A SMART FRAMEWORK FOR THE AVAILABILITY AND RELIABILITY ASSESSMENT AND MANAGEMENT OF ACCELERATORS TECHNICAL FACILITIES

L. Serio, A. Castellano[1], U. Gentile, CERN, 1211 Geneva 23, Switzerland
F. Antonello, P. Baraldi, E. Zio, Politecnico di Milano, 20100 Milano, Italy
[1]also at Politecnico di Milano, 20100 Milano, Italy

Abstract

CERN operates and maintains a large and complex technical infrastructure that serves the accelerator complex and experiments detectors.

A performance assessment and enhancement framework based on data mining, artificial intelligence and machine-learning algorithms is under development with the objective of structuring, collecting and analysing the operation and failure data of the systems and equipment, to guide the identification and implementation of adequate corrective, preventive and consolidation interventions.

The framework is designed to collect and structure the data and identify and analyse the associated driving events. It develops dynamically functional dependencies and logic trees, descriptive and predictive models to support operation and maintenance activities to improve the reliability and availability of the installations. To validate the performance of the framework and quality of the algorithms, several case studies are being carried out.

In this paper, we report on the design and implementation of the performance assessment and enhancement framework, and on the preliminary results inferred on historical and live stream data from CERN's technical infrastructure. Proposals for the full deployment and expected long-term capabilities will also be discussed.

INTRODUCTION

CERN's Technical Infrastructure (TI) is a large and complex system of systems providing essential services for the safe and reliable operation of its particles accelerators and experimental areas [1]. These essential services provided by the TI can be considered a critical infrastructure [2], since a failure impacts directly on the accelerators performance and overall availability for the physics experiments. Over the past few years, the LHC accelerator downtime due to TI equipment faults has been reduced to less than 10 % (inclusive of cryogenics system faults, which account for half of the downtime) of the physics operation time (which includes stable beams, operations, but excludes pre-cycles), but still accounts for 1/3 of the overall downtime [3]. The availability improvement obtained has been guided by the analysis, performed by an ad-hoc operation committee, of the data acquired during occurred major events that have affected the accelerators operation and the monitoring of the interventions undertaken to minimize the impact on the operating conditions. It is, therefore, paramount to identify clear and easy metrics and tools to monitor objectively the activities and guide the implementation

strategies to improve the infrastructure performances and overall availability in all configurations and operating scenarios. A smart framework for the availability and reliability assessment and management of accelerators technical facilities has been proposed to support the analysis, to guide the operation and improve the overall efficiency of the process. The framework provides an integrated environment to collect data from heterogeneous sources (sensors, alarms, logbooks), implements mining and machine learning techniques to infer functional dependency models and fault logic models, and dynamically updates them to follow the evolution of the TI [4].

WORK OBJECTIVE

CERN's TI is composed of a set of interconnected systems, performing different functions and based on technologies from various domains. Due to the complexity of the TI topology, its geographic distribution and the differences in the functionalities, the various systems are typically designed and built independently, taking into account only the direct physical interfaces and assuming a certain number of functional dependencies based on the assumed operational scenarios. Furthermore, the systems may change in time, e.g. grow in size, by including new components or updating old components as a result of technology advancements, consolidations and operation needs. Then, in general, the interconnections among the different systems and the functional dependencies between their components are in many cases modified with respect to the initial design, both at the physical and functional levels.

Given the importance of dependent failures in risk, reliability and availability analysis, the primary objective of this work is the identification of the functional dependencies among components of different systems and of sets of interconnected components.

Various benefits are expected from the capability of identifying functional dependencies and sets of interconnected components:

1. A more accurate estimation of the TI reliability and availability, which can be significantly underestimated were the dependencies not considered [5].
2. The decisions of control room operators for the TI daily operation and during the management of major events are expected to be facilitated by the knowledge of the functional dependencies between components of different systems.
3. The investigation of the causes of major events can benefit from the knowledge of the presence of set of interconnected components.

† luigi.serio@cern.ch

APPROACH

To support and simplify the work of systems experts and control room operators in the process of retrieving and keeping up-to-date the functional dependencies among components of different systems, two other possible sources of information are considered:

- Large datasets containing historical values of tens of thousands of signals collected from sensors measuring physical quantities from the TI components;
- Sequences of alarms provided by the TI supervision systems.

With respect to the first source of information, an approach for the identification of sets of signals whose correlated behaviors can explain major events (i.e. impacting the operation of the accelerator with the loss of the circulating beam) is being investigated. The idea is to deal with the very large number of available signals by selecting a small subset of them, which, in case of major events, become correlated. This will be performed by developing a feature selection wrapper algorithm based on the combined use of differential evolution and support vector machines [6, 7]. Then, the analysis of the correlations among the selected signals in cases of major events is expected to provide information on the unknown functional dependences among components of different systems.

In this work we focus on the use of the second source of information, i.e. the sequences of alarms provided by the TI supervision systems. To this purpose, we consider a TI formed by thousands of components, which, typically, generate tens of thousands of alarms every day. An alarm is generated when the measured value of a key physical quantity goes outside a predefined normal operation range. The alarm thresholds are set by experts of the technical infrastructure to prevent the occurrences of possible dangerous situations and to identify components malfunctions or degradation. Sequences of alarms are expected to contain information about the functional dependencies among components of different systems. In practice, the malfunction of a given component is expected to cause variations of correlated key quantities and, therefore, activate the corresponding specific alarms. If that component has functional dependencies with components of other systems, these latter are expected to trigger other alarms, as well.

The methodology proposed in this work for the identification of functional dependencies from the analysis of sequences of alarms is based on the following three steps:

1. Representation of the alarm sequences by means of Boolean vectors.
2. Generation of association rules.
3. Identification of groups of interconnected components and their functional dependencies.

Representation of the Alarm Sequences by Means of Boolean Vectors

We consider a TI made of L systems, with the generic l-th system formed by M_l components, each one with associated N_{ml} possible alarms. Time is discretized into a series of consecutive time intervals Δt_j, $j = 1, 2, …, q$ of the

same length Δt. The time length is based on expert's estimates of malfunctions propagation. It is further optimized to minimise computing time and maximise the number of true positives within each major event. Then, a Boolean variable, s_{lmn}, is associated to the generic alarm of type n, of component m of the the l-th system of the TI, with $l = 1, …, L$; $m = 1, …, M_l$ and $n = 1, …, N_{lm}$. The value $s_{lmn}(j)$ of the Boolean variable s_{lmn} in the time interval Δt_j is 1 if the corresponding alarm occurs at least once and 0 if it doesn't occur during the time interval Δt_j.

The state of the generic component m_l of the $l-th$ system in the time interval Δt_j is represented by the vector $\vec{c}_{lm}(j) = [s_{lm1}(j), s_{lm2}(j), …, s_{lmN_{lm}}(j)] \in [0,1]^{N_{lm}}$ where N_{lm} indicates the number of alarms associated to the component. Figure 1 and Table 1, show an example of a sequence of alarms generated by a generic component m_l with associated n alarms and the corresponding time evolution of the Boolean vector.
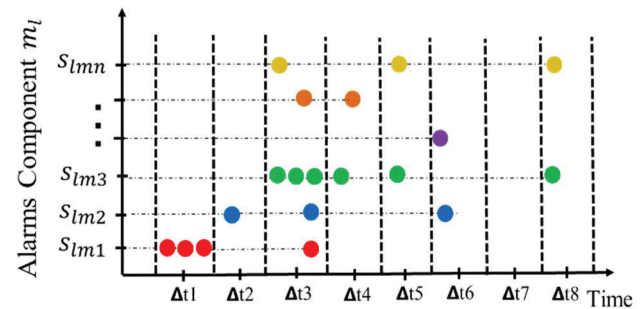

Figure 1: Example of sequence of alarms.

Table 1: Time Evolution of the Boolean Vector $\vec{c}$ Corresponding to the Sequence of Alarms

Δt	S$_{lm1}$	S$_{lm2}$	S$_{lm3}$	S$_{lm4}$	S$_{lm5}$	S$_{lm6}$
Δt1	1	0	0	0	0	0
Δt2	0	1	0	0	0	0
Δt3	1	1	1	0	1	1
Δt4	0	0	1	0	1	0
Δt5	0	0	1	0	0	1
Δt6	0	1	0	1	0	0
Δt7	0	0	0	0	0	0
Δt8	0	0	1	0	0	1

Since the generic l-th system is formed by M_l components, we can represent the system state, $\vec{s}_l$, by concatenating the corresponding vectors $\vec{c}_{lm}$, with $m = 1, … M_l$, i.e. $\vec{s}_l(j) = [\vec{c}_{l1}(j), \vec{c}_{l2}(j), …, \vec{c}_{lM_l}(j)]$. Finally, the overall state of the TI in the generic time interval Δt_j is obtained by concatenating the vectors $\vec{s}_l(j)$, i.e. $\overrightarrow{TI}(j) = [\vec{s}_1(j), …, \vec{s}_L(j)]$.

Generation of the Association Rules

We consider the sequence of vectors $\overrightarrow{TI}(j)$, $j = 1,\ldots,q$ obtained from the analysis of the alarms in a period of time of lengths $q\Delta t$. These vectors are organized in the dataset ETI, where each row is a vector $\overrightarrow{TI}(j)$. The methodology for the identification of functional dependencies among components of different systems is based on the analysis of the correlation among the alarms. This task is performed by using association rules.

Generally, correlations are implications from which it is possible to derive association rules, which are logical expression of the form $X \Rightarrow Y$, indicating the correlation between categorical objects, which, in this work, are the alarms represented by the Boolean variables s_{lmn} [8,9]. Informally, implication (and, therefore, association rules) tell us that whenever a given set of alarms X (antecedent of the rule) is verified at a given time Δt_j then, the set of alarms Y (consequent of the rule) is also verified at the same time [10]. Formally, the association rule $X \Rightarrow Y$ is created when $X \Rightarrow Y$, is verified if in at least c% of the ETI dataset events (time intervals) in which when X occurs also Y occurs, and if X and Y occur together in at least s% of the ETI events. The parameters c and s are called confidence and support, respectively [11].

It is worth noticing that implications identified by using association rules are not time-dependent rules, i.e., they do not allow inferring causal relations among the alarms.

Sets of Interconnected Components

Association rules allow identifying components whose malfunctions are correlated. In the cases in which the number of identified rules is very large, it can be useful to extract and visually represent sets of interconnected components and to separate groups of components which are not connected by rules. In this work, this has been done by applying a graph-oriented representation. In this kind of graph each vertex represents a component and the dependence between components is expressed by the edge connecting two vertices. Then, sets of interconnected components are identified in the graph by dividing groups of vertices which have no edges in common.

CASE STUDY

The proposed methodology has been applied to alarms generated by the CERN's TI. We consider a database containing alarms generated during the major events of the year 2016 by different supervision systems of LHC Zone 8, a specific zone of LHC that is representative of the complexity of the overall TI. The considered alarms are generated by the cryogenic, the cooling, the ventilation and the electric systems. The analysis of the dataset has shown that 253591 alarms have been generated during 2016. Considering the alarms description, we have found out that these alarms have been caused by 6800 different malfunctions which involved 2895 different components.

The one-year period (from January 1st 2016 to December 31st 2016) has been divided into $q = 17500$ time in-

tervals of duration of 30 minutes. The state of the considered part of the ICT in the generic time interval Δt_j is represented by the 6800-dimensional Boolean vector $\overrightarrow{TI}(j) = [\vec{s}_1(j),\ldots,\vec{s}_L(j)]$. Therefore, the dataset ETI is formed by 17500 rows and 6800 columns. The association rules have been generated by considering the support (s%) and the confidence (c%) values equal to 0.02 and 0.8, respectively. These parameters have been set by following a trial and error procedure to generate relevant and meaningful association rules. The method has found a total number of 1112 association rules from which 14 sets of interconnected components involved in the loss of the circulating beam have been identified. The rules involve components of different systems which are effectively correlated one with each other. An independent expert analysis confirmed that all the 14 identified sets are part of the chain of malfunctions of the considered systems, resulting in the major events of the year 2016.

It is important to underline that the association rules cannot be interpreted as causal rules. The causality will be extracted by the analysis of the operational data, related to the associated alarms.

CONCLUSIONS

In this work we have presented the smart framework under development, to be integrated with the existing environment to collect heterogeneous data, and implement mining and machine learning techniques to dynamically update predictive models.

The proposed smart framework is a valuable tool in support of systems experts and control room operators to identify functional dependencies between components of different systems of CERN's TI to improve availability and reliability assessment and facility management.

We have proposed a method for the extraction of association rules from alarms sequences. The obtained results show the capability of the method of identifying functional dependencies among components of different systems.

Future research will consider methods for the extraction of causal rules that take into account the time sequence of the events. Furthermore, we will investigate the integration of the identified functional dependencies between components of different systems with the available high-level models of the individual systems.

REFERENCES

[1] J. Nielsen and L. Serio, "Technical Services: Unavailability Root Causes, Strategy and Limitations", *in Proc. 7th Evian Workshop on LHC beam operation*, Evian Les Bains, France, December 2016.

[2] S. M. Rinaldi and J. P. Peerenboom and T, K, Kelly, "Identifying, understanding and analysing critical infrastructure interdependencies", *IEEE Control Systems*, 21.6 (2001).

[3] B. Todd *et al*, "LHC Availability 2017: Standard Proton Physics", CERN, Geneva, Switzerland, Rep. CERN-ACC.NOTE-2017-0063, December 2017.

[4] U. Gentile and L. Serio, "A Machine-learning based methodology for performance analysis in particles accelerators facilities", *First International Conference on Electrical Engineering & Computer Science (EECS), IEEE*, 2017, to be published.

[5] *NRC Nuclear Regulatory Commission, Probabilistic Risk Analysis: Procedure Guide*, NUREG/CR 2003, Washington, DC, 1983.

[6] K. C. Gryllias, I. A. Antoniadis, "A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments", *Eng. Appl. Artif. Intell.*, 25 (2), pp. 326–344, 2012.

[7] K. Zhu, X. Song, D. Xue, "A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm", *Measurement*, 47, pp. 669–675, 2014.

[8] C. Hu, Y. Wang, B. Zhang, Q. Yang, Q. Wang, J. Zhou, Y. Yan, "Mining quantitative associations in large database", *Lecture Notes in Computer Science*, 3399, pp. 405-416, 2005.

[9] R. Srikant, R. Agrawal, "Mining Quantitative Association Rules in Large Relational Tables", *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 25 (2), pp. 1-12, 1996.

[10] J. Baixeries, "Lattice characterization of Armstrong and symmetric dependencies". Ph.D. Thesis, Universitat Politècnica de Catalunya, Spain, 2007.

[11] A. Singh, M. Chaudhary, A. Rana, G. Dubey, "Online Mining of data to generate association rule mining in large databases", *International Conference on Recent Trends in Information Systems*, ReTIS 2011 - Proceedings, art. no. 6146853, pp. 126-131, 2011.