# A Method for Inferring Causal Dependencies Among Abnormal Behaviours of Components in Complex Technical Infrastructures

**Conference Paper** · November 2020

DOI: 10.3850/978-981-14-8593-0_5677-cd



Federico Antonello
Politecnico di Milano


Piero Baraldi
Politecnico di Milano


Ahmed Shokry
École Polytechnique


Enrico Zio
MINES ParisTech and Politecnico di Milano

# A Method for Inferring Casual Dependencies Among Abnormal Behaviours of Components in Complex Technical Infrastructures

Federico Antonello
*Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy.*

Piero Baraldi
*Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy.*
*E-mail: piero.baraldi@polimi.it*

Ahmed Shokry
*Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy*

Enrico Zio
*Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy.*
*MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France.*
*Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea*

Ugo Gentile
*Engineering Department, CERN, 1211 Geneva 23, Switzerland*

Luigi Serio
*Engineering Department, CERN, 1211 Geneva 23, Switzerland*

This work presents a novel method for inferring causal dependencies among abnormal behaviours of components in Complex Technical Infrastructures (CTIs) from large-scale databases of alarm messages. The proposed method extracts causal relationships from association rules performing a probabilistic analysis of the alarm occurrence times and applying a modified version of the quicksort algorithm. Its capability and effectiveness is illustrated by application to a real large-scale databases of alarm messages collected in the technical infrastructure of the European Organization for Nuclear Research (CERN).

*Keywords*: Complex Technical Infrastructures, Alarms, Association Rules, Dependent Abnormal Behaviours, Time-Dependent Analysis, Quicksort Algorithm, Causality.

## 1. Introduction

Complex Technical Infrastructures (CTIs) encompass sophisticated hierarchical architectures in which tens of thousands of components, which perform diverse functions and use different technologies (mechanics, electronics communication, etc.), are interconnected and distributed over vast geographic areas (Filip, 2008). The different systems and subsystems of CTIs are typically designed and built independently and are continuously modified though the time with respect to their initial design, both at physical and functional levels, by including new components and/or updating old components, due to technology upgrading, design retrofitting and consolidation requirements (Billinton and Allan, 1992; Filip, 2008; Serio, et. al., 2018; Antonello, et. al., 2019). As a result, hidden functional dependencies among different groups of components emerge and evolve during the CTI lifetime and local malfunctions or perturbations can easily propagate across and within the different CTI systems, originating unexpected cascades of failures. Thus, the identification of causal relationships among abnormal behaviors play a crucial role for CTI vulnerability and resilience analysis and for the management of the CTI operation.

Nowadays, the analysis of complex and large-scale systems can greatly benefit from the large

*Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*
*Edited by* Piero Baraldi, Francesco Di Maio and Enrico Zio

ISBN/DOI: TBA

amount of data and information collected from the CTI. In this work, we focus on the analysis of sequences of alarm messages. A method has been proposed in (Antonello, et. al., 2019) for the identification of groups of functionally dependent components in CTIs. It is based on: *i*) the representation of the heterogenous database of alarm messages in a binary form; *ii*) the use of Apriori-based algorithm to identify patterns of alarms that frequently occur together within a short period of time; *iii*) the analysis of the identified patterns for extracting association rules and *iv*) the analysis of the extracted rules to constitute groups of functionally dependent components. The method main limitation is that it does not provide causal relationships among the identified abnormal behaviours of the dependent components. This is due to the fact that the extracted association rules are probabilistic relationships in the form "*if* (*antecedent*) *then* (*consequent*)", which describe conditional co-occurrences of abnormal behaviors and cannot be interpreted as causal relationships among the alarms in the rule antecedent and consequent (Antonello et. al., 2019).

In this context, the present work proposes a method for causality inference based on the combined use of an Association Rule Mining (ARM) algorithm, the probabilistic analysis of the alarm occurrences times and the quicksort algorithm (Sedgewick, R. 1978). The causal relationships among malfunctions are assumed to be time-dependent relationships (Kühnert and Beyerer, 2014) and, hence, the causality between each couple of malfunctions is inferred by analysing their occurrence times in the available database of alarms. Given the large number of components in a CTI, it is computationally infeasible to perform the aforementioned analyses on each possible couple of component malfunctions. Therefore, the pairwise analysis is applied only to those alarms involved in the association rules extracted by the ARM algorithm. The causal dependencies among the alarms involved in the same rule are obtained by using a modified version of the quicksort algorithm (Sedgewick, R. 1978), which recursively sorts the alarms without the need of considering all the possible pairwise comparisons.

The effectiveness of the proposed method is shown through its application to a real large-scale database of alarms generated by the different supervision systems of the CTI of CERN Large Hadron Collider (LHC), composed by thousands of components and spread over a geographic area of about 27 km perimeter ring.

The remaining of the work is organized as follows. Section 2 describes the problem setting. Section 3 briefly introduces the ARM approach for the identification of functional dependencies in CTI. Section 4 describes the proposed method for the identification of the causal dependencies. Section 5 introduces the case study and discusses the obtained results. Finally, Section 6 draws some conclusions and recommends potential future lines of work.

## 2. Problem setting

We consider a CTI composed by a large number, $N_c \gg 1$, of components and we assume to have available a database containing a large number, $N^{al} >> N_c$, of alarm messages generated by the CTI during a long period of operation $[t_0, t_f]$. The generic *i-th* alarm message is represented by the pair $(t_i, m_i)$ which defines the time $t_i$ at which the alarm occurs and the label $m_i$ associated to the alarm type. Assuming that there are $M_j^{al}$ different types of alarms associated to the generic *j-th* component $c_j, j = 1,..,N_c$, the label $a_j^k$ refers to the *k-th* type of alarm message associated to component $c_j$. The set containing all the possible types of alarm messages in the database is:

$$A = \left\{a_1^1, \dots, a_1^{M_1^{al}}, , \dots a_j^{M_j^{al}}, \dots, a_{N_c}^1, \dots, a_{N_c}^{M_{N_c}^{al}}\right\} \quad (1)$$

and the total number of alarm message types:

$$M^{al} = \sum_{j=1}^{N_c} M_j^{al} \quad (2)$$

## 3. The Apriori-based algorithm for the identification of dependent abnormal behaviours

This Section briefly recall the method for the identification of dependent abnormal behaviours among components of a CTI introduced in (Antonello, et. al., 2019). Section 3.1 will describe the representation of the database of alarm messages in a binary form and Section 3.2 introduces the Apriori-based algorithm used for extracting association rules.

### 3.1 Alarm database representation

The entire time period $[t_0, t_f]$ during which the alarm messages of the database have been collected is subdivided into $Z$ consecutive small time intervals of the same length $\Delta t = \frac{t_f - t_0}{Z}$. A Boolean variable, $s_j^k(z)$, is associated to the occurrence of the alarm of type $a_j^k$ in the *z-th* time interval:

$$s_j^k(z): \begin{cases} 1 \text{ if alarm } a_j^k \text{ occurs at least once in} \\ \quad [t_0 + (z-1)\cdot\Delta t, t_0 + z\cdot\Delta t] \\ \quad 0 \text{ otherwise} \end{cases} \quad (3)$$

The state of the generic component $c_j$ in the generic *z-th* time interval is represented by the Boolean vector:

$$\vec{c}_j(z) = \left[s_j^1(z), s_j^2(z), \dots, s_j^{M_j^{al}}(z)\right] \in [0,1]^{M_j^{al}} \quad (4)$$

and that of the CTI by the concatenation of the component state vectors $\vec{c}_j(z)$:

$$\vec{\mathcal{T}}(z) = [\vec{c}_1(z), \dots, \vec{c}_{N_c}(z)] \in [0,1]^{M^{al}} \quad (5)$$

Finally, the raw database of alarms $(t_i, m_i), i = 1, \dots, N^{al}$, is transformed into the Boolean matrix:

$$T = \begin{bmatrix} \vec{\mathcal{T}}(1) \\ \dots \\ \vec{\mathcal{T}}(Z) \end{bmatrix} \in [0,1]^{Z \times M^{al}} \quad (6)$$

whose generic *z-th* row represents the state of the CTI during the *z-th* time interval. Therefore, $T$ provides a dynamic representation of the CTI state evolution in the time interval $[t_0, t_f]$.

### 3.2 Association rules mining

Considering a set of alarms $X \subseteq A$, an association rule is a probabilistic logical expression of the form $x^a \Rightarrow y^a, x^a \subset X,\ y^a = X - x^a$, representing the conditional co-occurrence of the two subsets, $x^a$ and $y^a$, of the set $X \subseteq A$, where $x^a$ and $y^a$ are referred to as "antecedent" and "consequent" of the rule, respectively (Srikant and Agrawal, 1996; Hui, et al., 2005). Let $n(X)$ be the counter of the number of vectors $\vec{\mathcal{T}}(z)$ of the database $T = \begin{bmatrix} \vec{\mathcal{T}}(1) \\ \dots \\ \vec{\mathcal{T}}(Z) \end{bmatrix}$ characterized by the occurrences of at least all the alarms of the set $X$ (i.e., $\forall\, a_{j,s}^k \subset X,\ s_j^k(z) = 1$). The expression $x^a \Rightarrow y^a$ is an association rule if:

*a*) the *support* of $X$ is larger than a minimum support $ms$:

$$Sup(X) = Sup(x^a \Rightarrow y^a) = n(x^a \cup y^a) \geq ms \quad (8)$$

*b*) the confidence of the rule $x^a \Rightarrow y^a$ is larger than a minimum confidence $mc\%$:

$$Conf(x^a \Rightarrow y^a) = \frac{n(x^a \cup y^a)}{n(x^a)} \geq mc\% \quad (9)$$

We consider the ARM procedure proposed in (Antonello, et. al. 2019), which is based on the two main steps:

1) Identification of frequent patterns of alarms, $X^{fp} \subseteq A$, characterized by a support larger than $ms$, i.e., $Sup(X^{fp}) \geq ms$.
2) Extraction of association rules $x^a \Rightarrow y^a, x^a \subset X^{fp},\ y^a = X^{fp} - x^a$ from each one of the frequent patterns of alarms identified in step 1), by identifying the sets $x^a$, $y^a$ satisfying the confidence condition $Conf(x^a \Rightarrow y^a) \geq mc\%$.

The final set of the obtained rules is referred to as $AR = \{x_l^a \Rightarrow y_l^a,\ l = 1, \dots, N_{rule}\}$, with $N_{rule}$ indicating the number of obtained rules.

### 4. Identification of the causality among the components malfunctions in the generated association rules

The proposed method is based on the two steps of: *i)* inferring the causal relationship between each possible couple of alarms involved in the same association rule, which is based on the statistical analysis of their times of occurrence (Section 4.1) and *ii)* deriving the causal relationships among the entire set of alarms involved in the same association rule using the quicksort algorithm (Section 4.2).

## 4.1 Causal relationship among two dependent alarms

Two components are functionally dependent if the operation of one is influenced by the operation of the other (Estesami et. al. 2016). In particular, since this work considers alarm messages, which are typically triggered when components have abnormal behaviours or malfunctions, we assume that a causal functional dependency, $c_1 \rightarrow c_2$, among two components of a CTI, $c_1$ and $c_2$, exist if a malfunction of component $c_1$, revealed by an alarm, $a_1^k$, causes a malfunction of components $c_2$, revealed by another alarm, $a_2^k$ (Antonello et. al., 2019). We assume that if the malfunction of component $c_1$ causes the malfunction of components $c_2$, then the alarm $a_1^k$ is likely to occur before alarm $a_2^k$, although there can be alarm configuration errors and delays between the times of occurrence of the malfunctioning and its record in the alarm database (C. Kühnert and J. Beyerer, 2014).

Considering the association rule $r_l = \{x_l^a \Rightarrow y_l^a\}$, the causal relationship among any couple of alarm $a_{j'}^{k'}$, $a_{j''}^{k''} \in (x_l^a \cup y_l^a)$ is inferred by estimating the probability that an alarm occurs before the other. To this aim, the cumulative density function of the difference between the times of occurrence of the two alarms $a_{j'}^{k'}$, $a_{j''}^{k''}$ is estimated using the following algorithm:

I. Identify the set of time intervals in which all the alarms involved in the rule $r_l$ occur together, $Z_l = \{z_i\} \mid \forall a_j^k \in (x_l^c \cup y_l^c), \forall z_i \in Z_l, s_j^k(z_i) = 1$.
II. For each time interval $z_i \in Z_l$, identify the times, $t_{j'}^{k'}(z_i), t_{j''}^{k''}(z_i)$, at which the two generic alarms $a_{j'}^{k'}$, $a_{j''}^{k''}$ start to occur. Then, calculate the difference between the exact time of occurrence of the considered couple of alarms, $\tau_{a_{j'}^{k'}, a_{j''}^{k''}}(z_i) = t_{j'}^{k'}(z_i) - t_{j''}^{k''}(z_i)$.
III. Estimate the cumulative density function $cdf(\tau_{a_{j'}^{k'}, a_{j''}^{k''}})$ over all the considered time intervals $z_i \in Z_l$.

Then, the following scheme is used to estimate the causal relationship among the two alarms:

a. If $cdf_{a_{j'}^{k'}, a_{j''}^{k''}}(\tau = 0) > tr$, $a_{j'}^{k'}$ is caused by $a_{j''}^{k''}$, where $tr \in [0,1]$ is the causality threshold.
b. Else If $cdf_{a_{j'}^{k'}, a_{j''}^{k''}}(\tau) < 1 - tr$, $a_{j''}^{k''}$ is caused by $a_{j'}^{k'}$.
c. Else If $t_{j'}^{k'}(z_i) - t_{j''}^{k''}(z_i) = 0$, $\forall z_i \in Z_l$, $a_{j''}^{k''}$ and $a_{j'}^{k'}$ occur simultaneously.
d. Else, there is no causal relationship between $a_{j''}^{k''}$ and $a_{j'}^{k'}$.

With respect to the choice of the threshold, it should be larger than 0.5, to guarantee that the probability of the identify causal dependency ($a_{j'}^{k'}$ is caused by $a_{j''}^{k''}$) is larger than its opposite ($a_{j''}^{k''}$ is caused by $a_{j'}^{k'}$), and lower than 1, to consider possible alarm desynchronizations or events not related to the causal dependency. Its setting depends from the characteristics of the system in terms of alarm desynchronization.
It is important to underline that point *c.* accounts for the fact that not all the dependencies are causal, e.g. caused by common cause failures, and, therefore, some association rules involve components which are not causally dependent.

Figure 1 presents an example of the proposed procedure by showing the occurrences of a couple of alarms, $a_{j'}^{k'}$ and $a_{j''}^{k''}$, along the time intervals $z_1 : z_5$ and the corresponding differences between the alarms starting times $\tau_i$. Notice that $cdf_{a_{j'}^{k'}, a_{j''}^{k''}}(\tau = 0)$ represents the probability that $a_{j''}^{k''}$ occurs before $a_{j'}^{k'}$, while $1 - cdf_{a_{j'}^{k'}, a_{j''}^{k''}}(\tau = 0)$ represents the probability that $a_{j'}^{k'}$ occurs before $a_{j''}^{k''}$.

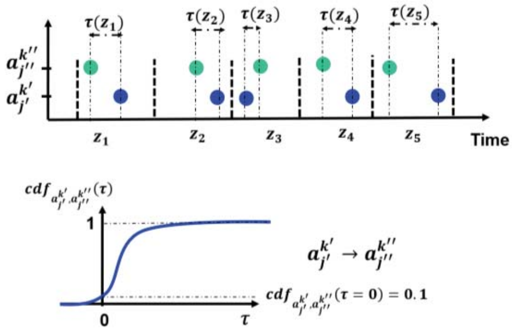


**Figure 1. Schematic representation of the proposed procedure for causality inference of the couple of alarms, $a_{j'}^{k'}, a_{j''}^{k''}$.**

## 4.2. Causality inference in association rules

The objective of this step is to define from each of the generated association rules $r_l = \{x_l^a \Rightarrow y_l^a\}$, $l = 1, \dots, N^p{}_{rule}$, the corresponding causal rule, $r_l^{\rightarrow} = \left\{ {}^1[a_{j'}^{k'}] \rightarrow {}^2[a_{j''}^{k''}] \rightarrow \dots \rightarrow {}^O[a_{j'''}^{k'''}] \rightarrow \dots \rightarrow {}^{n_l}[a_{j''''}^{k''''}] \right\}$, $a_{j'}^{k'} \in (x_l^a \cup y_l^a)$, $O = 1, \dots, n_l$. To assess the causal dependencies among alarms involved in a rule the method uses the algorithm introduced in Section 4.1 and the quicksort algorithm (Sedgewick, R. 1978). Quicksort is one of the most efficient sorting algorithms, in terms of computational time and required memory (X. Wang, 2011). In principles, given a set of elements (for Example the set of numbers *X*=[3, 4, 5, 8, 1, 9]) a quicksort algorithm ranks through a divide-and-conquer approach based on a specific sorting criteria (for example, sort from the lowest to the largest). The first step requires selecting a random element called 'pivot', i.e., number 5 of set *X*, and splitting the rest of elements of the set into two sub-sets according to whether they satisfy or not the sorting criteria, i.e., if the number is lower [1,3,4] or larger [8,9] than the pivot. Each of the sub-arrays is then sorted

0) Define the list the alarms involved in $r_l = \{x_l^a \Rightarrow y_l^a\}$.
1) Initialize the following 4 empty sub-lists: *previous,* $[]_{pre}$, *simultaneous,* $[]_{sim}$, *after,* $[]_{aft}$, and *uncorrelated,* $[]_{un}$.
2) Pick, randomly, an alarm (the pivot), $a_{j'}^{k'}$
3) For each of the remaining alarms, $a_{j''}^{k''}$, in the list compute its pairwise causal relationship with the pivot (algorithm in Section 4.1) and, then, assign the alarm to one of the predefined sub-lists:
a. If $a_{j'}^{k'}$ is caused by $a_{j''}^{k''}$, assign $a_{j''}^{k''}$ to the sub-list *previous*.
b. Else If $a_{j''}^{k''}$ is caused by $a_{j'}^{k'}$, assign $a_{j''}^{k''}$ to the sub-list *after*.
c. Else If $a_{j''}^{k''}$ and $a_{j'}^{k'}$ occur simultaneously, assign $a_{j''}^{k''}$ to the sub-list *simultaneous*.
d. Else, assign $a_{j''}^{k''}$ to the sub-list *uncorrelated*.
e. Finally, assign the pivot $a_{j'}^{k'}$ to the sub-list *simultaneous*: $[a_{j'}^{k'}, a_{j''}^{k''}, \dots, a_{j'''}^{k'''}]_{sim}$.
4) Compare each alarm in the sub-list *uncorrelated* with the alarms in the sub-list *previous*, and if it causes or is caused by any of those alarms (in *previous*), assign this alarm (of *uncorrelated*) to the sub-list *previous*.
5) In the case in which alarms remain in the sub-list uncorrelated, compare each one of them with the alarms in the sub-list *after*, and if it cause or is caused by any of the alarms in the sub-list *after*, assign the alarm of the sub-list *uncorrelated* to the sub-list *after*.
6) In the case alarms remain in the sub-list uncorrelated, add all of them to the sub-list *simultaneous* as a sub-list $[a_{j'}^{k'}, a_{j''}^{k''}, \dots, a_{j'''}^{k'''} + [\dots]_{un}]_{sim}$.
7) Rank the list as follows:

$$[\dots]_{pre} \rightarrow [\dots]_{sim} \rightarrow [\dots]_{aft}$$

8) Recursively sort each of the above list following the steps 2), 3), 4), 5), 6), 7) and 8).

**Figure 2. Algorithm used to infer the causal relationship among alarms involved in a rule.**

recursively following the same procedure until the final rank is found (i.e., [1], [3], [4], [5], [8], [9]).

In this work, we consider the causality as the sorting criteria and, for each of the generated association rules $r_l = \{x_l^a \Rightarrow y_l^a\}$, $l = 1, \dots, N^{\boldsymbol{p}}{}_{rule}$, we infer the causal relationship among its involved alarms by applying the algorithm in Figure 2 for each *l-th* rule:

The application of the algorithm ends when the sub-lists include one alarm, or a number of alarms equals to those included in the parent list, i.e. it is not possible to further rank the sub-list.

## 5. CERN complex technical infrastructure

The CTI of CERN LHC (the largest existing particle accelerator) is composed by several systems working together for the functioning of the accelerator complex (Antonello et. al., 2019). The LHC consists of a 27 Km ring of superconducting magnets and infrastructures, extending over the Swiss and French borders and located about 100 m underground. Since the analysis of the LHC operation performance in 2016 shows that technical infrastructure components faults contributed to more than 1/5 of the overall LHC downtime, the identification of unknown functional dependencies among the CTI components (and among their anomalies) is of paramount importance for improving the accelerators performance and availability for physics experiments.

We consider the alarm database generated during the period $[t_0, t_f]$=[January 1st, 2016; December 31st, 2016] by supervision of three systems of the LHC *point 8*, which is a part of the general infrastructure and, while scaled down, still representative of the overall CTI complexity. The considered alarms are generated by the cryogenic, the electric and the cooling and ventilation systems. During the considered period, $N_{al}$ =253591 alarm messages reporting $M^{al}$=6800 different types of malfunctions caused by $N_c$ = 2895 components, have been collected.

Relying on the technical infrastructure operators and system expert knowledge, the malfunction propagation time among the considered components and systems (Antonello, et. al.,

**Table 1. Example of extracted associating rules (left) and their causal counterpart (right) obtained by the proposed method.**

| | **Association Rule** | | | **Causal Rule** |
|---|---|---|---|---|
| | **Antecedent [System] {Alarm Identifier}** | $\Rightarrow$ | **Consequent {Alarm Identifier} [System]** | |
| | $\begin{bmatrix} Cryogenic \\ Cryogenic \\ Cooling\ and\ ventilation \end{bmatrix} \begin{Bmatrix} a_{123}^4 \\ a_{124}^4 \\ a_{1123}^2 \end{Bmatrix}$ | $\Rightarrow$ | $\begin{Bmatrix} a_{100}^3 \\ a_{101}^3 \\ a_{102}^3 \\ a_{1124}^2 \end{Bmatrix} \begin{bmatrix} Cryogenic \\ Cryogenic \\ Cooling\ and\ ventilation \\ Cooling\ and\ ventilation \end{bmatrix}$ | ${}^1[a_{1123}^2, a_{1124}^2] \rightarrow {}^2[a_{100}^3,\ a_{101}^3, a_{102}^3] \rightarrow {}^3[a_{123}^4, a_{124}^4]$ |
| ***Rule 3*** | $\begin{bmatrix} CV \\ CV \end{bmatrix} \begin{Bmatrix} a_{1324}^2 \\ a_{1374}^2 \end{Bmatrix} \Rightarrow$ | $\Rightarrow$ | $\{a_{1424}^2\}[CV]$ | ${}^1[a_{1324}^2] \rightarrow {}^2[a_{1424}^2] \rightarrow {}^3[\ a_{1374}^2]$ |
| ***Rule 4*** | $[CV]\{a_{1724}^2\} \Rightarrow$ | $\Rightarrow$ | $\{a_{1094}^2\}[CV]$ | ${}^1[a_{1724}^2] \rightarrow {}^2[a_{1094}^2]$ |
| ***Rule 5*** | $\begin{bmatrix} Electric \\ Electric \\ Electric \\ Electric \\ Electric \\ Electric \end{bmatrix} \begin{Bmatrix} a_{3223}^2 \\ a_{3613}^2 \\ a_{3569}^2 \\ a_{4445}^2 \\ a_{4789}^2 \\ a_{2991}^2 \\ a_{3367}^2 \end{Bmatrix} \Rightarrow$ | $\Rightarrow$ | $\begin{Bmatrix} a_{4523}^2 \\ a_{2451}^2 \\ a_{2937}^2 \end{Bmatrix} \begin{bmatrix} Electric \\ Electric \\ Electric \end{bmatrix}$ | ${}^1[a_{4523}^2, a_{2451}^2, a_{3367}^2] \rightarrow {}^2[a_{3223}^2, a_{3613}^2, a_{3569}^2] \rightarrow {}^3[\ a_{4445}^2] \rightarrow {}^4[\ a_{4789}^2, a_{2991}^2]$ |
| ***Rule 6*** | $\begin{bmatrix} Cryogenic \\ Cryogenic \end{bmatrix} \begin{Bmatrix} a_{723}^4 \\ a_{793}^4 \end{Bmatrix}$ | $\Rightarrow$ | $\begin{Bmatrix} a_{192}^3 \\ a_{561}^3 \\ a_{245}^2 \end{Bmatrix} \begin{bmatrix} Cryogenic \\ Cryogenic \\ Cryogenic \end{bmatrix}$ | ${}^1[a_{561}^3] \rightarrow {}^2[a_{192}^3] \rightarrow {}^3[a_{723}^4] \rightarrow {}^3[a_{245}^2] \rightarrow {}^3[a_{793}^4]$ |

2019), and therefore the time interval length is set to $\Delta t = 30$ min (Section 3.3). Therefore, the one-year period [January 1st, 2016, December 31st, 2016] has been divided into $Z = 17500$ time intervals. The transformed binary database of alarm $T = [0,1]^{17500 \times M^{6800}}$ is processed by the Apriori algorithm using a minimum support, $s$%, equal to 0.02 and a minimum confidence, $c$%, equal to 0.8 (Antonello, et. al.,2019) and 202 association rules are generated. Then, the causal rules have been generated applying the method proposed in Section 4, setting the threshold *tr=0.8,* to consider possible desynchronization among alarms.

Table 1 reports some examples of the extracted rules. Among them, Rule 1 is the most interesting, since it involves components of different systems of the CTI associating alarms generated by components from the Cryogenic and Cooling and Ventilation (CV) systems. Notice that the identification of causal dependencies among components of different systems is very useful given the difficulty of identifying them using individual system analysis techniques. Also, the subdivision of the alarms in the association rule antecedent and consequent (Table 1, left) does not imply causality among them, whereas the causal rule (Table 1, right) provides much more information on the malfunctions propagation mechanism, which represents valuable information for the CTI operators. The corresponding causal rule is also visualised in Figure 3, in terms of the components related to the alarms involved in it. This rule describes the propagation of a malfunction in a cooling tower of the CV system ($a^2_{1123}$, $a^2_{1124}$) that leads to a stop of the low pressure compressors of the Cryogenic system ($a^3_{100}$, $a^3_{101}$, $a^3_{102}$), which, finally, cause in the stop of the high pressure compressors of the Cryogenic system ($a^4_{123}$, $a^4_{124}$). It is importance to notice that the ordered sequence of events led by the functional dependency described in the association rule (Table 1, left) correctly identify the causal rule (Table 1, right).

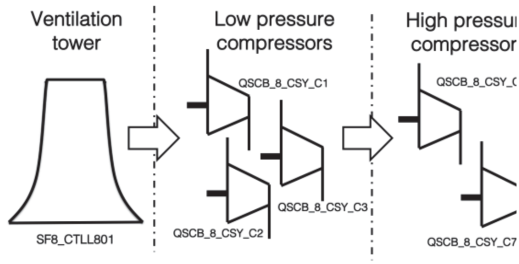


**Figure 3. Visualization of the first causal rule of Table1.**

## 6. Conclusions

We have presented a novel method for identifying causal functional dependences among groups of abnormal behaviours of components in CTIs. The proposed method is based on *i)* the alarm data representation in a binary form *ii)* the use of an ARM algorithm to identify association rules among dependent abnormal behaviour of CTI components and *iii)* the identification of the causal relationships among dependent abnormal behaviours in association rules.

The method is applied to a real large-scale database of alarms generated by the supervision of different systems at CERN Large Hadron Collider. The obtained results have been validated by operators and equipment experts.

## References

Agrawal, R., Imieliński, T., Swami, A. (1993). "Mining Association Rules Between Sets of Items in Large Databases", ACM SIGMOD Record, Volume 22, Issue 2, 6 January 1993, Pages 207-216.

Antonello, F., Baraldi, P, Shokry, A, Zio, E, Gentile, U., Serio, L. (2019). Data-Driven Extraction of Association Rules of Dependent Abnormal Behaviour Groups. *Proceedings of the 29th International European Safety and Reliability Conference*, ESREL 2019.

Antonello, F., Baraldi, P, Shokry, A, Zio, E, Gentile, U., Serio, L. (2019). Association Rules Extraction for the Identification of Functional Dependencies in Complex Technical Infrastructures. *Reliability Engineering & System Safety* Under review, (Submitted 2019).

Billinton R., Allan R.N. (1992) Network modelling and evaluation of complex systems. In: Reliability Evaluation of Engineering Systems. Springer, Boston, MA.

Etesami, J., Kiyavash, N. (2017), "Measuring causal relationships in dynamical systems through recovery of functional dependencies" (2017) *IEEE Transactions on Signal and Information Processing over Networks*, 3 (4), art. no. 7782866, pp. 650-659.

Filip, F. G. 2008. "Decision Support and Control for Large-Scale Complex Systems." *Annual Reviews in Control* 32(1): 61–70.

Kühnert and Beyerer. Data-Driven Methods for the Detection of Causal Structures in Process Technology Machines 2014, 2, 255-274.

Sedgewick, R. Implementing Quicksort Programs. (1978) *Communications of the ACM*, 21 (10), pp. 847-857.

Serio, L., Antonello, F., Baraldi, P., Castellano, A.,Gentile, U., Zio, E. (2018), "*Smart framework for the availability and reliability assessment and management of accelerators technical facilities*" 9th International Particle Accelerator Conference, IPAC 2018

Wang, X. Analysis of the time complexity of quick sort algorithm (2011) *Proceedings - 2011 4th International Conference on Information Management, Innovation Management and Industrial Engineering, ICIII 2011*, 1, art. no. 6115063, pp. 408-410.