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

CIGRE US National Committee
2013 Grid of the Future Symposium

The Implication of Faulty Sensors to Asset Analytics in Producing Reliable Indicators of Asset Health

Don Angell
Doble Engineering Company
United States

Lee Margaret Ayers*
Doble Engineering Company
United State

SUMMARY

Automating the determination of asset health using analytics that combine and analyze all available data sources is a challenge in itself, but errors coming from the field make calculating asset health an even greater challenge. Sensors and online monitors can degrade over time, or become unreliable with intermittent good and bad values or simply fail, leading to false positives. In the real world, crews must often respond to false positives generated by failing equipment, which redirects valuable Operations & Maintenance (O&M) dollars away from assets that need attention. When creating an automated application that determines asset health, it is important that the application not generate too many or too few alerts, or the users will never trust the system.

Humans typically learn to ignore bad sources of data, but how do we teach an automated system to do the same thing, and can the automated system debug questions about sensors humans cannot? When automating asset analytics, and before one can establish asset health, criticality, and risks, data errors must be evaluated through a validation process. Beyond monitoring issues, intermediate delivery systems such as gateways, relays, intelligent electronic devices, applications, and databases can introduce errors to otherwise good data sets along the way. Unless data errors are trapped, eliminated or managed, those developing asset analytics can never be certain if their results are valid. The authors will review some of the key steps for validating data, site case study examples and briefly discuss how asset analytics provide an excellent business case for Smart Grid standards of IEC 61850 and the Common Information Model (CIM)) (IEC 61970 and IEC 61968).

KEYWORDS

Analytics – Asset Analytics – Online Monitoring – Condition Assessment – Smart Grid Standards – IEC 61850 – CIM – Strategic Management of Assets - Asset Risk Management

INTRODUCTION

layers@doble.com

Over the last five years, the industry has seen an evolution in Asset Management technologies. With data volumes increasing—plus an ever-increasing demand to better manage assets with fewer resources—companies are looking toward applications that will allow its key people to make more informed decision. In this regard, we believe advanced asset analytics will provide a revolutionary leap that gives companies better insight into asset health and enables its people to be more effective in targeting asset issues for an entire fleet.

The problem is, not all results from the analytic processes are necessarily true or relevant. Automating the determination of asset health using analytics that combine and analyze all available data sources is a challenge in itself, but oftentimes the sensors and online monitors (temperature, oil temperature, etc.) behind a good data source like Supervisory Control and Data Acquisition (SCADA) Systems produce errors that make calculating asset health an even greater challenge. This is because field data from devices such as transducers, online monitors, communication equipment, and computers can fail or deteriorate over time—thus providing SCADA (and eventually an analytic engine) with variations of good and bad data, which makes identifying true asset issues more difficult.

Beyond transducer and monitoring issues, intermediate delivery systems such as gateways, relays, Remote Terminal Units (RTU), or Intelligent Electronic Devices (IED) can introduce errors to otherwise good data sets along the way. There may also be issues with configuration (or inadvertent reconfiguration) of devices such as relays—or how data are managed by interfaces, intermediate software, applications, and databases. In the worst case scenario, errors could be introduced by multiple components. No matter where an error is introduced, if bad or inconsistent data are delivered to the analytics, the output (as it relates to asset health) will be highly suspect.

From the outside, analytics may seem simple, but to the newly initiated, or even an experienced subject matter expert (SME) that understands the relevance of one asset-analytic over another in determining asset health (which in itself is extremely difficult), very few people in the world understand the full extent of data-behaviour as it relates to data-errors. In the everyday world of day-to-day operations and maintenance, personnel must respond to data errors from the field that produce false positives. Those with deep knowledge of in-place hardware, sensors, online monitors, supporting applications, communication hardware, and networks learn to ignore sources of bad data (which might present as an alert or alarm). Those with less knowledge about field errors, but with responsibility to respond to alerts must investigate the issue—which draws attention away from assets that need it and negatively impacts O&M budgets. So it is imperative that the Automated Asset Analytic Application (‘4-A’ for sake of conversation) not similarly redirect resources, or generate too many or too few alerts, otherwise users will never trust the system. Humans may learn how to filter out noise and bad data—but not always. So the question is: can we program/teach an automated system to manage these false positives and can the automated system be used to debug questions humans cannot?

‘Analytics and Big Data’ are the power industry’s new Holy Grail, but just like the search for the Holy Grail lead to an illusory and endless chase, so too can the chase for truth be during the analytics-build. Unless data errors are trapped, managed, or eliminated, those developing asset analytics will never be certain if results from the 4-A are true. Many common errors can be trapped up front, which allows those who truly understand asset health to focus on asset issues.

The field of data validation is not new, but it has seen limited deployment. Within the power industry, data validation is used to identify errors from customer meters. Because meter errors negatively impact customer billing, typical validation checks include clock drift, zero value, and spike. Meter validation is nontrivial, but it is a thousand times *easier* than validating data from online monitors and sensors at the substation. This is because one type of customer meter is typically connected through one common communication network to one database, so known errors are more easily validated. But when it comes to asset management, and we consider all the possible combinations of equipment, databases, networks and devices for assets in the substation, and all the spots where errors might be introduced, developers will find the search for true asset health as elusive as the search for the Grail.

CHALLENGES WITH MOVING DATA FROM THE FIELD

The validation process would be far simpler if all we had to do was write analytics against expected, incoming data from each field device. The problem is, we often times do not know what to expect. If a Dissolved Gas Analysis (DGA) Monitor is *directly* connected to the 4-A (Automated Asset Analytic Application) via a single interface and a cell router using a DNP3 interface, then data can be validated with a known set of parameters (DGA Monitor to cell router that talks DNP3 to an interface (three points of failure)). When there are a limited numbers of variables to manage, there is much higher confidence in the validation results. These results then feed the asset health analytics which results in much higher confidence in asset health scores.

In reality, there are a number of pathways used to route data from the field. These pathways include a number of components such as transducers, wires, online monitors, interfaces, gateways, Intelligent Electronic Devices (IED's), Remote Terminal Unites (RTU's), relays, communication devices, networks, applications, databases, etc. The more components used in the pathway, the more errors a single measurement point (such as Phase A Current) can accrue along the way.

Because of its dominance within organizations, and in-place infrastructure for the delivery and management of control data (voltage, current, MW, MVAR, breaker status) from the field, Supervisory Control and Data Acquisition (SCADA) Systems have been used as vehicle for also moving *non-control* data (temperature, fan bank status, etc.) from the field, and as such, have become one of these routers of information the 4-A relies upon. And why not? SCADA is typically considered a reliable data source. These systems go through rigorous Factory Acceptance Tests (FAT). Field data update every one- to two-seconds and the values are well tested before they are used for control—and anything used for control has to be reliable, right? Yes and no.

The problem is SCADA Systems were designed with control in mind and are often tightly coupled with a mix of RTUs, IEDs, relays, or gateways which made source-data and infrastructure predictable for data analysis, but adding non-control data was not part of its primary design. Once a SCADA system is built, bandwidth for new data sources are typically limited, which in turn limits how much new data can be added from a new source—meaning only one field is available of the fifty the online monitor might provide. In addition to having limited bandwidth, source details such as vendor and model are typically dropped when they are linked to SCADA. When this happens, known variables about the monitor the 4-A needs for validation are lost. SCADA might see a Dissolved Gas Analysis (DGA) alarm point, but the viewers of the alarm will no longer know if it is from a single value DGA or a multi-value DGA—where each monitor's alarm(s) means something different. When device characteristics are lost, only the crudest of validation checks can be made, after which only general asset analytics can be written. As a result, our confidence in producing a good analytic is diminished and asset health scores can only be broad rather than specific.

Note: SCADA is but one of many in-place infrastructures that companies use to move new data sources from the field today. As each company evolves its validation practices, data errors from each of these systems will have to be resolved.

CASE STUDY—DATA PATHWAYS

One of the greatest hurdles to creating asset analytics is simulating accurate lab data from field. It is even more impossible to simulate data errors that reflect true field conditions, so it is important that the analytics have healthy variations of real data. Initial steps in the validation process include: 1) identifying the variety of communication device and pathways data take before getting to an Asset Analytic Application (4-A); 2) determining common, and uncommon errors pathway components produce; and 3) developing a set of analytics to deal with these errors. If done right, a number of validation routines (such as communication checks) can be used to validate all assets at once.

Figure 1 depicts the way data are sent from customer and Consortium Member-sites to a Data Center. Data are delivered using standard interfaces such as DNP3, Modbus+, IEC 61850, PI-to-PI, OPC, RDBMS, and Web Services. These interfaces reside on a corporate gateway device at the utility site and work in conjunction with VPN links for the secure transfer of data. At some sites, online monitors are directly connected to the 4-A via secure, cellular devices using DNP3 and Modbus interfaces. Each customer is then given a secure portal view to the 4-A results within an asset risk management system.

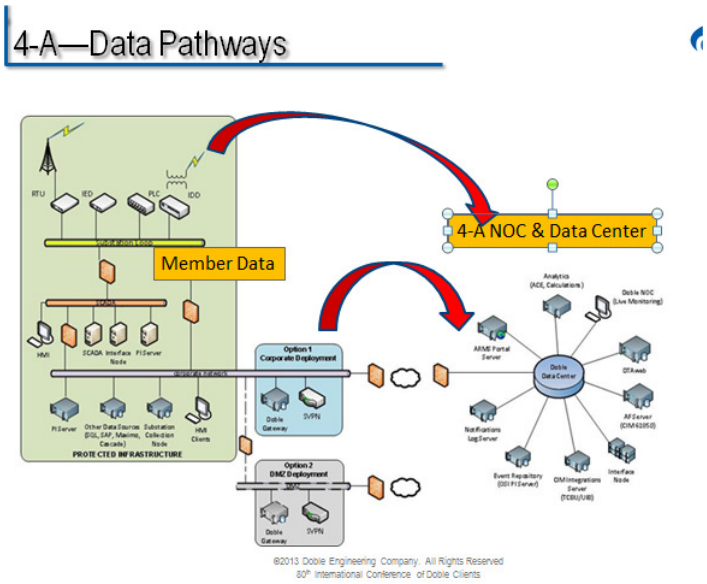

Figure 1: Secure data pathways from customer and consortium member sites via gateway to the 4-A Network operations Center (NOC) and Data Center (all part of the all part of the asset risk management system).

Before delving into the specific details of how and which data sources are being sent to the 4-A, it will be good to look at a simplified view of data pathways. Figure 2 shows three general configurations of how data can be moved to the 4-A:

- Direct connect, which is a one-to-one (1:1) relationship, whereby the DGA A Monitor is directly connected to the 4-A via a secure, cellular device using a DNP3 interface.
- Routed, which is a one-to-many (1:N) relationship. An example of this is to route the DGA A Monitor to a Remote Terminal Unit (RTU), then SCADA, then to an historical database, and then onto the 4-A. In addition to systems, note the number of interface touch-points—each a potential point of failure.
- Multiple data sources connect to one device or interface, which are then routed. This is a many-to-many (N:N) relationship. Multiple devices feed into one device (RTU) or interface (DNP3), after which the data are routed. A single DGA Monitor (DGA A) connects to SCADA at one sub, while a multi-value DGA Monitor (DGA B) may connect with one type of RTU at another substation, after which the data are sent to SCADA and then onto a database before being sent to the 4-A.

In reality, data pathways are more complex than shown. The DGA A Monitor could connect to an RTU in one scenario, to a relay in another and an IED in another. The protocols used to move data from one component to another are not shown, but will also need to be considered in the validation routines. Some of these communication devices have a shorter lifespan than others, so it is important that validation routines account for end-of-life patterns.

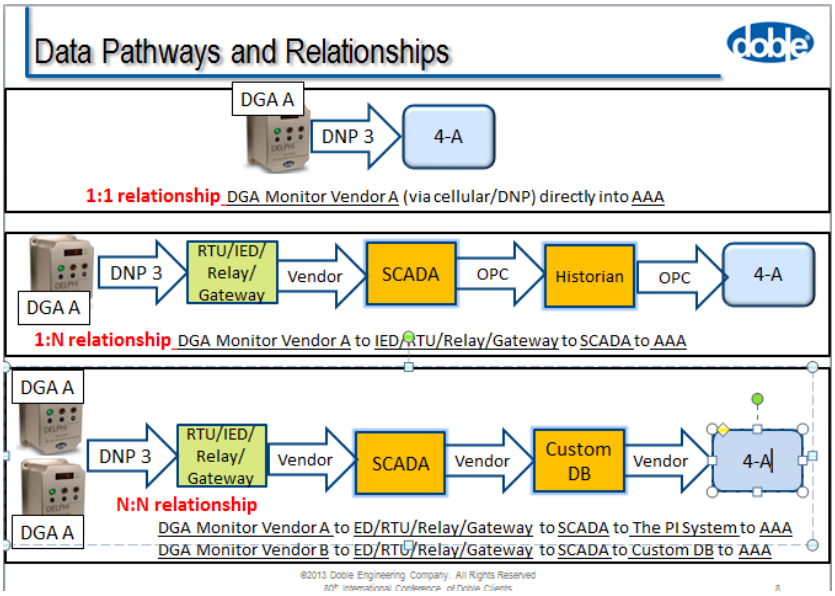

Figure 2: Data pathways and relationships showing 1:1, 1:N and N:N relationships.

Figure 3 shows the six different types of data sources currently being integrated into the 4-A and the arrows show a variety of ways in which these sources are routed. Sources of data include: offline diagnostic tests (bushing power factor), online diagnostics and monitors (DGA and Bushing), SCADA (real-time operational data), Insulated Fluids (oil test), Maintenance Inspections, and the an historical data base of apparatus test results, which contains over 80 years of apparatus diagnostic test results (and over 30,000,000 records) of assets by manufacturer from around the world. Red arrows and boxes indicate direct connections (1:1), while orange arrows and boxes indicated more complex pathways (1:N and N:N).

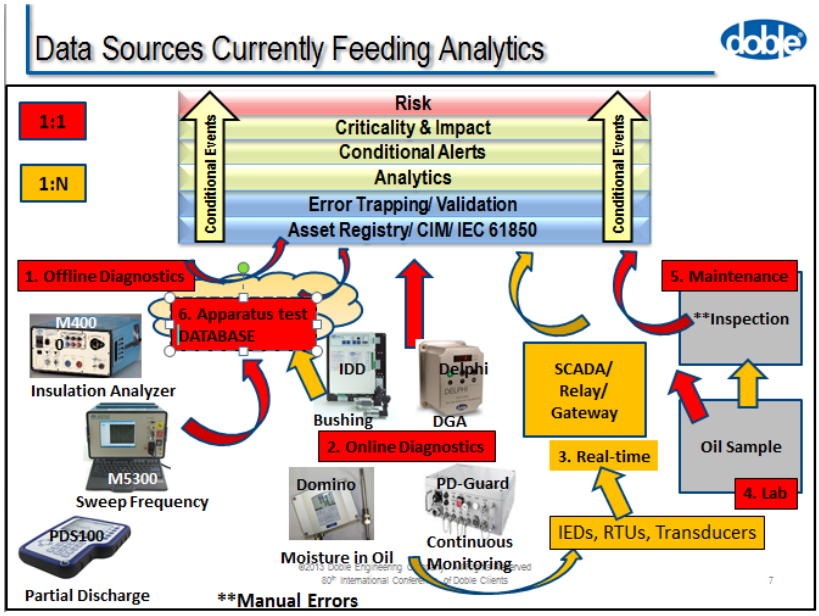

Figure 2: Data sources currently feeding the 4-A, depicting 1:1, 1:N and N:N relationships. Data sources include: 1) offline diagnostic tests, 2) online diagnostics and monitors, 3) SCADA/Relay/RTU/IED/Gateway, 4) Lab Tests, 5) Maintenance Inspections, and 6) an historical database of apparatus diagnostic test results (80 years of data and over thirty million records).

With the exception of the historical apparatus diagnostic test result database, any of the data sources can be routed via various pathways. Other delivery channels (not shown) include web services and integration bus technologies.

Beyond establishing data pathways, other things come into play that will help the validation phase. Some online monitors, devices and applications carry comms status values, quality bits, and in the case of SCADA, telemetry errors. Every measurement point that SCADA receives (like voltage) also carries one or many 'Quality Codes.' The number of Quality Codes can range into the hundreds; examples include: Generated an Alarm, Telemetry Error, Manual Entry, Good, etc. The challenge with using a quality bit is that a single measurement point can have multiple codes, but a SCADA engineer can provide validation process with some of the most valuable permutations. A simple validation check might be: if Phase A Current is out of range, then check if Quality Code = Telemetry Error. If this is true, and other Quality Codes for the same device can corroborate the error, then out-of-range Phase A Current is likely invalid.

The reasons for establishing data pathways are two-fold: one is for establishing points of failure, and the second is to highlight how current naming conventions fail us today. In the section on CIM/IEC 61850 harmonization, we will discuss the path forward.

THE ANALYTICS PROCESS—CASE STUDIES OF DATA VALIDATION

An area that adds further challenge is that monitored points might be the same between like assets (power factor), but the resulting values can be hugely different. Many things come into play, such as age, conditions under which the asset has been operated, factory baseline tests, etc. In this case, it takes a bit of science and then a bit of art on how to apply the science in order to determine what an assets condition really is. The reality is there are variations of good and bad data, so we need to establish confidence in the data sources, which means establishing your validation guidelines. Having manufacturer baseline test results for the asset is a plus.

The first step in the validation process is to establish expected ranges for values and behavior for incoming data. It is important to know how often the data present themselves—is it every two seconds or every hour? And if the data are not updating, then check the communications pathway. Data behavior fields include update rates (scan rates), expected ranges, engineering units, scaling, and what type of measurement points the data are—digital, integer, float, Boolean, etc.

Some of the analytics will rely upon high-speed data, but it is not economical to store all the values. Historical databases (or historians) typically have exception and compression features that must be well understood. This is where data tuning really kicks in. The goal in data tuning is to reduce the amount of data, using exception and compression, without eliminating details. The first few steps of the validation process tend to be iterative until a certain level of confidence in data tuning and validation are established. Until data behavior is established and data are tuned, there is no sense going further in the analytic process.

After data behavior is well established, it is important to verify there are no points of failure or delay between the data source and the 4-A .This includes intermediary pathway components previously discussed. One of the biggest culprits of comms failure is when a substation computer is rebooted, but the interface software is not. Lack of computer memory can also delay delivery. It is important to track the end-of-life traits of each component in the pathway (from monitor to relay) and characteristics of the communication network? Those who understand asset health will not typically have experience on the communication-side of things, so it is important that input from communication, engineering, interface, and database personnel are included in the analytics process.

Once the data are tuned, behavior is established, and all the comms-checks have been passed, data can then be further evaluated for missing points, spikes, out of range, or whether they are not changing.

Data can change and be within range but the readings and still be abnormal, such as when a transformer temperature-monitor sticks at a certain valid temperature. So it is important to incorporate checks that are specific to the source systems (or intermediary system's) behavior. Some of the data errors will be due to human error (especially from maintenance records), which are harder to track. Advanced algorithms will be able to leverage heuristics and pattern recognition of typical device failure scenarios.

If an asset has multiple data sources, it will be worthwhile to correlate analytic results with another data source. For instance, if SCADA sees a high number of Tap changes in thirty day period, this can be corroborated with the number of Tap changes recorded during inspection process. In this case, a questionable result would kick off the correlation—or it could simply be automated. If a breaker sees a fault, a correlation analytic can check to see if the transformer saw it too. Oil sample results can be compared to online DGA values.

YOUR BAD DATA POLICY

When your analytics suite first comes on line, be prepared to see more data errors than true asset issues. This is normal. Start-up is typically about correcting relay configuration and comms issues, or tuning data ranges and behavior, or correcting analytic tests, and it provides an opportunity to identify and track asset issues which result in false positives.

An important consideration is alerts and alarms from the field exist for a reason, so it is important that the automated system not pass these forward—even if the alert is considered a false positive. Some form of check-and-acknowledgement must take place. A scenario where this might happen is when a device produces both an alert and a telemetry error. The telemetry error makes the alert questionable. There is an ability to corroborate alerts with more complex analytics, so this is always an option, but as a practice, can we afford to ignore alarms? If so, then under what conditions can we hit the snore button, and when do we route alerts/alarms to someone who can take action?

Then, there is a question as to what we do with the bad data? Storing bad data will skew the analytic results for long-term health statistics. In the meter validation process, there are validate, edit and estimate (VEE)—whereby a single telemetry error is corrected so that holes do not exist in the data. This may be something that will be adopted for asset analytics—especially when loss-of-life calculations come into play.

One area most overlooked in the validation process is what happens if there is bad data, and they are reported to the front-end application? People tend to trust what they see in displays, so errors in data should be reported to the viewer. Companies' often leverage something called a "Watchdog" tag, which is an analytic that is checks the status of comms devices, and if there is an error, it is reported to the front-end application. The viewer knows not to trust the data.

One of the final big questions to answer is how and which alerts are moved toward a portal, and who sees them? And overall, a plan must be put into place for handling bad alerts and data (false positives), which then becomes part of company policy and the overall asset risk management application (of which a 4-A is a part). Do we send an alert to repair the issue, or ignore, or flag issue? And if the false positive is passed on, who sees it? A relay engineer is better equipped to resolve relay issues (versus asset issues).

IEC 61850 AND CIM HARMONIZATION

One of the most promising tools we have to facilitate the validation process and re-infuse the 4-A data model with the much needed source device/system details is to leverage a harmonized version of IEC 61850 and the Common Information Model (CIM) (IEC 61970 and IEC 61968). IEC 61850 naming conventions provide for source names (the names typcially dropped when data is linked to SCADA or other systems), while CIM provides an overarching data infrastructure for modeling assets and the

business. Combined together, these two smart grid standards almost provide a path for mapping data from a source to a target. Unfortunately, the two standards (to our knowledge) fall just shy of a full mapping. While IEC61850 gives us the source device, and CIM provides for the target (asset data base), naming conventions for the intermediary touch points between source and target are not available in the standards.

In a scenario where there is DGA Monitor A is tied to a Relay B, which is then tied to SCADA, and both source details (DGA and relay) are dropped, so that there are only the measurement data, the only way to compensate for the lack of information is to write a comprehensive set of validation routines that evaluate whether other data errors exist for both devices, and if those devices are seeing similar errors. Overall, the data validation process is better served if the missing source systems are captured in the asset model. If this occurs, then the validation checks can be more specific—but augmenting the details during integration puts a lot more work on the integration team. Preferably, one could augment the standard using conventions already in the standard (which will be discussed at the conference), however, if this is done today, it would no longer be a standard.

THE NEXT STEPS—ASSET ANALYTICS

So the true first step of the analytics process is *not* analyzing asset condition, but establishing and tuning data behavior, and then analyzing and validating the performance of monitoring equipment, communication processes (devices & networks) and intermediary applications. Once the validation process is well established, it is time to kick off analytics for asset health, criticality impact, and various risks. Data errors will continue to show up in the most unexpected places, so having a plan in place for handling false positives will speak to the success of your final product.