

Data Compression for Next Generation Phasor Data Concentrators (PDCs) in a Smart Grid

Erwan Olivo, Mitch Campion, Prakash Ranganathan

Department of Electrical Engineering, University of North Dakota, Grand Forks, ND, USA

Email: prakash.ranganathan@engr.und.edu

How to cite this paper: Olivo, E., Campion, M. and Ranganathan, P. (2016) Data Compression for Next Generation Phasor Data Concentrators (PDCs) in a Smart Grid. *Journal of Information Security*, 7, 291-296. <http://dx.doi.org/10.4236/jis.2016.75024>

Received: April 16, 2016

Accepted: September 11, 2016

Published: September 14, 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The storage space and cost for Smart Grid datasets has been growing exponentially due to its high data-rate of various sensor readings from Automated Metering Infrastructure (AMI), and Phasor Measurement Units (PMUs). The paper focuses on Phasor Data Concentrators (PDCs) that aggregate data from PMUs. PMUs measure real-time voltage, current and frequency parameters across the electrical grid. A typical PDC can process data from anywhere ten to forty PMUs. The paper exploits the need for appropriate security and data compression challenges simultaneously. As a result, an optimal compression method ER1c is investigated for efficient storage of IREG and C37.118 timestamped PDC data sets. We expect that our approach can greatly reduce the storage cost requirements of commercial available PDCs (SEL 3373, GE Multilin P30) by 80%. For example, 2 years of PDC data storage space can be easily replaced with only 10 days of storage space. In addition, our approach in combination with AES 256 encryption can protect PDC data to larger degree as per National Institute of Standards and Technology (NIST) standards.

Keywords

Compression, PDCs, Data Security, Smart Metering, Smart Grid

1. Introduction

The number of PMU deployment is growing exponentially in North America, and hence the amount of data to be stored even for a short-period is large [1]. The Phasor Data Concentrators (PDCs) have limited storage space (estimated 100 GB) and vary depending on vendors. So, data have to be stored in alternative formats such as DVDs or hard drives. The main idea of this paper is to optimize the storage of such data sets that has large data rates (~30 samples/second).

This optimal reduction in file or storage sizes of PDC data can help in reducing sto-

range cost, efficiently secure and organize SQL queries and retrieval with faster download-time. The parameters (compression rates, data retrieval time) have considered as a benchmark performance metrics at super PDC level.

This paper specifically uses a two-level of compression for PDC data with AES 256-bit encryption (see **Figure 1**). There are very limited literatures on PMU data compression. For example M. H. H. Wen and O. K. Li in [2] discuss a simple compression unit installation in the grid between the PMUs and a PDC with no security guaranteed. In [4], F. Zhang, L. Cheng, X. Li, Y. Sun, W. Gao and W. Zhao proposed a compression approach for real-time PMU data for Wide-Area Measurement Systems (WAMS). They used a swing door trending compression method at the PMU level for reducing the amount of data in the network. The security of PMU data has been not well studied and limited literatures do exist. In fact, the IEEE C37.118 standard does not give proper directions or solutions on how PMU data can be secured. In [3], authors discuss how decision trees approach for synchronized phasor measurements can improve the security of grid from voltage collapse. This paper discusses a novel and optimal compression algorithm known as Erwan-Ranganathan compression version (ER1c.v.1) method to process streaming PDC data sets. We define compression ratio (CR) by the following Equation (1):

$$\text{Compression Ratio (CR)} = \text{initial size} / \text{final size} \quad (1)$$

For example, a file with initial size of 1200 KB can be reduced to 300 KB through compression, and will have a compression ratio of 4. The higher the ratio is, the better the compression method and for efficient storage will be. The PMUs data are a measurement of voltage, current, frequency and phase angle of voltage and current marked with a timestamp. The timestamp follows rules as per IEEE C37.118 or IIRIG 200-04 standards. IIRIG-B Timestamp format is a standard for encoding timestamp information for PMU data. The specific standard followed for data is IIRIG STANDARD 200-04 [4]. An IIRIG-B frame uses 74 bits to encode information in Binary Coded Decimal (BCD). This frame is divided into 5 parts for the *date* (39 bits). In addition, there are 35 more bits for *time quality*, *leap second*, *leap year* and *local offset*. This data frame is longer than

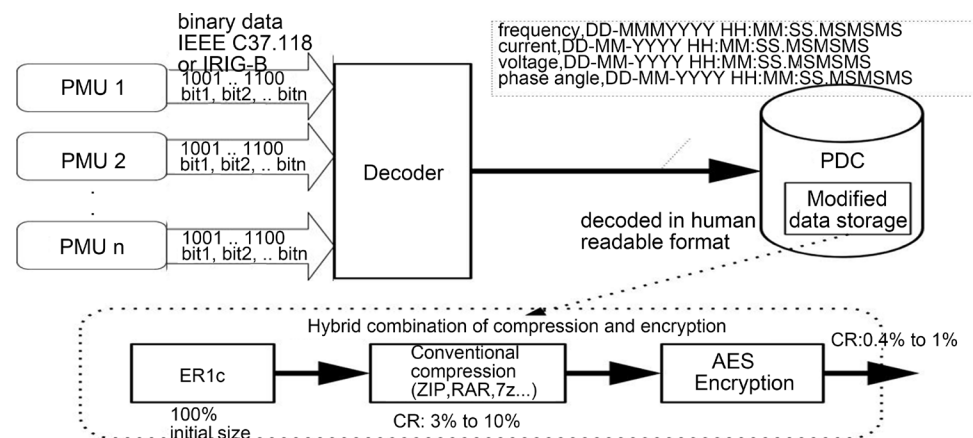


Figure 1. Secure Data Storage Architecture for Streaming PMU data in PDC's.

C37.118 standard, because of the index position marker, which use some bits too. This IRIG based PMU timestamp allows a better implementation for a human readable application, because the time corresponds to real-time.

2. Optimal Encoded PMU Measurements

All PMU measurements are encoded in ASCII characters. Mostly, the parameters in PMU measurements (f , v , and i) and some section of time-stamp (*date, month, or day*) remain same and redundantly repeated, and thus encoded value do not have to be changed frequently. For example, frequency (f) should be at 60 Hz, the voltage v at 300 kV and the current i at 500 A. Only during the event of any anomalies or changes in measurements, any altered values from this value need to be correctly encoded. Any such sudden change in data pattern can easily be tracked in our approach. Due to limited encoding fields (FRACSEC, SOC) on the time-stamp, the proposed compression technique is able to reduce the original file sizes of PDC's data. The results of our compression is discussed in the next section. PMUs considered for this investigation has a data rate of 30 samples per second. So, the number of samples remain constant for every second. These data rate patterns are not detected by the classical compression methods (Huffman coding, Dictionary coders or prediction by partial matching [5]-[7]), and thus compression of PMUs data can yield better lossless compression ratios compared to other classical software compression.

Figure 1 shows the proposed architecture for efficient data storage for PMU data at PDC level. The compression and encryption discussed here is for the PDC data sets. ER1c is a developed software program that compress PDC data in an un-encrypted format type (human readable form). It uses the strength of both the Huffman coding, dictionary coders, Prediction by Partial Matching (PPM) algorithm [7]. ER1c itself is an optimized program that yield savings on storage cost, and compression efficiency compared to commercial available PDCs. The program optimizes any redundant data fields, and automate sections of code for better compression ratio. ER1c program reads PMU data frame by frame, but not as a character by character. A frame is composed of both a timestamp and measurement as seen below. The measurement parameter shown here is a frequency of the grid, as an example.

Frame 1: 05-Dec-2015 17:41:36.666, 59.10 Hz.

Frame 2: 05-Dec-2015 17:41:36.700, 60.01 Hz.

The focus restricts only to redundant time-stamp information in the PMU data. The classical compression methods such as dictionary coders could compress all fields like date, hour, minute, seconds and milli-second. Thus, there exist a possibility of compressing the redundant data fields such as the date, day, hour, minute and seconds to certain period of duration. These are repeated information which are encoded again line by line. This is an un-necessary process and results in wastage of CPU time and storage costs. To avoid this problem, our ER1c program will capture the initial time-stamp information only once with its date, day, hour, minute, and second. Any repeated information will not be encoded or compressed again to save storage cost. The

only varying and non-repeated data field is the milli-second (*ms*) information, which will be encoded continuously. See the pseudo code shown in **Example 1**.

The ER1-c program can also detect and check whether the duration between each measurement is consistent or not. In other words, it would check the number of samples per second for data validation depending on PMU type. We assumed the PMU used in the PDC data set has a data rate of 30 samples per second. ER1c is a simple and optimal program to detect time-stamp errors. For example, if a second time-stamp (*frame 2*) following a first time stamp (*frame 1*) are not respecting duration between each measurement, the ER1c program will catch these duration or sampling errors.

3. Comparison of Compression Ratios (CR)

Table 1 and **Table 2** show the compression ratios using various methods for frequency (*f*), current (*c*), voltage (*v*) and phase angle (*ph*) data, and **Figure 2** show compression speeds as number of PMU scales. The results obtained using our approach 10× better than other conventional compression techniques. The more the data to compress, the better the utilization of ER1c compression is good. It is observed that ER1c offer better compression ratio (almost 3× better) than 7z, rar, zip, zipx, and uha methods.

The decompression process is shown in **Figure 3**.

```

For {
  • → all redundant time-stamp fields except milli second (ms) field
    check date (D), day (dy), hour (h), min (m), sec (s) for period Δ

  • → Perform the ER1c
    Skip the Data encoding for repeated fields
    Until non-repeated data fields; Compute CR, run-time
    Automate the process for repeated data fields

  • → Do AES-256

End }

```

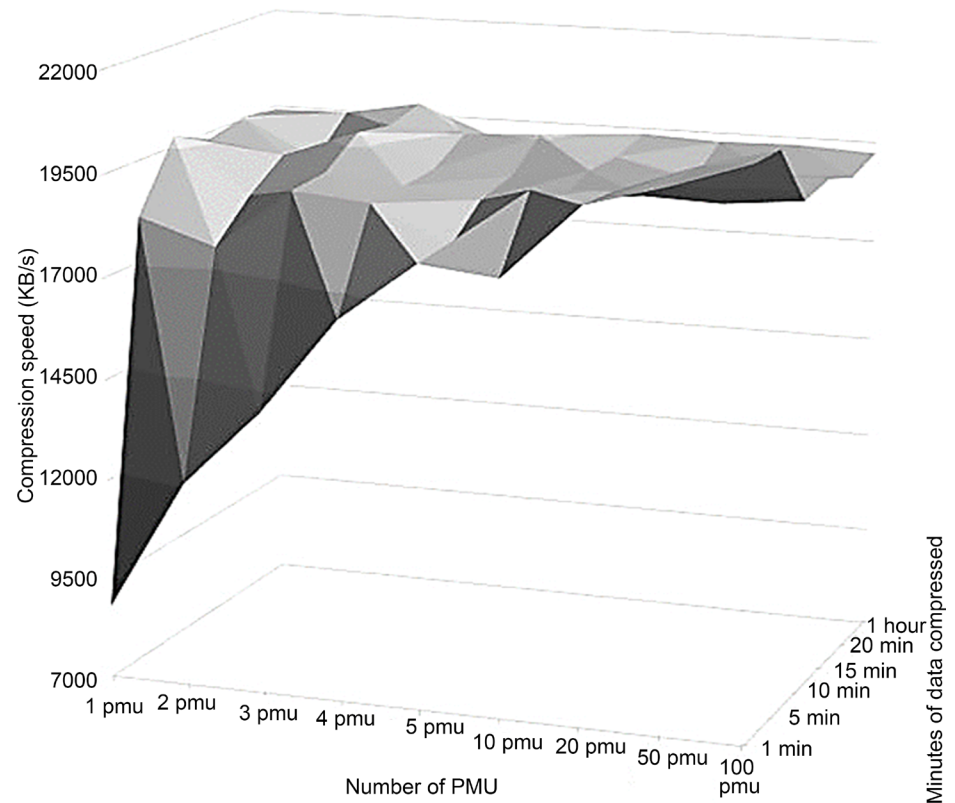
Example 1. Pseudocode for handling redundant time-stamps.

Table 1. Compression ratio for 5 minutes dataset.

Compression Type	CR _f	CR _c	CR _v	CR _{ph}
7z	17.15	20.04	19.88	19.88
rar	10.11	10.72	10.65	11.79
zip	10.16	6.25	7.13	5.83
zipx	18.66	11.72	12.84	11.43
uha	11.92	20.83	22.22	20.88
ER1c	30.69	11.40	10.97	11.69
ER1c + 7z	160.08	81.74	96.97	86.66
ER1c + rar	173.76	80.26	93.43	85.18
ER1c + zip	167.89	78.73	94.98	84.61
ER1c + zipx	184.49	81.33	100.89	90.43
ER1c + uha	180.63	83.43	109.48	90.93

Table 2. Compression ratio for 20 minutes dataset.

Compression Type	CR _f	CR _C	CR _v	CR _{ph}
7z	18.05	24.45	23.87	24.94
rar	9.99	13.53	12.18	15.02
zip	10.21	6.27	7.14	5.84
zipx	19.12	13.30	13.89	13.62
uha	12.02	26.18	27.03	29.24
ER1c	29.11	11.38	10.95	11.65
ER1c + 7z	271.50	192.74	223.13	199.50
ER1c + rar	270.49	181.61	197.09	190.01
ER1c + zip	241.18	176.51	193.00	182.05
ER1c + zipx	274.62	211.95	243.84	229.28
ER1c + uha	298.62	206.44	250.60	220.22

**Figure 2.** Compression speed with ER1c.**Figure 3.** Decoding process of PMU data using ER1-c.

4. Conclusion

An optimal compression method for streaming time-stamped data sets for PDCs is presented. The proposed approach is suitable at PDC level for efficient data storage, retrieval and post-event analysis. The preliminary results indicate that ER1c with combination from existing compression techniques can yield better compression ratio. We expect that our approach can greatly reduce the storage cost requirements of commercial available PDCs to 80%. For example, 2 years of PDC data storage capacity can be easily replaced by only 10 days of capacity. In addition, our approach with combination of AES 256 encryption can protect PDC data with a greater confidence and thus increase the security of growing big data sets in smart grid network.

Acknowledgements

This work is made possible through UND's RD & C (21418-4010-02294).

References

- [1] Bose (2010) Smart Transmission Grid Applications and Their Supporting Infrastructure. *IEEE Transactions on Power Systems*, **1**, 1949-3053.
- [2] Wen, M.F. and Li, V.O.K. (2015) Optimal Phasor Data Compression Unit Installation for Wide-Area Measurement Systems—An Integer Linear Programming Approach. *IEEE Transactions on Smart Grid*, 1949-3053. <http://dx.doi.org/10.1109/tsg.2015.2503425>
- [3] Khatib, A.R., Nuqui, R.F., Ingram, M.R. and Phadke, A.G. (2004) Real-Time Estimation of Security from Voltage Collapse Using Synchronized Phasor Measurements. *IEEE Power Engineering Society General Meeting*, **1**, 582-588.
- [4] IRIG Standard 200-04 2004. <http://www.irigb.com/pdf/wp-irig-200-04.pdf>
- [5] Kavousianos, X., Kalligeros, E. and Nikolos, D. (2007) Optimal Selective Huffman Coding for Test-Data Compression. *IEEE Transactions on Computers*, **56**, 1146-1152.
- [6] Bhadade, U.S. and Trivedi, A.I. (2011) Lossless Text Compression using Dictionaries. *International Journal of Computer Applications*, **13**.
- [7] Cleary, J.G. and Ian, H. (1984) Witten Data Compression Using Adaptive Coding and Partial String Matching. *IEEE Transactions on Communications*, **32**, 396-402. <http://dx.doi.org/10.1109/TCOM.1984.1096090>



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>