Advanced analytics at the edge

Andrej Čampa[1], Marko Hudomalj[1], Denis Sodin[1], Timotej Gale[1], Valentina Janev[2], Lazar Berbakov[2], Marko Batić[2],
[1] Comsensus d.o.o., Brezje pri Dobu 8, Dob, 1233, Slovenia
[2] Institute Mihajlo Pupin, Volgina 15, Belgrade, Serbia

andrej.campa@comsensus.eu

***Abstract*—This paper discusses the importance of Big Data processing and analysis at the edge of the electric power grid due to the increasing integration of renewable energy sources, electric vehicles, and new loads that enable a greener transition. The volume, velocity, and variety of data generated by the EPS is rapidly increasing and real-time processing and analysis is necessary to improve system efficiency, reliability, and security. This paper presents an advanced edge cloud computing framework that addresses quality of service challenges and tackles some of the Big Data challenges. The framework enables various instantiation scenarios and consists of open-source tools for managing and automating the Edge-Cloud infrastructure. A case study of a 50 kWp photovoltaic power plant is used to demonstrate the effectiveness of the framework in processing and analyzing data at the edge. Three different analytic tools are presented that address real-time and batch processing at the edge to offload data processing and data availability from the cloud to the edge. The paper concludes that edge computing plays a critical role in modernizing EPS and paving the way for a more sustainable and resilient energy future.**

***Index Terms*—Edge, Cloud, PMU, forecasting, predictive maintenance, Big Data**

I. Introduction

The electric power system (EPS) is undergoing significant changes due to the increasing integration of renewable energy sources (RES), electric vehicles (EVs), and the advance of larger new loads that enable a greener transition, such as heat pumps. As a result of the many new devices connected to the power grid, there has been a massive increase in the volume, velocity, and variety (3 V's) of data generated by the EPS that must be processed and analyzed in real-time to improve system efficiency, reliability, and security. This can be further extended to 5 V's with veracity and value [1]. Therefore, the role of Big Data has become very important and must be properly considered to improve the operation and protection of the system [2]. Consequently, the demand for reliable and efficient data processing and analysis at the edge of the grid is more important than ever.

The increasing use of electronic converters brings new challenges, especially in terms of stability, which need to be addressed [3]. Traditionally, stability is only evaluated through the parameters of the network, namely angle, frequency and voltage [4]. However, since the EPS system is live, a decentralized approach to Big Data processing and exploitation is desired. The results are then collected centrally and can be used to optimize the power efficiency and stability of the EPS. Therefore, edge computing becomes very important in a local micro-decision-making process, which can indirectly impact the whole EPS.

In previous work, we have already developed an advanced edge cloud computing framework [5] that addresses quality of service (QoS) challenges and tackles some of Big Data 5 V’s challenges. The main idea is to process as much data as possible at the edge and forward only the enriched data to the central or cloud computer. This data can then be used for better observability and decision-making on the EPS.

II. Edge-Cloud computing

By processing data at the edge of the EPS, advanced analytics and consequently operators can make faster and more informed decisions about grid operations, leading to greater efficiency, reliability, and resilience. Edge computing also enables new applications such as real-time monitoring of renewable energy sources, dynamic load balancing, and predictive maintenance of critical infrastructure. Edge computing plays a critical role in modernizing EPS, so we used it in our pilot and tested some analytical tools.

A. Pilot description

The pilot used for testing the created analytics at the edge and on the central computer belongs to the Institute Mihajlo Pupin (IMP). On the central computer for supervision and control, VIEW4 SCADA is used, which covers the entire energy value chain in Serbia and the wider region. However, our setup has focused on the observability and integration of edge nodes into the complete Edge-Cloud system.

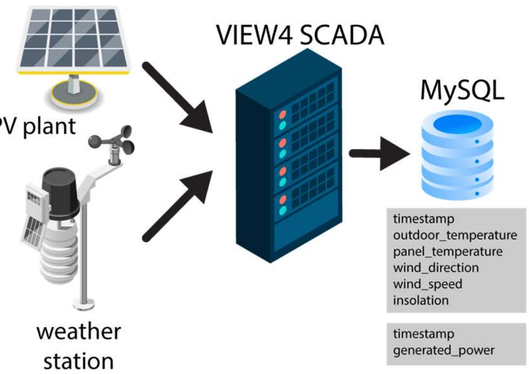

Figure 1. IMP pilot SCADA and MySQL database

At the edge, the 50 kWp photovoltaic (PV) power plant is installed on the roof of the building which was used for advanced analytics and Edge-Cloud case study demonstrations.

In the measurement cabinet next to the PV inverters, we installed an edge computer running Debian Linux and a Phasor Measurement Unit (PMU) to study grid behaviour and the impact of RES on the grid. The PMU is an instrument that measures voltages, currents, angles and frequency in real-time using a very accurate GPS clock. In our case, we used a PMU with a reporting rate of 50 Hz. This means that the Edge can process a lot of data that in many cases has no added value for the rest of the system. Due to the high frequency, accuracy and real-time value of the PMU data, the data should be processed at the Edge [6] to reduce delay and reduce the amount of data to be transmitted. This comes into consideration even more when there are many Edge nodes. The edge computer was located inside the IMP private network. However, the central computer in our case, had a MySQL database that was separate from the security sensitive VIEW4 SCADA system and was an intermediate point between the networks, as it is shown in **Error! Reference source not found.**. Namely, VIEW4 SCADA collects hundreds of datapoints from the pilot, where just few of them are relevant for the analysis presented in this paper. These data points, related to PV plant energy production and local weather measurements are fetched from SCADA and stored in MySQL database on a daily basis. From here on, the intermediate computer will be referred to as the central computer, since it has all functionality of the central computer and the Edge-Cloud framework was also installed on it.

B. *Edge-Cloud framework*

For the purpose of handling the algorithm development and deployment, we have developed an open-source Edge-Cloud framework[1] that enables various instantiation scenarios. The framework consists of well-maintained and open-source source tools such as Portainer[2], Rundeck[3], Munin[4] and Docker Swarm[5] that serve as tools for managing and automating the Edge-Cloud infrastructure. For security, the framework configures the tools and sets up SSH keys and an optional VPN.

Figure 2 shows the Edge-Cloud framework with analytics services at the edge. As shown in the figure, the PMU streams real-time data to the influx DB via the Phasor Data Concentrator (PDC). From the influx DB, the real-time data is available to the services at the Edge. As described in the next section, the services process the real-time data or data batches. The services communicate with MySQL DB, where they write

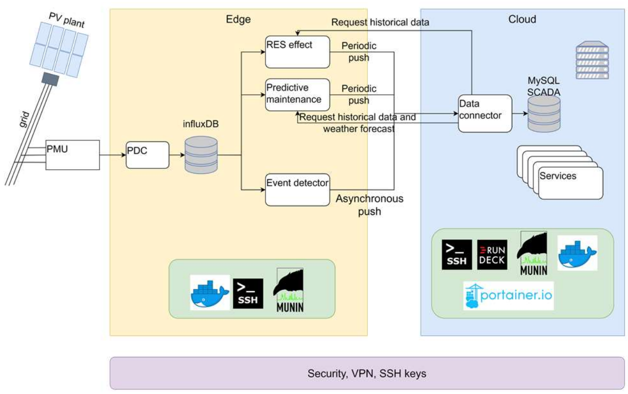

Figure 3. Edge-Cloud framework and analytic services at the edge.

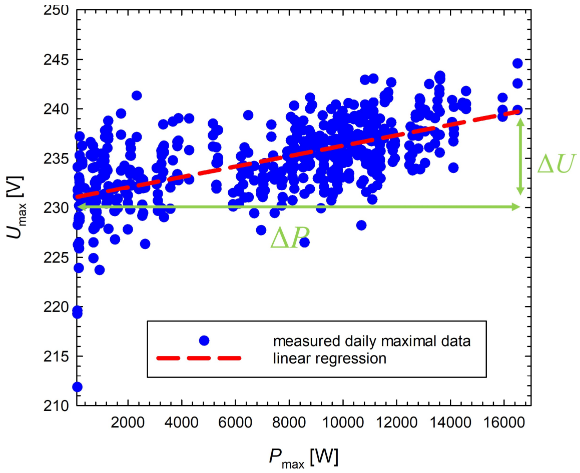

Figure 2. Measured daily maximal values of Voltages at with maximal daily PV production. edge.

the results and retrieve some of the historical values. Big Data analytics are performed on this data, which is outside the scope of this paper. It is worth noting that the live data at the edge can be stored only for a certain period of time. This adds an extra dimension to the analytics at the central layer as all data is available when needed. In most cases, once analytics detects an anomaly in the data, processing or performance, it can retrieve data for further processing, which can better support decision-making at the central level.

III. ANALYTICS

As mentioned in the previous section, we can divide analytics into analytics at the edge and central layer. Each layer has its advantages, especially at the edge real-time processing and control are preferable, while the central layer has good observability of the EPS and is able to process a large amount of data. With these considerations and capabilities of Edge-Cloud framework, the following services were developed:

- RES effective tool
- Predictive maintenance tool (with PV forecasting capabilities)
- Event detector tool

A. *RES effective tool*

One of the challenges today is the high penetration of RES in the low voltage (LV) grid. If not well designed, this can lead to power quality (PQ) issues. Unfortunately, to understand the LV grid, one must be able to fully describe its complexity (transformers, topology, lines and loads), dynamic behaviour due to real-time interactions between variables affecting the system, availability of real-time and historical data, and have sufficient computational resources for simulations. Unfortunately, this process must be repeated for each LV grid of interest.

1 https://github.com/PLATOONProject/edge-cloud-framework
2 https://www.portainer.io/
3 https://www.rundeck.com/
4 http://munin-monitoring.org/
5 https://www.docker.com/

Here we present a data-driven method that uses the PMU at the point of installation for grid and power injection observability. We used the PMU because it provides real-time data with a reporting rate of 50 Hz. Unfortunately, smart meters (SM) report aggregated data only hourly, 15 minutes or even 1 minute in some cases [7]. Already installed SMs usually do not have such a high reporting rate and should be reconfigured by DSOs to enable faster telegrams. However, even a granularity of 1 minute may not be sufficient to allow real-time observability. Therefore, we relied on PMU data for proof of concept.

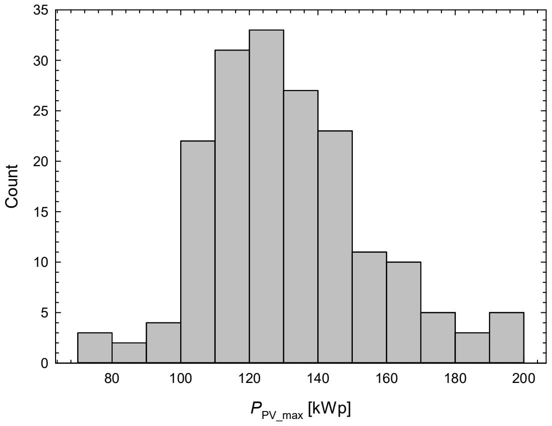

Figure 4. Histogram of maximal PV size for each day.

The main goal of the tool is to estimate the grid capabilities at the RES integration point using the grid observability data and production data collected by the PMU. The main idea is to monitor the grid and estimate each day the highest voltage (U_m) with and without the RES. Once these voltages are collected for each day, the worst-case scenario is plotted against the maximum RES production (P_{max}), Figure 3. Due to the stochastic nature of the low voltage grid, which is affected by both the situation on the LV and medium voltage (MV) side, we use only the daily maximum values. Applying a linear regression to the data, we obtain the slope (k):

$$k = \frac{\Delta U_m}{P_{max}} \tag{1}$$

which is related to resistance of line (R_{line}) according to (2).

$$R_{line} = \frac{\Delta U_m}{I} = \frac{\Delta U_m}{P_{max}} U_m = k U_m \tag{2}$$

In our case, the calculated R_{line} value is about 0.12 Ω, assuming nominal voltage (U_n) of 230 V for U_m. This is very close to the expected values that can be estimated from the characteristics and lengths of cables to the substation. This value and the maximum daily voltage without PV are used to estimate when we would reach the highest allowable voltage on LV grid by PQ standards ($1.1 U_n$). Using the histogram in Figure 4, we can see the typical values for the maximum PV size. If we are conservative and take the lowest value, the size of the PV power plant could be increased by 50%. There are many other measures, not considered here, that could allow us to use an even larger PV plant.

B. Predictive maintenance tool

The predictive maintenance tool consists of two parts. The first part is executed once per day and estimates PV module degradation. This can be used to create data-driven PV forecasts that change with the performance of the modules. The other part is a real-time monitoring and labelling tool described in Section III. *C*.

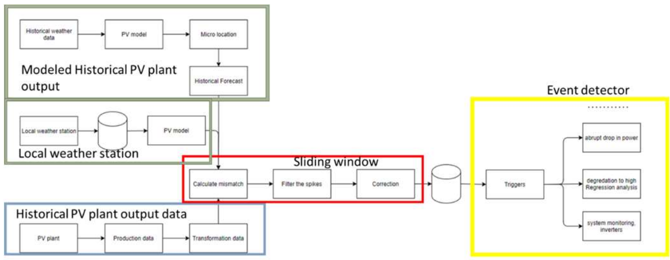

Figure. 5. Block diagram of PV predictive maintenance tool.

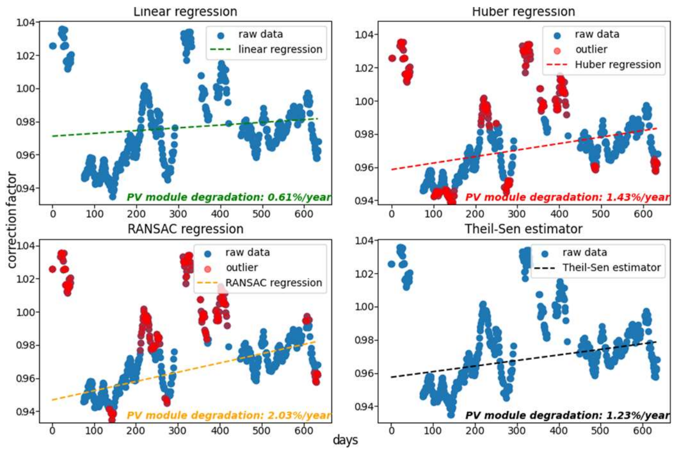

Figure 6. Estimation of PV modules degradation from c.f. using different linear regression estimators.

For good PV forecasting and predictive maintenance, the installation details such as nominal power, orientation and other design parameters such as the type of modules and inverters are needed. Usually, not all parameters are known, moreover PV modules degrade during the long operation time [8]. Therefore, a more robust data-driven model that can compensate for degradation and the absence of installation parameters is needed. Here we present such an approach.

The block diagram of the predictive maintenance tool is shown in Figure 4. It consists of two main branches, the modelled historical PV plant and the measured PV plant output. The branch that models the historical PV plant output is using simple PV model, enhanced with micro location adjustment. The adjustment is taken from the weather forecaster insolation data for specific location, which is adjusted to the micro location of the PV system, by using historical satellite values[6]. In this way shading in valleys, for example, can be effectively

6 https://joint-research-centre.ec.europa.eu/pvgis-online-tool/getting-started-pvgis/api-non-interactive-service_en

accounted for. By comparing the modelled output and generated output for some time using filters and weights we obtain the correction factor (*c.f.*) that is used in the PV model to improve the forecasting.

Since there are seasonal fluctuations in the power performance related to the change in spectrum, we would need a longer time period to better estimate the small changes in PV module degradation. Nevertheless, we can estimate the degradation for a period of less than 2 years, for which we used different regression estimators (Figure 6). Since *c.f.* contains the information on degradation of PV modules and location related power losses, it can be used to improve the simple PV forecast model. In Figure 7, we compare the forecasted values with the measured values. Good agreement is obtained, however due to local effects (e.g., obstruction of irradiance by local clouds) only averaged power performance (without spikes) is obtained with the use of general weather forecast for the pilot area.

C. Event detector tool

The Event detector tool is a real-time tool and is shown in the yellow box in Figure 5. The tool monitors in real-time the voltages, currents and powers of all three phases, which are received from the influx DB. The tool can detect anomalies in the production of the PV plant, it characterizes the anomaly, labels it and sends the alarm to the central MySQL DB, from where appropriate actions can be taken. It can detect power drop, unbalanced power production (e.g. inverter failure), lower production (module/string failure). Since no failure was detected during the runtime of the experiment, the PMU measurements were artificially distorted to test the detection of unbalanced power generation. The event detector tool generated an event with the label as shown in Figure 8. From the alarm we can see that inverter 1 has a problem in phase 2 (P2).

IV. CONCLUSIONS

With the increasing integration of renewable energy sources, electric vehicles, and larger loads such as heat pumps, the EPS is undergoing significant changes. The role of Big Data in improving the efficiency, reliability, and security of the power system is becoming increasingly important. The use of electronic converters has brought new challenges that need to be addressed, especially stability. Edge computing is becoming increasingly important in the local micro decision-making process, leading to a greater indirect impact on the power system. Edge-Cloud framework developed for this purpose enables various instantiation scenarios that algorithms can benefit during the design process. The pilot described in this paper demonstrates the successful integration of edge nodes into the complete Edge-Cloud system. By leveraging real-time data from PMUs and Big Data analytics, the EPS can be better optimized and maintained to create a more sustainable and resilient energy future.

We have presented three tools developed based on real-time data obtained at the edge. Data at the edge not only contains information about the grid, but is also heavily influenced by nearby connected assets. This has given us the opportunity to develop specialised tools that can give us advantages at the local level, at the community level, or at the DSO level.

With penetration of SMs into the LV grid, the feasibility test of replacing PMU measurements with the SM measurements should be done. This would make the tools much more accessible to a broader group of interested parties, as the investment costs would be lower. In the existing tools, we have interpreted the results very conservatively, therefore we expect performance should not deteriorate too much if a smaller amount of data is available from the SMs. For example, the SM based RES effect tool could be in the end used by DSOs to estimate the capability to integrate new RES into the grid without modelling and simulating the grid, which takes a lot of

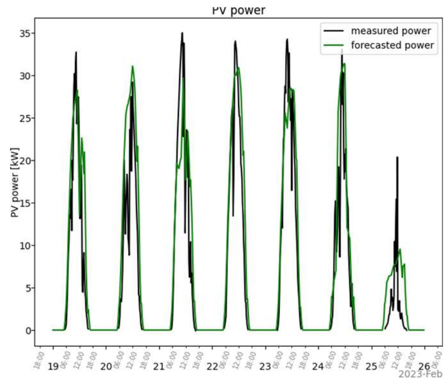

Figure 7. PV forecasted results (green curve) compared to measurements.

plant_id	timestamp	alarm
1	11.11.2022 15:16:14 000	P2 inverter issue

Figure 8. Alarm in the MySQL DB generated by event tool.

resources.

All three tools performed well and the developed analytics is not computationally or resource intensive, and could be easily deployed.

REFERENCES

[1] C. K. Leung, D. Deng, C. S. H. Hoi, in W. Lee, „Constrained Big Data Mining in an Edge Computing Environment“, v *Big Data Applications and Services 2017*, Singapore, 2019, str. 61–68. doi: 10.1007/978-981-13-0695-2_8.

[2] M. Kezunovic, L. Xie, in S. Grijalva, „The role of big data in improving power system operation and protection“, v *2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid*, avg. 2013, str. 1–9. doi: 10.1109/IREP.2013.6629368.

[3] D. Serrano-Jiménez, E. Unamuno, A. Gil-de-Muro, D. A. Aragon, S. Ceballos, in J. A. Barrena, „Stability tool for electric power systems with a high penetration of electronic power converters“, *Electr. Power Syst. Res.*, let. 210, str. 108115, sep. 2022, doi: 10.1016/j.epsr.2022.108115.

[4] P. Kundur *idr.*, „Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions“, *IEEE Trans. Power Syst.*, let. 19, št. 3, str. 1387–1401, avg. 2004, doi: 10.1109/TPWRS.2004.825981.

[5] D. Sodin, U. Rudež, M. Mihelin, M. Smolnikar, in A. Čampa, „Advanced Edge-Cloud Computing Framework for Automated PMU-Based Fault Localization in Distribution Networks“, *Appl. Sci.*, let. 11, št. 7, Art. št. 7, jan. 2021, doi: 10.3390/app11073100.

[6] „Learning-based Multivariate Real-Time Data Pruning for Smart PMU Communication“. https://ieeexplore.ieee.org/document/9700543/ (pridobljeno 7. marec 2023).

[7] M. Pau *idr.*, „Design and Accuracy Analysis of Multilevel State Estimation Based on Smart Metering Infrastructure“, *IEEE Trans. Instrum. Meas.*, let. 68, št. 11, str. 4300–4312, nov. 2019, doi: 10.1109/TIM.2018.2890399.

[8] D. C. Jordan, S. R. Kurtz, K. VanSant, in J. Newmiller, „Compendium of photovoltaic degradation rates“, *Prog. Photovolt. Res. Appl.*, let. 24, št. 7, str. 978–989, 2016, doi: 10.1002/pip.2744.