REALTIME-SUBSTITUTION MECHANISMS FOR MISSING FIELD MEASUREMENTS – FIELD-APPROVED METHODS AND EVALUATION RESULTS

Stefan KITZLER
TU Wien, AIT Austrian Institute of Technology GmbH
Austria
stefan.kitzler@ait.ac.at

Mark STEFAN
AIT Austrian Institute of Technology GmbH
Austria
mark.stefan@ait.ac.at

ABSTRACT

Measurement values become more and more important for the correct behaviour of automation systems in several domains. In low voltage grids, functions such as voltage control, topology identification or validation, as well as other monitoring functions demand for reliable measurement data. During field operation several failures might occur (e.g., communication errors, erroneous measurement devices) resulting in unavailable values. To ensure trustworthy functionality of the automation systems in case of failures, substitution values can be used instead of missing field values. In this paper, methods for calculating substitution values are presented, as well as the demonstration of the implemented method in the field, followed by an elaboration of the results.

INTRODUCTION

Due to the continuously increasing level of distribution and automation in the Smart Grid domain, safe and reliable systems become more and more important. Many automation systems must rely on a high quality of input data to fulfil their requirements such as control algorithms, monitoring tasks, or state estimation to name but a few in power networks. Furthermore, highly reliable and accurate measurement data is needed for settlement-processes to avoid inaccurate billing.

In general, physical data is measured by field measurement devices and transmitted via communication infrastructure to processing systems. During field operation, several failures might occur (e.g., faulty measurement devices, communication infrastructure failures) resulting in non-availability of measurement data. To improve the resilience of automation systems in times of missing field values, approaches for estimating these values can be applied.

In this paper, we present various approaches for determining substitution values in case of missing measurement values in distribution grids based on a communication middleware for intelligent secondary substations (iSSN). After deriving the requirements for a dedicated application in the field, we present details about the validation of the chosen method. The implementation was done in the research-project *Smart City Demo Aspern* (SCDA) and validated in a low voltage grid in Vienna (*ASCR Smart City testbed*).

State of the art

Since the application of Smart Meters in distribution grids is relatively new, activities on missing field values and the determination of substitution values are also relatively small. In the research project *"SmartCity Villach – Vision Step I"* [1], several methods for generating substitution values in the voltage domain are presented [2]. By using these methods well results could be achieved but they were not applicable in the concrete project due to minimal variations of voltage values.

In case of communication failures, pseudo-measurements can be used for state estimation with decreasing estimation quality when multiple meters fail at the same time [3].

Architecture

The application development and field operation environment bases on Gridlink – a communication middleware for intelligent secondary substations, already introduced in previous scientific publications [4-6]. Several applications, including a data storage module, Grid and Building Representation Modules (GRM, BRM), modules for topology identification and validation [7], and others have been successfully developed and tested in the field. Due to the demand of power measurement values to fulfil their functionality it is necessary to provide substitution values in case of failures and therefore, a new module called Unavailable Data Handler (UDH) was developed. This application is monitoring the data transfer between the so-called Grid Monitoring Devices (GMD) in the field and the storage module, where the measured power and voltage values are persisted. In the low voltage testbed, these values are measured every 150 seconds and forwarded to the storage module. If missing values are recognized – due to missing timestamps or too much time elapsed since the last recognized value – UDH starts calculating the substitution values immediately after the delay time. These values are already persisted when the next measurement values are expected. Details about different methods for creating these values are presented in the next chapter as well as quality criteria for the application in the field are derived.

METHOD DESCRIPTION

As already mentioned, the active power values in the dedicated testbed are measured by GMDs at several nodes within the network and persisted as timeseries in the storage module. To substitute missing power measurements, we elaborated *elementary concepts* and derived several methods. In the following section we give an overview about the concepts and the criteria for the evaluation of the selected methods, followed by a detailed description of the implemented method.

Concepts

In the following, the four elementary concepts are described in a rough way to get an idea of the basic principles for calculating substitution values in the application area or low distribution grids.

Nearest Neighbour Regression (NNR)

The primary idea was to substitute values by real measurements of other GMDs which had a similar behavior in the past. Therefore, we assumed that similar behavior in the past, implies a related one in the future, where similarity is referred to as closest gap (minimal mathematical difference of measurement values) between past measurements of two compared data profiles. This lead to the first concept of NNR, which can also be applied in various ways, leading to different interpretations of neighbors (nearest may refer to measurement, profiles, or network nodes). As already described, we utilized this principle to develop methods based on closest data profiles, by ranking the neighbors using the *Least Square Regression (LSR).*

Periodically Recurring Course (PRC)

In the analysis of historical low voltage grid measurements, we found another concept, by identifying patterns of recurring courses. More precise, the varying separates a day, most notably into a day-night-discrepancy. Therefore, we established the concept of a periodically recurring course. This approach assumes a trend in time series, a certain fluctuation is not predictable. To advance this method we distinguished between work days and weekends.

Periodically Recurring State (PRS)

Analog to the PRC, we introduced another concept by predicting a periodically recurring state (with reference to the power status) of the power grid. With the ambition to substitute missing values with recent measurements of the same profiles, we search for similar overall states. Hence we compare previous time steps with the actual one, using the LSR.

Linear Regression (LR)

Starting with the idea of NNR, we tried to improve the method by using several neighbors. Facing the uprising challenges of different power maxima and total energies of the profiles as well as the varying number of neighbors to consider for the calculation, we advanced the NNR to a Linear Regression in order to weight the profiles used for the substitution. A more detailed description is given in the implementation section.

Selection Criteria

Although there is a huge number of applications in the domain of collecting and using measurement data, some criteria regarding the quality of a substitution algorithm will stay the same within all of them – having a minimal deviation between the real value and the substitution value. Moreover, we introduced the following criteria for the *method evaluation* – ordered by their priority, highest as first one.

i. The primary goal is a minimal mean error between the estimated and the real values.
ii. Due to the fact that the arithmetical mean value is not sufficient in every case the distribution of the deviation should also be as small as possible.
iii. The number of outliers should be reduced.
iv. As some of the substitution methods are using additional information (e.g., topology information), the availability of this information and acquisition time must also be considered when choosing a method.
v. In addition to the last criteria point, with an equal priority, computational intensive algorithms might not be suitable for a real-time substitution, regarding the limited calculation capacity of the application architecture.

Based on the application (e.g., grid structure) the quality of the results might vary caused by (not-)existing relation between measurement devices, degree of grid coverage by GMDs, availability and amount of historical measurement data, or recurrence of system states, to name but a few.

Implemented Method

In this section we describe the implemented real-time substitution method for missing measurement values, which fulfilled the established criteria best. The basis for the procedure are the actual available measurements and a set of historical data processed with the concepts of nearest neighbor regression leading to a linear regression model.

The fundamental concept for the method derived of the grid's hierarchical topology and the arrangement of measurement points, which represent this structure. The electric power is measured with GMDs at the network nodes of the power supply system and the measurements are stored as a series in data profiles. By assuming that every node of the grid would be monitored by several values forming a profile, one missing measurement value could be reproduced almost exactly by the surrounded neighbor profiles. Note that on the application side an overall monitoring can't be assured, especially if more than one measurement is missing. From this consideration we derived the concept of nearest neighbor in a network sense for the substitution of missing values. Considering the concept of a nearest neighbor is not obvious anymore, because they do not need to be the closest node in the network structure and their weights might also be negative. An advanced search, rank and weight of nearest neighbors in an unknown network structure (at the application side) is implemented, using a linear regression analysis of historical profile data. This is particularly useful because grid nodes on top[1] of the topology might be higher in power consumption and total energy compared to lower ones, or vice versa in case of distributed generation. With the automatism of a regression analysis the handling of power difference as well as the number of nearest neighbors used for the calculation are self-controlled.

By using a *Linear Regression* model, the missing value at time t_i is substituted by a sum of weighted measurement values, calculated as follows.

$$\boldsymbol{y(t_i) = \sum_{j=1}^{n} w_j(t_i) x_j(t_i)},$$

where $y(t_i)$ is the substituted (missing) value, $w_j(t_i)$ are the weights, $x_j(t_i)$ is the measurement value for profile j and n is the number of available measurement values.

The weights were calculated by using the *Ordinary Linear Regression (OLR)*, which minimizes the equation above applied to the data profiles in the previous time range. To compute these fitting parameters a consistent dataset without gaps must be prepared.

POWER MEASUREMENT SUBSTITUTION

In the following, a brief overview about the previously presented application (UDH) – based on the selected method – deployed in the *Smart Grid testbed* of the *Aspern Smart City Research (ASCR)* in Vienna [8] is given and results from field evaluation are presented.

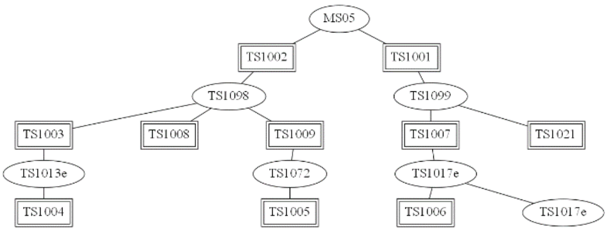

Figure 1 Grid topology graph (extract) used for validation of the implemented method. The nodes represent devices (e.g., transformer station, lines, busbars), whereas rectangular nodes indicate dedicated Grid Monitoring Devices (GMD) measuring the actual power.

Figure 1 shows an extract of the grid topology – presented as topology-graph – where UDH was deployed and validated in the field. It contains a substation node at the top (*MS05)* and other devices like lines, switches, or busbars. Components equipped with Grid Monitoring Devices are shown as rectangular nodes.

To validate the substitution method, the following scenarios were investigated within the field evaluation phase.

i. A *construction site* was assumed, causing an outage of two hours, repeated for several days to get a high amount of data.
ii. A *permanent failure* of a monitoring device was assumed resulting in a decreasing amount of measurement data available for the substitution.

During the time range of failure, no new incoming data can be collected for the substitution of the missing values. We use an internal data set of 5.000 recent time steps (150 seconds range between sequent steps), referred to the actual moment. Thus, especially for a permanent failure the amount of data is shrinking.

Construction site

In the first scenario a construction site for two hours at two particular nodes (TS1007 and TS1009, Figure 1) within the grid is assumed. To figure out the best interval for a construction site, the quality of the substituted method is evaluated during a day, by dividing the whole day into ratios of two hours.

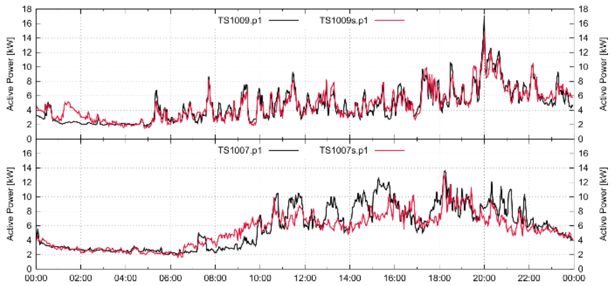

Figure 2 TS1009.p1: Active power (29th of June 2017), TS1009s.p1: Substitution values for TS1009.p1; TS1007.p1: Active power (10th of June 2017), TS1007s.p1: Substitution values for TS1007.p1

In Figure 2 the comparison of the active power and substituted values for phase one of two GMDs is illustrated. Due to the small electric power consumption the fluctuation is low in the first three stages of the day, which indicates a useful reparation time. On the downside, even a small variation in the course leads to a comparatively high error.

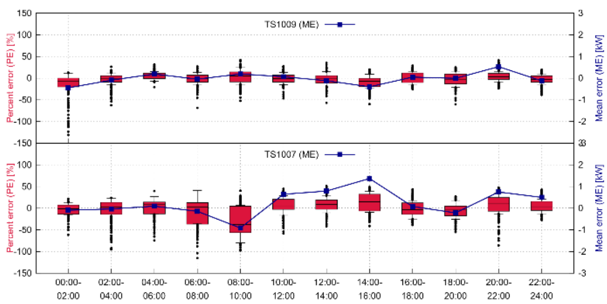

Figure 3 Permanent error (PE) in red, mean error (ME) in blue: Small deviation between power and substitution values for TS1009 with less than 1 kW mean error. An unmonitored network branch for TS1007 leads to higher altitudes.

This is cognizable in the outlier of the boxplots in Figure 3 (boxplot construction). For TS1009.p1 the mean error is constantly low with less than 1 kW. In the course of the day the quality of the substitution values distinguishes for

[1] Nodes closer to the substation than other nodes.

the two investigate GMDs. The deviation for TS1007.p1 is caused by one not monitored powerful network branch. Nevertheless, the mean trend of the curve is traced in both cases. These results leads to the conclusion that the best interval for a construction site depends on the magnitude of the error, which is highly related to the coverage of the high-powered network nodes. For poorly monitored networks, the construction time should be selected at night or in the early morning hours; for a total coverage by monitoring devices, the choice is uncritical.

Permanent failure

The ulterior motive of a permanent failure is to evaluate how a decreasing data set influences the quality of substitution values, till the internal data set of 5.000 timesteps is depleted after about eight days.

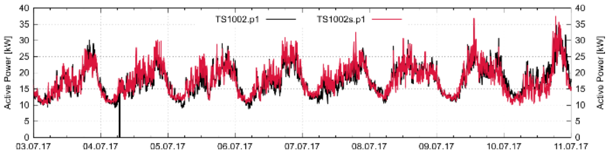

Figure 4 Comparison between measured power values (TS1002.p1) and substitution values (TS1002s.p1) when assuming a permanent failure

Figure 4 illustrates a comparison between the measured power and the substitution values during a week. For a more detailed analysis Figure 5 (boxplot permanent failure) provides boxplots for each day and mean error curve. We ascertain the trend of an increasing mean error, with a firstly unexpected reversal after the fifth day. But with the expertise that weekends perform different and the remaining reduced data set only covers this time section from the previous week, these proper results can be declared. Even so we conclude altogether a slightly increasing error for permanent failure, in a range almost less than 1 kW, referred to the mean error. As we can see from the boxplot the spread is quite small, with the borders of the box in the range of ± 10 %.
It should be mentioned that for long term failures, no substitution can be provided with this implementation after the data set has shrunken to zero.

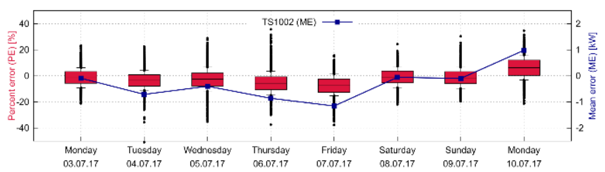

Figure 5 Permanent error (PE) in red and mean error (ME) in blue of the substitution values. Discern an increasing mean error because of the reducing data set at the permanent failure. Slump of trend concerning the small remaining data set and different behavior for the weekend progress.

CONCLUSION

In the consequence of increasing automation in distribution girds, reliable measurement values become more and more important. Due to communication problems or failed monitoring devices, these data might not be accessible. Therefore, algorithms calculating reliable substitution values are demanded. We presented some theoretical methods for determining substitution values as well as the deployment in the field, including die evaluation of the algorithm quality. As it turned out, our algorithm generates suitable substitution values for missing active power measurements in a low voltage grid. In general, it can be stated that a high coverage of network branches implies a higher quality of substitution values.

ACKNOWLEDGMENT

The presented work was developed in the *Smart Grid testbed* of the *Aspern Smart City Research (ASCR)* and conducted in the

i. *"iNIS"* project (849902), funded and supported by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) and the Austrian Research Promotion Agency (FFG), and in the
ii. *"SCDA – Smart City Demo Aspern"* project (846141), funded and supported by the Austrian Climate and Energy Fund (KLIEN).

REFERENCES

[1] R. Velik et al, 2013, "The Smart Energy Demo Project Vision Step I – SmartCity Villach".
[2] R. Schwalbe et al, 2016, „Smart City Villach: Final results of field test validation of a voltage estimation Application that supports distributed voltage control in times of communication loss*", CIRED Workshop 2016*, Helsinki, Finland.
[3] A. Abdel-Majeed et al, 2012, "Low voltage system state estimation using smart meters", *Universities Power Engineering Conference (UPEC)*, London.
[4] M. Faschang et al, 2016, "'iSSN Application Frame' — A flexible and performant framework hosting smart grid applications", *CIRED Workshop 2016*, Helsinki, 1-4.
[5] A. Einfalt et al, 2017, „Interaction of smart grid applications supporting Plug & Automate for intelligent secondary substations", *24th International Conference & Exhibition on Electricity Distribution (CIRED)*, 1257-1260.
[6] M. Faschang et al, 2017, „Provisioning, deployment, and operation of smart grid applications on substation level", *Computer Science - Research and Development vol. 32*, 117-130.
[7] M. Stefan et al, 2018, "Distribution Grid Topology Validation and Identification by Graph-based Load Profile Analysis", *IEEE International Conference on Industrial Technology (ICIT)*.
[8] Aspern Smart City Research GmbH & Co KG (ASCR), 2016, "SCDA – Smart City Demo Aspern," http://www.ascr.at/foerderungen/scda/, [Online; accessed 08-03-2018]