Signal-Preserving Compression for Large Synchrophasor Measurement Data Sets

Zeb Tate  
University of Toronto  
zeb.tate@utoronto.ca

Ray Klump  
University of Illinois  
klump@illinois.edu
Project Overview

• Goal
  – Signal-preserving compression of synchrophasor data archives

• Approach
  – Pre-compression steps tailored to power system domain

• Results
  – Using simulated and measured data
  – Compression ratios up to 20:1

• Future work
Why Bother Compressing Data?

• 1 PMU taking 20 measurements at 30 sps
  – 4,800 bytes / second
• 40 PMUs reporting to a PDC
  – 15.5 GB / day
• 10 PDCs reporting to a centralized data repository (e.g., SuperPDC)
  – 56 TB / year
• Less storage space means
  – Less hardware (reduced budget)
  – More redundancy (reduced data loss)
Key Principles for Synchrophasor Compression

• Knowledge of grid behavior should inform the compression of synchrophasor data
  – For example, under normal operating conditions, buses exhibit coherency in frequency
    • Similar to the coherencies in neighboring pixels of an image, which is the basis for PNG, JPG, etc.

• Compression of data archives must not reduce signal (information) content
Initial Approach

Two step process combining domain knowledge with high-performance, generalized compression tools

1. Pre-compression using knowledge of grid behavior
   - Exploit temporal and spatial correlation of signals

2. General compression tools
   - Apply state-of-the-art, common off-the-shelf (COTS) compressors such as bzip2, LZMA
Pre-Compression for PMU Data

• Idea: Use two types of signal coherency
  – Temporal
    • Cycle-to-cycle changes are usually small
  – Spatial
    • Signals from proximate buses evolve together

• Implementation: record sample-to-sample differences in each signal referenced to the sample-to-sample difference of a reference signal
Slack-Referenced Encoding

Input data ($p$ data sets, with $q$ data points per set)

$\mathcal{M} = \{m_1, m_2, \ldots, m_p\}$

$m_i = \{m_i(\Delta T), m_i(2\Delta T), m_i(3\Delta T), \ldots, m_i(q\Delta T)\}$

Difference Encoding (DE)

$m_i^{DE} = \{m_i^{DE}(\Delta T), m_i^{DE}(2\Delta T), \ldots, m_i^{DE}(q\Delta T)\}$

$m_i^{DE}(k\Delta T) = \begin{cases} m_i(\Delta T) & \text{if } k = 1 \\ m_i(k\Delta T) - m_i((k - 1)\Delta T) & \text{if } k \in [2, q] \end{cases}$

Slack-Referenced Encoding (SRE)

$m_i^{SRE} = \{m_i^{SRE}(\Delta T), m_i^{SRE}(2\Delta T), \ldots, m_i^{SRE}(q\Delta T)\}$

$m_i^{SRE}(k\Delta T) = m_i^{DE}(k\Delta T) - m_{REF}^{DE}(k\Delta T)$
Compression Stage

• Use well-known, high-performance lossless techniques
• Chosen implementations:
  – DEFLATE
  – BZip2
  – LZMA
# Test Cases

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
<th>Duration</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Sim</td>
<td>Simulation</td>
<td>37-bus system with line outage</td>
<td>10 minutes</td>
<td>432.7 KByte</td>
</tr>
<tr>
<td>Quiescent</td>
<td>TVA</td>
<td>Data from 8 PMUs during quiet period</td>
<td>10 minutes</td>
<td>351.6 Kbyte</td>
</tr>
<tr>
<td>Event</td>
<td>TVA</td>
<td>Data from 13 PMUs during an interval with a sizable line outage</td>
<td>60 minutes</td>
<td>5.36 Mbyte</td>
</tr>
<tr>
<td>Large Sim</td>
<td>Simulation</td>
<td>7,400 PMUs during a 3-phase-to-ground fault</td>
<td>10 seconds</td>
<td>67.9 Mbyte</td>
</tr>
</tbody>
</table>
### Compression Results

**Voltage Magnitudes**

**Compression Ratios of Voltage Magnitudes for Different Test Cases Using Various Compression Techniques**

<table>
<thead>
<tr>
<th>Case</th>
<th>Max Compression with Pre-Processing (SRE, DE, or SD)</th>
<th>Max Compression w/o Pre-Processing</th>
<th>Improvement by Pre-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Sim</td>
<td>14.35 (Bzip2 + SRE)</td>
<td>5.64 (LZMA)</td>
<td>2.54</td>
</tr>
<tr>
<td>Quiescent</td>
<td>2.54 (LZMA + SRE)</td>
<td>2.25 (BZip2)</td>
<td>1.13</td>
</tr>
<tr>
<td>Event</td>
<td>2.50 (LZMA + DE)</td>
<td>2.21 (BZip2)</td>
<td>1.13</td>
</tr>
<tr>
<td>Large Sim</td>
<td>3.03 (LZMA + DE)</td>
<td>2.06 (LZMA)</td>
<td>1.47</td>
</tr>
</tbody>
</table>
## Compression Results

### Phase Angles

<table>
<thead>
<tr>
<th>Case</th>
<th>Max Compression with Pre-Processing (SRE, DE, or SD)</th>
<th>Max Compression w/o Pre-Processing</th>
<th>Improvement by Pre-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Sim</td>
<td>10.37 (LZMA + SRE)</td>
<td>2.33 (LZMA)</td>
<td>4.45</td>
</tr>
<tr>
<td>Quiescient</td>
<td>2.97 (LZMA + SRE)</td>
<td>1.95 (LZMA)</td>
<td>1.52</td>
</tr>
<tr>
<td>Event</td>
<td>3.94 (LZMA + SRE)</td>
<td>2.47 (LZMA)</td>
<td>1.60</td>
</tr>
<tr>
<td>Large Sim</td>
<td>2.99 (LZMA + DE)</td>
<td>1.85 (LZMA)</td>
<td>1.62</td>
</tr>
</tbody>
</table>
Impacts of Random Noise on PMU Data Compression

• Bad news: random noise in signals not compressible
• Good news: we don’t care about storing/retrieving random noise
• Solution: Identify and remove noise before (pre-)compression
Signal-Preserving Data Compression

1. Read in raw PMU data sets
2. Extract bit streams from each data set and apply randomness tests
3. Apply SRE pre-compression
4. Mask off bits that meet/exceed threshold
5. Apply standard (e.g., bzip2, LZMA) compression
6. Store compressed data
Identification of Random Bitstreams

- Frequency
- Block frequency
- Cumulative sums
- Runs
- Longest run of ones
- Spectral test
Results of Randomness Tests

Phase Angles

Cumulative Sum Test Results for TVA Event Angle Data, 100 1000-bit sequences
Results of Randomness Tests

Voltage Magnitudes

Cumulative Sum Test Results for TVA Event Vmag Data, 100 1000-bit sequences
# Effect of Masking Noise Bits

## Phase Angles

### Compression Ratios of Voltage Angles for the Event Data Set After Masking the 4 Least Significant Bits

<table>
<thead>
<tr>
<th></th>
<th>Deflate</th>
<th>Bzip2</th>
<th>LZMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprocessed</td>
<td>2.02</td>
<td>1.75</td>
<td>3.58</td>
</tr>
<tr>
<td>SD</td>
<td>2.09</td>
<td>2.10</td>
<td>3.04</td>
</tr>
<tr>
<td>SRE</td>
<td>5.02</td>
<td>6.05</td>
<td>6.92</td>
</tr>
<tr>
<td>DE</td>
<td>4.49</td>
<td>5.71</td>
<td>6.38</td>
</tr>
</tbody>
</table>

### Compression Ratios of Voltage Angles for the Event Data Set After Masking the 8 Least Significant Bits

<table>
<thead>
<tr>
<th></th>
<th>Deflate</th>
<th>Bzip2</th>
<th>LZMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprocessed</td>
<td>3.84</td>
<td>4.72</td>
<td>11.00</td>
</tr>
<tr>
<td>SD</td>
<td>3.87</td>
<td>4.37</td>
<td>6.13</td>
</tr>
<tr>
<td>SRE</td>
<td>12.45</td>
<td>18.14</td>
<td>17.36</td>
</tr>
<tr>
<td>DE</td>
<td>13.38</td>
<td>20.81</td>
<td>18.82</td>
</tr>
</tbody>
</table>
Effect of Masking Noise Bits
Voltage Magnitudes

**Compression Ratios of Voltage Magnitudes for the Event Data Set After Masking the 4 Least Significant Bits**

<table>
<thead>
<tr>
<th></th>
<th>Deflate</th>
<th>Bzip2</th>
<th>LZMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprocessed</td>
<td>2.49</td>
<td>3.41</td>
<td>3.16</td>
</tr>
<tr>
<td>SD</td>
<td>2.02</td>
<td>2.91</td>
<td>2.82</td>
</tr>
<tr>
<td>SRE</td>
<td>2.58</td>
<td>3.41</td>
<td>3.38</td>
</tr>
<tr>
<td>DE</td>
<td>2.54</td>
<td>3.39</td>
<td>3.34</td>
</tr>
</tbody>
</table>

**Compression Ratios of Voltage Magnitudes for the Event Data Set After Masking the 8 Least Significant Bits**

<table>
<thead>
<tr>
<th></th>
<th>Deflate</th>
<th>Bzip2</th>
<th>LZMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprocessed</td>
<td>4.67</td>
<td>6.40</td>
<td>5.93</td>
</tr>
<tr>
<td>SD</td>
<td>3.63</td>
<td>5.17</td>
<td>4.93</td>
</tr>
<tr>
<td>SRE</td>
<td>4.99</td>
<td>6.21</td>
<td>6.35</td>
</tr>
<tr>
<td>DE</td>
<td>4.95</td>
<td>6.15</td>
<td>6.31</td>
</tr>
</tbody>
</table>
### Best Compression Ratios for Event Scenario Data

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Without Pre-Compression</th>
<th>With Pre-Compression</th>
<th>With Pre-Compression and 4-bit Noise Removal</th>
<th>With Pre-Compression and 8-bit Noise Removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage Mags</td>
<td>2.21</td>
<td>2.50</td>
<td>3.39</td>
<td>6.31</td>
</tr>
<tr>
<td>Phase Angles</td>
<td>2.47</td>
<td>3.94</td>
<td>6.92</td>
<td>20.81</td>
</tr>
</tbody>
</table>
Future Work

- Testing with more data sets
- Improved selection of reference signals
- Integration with commercial PDC
Contact Information

Zeb Tate
zeb.tate@utoronto.ca

Ray Klump
klump@illinois.edu

Acknowledgements

Funding
U.S. Department of Energy
Natural Sciences and Engineering Research Council of Canada

Sample data
Tennessee Valley Authority