Negative Impact of Lossy Compression of PMU Data

Md. Arif Khan and John W. Pierre
Dept. of ECE, University of Wyoming
and
Daniel J. Trudnowski
Dept. of EE, Montana Tech
May 17, 2018

Joint Synchronized Information Subcommittee Meeting - May 2018, Salt Lake City, UT
Compression overview
Swinging door compression algorithm
Impacts of compression
  - Time Domain
  - Frequency Domain
Flaws in decompression process
  - Interpolation error
  - Phase wrapping error
Mitigations
  - Should be corrected in compression/decompression code
  - Recovery algorithm for already corrupted data
Compression: reduces data volume → reduces storage cost and transmission time

Two types:
- Lossless: No loss of information due to the compression
- Lossy: Discards some information that cannot be recovered

Average compression:
- Lossless: 60—70% [1,2]
- Lossy: >90% depending on information loss [3—7]

Swinging Door Algorithm (SDA) used by some utilities in PI historians—lossy or lossless
• Swinging Door Algorithm (SDA) compresses data by discarding small variations in the piece-wise linear trends in the data [7]
• Discards samples that can be linearly interpolated from compressed samples
• Compression Deviation (CompDev), an SDA parameter, determines the level of compression
• Compressed data is decompressed using linear interpolation [8]
• Can provide real-time compression
Voltage Angle Measurements

- $\phi_i$ raw voltage angles captured at different locations in the WECC system at 30 samples/sec.
- $\tilde{\phi}_i$ decompressed voltage angle when $\phi_i$ was compressed using SDA and decompressed using linear interpolation.

Two different types of data:

- $\phi_1 - \phi_4$: Ambient voltage angle measurements where no significant events are evident.
- $\phi_5 - \phi_{17}$: Contain 1.28 Hz forced oscillation (FO) and the strength of the FO vary in the measurements from one location to another. All angles were captured during the same time period.
- SDA compression of real PMU data at different ComDev (degree)

\[
\%RMSE = 100 \times \sqrt{\frac{\sum_{n=0}^{N-1} |\phi[i,n] - \tilde{\phi}[i,n]|^2}{\sum_{n=0}^{N-1} |\phi[i,n]|^2}}
\]

\(\phi\) raw angle
\(\tilde{\phi}\) decompressed angle

Time Domain Impact
Time Domain Impact
Relative frequency derived from decompressed angles – Ambient Data

Welch estimate, 30 min. measurements, 50 sec. Von Hann window, 80% overlap
Relative frequency derived from decompressed angles

- Compression deviation 0.01 degree
- WECC PMU data captured in February 2010 containing FO

Compression:
\[ \phi_7 = 87.63\%, \phi_8 = 76.98\% \]

%RMSE:
\[ \phi_7 = 1.21 \times 10^{-4}\%, \phi_8 = 1.16 \times 10^{-4}\% \]

Welch estimate, 30 min. measurements, 50 sec. Von Hann window, 80% overlap
Compression: $\phi_9 = 89.86\%, \quad \phi_{10} = 83.14\%$

$\%$RMSE: $\phi_9 = 1.23 \times 10^{-4}\%, \quad \phi_{10} = 1.18 \times 10^{-4}\%$

Compression: $\phi_{11} = 41.93\%, \quad \phi_5 = 44.79\%$

$\%$RMSE: $\phi_{11} = 0.92 \times 10^{-4}\%, \quad \phi_5 = 0.97 \times 10^{-4}\%$
Compression: $\phi_9 = 89.86\%, \phi_{12} = 90.47\%$

%RMSE: $\phi_9 = 1.23 \times 10^{-4}\%, \phi_{12} = 1.23 \times 10^{-4}\%$

Compression: $\phi_{13} = 92.1\%, \phi_{14} = 43.25\%$

%RMSE: $\phi_{13} = 1.31 \times 10^{-4}\%, \phi_{14} = 0.91 \times 10^{-4}\%$
Compression: $\phi_{15} = 91.93\%, \phi_{7} = 87.63\%$

%RMSE: $\phi_{15} = 1.19 \times 10^{-4}\%, \phi_{7} = 1.21 \times 10^{-4}\%$

Compression: $\phi_{15} = 91.93\%, \phi_{16} = 92.3\%$

%RMSE: $\phi_{15} = 1.19 \times 10^{-4}\%, \phi_{16} = 1.32 \times 10^{-4}\%$
Compression: $\phi_5 = 44.79\%$, $\phi_6 = 47.04\%$

%RMSE: $\phi_5 = 0.97 \times 10^{-4}\%$, $\phi_6 = 0.97 \times 10^{-4}\%$

%RMSE: $\phi_{17} = 1.06 \times 10^{-4}\%$, $\phi_{15} = 1.19 \times 10^{-4}\%$

Compression: $\phi_{17} = 79.94\%$, $\phi_{15} = 91.93\%$
False harmonics and aliasing simulated data

FO detection algorithm output
- **Raw**: original measurements
- **Decompressed**: the raw measurements went through the compression and decompression processes
- Both of the data sets are from the utility
- **Raw**: original measurements
- **Decompressed**: the raw measurements went through the compression and decompression processes
- Both of the data sets are from the utility
- **Raw**: original measurements
- **Decompressed**: the raw measurements went through the compression and decompression processes
- **Both of the data sets are from the utility**

Welch estimate, 30 min. frequency data, 50 sec. Von Hann window, 80% overlap
Two types of errors were observed in the decompressed data

- Interpolation error
- Angle wrapping error
\[ y(n) = \frac{(n_2 - n)y(n_1)}{n_2 - n_1} + \frac{(n - n_1)y(n_2)}{n_2 - n_1} \]

\[ y(n) = y(n_1) + (n - n_1)m \]
\[ y(n) = (n_2 - n)y(n_1) \frac{1}{n_2 - n_1} + (n - n_1)y(n_2) \frac{1}{n_2 - n_1} \]

\[ y(n) = y(n_1) + (n - n_1)m \]

\[ y(n) = y(n_1) + (n - n_1 + 1)m \]
\[
y(n) = \frac{(n_2 - n)y(n_1)}{n_2 - n_1} + \frac{(n - n_1)y(n_2)}{n_2 - n_1}
\]

\[
y(n) = y(n_1) + (n - n_1)m
\]

\[
y(n) = y(n_1) + (n - n_1 + 1)m
\]
\[ y(n) = \frac{(n_2 - n)y(n_1)}{n_2 - n_1} + \frac{(n - n_1)y(n_2)}{n_2 - n_1} \]  

**Equation 1**

\[ y(n) = y(n_1) + (n - n_1)m \]

**Equation 2**

\[ y(n) = y(n_1) + (n - n_1 + 1)m \]
\( N_d = \{3, 6, 7, 11, 12, 13\} \)

**Equation 1**

\[
y(n) = \frac{(n_2 - n)y(n_1)}{n_2 - n_1} + \frac{(n - n_1)y(n_2)}{n_2 - n_1}
\]

**Equation 2**

\[
y(n) = y(n_1) + (n - n_1 + 1)m
\]
Phase Wrapping Error
Ideally, the utilities would correct the errors in the decompression code

For already corrupted data—two-stage three-sample algorithm (IEEE PES GM 2018 [9])

Stage 1: In this stage all artifacts due to the flawed phase wrapping are removed. This can be done by taking absolute value of forward derivative of decompressed wrapped voltage angle. Then find set of all indexes $N_w$ where the absolute value is greater than some threshold. The value of the threshold can be set slightly less than $2\pi$. Then discard all $(n + 1)$th sample such that $n \in N_w$ from the decompressed angle, and interpolate the discarded samples using eq. (1). Here $a \in A$ refers $a$ is an element of set $A$.

Stage 2: In this stage all artifacts due to the flawed decompression algorithm are removed using a three-sample algorithm. Let three consecutive samples in corrupted decompressed measurements are $y(n)$, $y(n + 1)$, and $y(n + 2)$. Starting with an empty set of sample indexes $N_e$, perform the following for all samples in the decompressed angle for some small positive value $\varepsilon$ close to zero.

1) Calculate $s_1 = \frac{y(n+1) - y(n)}{2}$, $s_2 = \frac{y(n+2) - y(n)}{2}$, and $s_3 = \frac{y(n+2) - y(n)}{3}$.

2) If $|s_2 - s_1| < \varepsilon$ add $n + 1$ to $N_e$, set $n = n + 2$ and go to step 1. Otherwise, go to step 3. Here the symbol “$\cdot$” represents the absolute value of underlying quantity.

3) If $|s_3 - s_1| < \varepsilon$, set $n' = n + 2$ and go to step 4. Otherwise, set $n = n + 1$ and go to step 1.

4) Check if $|y(n'+1) - y(n') - s_1| < \varepsilon$. If true, increase $n'$ by one, repeat this step until the condition is false (if $n'$ becomes equal to the measurements length $N$, this step also returns false). If false, check whether $|y(n'+1) - y(n')| < \varepsilon$ is true. If yes, add indexes $n + 1$ through $n'$ to $N_e$, set $n = n' + 1$ and go to step 1. Otherwise, set $n = n'$ and go to step 1.
- Two-stage three-sample algorithm
Two-stage three-sample algorithm

Compression achieved 1.87%

swinging door algorithm can achieve 90+% compression with significant loss of information, and the type of loss cannot be predicted precisely
- SDA lossy compression is highly nonlinear and impacts of the compression algorithm are erratic and how a measurement will be affected due to the compression is difficult to predict.

- SDA introduces false harmonics of line spectra and aliasing of false higher order harmonics of the line spectra.

- Interpolation and phase wrapping errors were observed in a decompressed data set.

- Already corrupted decompressed data can be recovered using the two-stage three-sample algorithm.

- Ideally utilities should correct their decompression process.


Thanks!
Questions?
- Compression in PI PMU historian
- A measurement has to pass two tests to be archived
  - Exception test
  - Compression test

Yes! Real time operation.

Lossy!
- **Exception Test**
  - Takes place on the PI interface node before the value is ever sent to the PI server
  - Discards small variations due to noise

![Diagram](image)

- **ExcDev**
  - Snapshot (current value sent to the PI server)
  - Raw value
  - Discarded
  - Stored in Snapshot Table

Exception Test: Takes place on the PI interface node before the value is ever sent to the PI server. Discards small variations due to noise.
- **Exception Test**
  - Takes place on the PI interface node before the value is ever sent to the PI server
  - Discards small variations due to noise

---

Value passed Exception test and sent to Snapshot table

Discarded

Snapshot (current value sent to the PI server)
- **Exception Test**
  - Takes place on the PI interface node before the value is ever sent to the PI server
  - Discards small variations due to noise

![Graph showing value and time relationship with exception points](image-url)

- **Value passed Exception test and sent to Snapshot table**
- **Discarded**
Compression Test

- Swinging door algorithm—by Edger H. Bristol in 1990. Yes! One of the owners of “The Foxboro Company.”
Compression Test

- Swinging door algorithm (SDA)— by Edger H. Bristol in 1990. Yes! One of the owners of “The Foxboro Company.”

Recently archived
Current snapshot
Incoming value
Discarded

CompDev ≈ 2* ExcDev

value

Deviation “blanket”

time

PI System Compression
- **Linear interpolation**

\[ y(n) = \frac{(n_2 - n)y(n_1)}{n_2 - n_1} + \frac{(n - n_1)y(n_2)}{n_2 - n_1} \]

![Graph showing linear interpolation with markers for archived, compression test discarded, and reconstructed data points.](image)