)$ are calculated at the PMU measuring device and communicated, along with an accurate measurement time T_s. The quasi-stationary representation ignores signal dynamics not of interest for many common applications. This allows for restrictive filtering and reduces communications bandwidth and storage requirements. At the application level, when data are received, this representation also helps operators focus on the most important signal characteristics.

Fig. 1 shows the signal processing stages of a typical PMU [2]. The PMU measures power system signals, converts them to synchrophasors using (3), and transmits the resulting data. The data are initially sampled at a thousands-of-measurements-per-second rate with analog-to-digital conversion (ADC). Subsequent correlation removes the fixed nominal power system frequency component of the sampled signal. The data are then passed through a low-pass filter

(LPF) to a Nyquist-limited bandwidth in order to be transmitted at a desired message rate without aliasing. For the common 60-messages-per-second rate in North America, the filter has a bandwidth of 30 Hz or less. Finally, phasor quantities $A(T_s)$ and $\delta(T_s)$ are calculated. These are communicated, along with the measurement time T_s, through the network interface. The signal processing performance and a definition of the communicated signals are specified in the IEEE C37.118 standard [3].

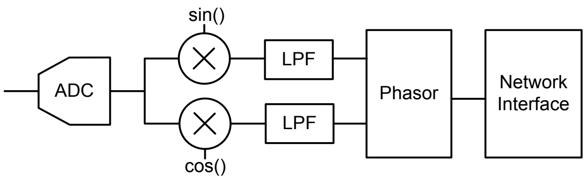

Fig. 1. Basic PMU signal processing.

B. *Wide-Area, Wide-Spectrum System*

For proper operation and analysis of a modern power system, it is becoming more important to keep dynamics that the narrow PMU filtering removes while maintaining an accurate time stamp, the Nyquist filtering, and the wide-area nature of the measurements. Historically, the limitations of both SCADA and synchrophasor calculations were not as important because of the slower dynamics that characterized previous power systems. Today, with a strong renewable generation growth rate, an increasing number of power electronic loads, and more complicated interconnections, the system dynamics are changing faster. Reliance on electric power, as well as the demands for reliability and power quality, continues to increase. Regulatory guidelines are necessitating more thorough recording and analysis of system performance and events [4][5][6]. Therefore, in addition to SCADA and synchrophasor data, applications are beginning to need untransformed wide-spectrum signal data from across a wide area.

This is partially available today through intelligent electronic device (IED) oscillography recordings. These recordings can be found in digital fault recorders (DFRs), protective relays, and meters. Solutions are available that can automatically retrieve these stored files and make them available for analysis [7]. However, this information is presently gathered in a triggered and buffered file-based manner. Under prespecified conditions, such as a power system event, the IED collects and stores a window of x(t) data. This window of data becomes available as an oscillography file for either manual or automated collection. Once collected, the data are managed as individual collection sets. Organizing all of the various files together into a coherent view of the power system is a complicated task. The individual files often become stored in a database structure, but the ability to search, analyze, and use these data across time and devices is limited.

The data system discussed in this paper measures, streams, archives, and presents a wide bandwidth version of x(t) suitable for many power system operation and analysis applications. The communications architecture parallels existing SCADA and synchrophasor systems in that it is designed for wide-area measurements. This new approach moves the application-specific signal processing from the measuring device to the applications at the receiver. So, for operational or analysis applications that require the quasi-stationary approximation, it is at the application that these calculations are performed. For operational or analysis applications that need other signal characteristics, the filtering and calculations necessary are performed as needed at the application. A larger variety of data becomes available.

III. DATA SYSTEM ARCHITECTURE

A block diagram of the data system is shown in Fig. 2. Measurement devices are distributed throughout the power system. These IEDs can be part of recloser cabinets or located in substations, microgrids, and other locations. At each distributed site, a local communications network can exist for local human-machine interfaces (HMIs), for local protection and control measurements, such as IEC 61850-9-2 [8], and for communication between IEDs. There can also be communications interfaces between substations for regional protection and control, such as line current differential.

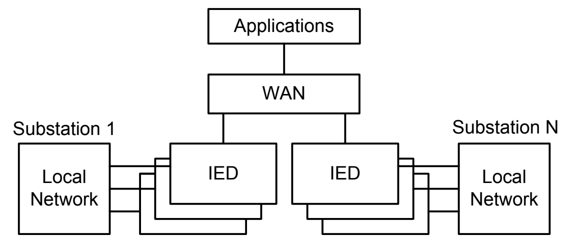

Fig. 2. System for data visualization and analysis applications with wide-area, wide-spectrum measurements.

Separate from these local or regional networks, the data system described in this paper communicates over a wide-area network (WAN). The WAN quality of service is appropriate for situational awareness and analysis applications [9]. These requirements are different than what is needed for IEC 61850-9-2 protection and control applications.

A. *Measurement Device Processing*

The processing at each IED is shown in Fig. 3. The first step is ADC of the power system signals. The LPF limits the data as required to meet Nyquist rate requirements. The upper f_{max} can be much wider than a PMU limit. The source coding is suitable for the signal quality desired up to frequency f_{max}. In the simplest case, the data are not compressed. This paper does not address the data compression step, other than to note that the filtering associated with the quasi-stationary representation results in signal loss and compression. In the system described in this paper, it becomes possible to apply other data compression methods, with new signal quality tradeoffs.

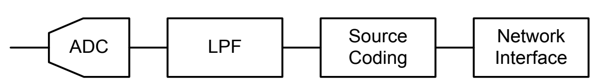

Fig. 3. Portion of IED signal processing applied for the wide-area, wide-spectrum system.

B. Application Processing

Fig. 4 shows a diagram of the processing at each receiver. The data are optionally decompressed prior to storage in the historian and then recompressed in a form suitable for storage.

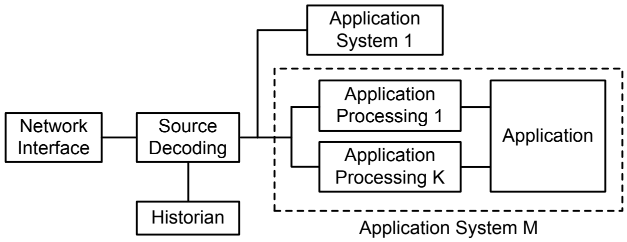

Fig. 4. Receiver and application signal processing.

Each of the *M* application systems time-aligns and processes data according to its own needs. A large variety of processed data becomes available with this approach. The following are some examples:

- A simple HMI application requires phasor data and a relatively low update rate. The sequentially sampled values are filtered and downsampled to a rate of one update every few seconds. Phasor quantities for the voltages and currents, along with power quantities, are calculated for display.
- A wide-area situational awareness application requires phasor data at a higher rate, along with live streaming charts of synchrophasor samples [10]. The processing for synchrophasor data is similar to Fig. 1, but with the bandwidth selected for the needs of the situational awareness application unconstrained by channel bandwidth limitations.
- Offline analysis or modeling applications process and display the full sample rate data. Data are analyzed from across the power system, and during times well before and after the main duration of any events.

In order to conserve computing resources, it is possible to share processing between application systems (for example, calculating and archiving phasors as an independent step, with the resulting data communicated between application systems).

C. Communications and Storage Requirements

The network bit rate requirements are high with wide-area, wide-spectrum measurements but not outside of the capabilities of a modern communications system. Consider transmitting positive-sequence voltages in a synchrophasor system. For ease of comparison, IEEE C37.118 is used as a reference for both a PMU system and the new system. The standard requires B_{fix} = 18 bytes of fixed overhead. The phasor magnitude and angle require another B_{data} = 8 bytes combined. Also, the standard requires sending B_{freq} = 4 bytes for frequency. The resulting bit rate (f_{bit}) is approximately 14.4 kilobits per second at a transmit rate of f_r = 60 samples per second.

$$f_{bit} = \left(B_{fix} + B_{data} + B_{freq}\right)\frac{8 \text{ bits/byte}}{1{,}000 \text{ bits/kbits}} f_r \qquad (4)$$

When sending wide-area, wide-spectrum measurements, it is necessary to communicate all three phases instead of only the positive-sequence phase. However, the magnitude and angle (or real and imaginary) pair is not needed. So, B_{data} increases only by a factor of 1.5. The data rate depends on the upper limit f_{max} of interest. Meeting the Nyquist rate requires a transmit sample rate of $f_r = 2f_{max}$. For example, when f_{max} = 600 Hz, f_r = 1,200 samples per second and (4) results in a bit rate of 327 kilobits per second. The total bit rate as a function of f_r is shown in Fig. 5.

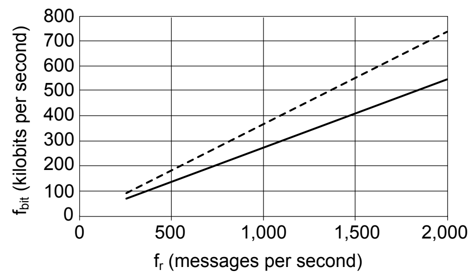

Fig. 5. Required data rate as a function of the transmitted sample rate for three-phase voltages (solid) and also with three-phase currents (dashed).

The storage requirements are also higher for the present method than for SCADA and synchrophasors, but they are still within the capabilities of modern big data enterprise systems. The storage requirements of data at the f_r = 1,200 sample-per-second rate as a function of storage history is shown in Fig. 6.

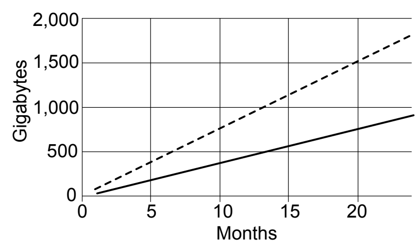

Fig. 6. Storage per IED for three-phase voltages (solid) and storage with currents added (dashed).

Technology continues to increase both the available channel bit rate and the amount of storage available for a fixed cost. The trends are in favor of the full-signal wide-area system becoming increasingly viable.

IV. Test System Implementation

A wide-area, wide-spectrum operation and analysis big data test system has been implemented in North America (as shown in Fig. 7). Initially, three measuring sites were installed. One is located in Charlotte, North Carolina; one in Columbus, Ohio; and one in Pullman, Washington. Data are communicated to the processing location in Pullman. For demonstration purposes, each measuring device is simply connected to a wall outlet distribution voltage signal. The devices are also connected to the Global Positioning System (GPS). The communications network has a capacity bit rate of 50 megabits per second. Network measurements show a peak rate of just over 300 kilobits per second. Sending a single voltage in the test system reduced the bandwidth slightly compared to (4), but packet overhead due to sending small packets increased the bandwidth. All data are archived in a time-series historian. From the historian, data are streamed for visualization. The historian supports data queries for the offline analysis of the sampled data.

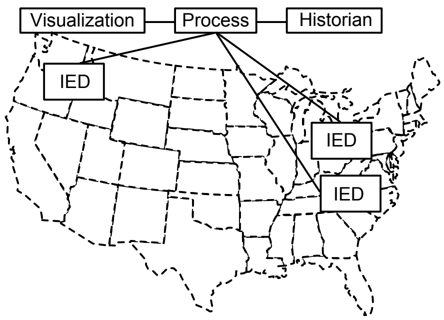

Fig. 7. Fully operational test system deployed in North America.

The complete system has been operational for several months. The intent is to add more measurement devices. Data are transmitted at 1,200 samples per second. Storage requirements at the receiver are approximately 1.5 gigabytes per day. Adding the open-source 7-Zip compression algorithm can reduce these storage requirements.

A visualization display, demonstrating the variety of data possible with this system, is shown in Fig. 8. In the upper portion are synchrophasor frequency data, sampled 60 times per second. In the lower portion are the new wide-spectrum measurements, sampled 1,200 times per second. Both screens display time stamps with 1-microsecond accuracy. The lower display zoom level is in the 3:55:33.60-second to 3:55:33.72-second interval. The solid line shows Pullman measurements. The thin and dashed lines are Columbus and Charlotte measurements (the power systems in the Western and Eastern United States operate at different frequencies). An additional fixed phase angle difference in the lower plot is due to the distribution level measurements of the test system being on separate three-phase terminals. In real-time mode, the data stream continuously. The lower screen is set to update at a threshold trigger, similar to an oscilloscope function. In offline mode, it is possible to scroll, zoom, and pan history for analysis of the signal content.

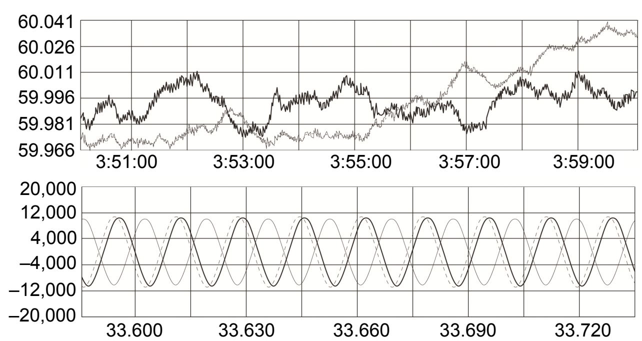

Fig. 8. Real-time visualization of synchrophasors with wide-spectrum data.

V. Select Applications

This section briefly outlines several applications of the wide-area, wide-spectrum system. The following examples are not intended to be an exhaustive list.

A. Subsynchronous Oscillations

Generators are some of the most expensive equipment in the utility asset base. Protecting generators from damage is a high priority. During certain situations, it is possible for a resonance condition to occur between generation and the network impedance. Perturbations can result in low-frequency oscillations. These subsynchronous oscillations have the potential to damage the generator. The large penetration of renewables in recent years is increasing conditions that can cause subsynchronous oscillations. It has become difficult to predict when these oscillations will occur.

Existing monitoring methods for subsynchronous oscillations (SSOs) consist of special purpose devices, custom-configured local relays, or local recorders [11]. Obtaining a system-level view with continuous recording and convenient access to the data is needed to help engineers better understand these oscillations and design the power system to minimize impact. Synchrophasor data can provide this view. However, PMU quasi-stationary class filtering is not compatible with the SSO spectral content, as shown in Fig. 9.

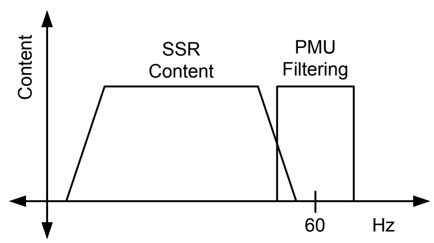

Fig. 9. Approximate comparison of SSO signal content and PMU filtering.

The difficulty in applying synchrophasors for this application demonstrates the problem with applying so much signal processing in the measuring device. It becomes challenging to use the data for applications other than those for which the original system is specifically designed. Moving the signal processing from the measuring device to the analysis application, in this case SSO, enables the application of processing specific to the application. For the case of transmitting at 1,200 samples per second, Fig. 10 shows how SSO signal content is preserved.

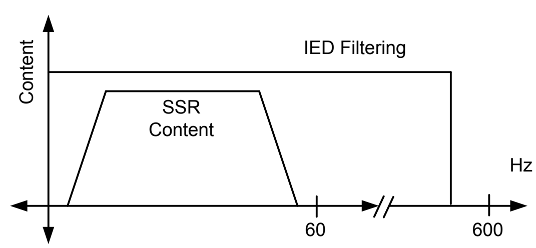

Fig. 10. Signal content comparison between SSO and IED signal processing for the wide-area, wide-spectrum system.

B. System Disturbance and Model Validation Analysis

Getting to the root cause of system-wide disturbances is important in order to apply lessons learned and improve the power system design. Typically, the data for analysis are collected at specific triggers and saved locally in IEDs. This division of functionality complicates data access, storage, organization, and searchability. Synchrophasor data have become an additional data source for analyzing events. The streaming, real-time nature of synchrophasors reduces the delay time from an event that occurs until it is available for analysis. However, the heavy filtering applied in the measuring device, such as a PMU, limits the types of events for which these data are useful. SCADA and synchrophasor data suffer from similar preprocessing problems.

The data system described in this paper provides a simple solution for system analysis. Measurements cover a wide area and are continuously available. There is no possibility of missing information from devices that did not trigger. Data are available at any time interval, not just within narrow windows around the event duration. Because all of the data are together and in a single database, it is easy to search and share the data.

C. *Power Quality*

Meters placed throughout a power system collect specialized subsets of data, such as significant voltage changes. They also collect profile data on loads. Each of these types of data requires specialized systems for collection and management. With the wide-area, wide-spectrum system, a utility metering department can apply signal processing as needed for their application needs. Furthermore, additional uses of the data become available. The metering department may receive unexpected concerns from a customer about excessive high-frequency noise. With only preprocessed data from existing systems, there is no means to further investigate problems like this without installing additional infrastructure.

D. *Geomagnetic Disturbances*

Low-frequency currents induced by solar events are a concern because of the potential for outages or equipment wear. It is known that increased harmonic signal content over a wide area can provide a leading indication of these conditions [12]. The system discussed in this paper provides the ability to analyze and monitor the full spectrum, from low frequencies to the high-frequency harmonics, over a wide geographical area and gives operators extra time to implement mitigation procedures.

VI. Conclusion

A big data system is defined in this paper, and the design and implementation of such a system to provide a means of transmitting and using sequentially sampled wide-spectrum power system measurements are discussed. Data are collected over a wide area, time-stamped, and stored for shared access. The signal processing is shifted from the measuring device to the data utilization point in the receiver. Everyone within an organization accesses the same data set. This provides consistency when working between divisions. It is no longer necessary to set up individual measuring devices and physical systems that are each tuned to the needs of an individual application.

Future work includes extending the test system to gather more information on performance, quality of service, and usability. The extension of the test system to transmission measurements, as are available at a utility, is under consideration. Finding the best data compression methods for transmission, reception, and storage is important work that will continue. Another potential area of investigation is the limitations on parallelization in application processing. Further applications are under development.

VII. Acknowledgment

The author would like to thank Matt Halladay, Daniel Miller, Nate Lovallo, Chad Yates, and Trevor Davenport for helping set up the test system.

VIII. References

[1] M. Ilić and J. Zaborszky, *Dynamics and Control of Large Electric Power Systems*. John Wiley & Sons, Inc., 2000.

[2] G. Benmouyal, E. O. Schweitzer, III, and A. Guzmán, "Synchronized Phasor Measurement in Protective Relays for Protection, Control, and Analysis of Electric Power Systems," proceedings of the 57th Annual Conference for Protective Relay Engineers, College Station, TX, March 2004.

[3] IEEE Standard C37.118.1a-2014, IEEE Standard for Synchrophasor Measurements for Power Systems – Amendment 1: Modification of Selected Performance Requirements.

[4] NERC Standard PRC-004-2.1a – Analysis and Mitigation of Transmission and Generation Protection System Misoperations. Available: http://www.nerc.com.

[5] NERC Standard EOP-004-2 – Event Reporting. Available: http://www.nerc.com.

[6] North American Electric Reliability Corporation, "Electric Reliability Organization Event Analysis Process – Version 2," July 2013. Available: http://www.nerc.com.

[7] T. Rosenberger, D. Prestwich, M. Watkins, and M. Weber, "Automated Event Retrieval Reduces Operating Costs," proceedings of the 61st Annual Conference for Protective Relay Engineers, College Station, TX, April 2008.

[8] IEC 61850-9-2, Communication Networks and Systems in Substations.

[9] D. E. Bakken, A. Bose, C. H. Hauser, D. E. Whitehead, and G. C. Zweigle, "Smart Generation and Transmission With Coherent, Real-Time Data," *Proceedings of the IEEE*, Vol. 99, Issue 6, June 2011, pp. 928–951.

[10] R. Moxley and G. C. Zweigle, "Real-Time, Wide-Area Measurements for Improved Operator Response," proceedings of CIGRE Session 2010, Paris, France, August 2010.

[11] L. C. Gross, Jr., "Sub-Synchronous Grid Conditions: New Event, New Problem, and New Solutions," proceedings of the 37th Annual Western Protective Relay Conference, Spokane, WA, October 2010.

[12] V. Albertson and B. Bozoki, "Geomagnetic Disturbance Effects on Power Systems," *IEEE Transactions on Power Delivery*, Vol. 8, Issue 3, July 1993, pp. 1206–1216.

IX. Biography

Gregary C. Zweigle (SM '13) serves as a Schweitzer Engineering Laboratories, Inc. fellow engineer and leads a research team and software development department. He holds a doctorate in electrical engineering and computer science, a master of science degree in (physical) chemistry, and a master of science degree in electrical engineering from Washington State University. He also has a bachelor of science degree in physics from Northwest Nazarene University.

Previously presented at the 2015 IEEE PES
General Meeting, Denver, CO, July 2015.

20150202 • TP6675