21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2015 Grid of the Future Symposium

Data Quality Considerations for Waveform Analytics

J. WISCHKAEMPER
C.L. BENNER
B.D. RUSSELL
K. MANIVANNAN
Texas A&M University
USA

SUMMARY
"Analytics" has achieved buzzword status across multiple industries over the past decade. Generally speaking, analytics refers to the acquisition and analysis of large data sets, from which patterns and information are extracted, and which in turn allow personnel to make more effective decisions.

Utilities increasingly use analytics in the operation of their systems, with many utilities acquiring data from multiple sources, such as smart meters and distributed protection devices. Additionally, many utilities employ devices that record brief periods of high-speed waveform data, including but not limited to relays, digital fault recorders, and power quality monitors. Automated analysis of high-speed waveform data from these devices, known as waveform analytics, can provide substantial benefits to utility companies when integrated into their operational practice [1].

As utilities attempt to integrate additional sources of data, engineers and operations personnel become cognizant that 1) more data does not automatically imply more useful information and 2) data themselves are of varying quality. Engineers hoping to gain operational improvements through the use of waveform analytics may be aware that not all devices produce equally useful data, but are often unaware of and unable to judge the nexus of factors that contribute to the suitability of a particular device for use as a waveform analytics platform.

This paper discusses some of the major factors that impact data quality as it relates to to the application of waveform analytics.

KEYWORDS
Waveform analytics; smart grid; advanced monitoring; vegetation management; fault anticipation; incipient faults

jeffw@tamu.edu

Introduction

In industries as diverse as marketing and baseball, "analytics" has attained buzzword status in the past decade. In general, "analytics" refers to the process of acquiring and analysing large datasets, with the intent of extracting patterns and information that are not immediately obvious to a casual observer. In sport of baseball, as an example, analytics might reveal that a particular player has a tendency to perform worse than expected in games where his team returned home from a road trip the previous day. Knowledge acquired from analytics – whether in baseball or the energy sector – allows personnel to make more informed decisions, ideally leading to better outcomes.

Modern utilities face a deluge of data, from an ever-increasing variety of sources, including AMI, RTUs, smart sensors, or self-healing systems. The implementation of a "smarter" grid has expanded the number and types of data sources available to utilities, but the corresponding intelligence needed to extract useful information from these sources is often less developed. Many utilities collect terabytes of data, but have difficulty articulating the impact of such data collection on actual improvements to system performance. Conversations with utilities often reveal that they *believe* there is tremendous value in the data they collect, but also underscore the challenge of transforming that data into actionable information.

Waveform Analytics

One particular aspect of utility data analytics under active research and development uses high-speed waveform data to assess the health and status of both individual power system apparatus and the system as a whole [1]. This field, known as waveform analytics, utilizes a variety of techniques that enable a broad range of functionality, including fault location, incipient fault detection, and condition-based maintenance [2].

Data used for waveform analytics differ from data obtained by AMI/SCADA systems, which are typically single data points collected over a period of minutes. Instead, waveform analytics operate on a "snapshot" of high-speed waveform data obtained from waveform recording devices, for example, a digital fault recorder, power quality monitor, digital relay, or other dedicated recording device. Sophisticated algorithms process these "snapshots" and attempt to extract information from the high-speed waveforms, typically expressed as a description of the event (e.g. "Capacitor Switch", "Motor Start", "Failing Switch", etc.), along with useful features (e.g. "Phase B", "214 Amperes", "14 Cycles", etc.). Waveform analytics algorithms rely on the fundamental principle that electrical waveform activity recorded by monitoring devices is a representation of actual events occurring on the power system. Stated differently, when power system apparatus operate, both normally and abnormally, or begin to fail, they produce electrical transients on the power system that are detectable by waveform recording devices. By applying sophisticated algorithms to waveform data recorded with sufficient fidelity, utilities are able to extract actionable information, which in turn can improve the overall operation of their systems.

Data Quality for Waveform Analytics Inputs

Saying that inputs to a waveform analytics algorithm must be of sufficient quality to characterize the recorded event is a tautology. If input data does not contain useful information, then it is obviously not possible to extract useful information from said input data. It is the authors' experience, however, that many engineers are unaware of the range of factors which can impact the quality of data used as an input for waveform analytics, or how those factors interact with each other. When comparing particular recording devices, engineers are generally aware of key points on a product brochure – most notably sample rate and bits of resolution. Unfortunately, these are often the *only* technical characteristics considered by engineers when assessing the performance of a waveform recorder. While sample rate and bits of resolution are important factors to consider, they do not provide, by themselves, a sufficient basis for assessing the quality of a waveform input.

Data Quality: A Holistic Concept

When most engineers think about the quality of data recorded by a particular device, they often consider only a limited number of characteristics of the final, digital output file – most notably, as stated above, the sample rate and bits of resolution. In reality, these characteristics assess only one of the steps in the process of representing a physical electrical event as a digital waveform file.

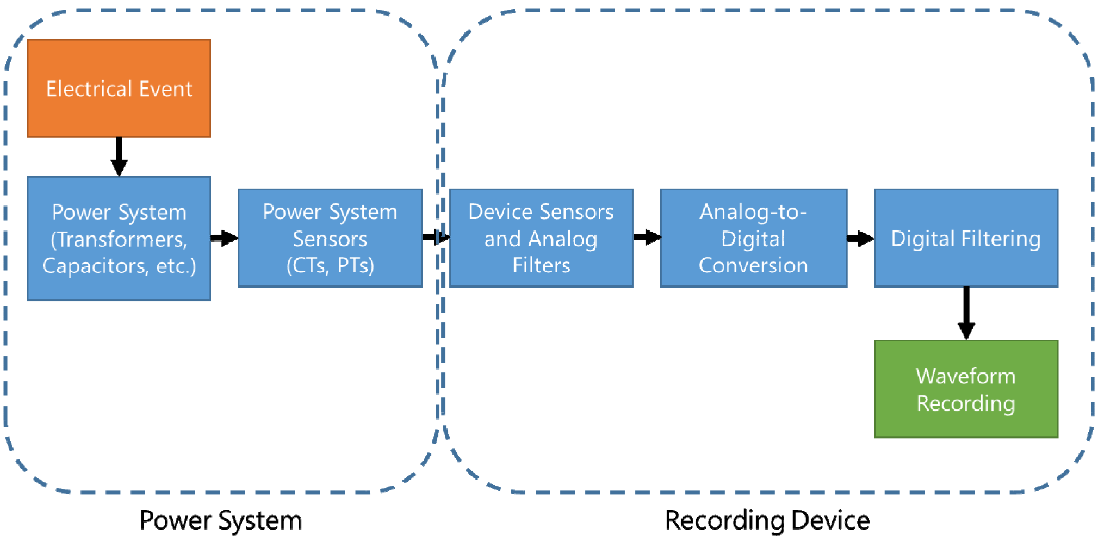

Figure 1: From electrical event to waveform recording

Error! Reference source not found. shows five functional blocks between a system event and waveform recording, each of which can have a substantial impact on the type and quality of data eventually represented in a file. The following sections expand on each of these blocks, and describe important considerations for each.

Data Quality: Power System Considerations

Waveform recordings are typically triggered due to the presence of a transient detected at the input terminals of the device. The presence of such transients, however, assumes that they have not been filtered or distorted by 1) other power system apparatus located between the event and the device (e.g. capacitors, power transformers, or even conductors), or 2) the power system sensors connected to the device itself (e.g. CTs, PTs, or alternative sensors).

Power system apparatus located between the physical location of the electrical event of interest and the waveform recorder have the potential to filter transients produced by the electrical event in question. This is an important consideration, because these filtering effects generally reduce the frequency bandwidth observed at the terminals of the device itself. Today's high-end power quality meters boast sample rates of up to 1,024 samples per cycle for current signals, and some meters sample at up to 100,000 samples per cycle for voltage signals. While such sample rates may (or may not) be appropriate if the recording device is measuring at the point of the actual disturbance, they are almost certainly excessive if the device is at a significant distance from the event (e.g. the substation). Indeed, even for events which contain significant high-frequency transients (e.g. capacitor switching), the vast majority of frequency content observed at a remote location is contained below 2kHz. As a result, arbitrarily increasing the sample rate of a recording device (e.g. from 256 to 1,024 samples per cycle) may result in four times as much recorded data but *no additional information*, if the frequency content of the disturbance – at the terminals of the device – would have been fully captured by the 256 sample per cycle converter. Moreover, if increasing the sample rate of the device has a negative impact on other aspects of the data (e.g. increased digital noise, reduced bits of converter resolution, etc.), increasing the sample rate can actually result in *less* useable information.

This concept is important enough and overlooked enough that it bears repeating. The sample rate of a waveform recorder must be high enough to characterize sufficiently the event in question. Arbitrarily increasing the sample rate of a device beyond this point, however, will have little if any positive effect, and indeed may have a negative effect on the overall quality of the recorded waveform.

Someone might respond that, "In general, the faster the sampling rate, the more accurate representation (waveform display) of the measured input signal." This assumes, however, that all other factors (most notably bits of resolution and bits of noise) remain unchanged as the sample rate increases. As a practical matter, this cannot be assumed. A 1GHz, 1-bit converter would not produce a "more accurate representation of the measured input signal" – assuming the measured input signal is a typical power system waveform – than a 960Hz, low-noise, 16 bit converter. Again, it cannot be overstated: the sample rate of a waveform recorder is only one factor that influences the accurate representation of a measured input signal.

The power system sensors used to connect the recording device to the system also merit consideration. The de facto standard for power system sensing has been and continues to be conventional current and potential transformers (CTs and PTs), but a variety of alternative sensing technologies have been developed in recent years which offer lower cost and claim equivalent sensing performance. While few comprehensive studies exist, limited research suggests that different sensing technologies should not be assumed to have similar performance [3, 4]. Figure 2 shows a simultaneous recording of current waveforms from a capacitor switching event, as rendered by both a traditional current transformer, and a popular alternative sensor. It is important to note that the same analog front-end was used in both recordings (i.e. they have the same sample rate and bits of resolution). This particular alternative sensor tends to amplify high-frequency transients as compared to the traditional CT, resulting in a substantially different waveform recording. While the traditional CTs output cannot be considered "correct" per se, the more important point is that waveform analytics algorithms designed to detect the transient as rendered by the traditional CT would likely behave in an unpredictable manner if they were given as an input the same transient rendered by the alternative sensor, and vice versa.

Again, it seems obvious to say that the quality of the digital output from a waveform recorder will only be as good as the quality of the analog signal at its input terminals. The authors' experience, however, suggests that engineers often assume that sensors will be "good enough" for a particular application, without fully understanding how different sensing technologies may impact results.

Data quality: Inside the Device

Inside the device itself, the analog signal path continues to be an important consideration. Unfortunately, users should be aware that stated compliance with IEC and IEEE standards is no guarantee of similar performance. As Delle Femine et. al. detail in [5], “different implementations of PQ monitors that fully meet definitions reported in [a list of IEC standards] can still disagree significantly in some actual measurements.” This occurs, the authors of [5] assert, because, “standards include performance specifications, but without a well-defined procedure for their verifications. [sic] The ambiguity contained in these procedures … [leaves] a certain degree of freedom to manufacturers.” In short, users should not simply assume that devices that meet identical IEC standards produce identical waveform recordings.

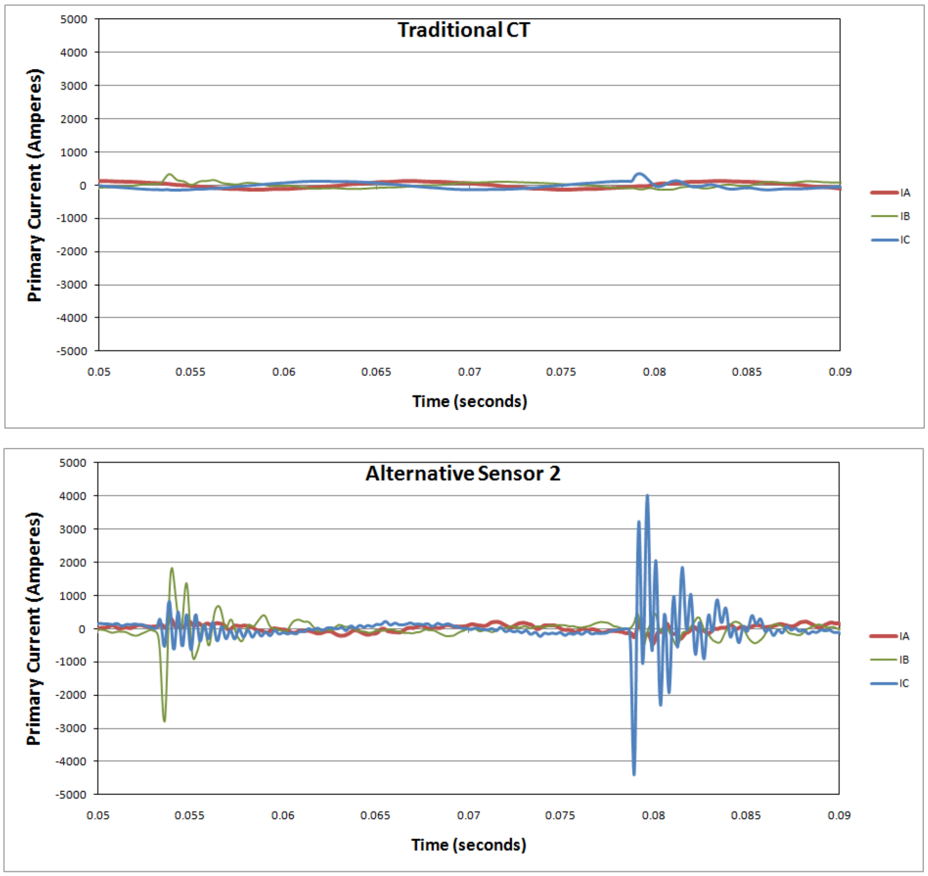

Figure 2: Simultaneous recordings of capacitor switching event from traditional CT and alternative current sensor (same analog front end was used – i.e. the graphs have the same sample rate)

Users assessing the viability of devices for use as waveform analytics platforms should also be aware of any intentional filtering in the analog section of the device. Some devices, most notably relays, are primarily interested in the 60Hz component of the signal, and may apply steep low pass filters before sampling. Obviously, any intentional filtering – whether analog or digital - may reduce the ability of a device to characterize an event sufficiently and accurately.

The functional block from Figure 1 most engineers are familiar with is the analog-to-digital conversion process. Even here, however, there are several important nuances that go beyond picking the highest sample rate and bits of resolution. As discussed previously, arbitrary increases in sample rate do not necessarily result in more information. In fact, sample rate is often *not* the most important consideration when evaluating a device, in the authors' opinion.

Engineers are also usually familiar with the bits of resolution provided by the ADC in a device (e.g. 12-bit, 16-bit, 24-bit), but may not appreciate that a device's *resolution* needs to be combined with its *range* to understand its ability to render waveforms accurately. Consider, for example, a waveform recorder with 16-bits of resolution scaled to measure 0-200 amperes of current (as measured at its terminals – i.e. the output of a 5A-nominal CT). Such a device will have sufficient *range* to accurately render high current faults on the power system, but may not have the *resolution* to accurately render minor, incipient failures such as the signature produced by a failing switch. A similar device with the same 16-bits of resolution scaled to measure 0-10 amperes of current would have the opposite problem: it would likely have the *resolution* to measure the small incipient faults, but would not have the *range* to measure high-magnitude faults without saturating. Ideally, of course, a device would have both the range to measure high current faults, and the resolution to detect minor power system variations. Unfortunately, such capability is rare in commercially available devices.

Additionally, digital and analog noise, introduced in a variety of mechanisms, reduces the number of effective bits of resolution for all analog-to-digital converter outputs. Users should be aware that a 16-bit converter will not provide 16-bits of noise-free output, and as such, will decrease the ability of the device to accurately render small signals.

Data Quality: Additional Concerns

The considerations mentioned up to this point have largely involved the quality of the recorded signal, but at least two other factors should be considered. First, many normal and incipient power system events (e.g. downed conductors, clamp failures, etc.) take place over an extended period of time – sometimes many tens of seconds. Many waveform recorders, by contrast, have maximum recording lengths for any given event typically measured in cycles, or at most a few seconds. For many event types, this limitation prevents these waveform monitors from fully characterizing the event, even if the device's signal path meets all necessary conditions.

Additionally, many incipient failure events produce only slight variations in current, and almost no perceptible variation in voltage. Because many normal power system events, such as motor starts, produce much larger variations of current and voltage, many waveform recording devices have triggering thresholds of several percent (e.g. 3%) on voltage, and large thresholds (e.g. 100+ amperes) on current, to reduce nuisance triggers. Waveform recorders not designed to cope with very sensitive triggering – either because they lack sufficient storage or processing power – will be limited in their ability to capture incipient failures.

Conclusion

Waveform analytics research has demonstrated substantial benefit for utilities and customers. These benefits, however, require competent platforms for data acquisition and processing. Many engineers are familiar with spec-sheet numbers including sample rate and bits of resolution for individual devices, but may be unfamiliar with the nexus of factors that contribute to overall data quality. Neglecting any aspect of data quality can result in waveform inputs that are unsuitable for waveform analytics algorithms. Engineers and researchers looking to utilize existing platforms for waveform analytics should consider all relevant factors in the selection of recording devices.

BIBLIOGRAPHY

[1] J. A. Wischkaemper, C. L. Benner, B. D. Russell, and K. Manivannan, "Application of Waveform Analytics for Improved Situational Awareness of Electric Distribution Feeders," *Smart Grid, IEEE Transactions on,* vol. 6, pp. 2041-2049, 2015.

[2] "Distribution Fault Anticipation: Phase III System Integration and Library Enhancement," Electric Power Research Institute (EPRI), Document #1016036, Palo Alto, CA. July 2009.

[3] "Comparative Characterization of Parallel Distribution Sensors Under Field Conditions," Power Systems Engineering Research Center (PSERC). August 2010.

[4] "Sensor Technologies for Real-Time Distribution System Monitoring," Electric Power Research Institute (EPRI), Document #1017832, Palo Alto, CA. December 2009.

[5] A. Delle Femine, D. Gallo, C. Landi, and M. Luiso, "Performance Analysis of Power Quality Monitoring Instruments," in *Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE*, 2008, pp. 2026-2031.