2022 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA) | 978-1-6654-9823-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/SGSMA51733.2022.9806010

Using Synchrophasor Status Word as Data Quality Indicator: What to Expect in the Field?

Zheyuan Cheng, Yi Hu
Quanta Technology, LLC
Raleigh, NC, USA
{zcheng, yhu}@quanta-technology.com

Zoran Obradovic
Temple University
Philadelphia, PA, USA
zoran.obradovic@temple.edu

Mladen Kezunovic
Texas A&M University
College Station, TX, USA
kezunov@ece.tamu.edu

Abstract— **Data quality plays a crucial role in successful applications of synchrophasor data in power system operation and control. This paper presents the results of a data quality analysis of a multi-year field-recorded synchrophasor dataset. The analysis has identified several typical data quality issues encountered in the field data. An examination of the PMU status words included with the dataset has revealed several inconsistent implementations and the lack of correlation between the PMU data quality and the status word, which impacts the usefulness of such information. Our investigation has concluded that the status word alone as found in the recorded field dataset could not be used as a reliable indicator of data quality for field-recorded data. Several recommendations are proposed to improve the usefulness of the PMU status word.**

Keywords—Data Quality, Phasor Measurement Unit, Status Word, Synchrophasor

I. Introduction

Synchrophasor technology and systems use Phasor Measurement Units (PMUs) to record electrical quantities, e.g., voltage, current, and frequency, at a specific location in an electric power system. These electrical quantities are typically measured at a fixed reporting rate, e.g., 30 or 60 Frames Per Second (FPS). They typically are time-stamped using Global Positioning System (GPS) signal as the time reference [1]. The synchronized measurements technology was quickly recognized as an enabler of new applications such as situational awareness, event detection, stability monitoring, fault location, etc. [2]–[5]. While over 2,500 PMUs were reported as being deployed in the USA and Canada as of 2017 [4], this number has been rapidly growing in the transmission and distribution networks ever since [5].

Data quality is crucial for any data-driven analytics [6]. Identifying and mitigating data quality issues, e.g., missing data, overlapping data, measurement errors, etc., is integral to any successful algorithm development and practical applications. Huang *et al.* conducted a detailed review of the data quality issues for synchrophasor applications in [7]. A similar investigation on the data quality has been performed by Kirihara *et al.* in [8]. In addition to data quality, cybersecurity risk associated with data quality is also evaluated by Sundararajan *et al.* in [9]. To assess and mitigate various data quality issues, methods and guidelines are proposed in a NASPI white paper [10].

Synchrophasor status word, defined in standards [1], [11], contains16-bit Boolean bit mapped flags located at the beginning of each PMU data block. It indicates various conditions of the PMU data, e.g., time synchronization error, the validity of the erroneous data, etc. These bits are set initially by the PMU that generates the data and can be altered by other processors in the data transportation chain, e.g., Phasor Data Concentrator (PDC). The status word is meant to offer valuable information related to the usefulness of the PMU measurement data. For example, one can quickly tell whether a PMU is out-of-sync by interpreting bit-13. Similarly, one should be able to determine whether the PMU is experiencing errors by interpreting bit-14 and bit-15. Hence, it appears conceivable to use the status word, especially two of the most significant bits, as the indicator of the PMU data quality [12].

We perform detailed data quality and PMU status word analysis on 11 terabytes of data, including 188 PMUs recorded in Eastern Interconnection of US from 2016 to 2017 [13]. This is one of the very few instances when field-recorded data from hundreds of PMUs includes different noise patterns, normal fluctuations, errors, inconsistencies, and other data quality issues is made available to the researchers by the data owners.

Data quality and PMU status word analysis performed in this paper are outlined in Fig. 1. As a result of our detailed study of this problem, our technical contributions are: (1) identified field PMU data quality issues and provided feedback that positively impacts the algorithm development and validation work for the machine learning applications [14]; (2) confirmed that the status word as provided in the dataset cannot be used as a reliable data quality indicator; and (3) made several recommendations for improving the data quality and the usability of the status word including suggestions for the standard revision.

The rest of the paper is structured as follows: Section II presents the data quality analysis results. Section III summarizes the overall status word behavior observed in the field-recorded dataset. Section IV analyzes the likely causes of several status word bit assignment inconsistencies. The recommendations and conclusions are summarized in Section V and VI, respectively. References are given at the end.

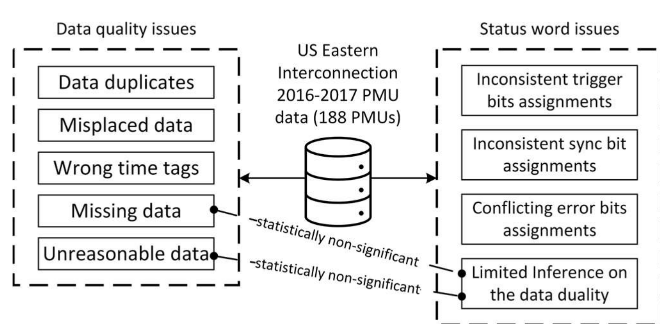

Fig. 1. Data quality and PMU status word analysis overview.

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000913.

II. Findings of Field PMU Data Quality Analysis

A data quality analysis is performed by automated scanning through the entire dataset. The key findings of the analysis, i.e., several issues, such as missing data, unreasonable data values, wrong time tags, etc., are summarized below.

A. Missing Data

Among the 188 PMUs, 99 do not have three-phase data from 2016 to 2017, i.e., they only have positive sequence voltage and current phasor data. This is possibly due to different PMU data use and storage practices. Some utilities or independent system operators may only use/archive positive sequence data. To conduct a fair missing data comparison in this dataset, a missing data check is performed on all key measurements, i.e., positive sequence voltage magnitude denoted as "vp_m", positive sequence voltage angle denoted as "vp_a", positive sequence current magnitude denoted as "ip_m", positive sequence current angle denoted as "ip_a", frequency denoted as "f", and Rate-Of-Change-Of-Frequency (ROCOF) denoted as "df". The exact missing data percentage of each PMU and variable is summarized using a boxplot in Fig. 2. The mean is labeled with the cross "x", while the median is labeled with the horizontal line inside the box (whisker). The first (Q1) and third quartiles (Q3) of the data are plotted as two upper, and lower bounds of the boxes, the two tails extended from the box are the distances of 1.5 times the interquartile range (Q3-Q1). The outliers are plotted as individual points outsides the interquartile range. According to Fig. 2, the majority of the PMUs are missing 5-20% of the key measurements from 2016 to 2017, whereas a few PMUs are missing 100% of specific measurements. The frequency and ROCOF data have fewer missing data occurrences among all key measurements. It may be concluded that missing data generally is a system implementation and configuration issue which can be improved by identifying the root causes of why the data is missing and taking proper corrective actions.

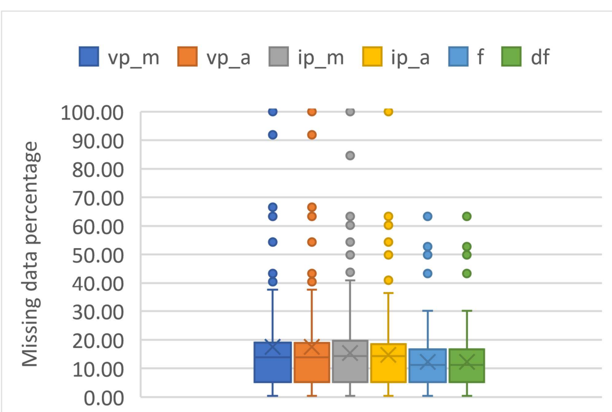

Fig. 2. Boxplot of the missing data percentage of key measurements.

Table I. Criteria used to filter unreasonable data.

Measurements	Criteria
Voltage magnitude V (V)	$V > 2 \times V_{Rated}$ or $V < 0$
Current magnitude I (A)	$I > 1{,}000{,}000$ or $I < 0$
Angle value θ (Degree)	$\theta > 180$ or $\theta < -180$
Frequency f (Hz)	$f > 70$ or $f < 0$
ROCOF $\frac{\partial f}{\partial t}$ (Hz/sec)	$\frac{\partial f}{\partial t} > 100$ or $\frac{\partial f}{\partial t} < -100$

B. Unreasonable Data

A thorough check on the validity of the measurement values reveals some extremely large and small values reported by some PMUs. The criteria defined in Table I are used to detect extreme unreasonable values. An example of captured unreasonable data in "vp_m" is presented in Fig. 3, according to which extremely large and small measurements have been consistently found from time to time throughout these two years. Similar unreasonable data is also detected in other key measurements.

In addition to extreme values, we also notice 18 PMUs with flat 60 Hz frequency measurement values. Among these 18 PMU, there are 4 PMUs that report constant 60 Hz frequency value, whereas the other 14 PMU exhibit this flat 60 Hz behavior periodically throughout 2016-2017. As real power system frequency constantly changes around the nominal frequency, a flat 60 Hz is equivalent to having no frequency measurement at all. We conclude that such flat 60 Hz measurements are most likely due to configuration errors, e.g., wrong frequency signal assignments.

C. Wrong Time Tags

After closely examining the field data, we find that the PMU time tags have different precision and rounding mechanisms. We classify all 30 FPS PMU time tags into three types. The Type-I PMU time tag preserves five digits after the decimal point. Type-II PMU time tag rounds up and preserves three digits after the decimal point, and some time stamps end with x.xx3 and x.xx7, e.g., 0.133 and 0.167 sec. The Type-III PMU time tag also preserves three digits after the decimal point, but some time stamps end with x.xx2 and x.xx5, e.g., 0.132 and 0.165 sec. This could be caused by differences in PMU vendor implementations. In addition to the time tag classification, wrong time tag recording is discovered in two PMUs' data. Table II provides an

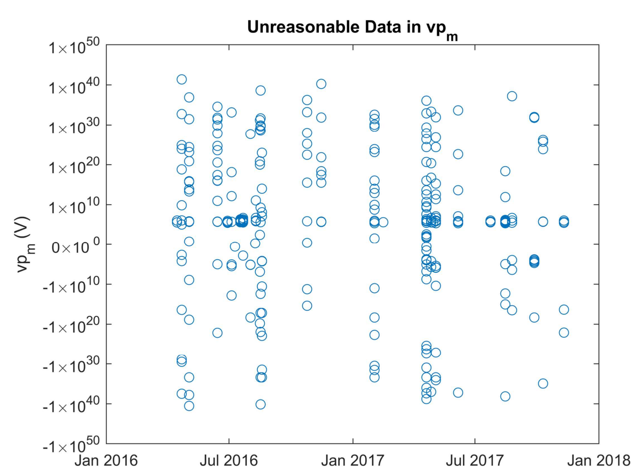

Fig. 3. Unreasonable data in the positive sequence voltage magnitude.

Table II. Wrong time tag in PMU C569's data.

UTC	vp_m	vp_a	ip_m	ip_a	f	df
13:15:17.76667	208258.95	49.00	387.79	53.00	60.02	0.002
13:15:17.76676	NaN	NaN	NaN	NaN	NaN	NaN

example of the detected wrong time tag by showing the correct time tag in the first row and the captured time tag in the second row. It is also noticeable that all analog measurements associated with the wrong time tag are listed as Not-a-Number (NaN). In total, these two PMUs reported 202 measurements with wrong time tags during these two years.

D. Misplaced Data

After a review of excessive captured unreasonable data, we discovered that in July 2016, there were 57 PMUs may have misplaced data in their “ip_m” or “f” measurements. An example of the detected misplaced data of a 200 kV PMU is presented in Fig. 4. It appears that the positive sequence current magnitude “ip_m” and frequency “f” data are replaced with voltage data. However, these three sets of voltage measurements have different DC biases when compared to each other. We have identified three 80 kV, one 96 kV, fifty-one 200 kV PMUs, and one 300 kV PMUs that have suspected voltage magnitude measurements mixed in the positive sequence current magnitude measurements. It was inconclusive what exactly may have caused such data misplacement.

E. Data Duplicates

We have identified several instances of the duplicate data issue in the field PMU dataset. On 2017/05/15 hour 21, 79 PMUs’ data in this hour were copied twice, and one 5-sec period data (21:55:00 – 22:00:00) was replicated three times. For any of these 30 FPS PMU, the total number of duplicated data points in this hour is 10815. Notably, all the 79 PMUs exhibit the exact same data duplicates behavior and all 79 PMUs have the same time tag type, i.e., Type-II time tag. This may be caused by improper data handling while storing/packaging process of the dataset.

F. Data Quality Analysis Summary

Among the findings of our data quality analysis, missing data and unreasonable data have the most impact on subsequent data uses. For PMUs that miss a big block of data, it is almost impossible to recreate the original field data. Interpolation-based missing data filling methods may help deal with magnitude quantities with less variability, e.g., voltage and frequency. But it is incredibly challenging to repair angle data, especially unwrapped angle data, as the wrapping points are permanently lost. Such unreasonable data should be removed from any subsequent uses because it may negatively impact threshold-based or gradient-based analytics.

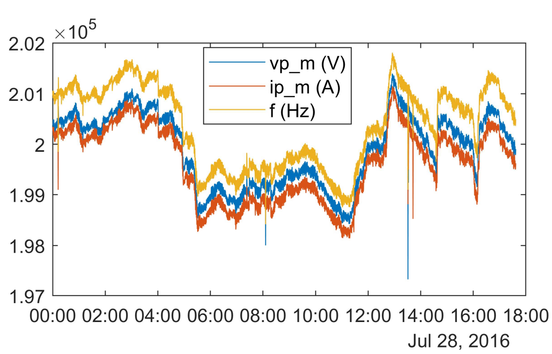

Fig. 4. Example of misplaced data in PMU measurements.

III. Interpretation of Status Word in Field PMU Data

An examination of PMU status word is performed to assess whether the information provided by the status bits in this field-recorded dataset could be used as a reliable data quality indicator. The results have shown that this is not possible, as described below.

A. Differences Between Standards

The provided PMU data includes a status word that is either defined in IEEE Std C37.118-2005 or IEEE Std C37.118.2-2011 for each block of PMU data, as shown in Table III. The bits in the status word are defined to indicate several data quality related conditions. To properly decode the data quality related information from the received status word, it is essential to know the correct standard version that the PMU complies with. Although the newer 2011 standard [10] had attempted to be backward compatible with the earlier 2005 standard [1], some differences still exist for specific bits definitions.

In conclusion, interpreting the meaning of the status word bits without knowing the version of the standard being implemented could lead to misinterpretation if a wrong version has been assumed. Finding out which version of the standard is implemented in a PMU could be a simple task by checking with the supplying vendor. However, this is impossible when specific PMU model information is not provided with the corresponding dataset. Furthermore, bits in PMU status words could be modified by other processors (e.g., PDCs) without any recorded

Table III. Differences between C37.118-2005 and C37.118.2-2011.

Bits	C37.118-2005	C37.118.2-2011
15-14	Bit 15: Data valid, 0 when PMU data is valid, 1 when invalid or PMU is in test mode Bit 14: PMU error including configuration error, 0 when no error	00 = good measurement data, no errors. 01 = PMU error. No information about data 10 = PMU in test mode (do not use values) or absent data tags have been inserted (do not use values) 11 = PMU error (do not use values)
13	PMU sync, 0 when in sync	PMU sync, 0 when in sync with a UTC traceable time source
12	Data sorting, 0 by time stamp, 1 by arrival	Data sorting, 0 by time stamp, 1 by arrival
11	PMU trigger detected, 0 when no trigger	PMU trigger detected, 0 when no trigger
10	Configuration changed, set to 1 for 1 min when configuration changed	Configuration change, set to 1 for 1 min to advise configuration will change, and clear to 0 when change effected
09-06	Reserved for security, presently set to 0	Bit 09: Data modified, 1 if data modified by post processing, 0 otherwise Bits 08-06: PMU Time Quality. Refer to codes in Table 7 in [11]
05-04	Unlocked time: 00 = sync locked, best quality 01 = Unlocked for 10 s 10 = Unlocked for 100 s 11 = Unlocked over 1000 s	Unlocked time: 00 = sync locked or unlocked < 10 s (best quality) 01 = 10 s ≤ unlocked time < 100 s 10 = 100 s < unlock time ≤ 1000 s 11 = unlocked time > 1000 s
03-00	Trigger reason: 1111–1000: Available for user definition 0111: Digital 0110: Reserved 0101: df/dt high 0100: Frequency high/low 0011: Phase-angle diff 0010: Magnitude high 0001: Magnitude low 0000: Manual	Trigger reason: 1111–1000: Available for user definition 0111: Digital 0110: Reserved 0101: df/dt high 0100: Frequency high/low 0011: Phase-angle diff 0010: Magnitude high 0001: Magnitude low 0000: Manual

indications. The PMU status words of this dataset may not all be set by PMUs, and PDCs may have implemented a version of the standard different from what is used for PMUs.

B. *Overview of Field Status Word*

In the dataset we studied, we have identified 1354 unique nonzero status word bit patterns, which account for 2% of the entire dataset. To summarize the status word bit patterns, we have grouped all 16 bits into six categories: (1) Error (ERROR) flag bits: bit 15 and 14; (2) Sync (SYNC) related bits: bit 13, 8, 7, 6, 5, and 4; (3) Data sorting (SORTING) bit: bit 12; (4) Trigger (TRIGGER) related bits: bit 11, 3, 2, 1, and 0; (5) Configuration change (CONFIG.) related bit: bit 10; (6) Data modification (MODIFY) related bit: bit 9. Table IV summarizes the top 10 most seen field status word patterns using these categories. These top 10 status word patterns already cover 99.84% of all the nonzero status words. According to Table IV, the most common pattern is PMU not in sync; data sorted by arrival; and PMU in error. 97 PMUs have reported this type of pattern in these two years. The second most common pattern is also about the unsync issueThe PMU data is sorted by time, and in this case, the error bits are not set. Overall, the majority of the nonzero status words are related to time synchronization issues.

IV. Inconsistencies in Field Captured Status Word

A. *Inconsistent trigger related Status Bits Assignments*

We have observed some inconsistencies in terms of PMU trigger bit assignment, i.e., bit-11. A detailed interpretation of the status word can be found in III. Generally speaking, the bit-11 should be set as one if any of the trigger conditions are satisfied. The trigger condition bits can be set by either any one of the standard-defined default magnitude related conditions or user-defined conditions. When one or more than one of the conditions are met, non-zero bits among the bits 03-00 in the status word should be observed, and the bit-11 should be 1 as well. However, some PMUs flagged trigger conditions without setting the bit-11 as 1. The left-hand side pie chart in Fig. 5 summarizes the PMU bit-11 assignments when one or more than one of the trigger conditions are met. It is evident that ~64% of the PMUs would set bit-11 to 1, but there is a fair number of PMUs who do not follow this pattern. Such inconsistencies in the trigger bit assignment could result from a PMU vendor's misinterpretation or incorrect implementation of the standard, or a user misconfiguration error if bit-11 is configurable by users.

TABLE IV. Summary of status word pattern

ERROR	SYNC	SORTING	TRIGGER	CONFIG.	MODIFY	PMUs
Error	Unsync	ByArrival	NoTrigger	NoChange	NoMod	97
Good	Unsync	ByTime	NoTrigger	NoChange	NoMod	111
Good	Sync	ByTime	Triggered	NoChange	NoMod	92
Good	Unsync	ByArrival	Triggered	NewCnfg	NoMod	97
Error	Unsync	ByTime	NoTrigger	NoChange	NoMod	11
Good	Unsync	ByArrival	NoTrigger	NoChange	NoMod	50
Error	Unsync	ByArrival	NoTrigger	NoChange	Modified	97
ErrNoInfo	Unsync	ByTime	NoTrigger	NoChange	NoMod	77
Good	Sync	ByTime	NoTrigger	NewCnfg	NoMod	106
ErrNoInfo	Unsync	ByArrival	NoTrigger	NoChange	NoMod	12

It should be noted that additional information will be needed that could make this trigger-related bit information more useful. The condition upon which these bits are set is the most important one. Without this information, it cannot be cross-checked with the PMU data whether the bits were set correctly according to the standard. Providing measurement point/phase indication for a triggered condition when multiple measurement points (e.g., multiple current measurement points) and three-phase measurement data are involved would also be instrumental. Currently, neither version of the standard requires such information to be provided.

B. *Inconsistent Sync Related Status Bits Assignment*

Here, we want to examine the assignment of bit-13 when any of the sync bits, i.e., bit-4 and bit-5, is not zero. According to the standard, when PMU's clock is unlocked for over 10 seconds, the PMU would set bit-4 and bit-5 accordingly to reflect how long the clock is unlocked. Typically, when the clock is unlocked for more than 10 seconds, the PMU would also set bit-13 to 1, indicating if the PMU is not in sync with UTC. The right-hand side pie chart in Fig. 5 summarizes the PMU bit-13 assignments when PMU's clock is unlocked for over 10 seconds. Based on Fig. 5, a fair number of PMUs would not set bit-13 to 1 when unlocked for over 10 seconds. Without additional information, it is unclear whether such inconsistencies are due to the PMU that may have a second timing sync source, a better clock that drifts very slowly, the differences in vendor interpretation/implementation of the standard, or a combination of all. Regardless, this inconsistency would directly impact the usefulness of this information in determining the PMU data quality. In summary, such conflict in the sync bit assignment can be a PMU vendor or setting-specific problem.

C. *Conflicting Error Bits Assignment When Not in Sync*

When the PMU is not in sync (bit-13 = 1), it is logical according to the standard to set the error bits, i.e., bit-14 and bit-

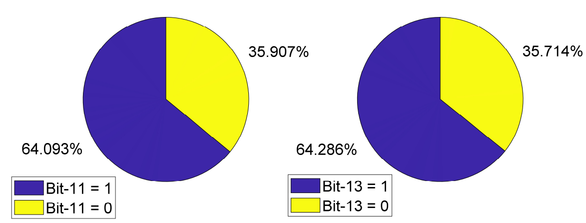

Fig. 5. Inconsistencies in status bit assignments

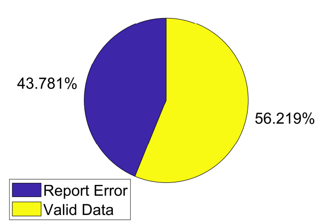

Fig. 6. Conflicting error bits assignment when PMU not in sync.

15, indicating an error in time synchronization and the resultant error in data. However, this is not a universal practice in the dataset we examined. In our analysis, we first collect all the status words with only unsync issues (i.e., other bits are zero), then examine whether the error bits are set. Fig. 6 summarizes the PMU error bits assignment behaviors during unsync conditions. We can see that less than half of the PMUs had set the error bits when not in sync, whereas the majority of the PMUs did not set the error bits. The conflicting error bits assignment under the unsync conditions may be attributed to PMU vendors implementing the standards differently, although more information is needed to confirm the causes. This conflict has dramatically reduced the usefulness of the error bits in the status word.

D. Limited Inference on the Data Quality

It is conceivable to use the status word, especially the two error bits 15 and 14, as the indicator of the PMU data quality. Ideally, one could rely on the status word to identify bad data. Our study further investigated the correlation between zero and nonzero status words and two data quality issues, namely missing data and unreasonable data. This investigation aims to evaluate the potential of a status word being used as a reliable data quality indicator. Since the status word is not defined to address other data quality analysis identified issues such as wrong time tag and duplicated data, these issues are excluded from the investigation.

1) Correlation with Missing Data: We first identify all the PMU data with missing key measurements in this two-year-long dataset. We then examine how the status word correlates to PMU missing data conditions. We further categorize the missing data condition into two scenarios: partial key measurements missing and all key measurements missing. According to Table V, when missing part of the key measurements, 84.75% of the time, the corresponding status word is zero indicating good data quality. When a PMU is missing all the key measurements, only 48.69% of the time, this PMU reports nonzero status word signifying problems. In this regard, the PMU status word is better at reporting all data missing conditions. However , it cannot be used as a reliable indicator for missing data conditions in either case.

2) Correlation with Unreasonable Data: As discussed in Section II, there are a lot of unreasonable data reported by PMUs in this dataset. We collect all the unreasonable data occurrences in these two years and check its corresponding status word out of curiosity. As summarized in Table V, when the PMUs have unreasonable data, 99.76% of the time, the status word is zero indicating good data quality is. Again, it is evident that status word cannot be used as a reliable indicator for unreasonable data for this field-recorded dataset.

TABLE V. STATUS WORD POPULATION IN THE PRESENCE OF MISSING AND UNREASONABLE DATA.

Status Word	Partial Missing	All Missing	Unreasonable
Zero	84.75%	51.31%	99.76%
Nonzero	15.25%	48.69%	0.24%

V. RECOMMENDATIONS

A. Recommendations for Application Developers Based on Status Word

1) Actual PMU Configuration Matters: PMUs may have some parameters related to the status bits that are configured by the end-users, such as the user-defined trigger bits. Without knowing what is configured in the PMU by the users, it will not be possible to interpret the user-defined status bits. Hence, it is recommended that the user should provide the actual PMU configurations in addition to the PMU data.

2) Taking PDC Action into Account: PMUs are not the only ones that set and reset the status word bits. As PMU data travel through a synchrophasor system and get processed by one or more intermediate functions or devices such as PDCs or phasor gateways, some of the status bits may be altered or re-assigned for various reasons: (1) If the PMU data arrival is not correctly aligned since their clock is not synchronized to UTC. (2) If the PMU data frame checksum failed to indicate that data is corrupted. (3) If the intermediate functions or devices did not receive the PMU data in time to send the aggregated data out and have filled the missing data for the PMU with either NaN or interpolated/extrapolated data. (4) If intermediate functions or devices have made some changes to the data, such as data format conversion (e.g., fixed point ↔ floating-point, rectangular ↔ polar, etc.).

The recommendation is that a PMU dataset should also provide information for PDCs/gateways that handle the PMU data and their configuration information.

3) Unable to Tell Measurement Signals Apart: Even if all the information mentioned above is available, it may not be enough to properly interpret the exact meaning of certain status bit(s) because status bits are defined for all measurement signals within one PMU data block, not for individual measurements. It is recommended that application developers should not attempt to tie any particular measurement signal with status bits, and the standard should consider defining additional status bits for individual signals.

4) Need for Establishing Evolving Best Practices: Some data quality issues, e.g., misplaced data and wrong time tags, are not very intuitive and could be easily overlooked. This paper provides a good starting point for other researchers/engineers to examine at their PMU data quality, and we recommend that existing best practices be updated as new discoveries are made. We are not aware of any "off-the-shelf" automated data quality analysis tools that can provide similar comprehensive data quality analysis.

B. Recommendations for IEEE Standard Revision and Industry Practice

1) Clarification on When to Set/Reset Certain Status Bit(s): As already mentioned, the status word is generally defined for an entire PMU data block, not for a particular measurement signal or any specific condition. The standards did not define the exact conditions under which one of the status bits or a block of bits should be set and reset. Taking the error bit(s) as an example, when it is set, it indicates there is(are) some type(s) of error(s) in the PMU data block, which could be that all data are in error, a single measurement signal data is in error, or the entire PMU function is experiencing some hardware/software error, etc. It is impossible to know which of the listed errors might have occurred without knowing the implementation details. It is recommended that the standard should clearly define how the status bits can be altered or assigned under various conditions, such as PMU data not received, PMU error detected, etc.

2) Need for a Standard Compliance Certification Program: Once the standard defines the detailed conditions when one of the status bits or a block of bits should be set and reset, it is recommended to have a standard compliance certification program to ensure a consistent standard interpretation and implementation by all vendors. Utilities are recommended to only use certified PMUs/PDCs/Gateways/etc. to ensure a consistent standard implementation among them.

3) Need for dedicated specialists: Assigning dedicated specialists in an organization to clean up the problems with existing PMUs/PDCs/Gateways and properly commission new PMUs/PDCs/Gateways would be very helpful.

VI. Conclusions

A. Data Quality

A thorough data quality analysis on the 11 terabytes field-recorded dataset has revealed several issues on a large percentage of data such as missing data, unreasonable data, data with wrong time tags, misplaced data, and data duplicates, that affected the effective use of such data. The identified data quality issues reported here could be improved by the data contributors through appropriate data quality improvement actions.

B. Status Word Usefulness

In the dataset, (1) nonzero status word only appeared in a small portion (2%) of the dataset while a much larger portion of the data had data quality issues; (2) the most common cause of the nonzero status bits is the time synchronization issue; and (3) some inconsistencies and conflicts of status bit assignments are observed, which could be attributed to the IEEE standards not being specific causing differences in vendors' implementations. The presented statistical data quality comparison suggests that one should not treat the status word as a reliable data quality indicator, at least in its current implementations.

C. Recommendations

We offer several recommendations to help the research community and industry taking actions to improve the overall PMU data quality, better utilize the PMU status word, and improve the standardization of PMU status word implementation in the future.

VII. Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] "IEEE Standard for Synchrophasors for Power Systems," *IEEE Std C37118-2005 Revis. IEEE Std 1344-1995*, p. 65, Mar. 2006, doi: 10.1109/IEEESTD.2006.99376.

[2] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized Phasor Measurement Applications in Power Systems," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, Jun. 2010, doi: 10.1109/TSG.2010.2044815.

[3] F. Aminifar, M. Fotuhi-Firuzabad, A. Safdarian, A. Davoudi, and M. Shahidehpour, "Synchrophasor Measurement Technology in Power Systems: Panorama and State-of-the-Art," *IEEE Access*, vol. 2, pp. 1607–1628, 2014, doi: 10.1109/ACCESS.2015.2389659.

[4] M. U. Usman and M. O. Faruque, "Applications of synchrophasor technologies in power systems," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 2, pp. 211–226, Mar. 2019, doi: 10.1007/s40565-018-0455-8.

[5] M. Hojabri, U. Dersch, A. Papaemmanouil, and P. Bosshart, "A Comprehensive Survey on Phasor Measurement Unit Applications in Distribution Systems," *Energies*, vol. 12, no. 23, p. 4552, Nov. 2019, doi: 10.3390/en12234552.

[6] Farantatos, Evangelos and Amidan, Brett, "Data Mining Techniques and Tools for Synchrophasor Data," North American SynchroPhasor Initiative (NASPI), White Paper NASPI-2018-TT-007, Jan. 2019. [Online]. Available: https://www.naspi.org/node/743

[7] C. Huang *et al.*, "Data quality issues for synchrophasor applications Part I: a review," *J. Mod. Power Syst. Clean Energy*, vol. 4, no. 3, pp. 342–352, Jul. 2016, doi: 10.1007/s40565-016-0217-4.

[8] K. Kirihara, K. E. Reinhard, A. K. Yoon, and P. W. Sauer, "Investigating Synchrophasor Data Quality issues," in *2014 Power and Energy Conference at Illinois (PECI)*, 2014, pp. 1–4. doi: 10.1109/PECI.2014.6804563.

[9] A. Sundararajan, T. Khan, A. Moghadasi, and A. I. Sarwat, "Survey on synchrophasor data quality and cybersecurity challenges, and evaluation of their interdependencies," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 3, pp. 449–467, 2019, doi: 10.1007/s40565-018-0473-6.

[10] "Synchrophasor Data Quality Attributes and a Methodology for Examining Data Quality Impacts upon Synchrophasor Applications," North American SynchroPhasor Initiative (NASPI), White Paper NASPI-2016-TR-002, Mar. 2016.

[11] "IEEE Std C37.118.2-2011 (Revision of IEEE Std C37.118-2005) IEEE Standard for Synchrophasor Data Transfer for Power Systems," IEEE. doi: 10.1109/IEEESTD.2011.6111222.

[12] J. Follum and B. Amidan, "A data quality filter for PMU measurements: Description, experience, and examples," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, USA, Jul. 2016, pp. 1–5. doi: 10.1109/PESGM.2016.7741378.

[13] J. Banning, E. Andersen, and J. Follum, "Data for FOA 1861 – Bringing It All Together," 2021 NASPI Work Group Meeting, Presentation, May 2021.

[14] A. A. Hai *et al.*, "Transfer Learning for Event Detection From PMU Measurements With Scarce Labels," *IEEE Access*, vol. 9, pp. 127420–127432, 2021, doi: 10.1109/ACCESS.2021.3111727.