The Digital Tsunami

Fred L. Elmendorf and F. Russell Robertson
Grid Protection Alliance, Inc. (GPA)

ABSTRACT

This paper discusses the difficulties and effective strategies for managing Phasor Measurement Unit (PMU) data flows in real-time and for storing synchrophasor data. As many hundreds of PMUs are being deployed across the country, huge volumes of data are being generated on a continuous basis. Aggregating this streaming data from substations, to control centers, to coordinating entities, requires consideration of design factors not encountered with traditional Supervisory Control and Data Acquisition (SCADA) systems.

The U.S. Department of Energy (DOE) and the North American SynchoPhasor Initiative (NASPI) proposed a gateway device that offers a solution to the big data problem presented by synchrophasor data. This paper provides examples of data volumes from existing PMU installations and present estimates of potential volumes of data that will be realized through the installations now underway in Smart Grid Investment Grant (SGIG) projects. This paper presents the NASPInet Gateway Key Requirements and how each requirement contributes to the effective management and transport of this massive synchrophasor data resource. A detailed example of software that implements the NASPInet gateway, openPG, is included.

There are a number of options for the eventual destination of the data; forward to coordinating entities, archive for future reference, down-sample and archive, down-sample and discard, or simply discard, among other possibilities. This paper will explore a number of the most likely destinations for the data and discuss the long-term implications of each. Since data archival is an inevitable piece of the puzzle, a variety of approaches and considerations for making archival decisions will be presented.

OVERVIEW OF THE PROBLEM

Phasors represent a big data problem for grid operators – both in real time and for historical data storage, as synchrophasors are entering the production phase of technology development.

The design standard for real-time synchrophasor data systems is support for a 1/60 of a second data sampling rate. However, PMUs are typically configured to record and transfer data 30 times per second using the IEEE C37.118 protocol.

The C37.118 was designed to be efficient and minimize the burden on communication systems, which is particularly important for substation-to-control center communication. As shown in Table 1, the bandwidth required to transport phasor data in C37.118 from a substation is small – less than 60 kbits/sec for 2 PMUs.

However, use of C37.118 is problematic for large numbers of PMUs, as is the case within a control center or among multiple control centers where data is aggregated from many PMUs. Latency increases from parsing large protocol frames. In addition, there are limits to frame sizes within the commonly used IP/UDP stateless protocol.

Samples per Second	Number of PMU's			
	2	**10**	**40**	**100**
30	57	220	836	2,085
60	114	440	1,672	4,170
120	229	881	3,345	8,340

Values in kilobits per second (kbits/s)
Assumes 20 measuremetns per PMU (16 are used for 8 phasors)

Table 1. Approximate Bandwidth (Kbits/sec) as a Function of PMUs and Sampling Rates[1]

Storage of phasor data also represents a big data problem. SCADA data archival techniques cannot easily be extended to phasor data. Most SCADA historians are efficient due to compression techniques such as swinging gate compression where data is only stored when it changes significantly. It is not uncommon for SCADA data retention periods to be 7-years. Assuming 16 measurements per PMU, 7 years of data storage results in 6.6 trillion points or about 1.5 TB of storage for each PMU. For a typical large deployment (100 PMUs) this can result in a requirement for 200 TB of storage over the 7 years – more than two orders of magnitude more storage than required for SCADA data.

[1] *Real-Time Application of Synchrophasor for Improving Reliability* (RAPIR), NERC Operating Committee, October 18, 2010

The Grid Protection Alliance (GPA) is a not-for-profit corporation formed to facilitate and support the development and deployment of comprehensive electric energy system-wide solutions. GPA focuses on improvements to electric energy resiliency and assurance of grid reliability and availability. GPA develops free, open source software applications to meet the specific demands of real-time data collection and storage to support electric grid operations and planning. For more information, see: http://www.gridprotectionalliance.org/

BACKGROUND

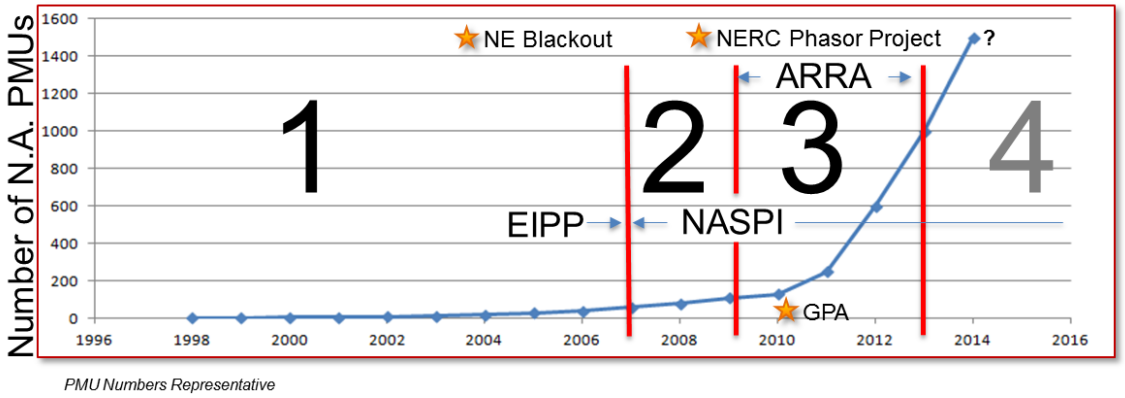

Figure 1. The Epochs of Synchrophasor Deployment

Use of wide-area measurements, where the footprint of measurements can easily double the number of phasor measurements made internally, compound both the real-time and historical big data problem. As seen in Figure 1 above, the number of phasors installed and available for use has grown rapidly over the past 3 years. This growth curve has been divided in four phases.

1. **R&D Phase** – The early years included a small group of utility innovators such as the Bonneville Power Administration (BPA) and the Tennessee Valley Authority (TVA), in addition to directed DOE research funding. The academic community played a key role in the advancement of the synchrophasor technology, and the Eastern Interconnection Phasor Project (EIPP) was created by DOE and lead by Lawrence Berkeley National Laboratory to champion the adoption of the synchrophasor technology.
2. **Consensus Building Phase** – Following the North American Energy Reliability Corporation (NERC) technical analysis of the August 14, 2003 blackout in the Northeast[2], the value that synchrophasors bring to the grid became apparent, as a redundant tool for wide area visibility and as a data source to conduct forensics following disturbances. NERC assumed ownership of the efforts to accelerate adoption of phasor technology and provided funding support for NASPI through a 5-year project (2009-2013). Efforts in this phase where focused on building consensus for implementation standards and the availability of better, more rigorously tested vendor products.
3. **Stimulus Phase** – SGIGs where made under the American Reinvestment and Recovery Act in 2009. These grants called for the installation of 950 PMUs by the end of fiscal year 2013, through 12 different projects, with 341 being located in the Western Interconnection. During this period, grid operators not receiving SGIG funding also began making significant investments in synchrophasors. These investments were either to piggy-back on SGIG projects or as separate projects to address emerging power system operations issues resulting from a rapid increase in dispersed generation penetration.
4. **Production Phase** – The SGIG funding has accomplished the objective of "acceleration of the adoption of phasors." In addition, during the Stimulus Phase, NERC has strengthened its role as a compliance monitor, while performing its duties as the Electric Reliability Organization (ERO)[3], and is working to divest itself from the ownership of systems and tools that are necessary for grid operations. At the end of 2013, NASPI leadership and support will be transitioned to the Consortium for Electric Reliability Technology Solutions (CERTS) organization, which is sponsored by DOE. NERC staff will be working to make their standards setting committees aware of the value of including synchrophasor data use requirements.

[2] *Technical Analysis of the August 14, 2003 Blackout: What Happened, Why, and What Did We Learn?*, Report to the NERC Board of Trustees, July 13, 2004

[3] The Energy Policy Act of 2005 directed FERC to establish an Electric Reliability Organization to establish and enforce electric reliability standards

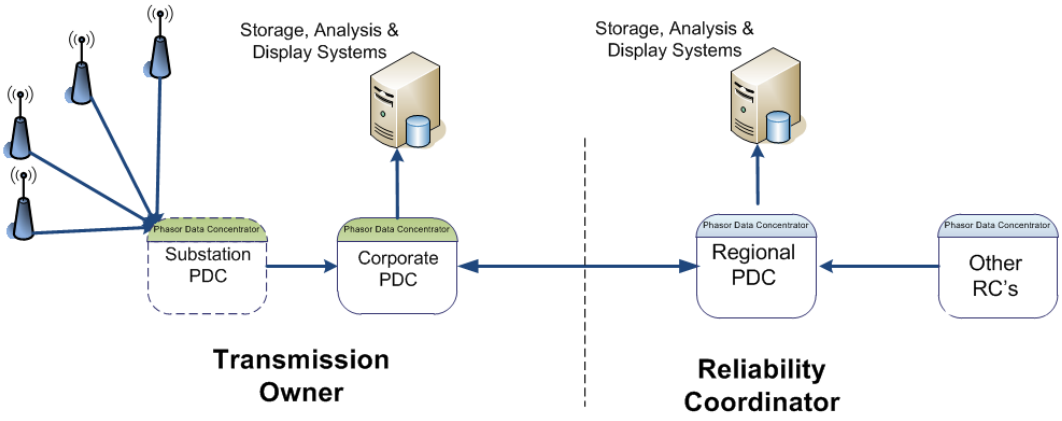

Figure 2. Typical Synchrophasor Data System Design

There are currently many thousands of relays and digital fault recorders deployed on the electric grid that can be used to provide PMU functionality. In the most modern of these devices, phasor measurements are already enabled and older devices can be updated to add PMU functionality through firmware upgrades. It is estimated that from one vendor alone[4], more than 183,000 devices are available to provide phasor data in North America today.

SYNCHROPHSOR DATA ARCHITECTURE

The core technology in today's synchrophasor data architecture is the phasor data concentrator (PDC). A PDC correlates phasor data by its time tag and then broadcasts the combined, or synchronized, data to other systems. A PDC must buffer its input data to accommodate the differing times of data delivery from individual PMUs.

As shown in Figure 2, there can be up to three levels of data concentration in typical synchrophasor data architecture. Each layer provides the service of concentrating phasor data for use by applications at that layer. However, each PDC also adds latency – at increasing amounts as wait times are added for data delivery across larger and larger footprints.

In this typical architecture, the frame-based IEEE C37.118 protocol is used at all levels. Security is currently provided through encryption at the network layer.

Most implementations of synchrophasors today are not considered to be Critical Infrastructure and therefore are not subject to requirements for strong isolation within electronic security perimeters. Even so, many entities are adding PDCs to this architecture to provide an isolation layer between their internal systems and external ones.

The transport layer between a regional PDC and other RC's is targeted to be provided by the new NERCnet MPLS cloud.

NASPInet AS A SOLUTION

NASPInet [5] was envisioned in 2009 to provide the protocols and services for effectively exchanging real-time and historical phasor data among data providers and data consumers. NASPInet was not designed as a physical network. Rather, it is an architecture for exchanging data once transport is provided, for example using NERCnet[6]. At the core of NASPInet is its data bus, which was envisioned to link measurement devices directly with applications through the use of a "Phasor Gateway" that would place data on the bus and retrieve data from it. The conceptual architecture for NASPInet is shown in Figure 3,

The NASPInet gateway specification defined the latency requirements for the five data classes which must be accommodated by synchrophasor data systems as shown in Figure 4. While feedback control (Class A) is not yet on the horizon, work is progressing on systems that use phasor measurements for feedforward control (Class B) where the overall latency requirement from measurement to control action is specified at 100 milliseconds. Meeting the NASPInet latency requirements for both the Class B and Class C data types becomes problematic without the implementation of gateways to parse data from PMUs and quickly route it to the consuming applications.

[4] Dr. Edmund O. Schweitzer, NASPI Meeting Presentation, February 29, 2012

[5] Phasor Gateway Specification for NASPInet, Yi Hu, Quanta Technology, May 29, 2009

[6] As required for data exchange of SCADA data by NERC Standard COM-001-1.1, Telecommunications

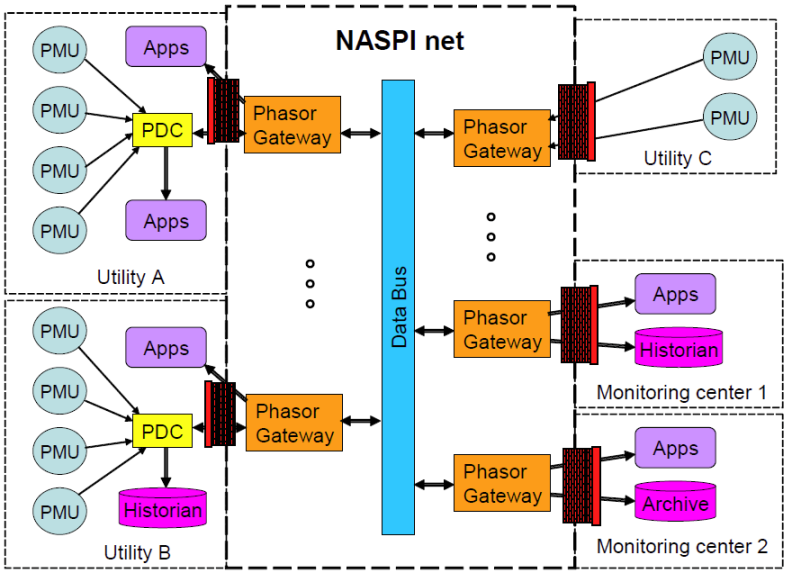

Figure 3. NASPInet Conceptual Architecture

Limitations of NASPInet - The NASPInet technical specification did not address compliance issues unique to maintaining certification as a NERC operating entity. Therefore, while the NASPInet technical specification provides valuable insights for implementation of synchrophasor systems within an individual operating entity's compliance footprint, it is not easily implementable to exchange data among multiple operating entities given current compliance requirements.

To work around this problem, in 2010 the University of Illinois at Urbana-Champaign(UIUC)[7] proposed a tiered NASPInet architecture that preserved the conceptual architecture of NASPInet shown in Figure 3, while abandoning the concept of a "universal bus" and allowing data to be retained and managed in regional hubs by the Transmission Operator and/or Reliability Coordinator. In this Illinois model, gateways serve as access points for exchange of data among these phasor data hubs.

The openPG AS A SOLUTION

The openPG solves the problem of large real-time phasor data exchange. Funded by NERC and the Entergy SGIG project, the openPG has been under development at GPA since January 2011 as a device that is designed to have value for immediate implementation, while it also meets many of the NASPInet technical requirements. The production version of the openPG 1.0 has been released and is in production service exchanging phasor data between TVA and Entergy. Later this year, the openPG will undergo extensive security testing by UIUC as part of the Entergy SGIG project.

	Description	Qty kB/rqst	Frequency qty/sec	Latency	Geography
A	Feedback Control	0.2kB	30 to 60	0.05s	Sub-regional
B	Feedforward Control	5kB	1 to 30	0.1s	Regional
C	Visualization	10^2kB	10^{-1} to 1	0.25s	Super-regional
D	Post-Event	10^6kB	10^{-6} to 10^{-4}	10^2s	National
E	Research	10^8kB	10^{-7} to 10^{-5}	10^4s	National

Figure 4. NASPInet Data Class Definitions

A typical openPG installation is shown in Figure 5, with the openPG serving as a security device on the perimeter of the control center phasor data infrastructure. The openPG is designed to integrate easily with the openPDC but can function with any vendor's PDC.

As seen in Figure 5, the openPG accepts inputs from multiple devices such as PMUs, PDCs, or Frequency Data Recorders (FDRs). It produces outputs for use by applications and systems with the internal infrastructure. Finally, the openPG exchanges data securely with other trusted gateways.

[7] Exploring a Tiered Architecture for NASPInet, Bobba, Heine, Khurana, and Yardley, IEEE PES Conference on Innovative Smart Grid Technologies, 2010

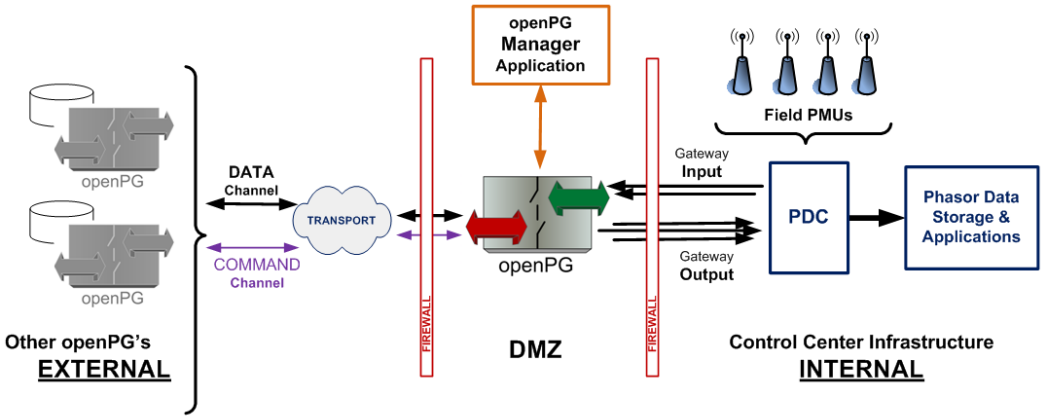

Figure 5. Typical Phasor Gateway Implementation

The openPG breaks the frame-based IEEE-C37.118 protocol down into small measurement-point based packets and makes them available for publication by data owners. The benefits of the openPG include significantly improved security, lower latency, greater scalability, ease of configuration, and lower configuration costs.

Security - The openPG is purpose-built with a low attack surface for the security perimeter. As such, it is designed to be deployed within a network DMZ. The highest level of data confidentiality is achieved through best practice encryption. Data is only exchanged with other gateways that have established trusted relationships, using out-of-band key exchange to protect the security of reliability data, and market sensitive bulk electric system (BES) data.

Low Latency – The openPG forwards phasor data to other authorized gateways on receipt, without time-alignment or other concentration delays.

Scalability – Through the use of a small, point-based data packet (9-bytes), the openPG's data exchange protocol easy scales up without limitations imposed by IP protocols. In addition, it is much more efficient in TCP/IP deployments since retransmission only involves small packets.

Easy Configuration at Low Cost – The dominate business driver for installation of the openPG is lowered configuration costs.

There are low operational costs to publish phasor data since an openPG owner authorizes phasor data measurements as available for subscription by other trusted gateways through a simple check-box process. The authorized subscriber can then pick and choose among these points to select the individual points they will receive without any action required by the publisher.

There are low costs to subscribe to phasor data through a gateway since all meta data about the points available for subscription are automatically made available to the subscriber. There is no need for spreadsheets to manage multiple keys as in the case for ICCP today.

The costs are low to integrate data from one entity's namespace to other's, since the openPG can perform namespace translation. Every entity establishes its own rules for naming and identifying measurement points. Generally these rules are driven by the requirements of major tools in the infrastructure which can be the SCADA/EMS, work management, and planning tools, among others. In the openPG, multiple common names can be used to reference phasor measurements and so that these names can be assigned and changed as needed by each gateway user. As a gateway owner makes changes to measurement metadata, these changes are made known to subscribing gateways.

The operating costs are low to meet CIP requirements since the openPG logs all configuration changes and provides comprehensive operating performance logs and alarms.

The high-level functional requirements of the openPG are to:

- Reliably exchange high-sample rate signal values and timestamps (measurements) with other gateways so that this information moves between each owner's PDCs with minimum time delay.
- Enable gateway administrators to easily select the measurement points which are to be made available to owners of other gateways.

NASPInet Gateway Key Requirement	openPG
Serve as the sole access to the Data Bus	**Aligned.** The openPG is the sole access point for exchange of phasor information with other openPGs with establishing trusted gateway unions as a key design element.
Facilitate and administer registration of PG_REQUESTER's PMU, PDC, and signals	**Aligned.** The openPG configuration database contains metadata on the phasor measurements. An openPG downloads the metadata for points it is authorized to receive from other openPGs.
Facilitate and administer the subscription and publishing of phasor data	**Aligned.** The openPG supports both publishing and subscription of phasor data.
Monitor data integrity	**Aligned.** Each block of points transferred by the openPG contains checks to assure data integrity.
Manage traffic priority through the PG according to service classes	**Not Aligned.** All data exchanged by the openPG is in the single service class of "critical real-time data".
Provide logging of data transmission, access controls, and cyber security for analysis of all anomalies	**Aligned.** The openPDC produces both configuration and operational logs
Provide application programming interfaces (APIs) for interfacing with PG_REQUESTER systems and applications.	**Aligned.** The openPG is extensible by the user through development of new input and output adapters.

Table 2. NASPInet Key Requirements vs. openPG

- Support encrypted communication among gateways as well as implement features to minimize bandwidth requirements for gateway-to-gateway data exchange.
- Enable gateway administrators to easily select the points that they choose to consume; for example, a subset of the points made available to them from other gateways.
- Utilize standard communications, networking, and server hardware.
- Be easily extensible to support the development of custom interfaces to the gateway owner's internal infrastructure and/or new phasor data protocols.
- Detect, log, and alarm on communications issues.
- Be implementable as a high-availability solution that can meet NERC CIP compliance requirements and serve as an aid to minimize CIP non-compliance risk.

These functional requirements do not include data concentration since a gateway must move data with minimal latency. Concentration is provided by a PDC that time-aligns data from both internal PMUs and external data sources via the gateway.

openPG and NASPInet - Table 2 provides the NASPInet Phasor Gateway key requirements. The openPG meets all of these requirements but one – "Manage traffic priority through the PG according to service classes." The openPG supports just one service class – "critical real-time data." This means that the openPG cannot currently exchange lower priority historical phasor data simultaneously with real-time data. The openPG is optimized for real-time data exchange.

PROBLEM: HISTORICAL BIG DATA

The problem of management of large volumes of historical phasor data is solved through:

- A tiered data retention strategy
- Efficient storage of high-resolution data

Three tiers of phasor data storage are recommended: (1) full-resolution data, (2) event data, and (3) phasor data down-sampled to SCADA periodicity.

Tier 1 - There is limited business value of saving large quantities of high-resolution data for a "blue sky" power system. However, there is value in the long-term storage of high-resolution data from interesting periods. These periods could be short, such as the few seconds surrounding a fault, or long, such as the many hours leading up to a major blackout or disturbance.

Therefore, it's recommended that blue sky full-resolution data only be saved sufficiently long to assure that it's not interesting. This storage layer is classified as Tier 1 with a recommended retention period of 1 year. Tier 1 storage has the following attributes:

- Fast access for investigation of disturbances
- Immediately available following an event
- Available for disturbance analysis within the days following the event

Tier 1 data must be stored efficiently. Historians provide this service for process control data. GPA is

working on a free, open-source version of the openHistorian (targeted for release in late 2012) that uses lossless compression to reduce the physical storage requirements by a factor of 2 or more as compared to commercial historians.

Tier 2 - The interesting data from Tier 1 would be migrated to Tier 2 data storage that would house event data. This data often has long-term value and can be useful for a decade or more.

Tier 3 – As a long-term data store of phasor data for use by planners, it is recommended that an entity's current SCADA data archive be utilized to store down-sampled phasor data (e.g., every 4 seconds) using the same compression techniques and retention period as for SCADA data.

CONCLUSION

Gateways offer value when implemented today in moderate-sized synchrophasor systems and will be necessary in the future, as synchrophasor data systems grow. Gateways provide required security isolation and reduce operational costs by creating a hardened security buffer between critical internal and external systems and encryption to protect the confidentiality of the data. Ideally, a gateway will only exchange data with other gateways where a trusted union has been established, automatically discover available measurements from other gateways and allow selective subscription, provide effective administrative tools to manage published points, and support strong encryption for data exchanged.

Using a gateway, rather than standard protocols, such as IEEE C37.118, has the advantage of improving security and reducing bandwidth by only exchanging needed measurement points, simplified configuration management through automated metadata exchange and the ability to easily rename phasor data points, reduced latency for most phasor data since the concentration step of a PDC is not required, and increased scalability and extensibility.

Using a carefully designed and tiered phasor data retention strategy that addresses the needs of each of the potential users--such as system operators, operations engineers, and system planners--and employing appropriate data stores to meet each category of need, makes it possible to extract the most value from the data without being overwhelmed by the volume.

By understanding the concepts described in this paper and employing them through the use of relatively new tools such as gateways and historians, you can turn the potential "Digital Tsunami" of phasor data into the "Perfect Wave" of new information to support a smarter grid.