FORMALIZING EXPERT KNOWLEDGE IN ORDER TO ANALYSE CERN'S CONTROL SYSTEMS

A. Voitier, M. Gonzalez-Berges, F.M. Tilaro, CERN, Geneva, Switzerland
M. Roshchin, Siemens AG, Corporate Technology, München, Germany

Abstract

The automation infrastructure needed to reliably run CERN's accelerator complex and its experiments produces large and diverse amounts of data, besides physics data. Over 600 industrial control systems with about 45 million parameters store more than 100 terabytes of data per year. At the same time a large technical expertise in this domain is collected and formalized. The study is based on a set of use cases classified into three data analytics domains applicable to CERN's control systems: online monitoring, fault diagnosis and engineering support. A known root cause analysis concerning gas system alarms flooding was reproduced with Siemens' Smart Data technologies and its results were compared with a previous analysis. The new solution has been put in place as a tool supporting operators during breakdowns in a live production system. The effectiveness of this deployment suggests that these technologies can be applied to more cases. The intended goals would be to increase CERN's systems reliability and reduce analysis efforts from weeks to hours. It also ensures a more consistent approach for these analyses by harvesting a central expert knowledge base available at all times.

INTRODUCTION

CERN employs about 600 industrial Supervisory Control and Data Acquisition (SCADA) systems for the supervision and monitoring of its accelerators, detectors and infrastructure machines. While the day-to-day operations are running smoothly, a growing need appeared to exploit the data generated by these applications. Indeed, collectively they produce more than 100 terabytes of control data over 45 million parameters. The gathered data could be seen as a deep reflection of the current state of the processes under control. A lot of information about the performance, stability and overall behaviour of the machines resides within these data. Today, an expert can manually follow the signals deemed important and apply his/her knowledge to maintain a good level of service. However, current tools for industrial control systems are not properly designed for doing such dedicated analysis. The external analysis tools used today are not well integrated with the operator applications. Moreover the size and complexity of some systems does not allow running advanced data analytics methods in normal office computers. In addition, this way of working is a non-sense for analyses that have to be part of the operational tools themselves. Lastly, while the experts are knowledgeable about their domain, they are not properly skilled to perform these data analysis or computing problems tasks.

In this context, formalizing expert knowledge means capturing the methods and knowledge used by experts and transforming them into analyses that can be scaled out to all similar systems without burdening the users with a vast amount of manual operations to carry out.

From a computer science point of view, this problem is part of the Data Analytics, or Big Data, field which combines technologies for processing vast amount of data carrying out analytical tasks tailored in our case to industrial control system needs

Once an analysis can be applied to solve a control problem, it becomes part of the control system itself making possible to raise awareness of operators on specific issues from the supervision application they are accustomed to.

This paper will present the knowledge capture and the methods used to detect specific conditions traditionally done "by hand" (exporting data, importing them into a spreadsheet software or writing an adhoc script) or by looking at trends. Then, we discuss briefly their implementation as analytics tools that were then scaled out and applied automatically with the help of readily available software solutions.

CONTROL SYSTEM DATA ANALYSIS

Control systems play an essential role to run the CERN accelerators complex and generate a huge amount of data that can be used to analyse the behaviour of these systems and to find insights useful to operators and experts. The data analytics activities have been divided into three different categories:

- Online monitoring
- Fault diagnosis
- Engineering design

These three families of analysis focus on control data to offer analytical services as added value on top of the traditional industrial services.

Online Monitoring

Currently the monitoring and operation of CERN industrial systems is – for most parts – achieved through the deployment of specific applications based on the commercial SCADA software WinCC Open Architecture [1].

Therefore, these data analytics activities aim to add new features and services to this monitoring layer. This close integration allows reaching a high level of automation

ISBN 978-3-95450-148-9

necessary to provide both real-time and historical information on their performances.

Events Threshold Learning Analysis

Most events produced by the control systems are already filtered out and only the most relevant are checked by the operators in order to take appropriate actions. Nevertheless, the rejected events that cannot be handled manually may contain valuable insight on the status of the running systems [2].

The main objective here was to analyse these hidden data to provide operators with valuable information improving their understanding of the systems.

An online analysis system - based on Complex Event Processing (CEP [3]) engines - has been designed and developed to continuously collect events generated by each device. These messages have been clustered by type to avoid replicas and to produce structured data

Then, the generated data stream has been parsed by an online learning algorithm in order to discover behavioural patterns [4][5].

Specifically, the analysis goes through two different steps. Firstly, the stream of events is used to learn the amount of messages produced by single devices in normal conditions. Secondly, the algorithm uses the learnt behaviours to detect anomalies [6], with the assumption that faults in the system would result in the generation of an increasing number of messages (as shown in Fig. 1).

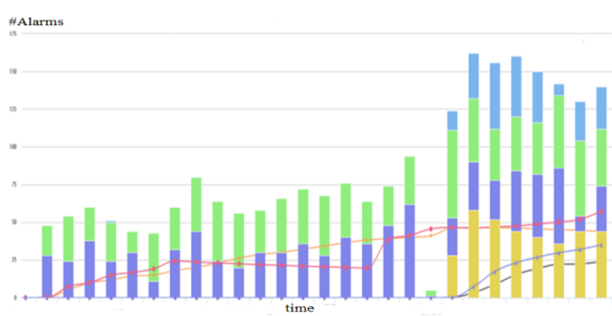

Figure 1: Cumulative alarms number and learnt threshold.

The nature of CERN systems, continuously updated both in hardware and software, imposes that the learning phase runs continuously, trying to detect new behaviours.

Signal Oscillation Analysis of the LHC Cryogenics System

The signal oscillation detection analysis [7] represents a second example of online monitoring activity [8]. Unlike the previous scenario where textual events were analysed, this time it makes use of numerical data (such as pressure values or valve position). The analysis can be applied to any control system where signal oscillation detection [9] is of interest.

For instance, in the cryogenics systems several abnormal behaviours have been identified by inconsistent sensor readings when the system was running smoothly. In this specific analysis the attention was focused on the detection of the oscillation of control valves. Under nominal conditions the process values oscillate, causing the valves to open or to close as expected. However, for various reasons these valves may start oscillating with an unexpected frequency or amplitude causing hardware damage.

The developed algorithm follows the analysis flow shown in Fig. 2. It consists first in a univariate signal analysis with a sliding time window. Then, the discrete Fourier transform is calculated to detect possible peaks in the spectrum [10][11][12] comparing each component against a given threshold. This threshold is calculated on the base of expert knowledge with a shape of a logarithmic function, which best fits the frequencies component amplitudes.

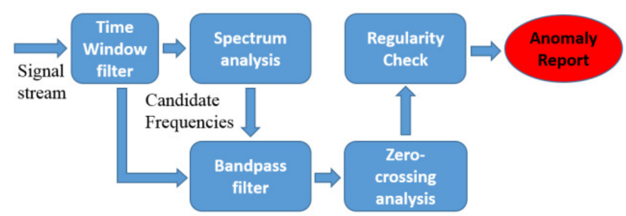

Figure 2: Analysis flow of the oscillation detection algorithm on a sample signal.

To verify it the initial signal is given as input to a band pass filter built on the previously discovered candidate frequencies. The frequencies that have passed the precedent conditions are further analysed. A zero-crossing analysis [13] is applied to the demeaned version of the filtered signal to check the presence of oscillations. This condition is verified if the number of zero-crossing is higher than a parameterised threshold, the zero-crossing rate. This rate is proportional to the frequency under analysis: lower frequencies will be associated to lower zero-crossing rates. As a final step the regularity of both the oscillation period and amplitude is checked by comparing the standard deviation of the period/amplitude with its mean. After an initial tuning the above algorithm has been able to detect the presence of multiple oscillations [14] and found their relative oscillation periods by recognizing regular patterns in the analysed data.

Parallelization of the Control Analysis

The huge amount of data produced by the CERN control systems will make necessary to run the developed algorithms against a large dataset of control signals and events. A cluster-based computing approach would be able to handle the enormous computation needs. This is the reason why a Docker-based cluster solution has been designed to parallelize the execution of such algorithms, showing the positive benefits of scaling the analysis across multiple nodes. Moreover, the lightweight portability of Docker containers minimizes the deployment issues linked to differences in execution environments.

ISBN 978-3-95450-148-9

Fault Diagnosis

This category includes the activities which perform an analysis posteriori to a fault occurrence. Therefore, it mainly aims at finding out the possible reasons for malfunctions looking into historical data. However, due to the complexity of the analysed systems it is generally not possible to identify the real initial cause factors. Nevertheless, it is necessary to reach a convenient degree of accuracy (diagnostic resolution) to which faults origin can be located to guide operators to relevant events/anomalies which need to be further investigated. As an example of this analysis category the GAS system alarms avalanche is presented.

Use Case: Gas System Alarms Avalanche

Supervision systems trigger alarms when process variables are out of their acceptable ranges. Alarms are the visible symptoms of an anomaly that may have occurred even several hours before, but they do not pinpoint to the exact root cause of the problem.

When a fault appears many alarms and events are raised rapidly by the system due to interaction between the different sub-systems. Moreover, the 1st raised alarm is not necessarily the most relevant to identify the root cause of the problem. Thus, alarm and event sequences must be analysed in depth by the operators to deduce a correct diagnosis before taking the appropriate actions [15]. This flood of information often overloads operators, generated by slowing down the diagnosis process.

Our analysis consists of a fault isolation method based on event pattern matching. As initial step all the events generated by specific faults are collected. These information are then processed to detect events that were always present in a specific fault list, even with different orders. As a last step, these fault signatures are injected into an expert knowledge database which can be used to detect occurrences of similar faults in the future (Fig. 3). It must also be pointed out that the method proposed here is quite generic and not dedicated to a specific process. For the purpose of this work a simulator has been used to replicate the faults and extract accordingly the generated event list.

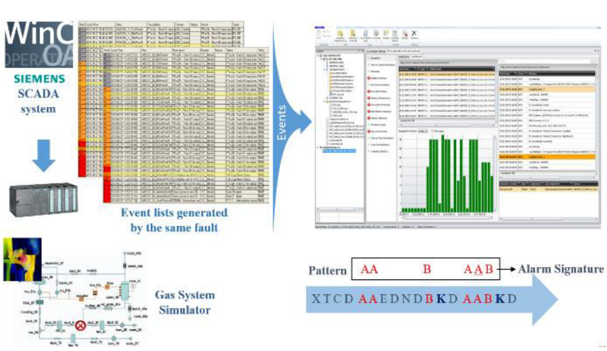

Figure 3: Analysis flow of the gas system alarms.

Engineering Design

The last family of activities is related to the analysis of historical data to draw conclusions about systems behaviour. The results can then help engineers to optimize specific system aspects like control regulation, parameterisation, or even drive the design of new parts of a system. Several use-cases, such as the analysis of historical data related to the CERN electric network which belong to this category, are under investigation and their results will be published later. Then, future consumption of different CERN areas can be predicted according to external factors like the accelerators schedule, weather conditions, technical interventions, and so on.

CONCLUSION

The control system of a facility produces a large amount of data and the regular extraction of insights from these data is a burdening task for experts, especially in the context of a very large and complex machine.

Capturing expert knowledge and formalizing how it can be used to deter issues is a first and essential step to provide assisting tools.

Once an abstract knowledge is converted into a tool it can be connected to the control system. This tool allows automation of advanced monitoring and pre-emptive maintenance, tasks that were previously triggered mostly by manual operators and experts actions. It helps them to look for abnormal conditions leveraging both their knowledge and scalable computing resources. It increases efficiency and availability of a machine, and lowers risks of disastrous events.

Knowledge capture and analysis is not always a straightforward task. Firstly, the expert needs a good understanding of the potential of an analysis tool in order to express problems that are addressable. Then, a mathematical approach is needed to transform a problem statement into a program able of discovering the actual issues from the data. Also, often an intermediate data science expert is needed as the machine experts are not fluent in the data analysis tools available at hands.

In addition, the set of available data analysis tools is limited by the interoperability capabilities of industrial control systems. These tools also have to be a good match for computing parallelisation frameworks. This includes the capacity of their algorithms to be parallelised, as well as for the software itself to be scheduled in a greater infrastructure.

Despite these limitations we have seen that these projects led to improved confidences into analyses on control systems. Involving machine experts for their knowledge is a prerequisite. Naïve approaches such as large-scale cross-correlations are not enough to extract meaningful insights. Further work is planned on different use cases of online monitoring, fault diagnosis and engineering design.

Furthermore, to address these limitations, we plan to provide a data analysis service (DAaaS) adapted to

ISBN 978-3-95450-148-9

control systems needs and their classical infrastructure, available to the experts directly.

ACKNOWLEDGEMENTS

A particular thanks to B. Bradu, P. Gayet, A. Rijllart, N. Bouchair, E. Blanco Vinuela and C.R. Vintila for their helpful contribution to the presented work.

REFERENCES

[1] ETM.at website: http://www.etm.at/index_e.asp

[2] R. de Haan, M. Roshchin, Detecting Temporally Related Arithmetical Patterns, 2012.

[3] E. Wu, U. Berkeley, High-performance complex event processing over streams, 2006.

[4] V.J. Hodge, J. Austin, A Survey of Outlier Detection Methodologies, 2004.

[5] S. Ramaswamy, R. Rastogi, K. Shim, Efficient Algorithms for Mining Outliers from Large Data Sets, 2000.

[6] Knorr, Algorithms for Mining Distance-Based Outliers in Large Datasets, 1998

[7] L. Wilhelm, M. Proetzsch, K. Berns, Oscillation Analysis in Behavior-Based Robot Architectures, 2010.

[8] T. Miao, D.E. Seborg, Automatic detection of excessively oscillatory feedback control loops, 1999

[9] L. Ettaleb, M.S. Davies, G.A. Dumont, E.E. Kwok, Monitoring oscillations in multi-loop systems, 1996.

[10] C. Pryor, Auto-covariance and power spectrum analysis, Control Engineering, 1982.

[11] G. Chowdhary, S. Srinivasan, E.N. Johnson, A Frequency, Domain Method for Real-Time Detection of Oscillations, 2010.

[12] N.F. Thornhilla, S.L. Shah, B.Huang, A.Vishnubhotla, Spectral principal component analysis of dynamic process data, 2002

[13] B. Kedam, Time series analysis by higher order zero -crossings, 1993.

[14] N.F. Thornhill, B. Huang, H. Zhang, Detection of multiple oscillations in control loops, 2003.

[15] S. Charbonnier, N. Bouchair, P. Gayet, Fault isolation by comparing alarm lists using a symbolic sequence matching algorithm, 2014.

ISBN 978-3-95450-148-9