THE ISSUE OF SYMPTOMS ARISING DELAYS DURING DIAGNOSTIC REASONING

Jan M. Kościelny, Michał Syfert, Bolesław Dziemowski

Institute of Automatic Control and Robotics Warsaw University of Technology
ul. św. A. Boboli 8, 02-525 Warszawa, Poland
e-mail: {jmk, m.syfert, b.dziembowski}@mchtr.pw.edu.pl

Summary

The paper shows the influence of the dynamics of the symptoms forming on the correctness of generated diagnoses. There are given a few approaches that allow to take into account the symptoms delays in the algorithms of diagnostic reasoning. Finally, there is presented the algorithm of proper reasoning while the information about symptoms delays is omitted. Key issues are illustrated with simple examples.

Keywords: fault isolation, diagnostics, dynamic systems.

1. PROBLEM FORMULATION

In most of the diagnostic strategies, to be able to proceed with diagnosis, the mapping of the diagnostic signal space (residual values) onto the fault space is necessary. Different forms of this representation are known [3, 4, 8]. Most of them have static nature. However, the diagnosed processes are dynamical systems. Therefore, from the moment of fault occurrence to the moment when one can obtain measurable symptoms the particular time period elapses. In general, this time period is different for each fault and each diagnostic signal which detects that fault. Only after some period of time all symptoms are observed. Just a few approaches reference this problem [2].

The wrong diagnosis could be generated if one didn’t take the dynamic of symptoms into consideration. This is illustrated by the following example.

Example 1. Let us consider the diagnostic binary matrix presented in Fig. 1. Let us assume, that fault \( f_3 \) occurred. It is detectable by diagnostic signals \( s_2, s_3 \) and \( s_4 \). Assume that symptoms arising times are different for each diagnostics signal and equal respectively: \( \theta_1=1, \theta_2=2, \theta_3=4, \theta_4=6 \) [s]. Parallel diagnostic reasoning runs as follows:

- **Time 0-2 [s]:** the fault is not detected.
- **Time 2-4 [s]:** achieved diagnostic signals - \( S=\{0,1,0,0\} \); diagnosis - \( DGN_{2-4}\{f_6\} \).
- **Time 4-6 [s]:** achieved diagnostic signals - \( S=\{0,1,1,0\} \); diagnosis - \( DGN_{4-6}\{f_2\} \).
- **Time \( \geq 6 \) [s]:** achieved diagnostic signals - \( S=\{0,1,1,1\} \); diagnosis - \( DGN_{\geq6}\{f_3\} \). Only diagnosis \( DGN_{\geq6} \) is proper and has final nature. The earlier ones were false.

The following problems appears: How to take into account the symptoms delays in the fault isolation algorithms in order to eliminate possibility of formulating false diagnosis? Is it possible to develop faults isolation algorithm insensitive to symptoms delays? The problems presented above are the subject of this paper. Firstly, the formal description of symptoms delays is introduced together with the theoretical way of its calculation. Then, several diagnostic reasoning strategies that take these delays into account are presented. There are shown complex as well as simplified approaches.
2. SYMPTOMS DELAYS

The delays of symptoms forming depend on the dynamic characteristic of the process, fault type (abrupt, incipient), its time development characteristic, the applied method and detection algorithm parameters. It’s possible to calculate analytically these times basing on the dynamic description (e.g. transmittance) of the controlled part of the process (where the fault is an input and the process value is an output) and the transient response of fault appearance. We make an assumption that the limitation function parameters are known and there is no influence of the diagnostic test methods on the process operation.

The mathematical process description can be achieved based on the equations describing the physical effects taking place in the process. In this case, it is necessary to treat all the possible faults as separate inputs in the system of equations. After the linearization at the operating point and applying Laplace transformation one achieves linear model in the form [3, 4, 8]:

\[ y(s) = G(s)u(s) + G^F(s)f(s) \]  

Each constituent equation has the following form:

\[ y_j(s) = G_j(s)u(s) + G^F_{j}(s)f(s) \]  

while \( G_j(s) \) (j=1,…,J – number of diagnostic signals) denotes input-output transmittance:

\[ G_{p_j}(s) = y_j(s)/u_p(s); \quad p = 1,...,P, \]  

and \( G^F_{j}(s) \) denotes the transmittance for each fault-output couple:

\[ G_{kj}(s) = y_j(s)/f_k(s); \quad k = 1,...,K. \]  

If there are no existing faults in the process, then the following dependence is satisfied:

\[ y_j(s) - G_j(s)u(s) = G^F_j(s)f(s) = 0. \]  

The residuals are calculated based on the following equation called calculation form:

\[ r_j(s) = y_j(s) - G_j(s)u(s). \]  

Equation (7) (internal form) reflects general relation between particular residual and faults:

\[ r_j(s) = G^F_j(s)f(s) = G^F_{j1}(s)f_1(s) + ... + G^F_{jk}(s)f_k(s) + ... + G^F_{jK}(s)f_K(s). \]  

If \( r_j \) is sensitive for fault \( f_k \) and no other faults are present (\( f_m = 0 \)) then one achieves:

\[ r_j(s) = G^F_{jk}(s)f_k(s); r_m = 0 ; m = 1,2,...,K, m \neq k. \]  

For so defined residual its time development function is defined by the following relation:

\[ r_j(t) = L^{-1}(G^F_{j}(s)f_j(s)). \]  

The analytical calculation of the symptom forming times is difficult in practice because it requires the modeling of fault influence on measurable outputs. The fault development function as well as residual threshold value must be assumed arbitrary. Because of modeling errors the precision of analytical estimation of symptom times is poor.

In practice, based on the knowledge about the process and detection algorithms, it is possible to estimate symptom times by giving their minimum and maximum values [4, 6]. Let us use the following notation:

\[ \theta^1_{kj} - \text{minimal time period from } k^{th} \text{ fault occurrence to } j^{th} \text{ symptom appearance}, \]  

\[ \theta^2_{kj} - \text{ maximal time period from } k^{th} \text{ fault occurrence to } j^{th} \text{ symptom appearance}. \]  

These parameter can be expressed in seconds or dimensionless units equal multiple of the smallest process value sampling time. They are assigned to each ordered pair (fault, diagnostic signal) \( <f_k, s_j> \) satisfying the relation \( S \Rightarrow F \). The actual symptom time belongs to interval \( <\theta^1_{kj}, \theta^2_{kj}> \).

During process state monitoring the diagnosis should be formulated after all of the symptoms are time invariant. The approach that takes into account only the maximum symptom times \( \theta^1_{kj} \) allows to avoid generating false diagnosis.

The problem can be additionally simplified, by assigning the cumulative symptom time \( \theta_j \) to each diagnostic signal [5, 7]. The cumulative symptom time is defined as the maximum interval from the appearing of any of the faults controlled by this test to the moment when the symptom is detected:

\[ \theta_j = \max \left\{ \theta^1_{kj} \right\}; k : f_k \in F(s_j) \]  

The use of cumulative symptom times simplifies the way of describing the process dynamic properties and, especially, the reasoning algorithm. It is easier to define these parameters, however, it is still not an easy task.

The use of cumulative symptoms times was implemented in DTS and F-DTS methods presented by Kościelny (1995). The full description of the dynamic properties was applied in method i-DTS [7].

3. DIAGNOSTIC REASONING TAKING INTO ACCOUNT CUMULATIVE SYMPTOMS TIMES

Presented below diagnostic reasoning is based on the analysis of successive diagnostic signals and its cumulative symptoms times \( \theta_j \) introduced in section 2. The diagnosis is formulated in several steps, in which the set of possible faults is gradually constrained [4]. In the case of such reasoning (serial approach), the diagnostic relation \( R_{FS} \) is defined by attributing to each diagnostic signal the subset of faults detectable by this signal:

\[ F(s_j) = \left\{ f_k \in F : f_k \in R_{FS}s_j \right\} \]  

(11)
The isolation procedure is started after the first symptom is observed. Its occurrence indicates that one of the faults from the set \( F(s_j) \) of the faults detectable by that diagnostic signal had arisen. Such a subset of possible faults is indicated in the primary diagnosis:

\[
(t = t^1) \land (s_j = 1) \Rightarrow \text{DGN}_1 = F^1 = F(s_j = 1). \quad (12)
\]

The subset of diagnostic signals \( S^1 \) useful for isolation of faults from the set \( F^1 \) is created:

\[
S^1 = \left\{ f_j \in S : F^1 \land F(s_j) \neq \emptyset \right\}. \quad (13)
\]

The values of the diagnostic signals from the set \( S^1 \) are interpreted, step-by-step, according to the sequence determined by the attributed symptoms times. The \( j \)th diagnostic signal is used under the following condition that protects against formulating false diagnosis:

\[
(t - t^1) > \theta_j. \quad (14)
\]

The time instants of consecutive diagnostic signals interpretations are determined. They create the following series:

\[
0^1 \leq 0^2 \leq \ldots \leq \theta^1 \leq \ldots \leq \theta^p,
\]

where \( \theta^r \in \{ \theta_j : s_j \in S^1 \} \), while \( r \) defines the sequence of analysis of the diagnostic signals from the set \( S^1 \).

Successively, in the time instant \( t = t^1 + \theta^r \) for \( r = 1, \ldots, p \), the values of particular diagnostic signals are analyzed and the reduction of the set of possible faults takes place. The process state \( z(f_k) \) is attributed to each of the faults \( f_k \) from the set \( F \). It is defined in the following way:

\[
z(f_k) = \begin{cases} 0 & \text{the state without fault } f_k \\ 1 & \text{the state with fault } f_k \end{cases}
\]

(16)

The “0” value of the diagnostic signal testifies, that none of the faults controlled by that diagnostic signal had occurred:

\[
s_j = 0 \Leftrightarrow \bigwedge_{k : f_k \in F(s_j)} z(f_k) = 0. \quad (17)
\]

The “1” value testifies, that at least one of the faults from the set \( F(s_j) \) had occurred:

\[
s_j = 1 \Leftrightarrow \bigvee_{k : f_k \in F(s_j)} z(f_k) = 1. \quad (18)
\]

When single fault occurrence is assumed, the following rules of reducing the set of possible faults indicated in the consecutive steps of diagnosis formulation are used:

- The value of “0” of the diagnostic signal causes the reduction of the set of possible faults by the faults detectable by that signal:
  \[
s_j = 0 \Rightarrow \text{DGN}_r = \text{DGN}_{r-1} \cap F(s_j). \quad (19)
\]

- The value of “1” of the diagnostic signal causes the reduction of the set of possible faults by the faults undetectable by that signal. The new set of possible faults is a product of past possible faults and the set of faults detectable by that signal \( F(s_j) \):
  \[
s_j = 1 \Rightarrow \text{DGN}_r = \text{DGN}_{r-1} \cap F(s_j). \quad (20)
\]

During the diagnostic reasoning the preliminary diagnosis is formulated after the first symptom is observed and then constrained when further, consecutive diagnostic signal values are taken into account. Usually, there is no need to analyze all the signals to be able to formulate the final diagnosis. Such situation takes place when the diagnosis consists of only one fault or the set of indistinguishable faults.

**Example 2.** The serial reasoning in the case of fault \( f_6 \) appearance (example form Fig. 1) is show below. The first observed symptom is \( s_2 = 1 \). As a result, the following sets are created: the subset of possible faults \( F^2 = \{ f_1, f_2, f_3, f_5, f_6 \} \) and the subset of useful diagnostic signals: \( S^1 = \{ s_1, s_2, s_3, s_4 \} \).

The time instants when the successive diagnostic signals should be analyzed are determined:

\[
\theta^2 = \theta_1 = 1; \quad \theta^3 = \theta_3 = 4; \quad \theta^4 = \theta_6 = 6.
\]

Then, in the following steps, the diagnosis is constrained:

- \( t = t^1 + 1; s_1 = 0 \Rightarrow \text{DGN}_2 = \{ f_2, f_3, f_6 \} \)
- \( t = t^1 + 4; s_3 = 0 \Rightarrow \text{DGN}_3 = f_6. \)

After the second step the process of diagnosing is stopped. Finally, the same diagnosis as in the case of parallel reasoning is assumed but it is concluded basing on only three diagnostic signal values after 4 seconds when the first symptom was observed. The value of diagnostic signal \( s_2 \) was not needed for final diagnosis formulation.

### 4. SYMPTOMS BASED REASONING

In the above described diagnostic reasoning methods the information about the symptoms delays was used to avoid formulating false diagnosis before all the symptoms occur. However, achieving the data concerning the times of symptoms arising is not easy. The following question appears: Is it possible to formulate proper diagnosis without taking into account the symptom arise times? It is shown below, that is it possible.

In the described reasoning rules the information about the appearance of the particular symptoms, in the predefined interval, as well as the lack of other ones was used during diagnosis formulation. While the symptom appearing is easy to observe, one must wait for proper time period, when the symptom should appear, to be able to take into account its lack. It is possible to simplify the reasoning procedure by taking into account only the observed symptoms and rejecting the information carried out by the lack of symptoms. It means the use of the rule (18) and rejecting the rule (17). The set of possible faults is reduced only according to the rule (20). The
diagnosis, in each reasoning step, is proper and points out such faults, for which the observed symptoms are consistent with those ones defined in the signatures. However, the fault isolability can be lower.

Example 3. Let us assume, that the fault $f_2$ appears (example form Fig.1). It is detected by the symptom $s_3=1$, so: $t^*=1$.

$$\text{DGN}_v = \{f_1, f_2, f_3, f_6\}, \quad S^1 = \{s_1, s_2, s_3, s_4\}.$$ 

The diagnosis is modified after each new symptom appears. In this case, only one symptom appears: $s_3=1$. According to (20) one achieves: $s_3=1 \Rightarrow \text{DGN}_v = \{f_2\}$. It is a final diagnosis, because none other symptoms will appear. In comparison, the serial reasoning which takes into account symptom times was finished after taking into account the diagnostic signal $s_4=0$, in time moment $t=t^*+6$. It leads to more precise diagnosis: $\text{DGN}_v = f_2$.

One must also notice, that in the case of fault $f_6$ considered in Example 2, the serial reasoning based on symptoms is finalized in the first step: $s_2=1 \Rightarrow \text{DGN}_v = \{f_1, f_2, f_3, f_6\}$ with the diagnosis that pointing out four faults that are unisolable in respect to only one observed symptom (so far).

5. DIAGNOSIS BASED ON THE SYMPTOMS SEQUENCE

The sequence of symptoms arising is an important information which is worth of using in the diagnostic process. The different sequence of symptoms arising can allow to isolate undistinguishable faults with identical fault signatures.

The symptoms sequence for particular fault does not depend on the fault time development characteristic $f_x(t)$. Based on (9), assuming particular form of a function $f_x(t)$ (e.g. step function) and the threshold residual value it is possible to calculate the time, after which the symptom of a $k^{th}$ fault will appear. Such calculations must be done for all the residuals sensitive for $k^{th}$ fault.

The sequence of symptoms forming can be done by arranging the values of symptoms delays for particular set of residuals for each fault is achieved:

$$\text{SK}(f_k) = <s_1, s_2, s_3, s_4, \ldots>.$$  

The signature consists of the series of symptoms $s_j$ for particular faults $f_k$ written down in the order of appearing.

The different symptoms sequence can characterize faults that are unisolable based on binary diagnostics matrix (fault with the same signatures). The symptoms are not isolable (in respect to the relation $R_S$) based on symptoms sequence if their sequence signatures (21) are identical:

$$f_k R_S f_6 \Leftrightarrow \text{SK}(f_k) = \text{SK}(f_6).$$  

In this case, the reasoning consists of comparing registered symptom sequence with pattern ones describing particular faults:

$$\text{DGN} = \{f_k : \text{SK}(f_k) = \text{SK}\},$$  

where $\text{SK}$ denotes currently registered symptoms sequence.

It is sufficient to be able to isolate any pair of faults for which the sequence of any pair of fault symptoms is different:

$$\text{SK}(f_k) =< s_j, s_p >, \text{SK}(f_n) =< s_j, s_p >.$$

6. DIAGNOSTICS BASED ON THE KNOWLEDGE ABOUT SYMPTOM INTERVAL DELAYS

This section presents the fault isolation algorithm that utilizes the knowledge about the diagnostic relation and the values of the minimal and maximal symptoms forming delays. It assumes single fault scenarios, however, the multiple faults issue is also addressed. The algorithm implements serial diagnostic reasoning. The following notation is used: $\text{DGN}_v$ – final diagnosis elaborated in $r^{th}$ step of reasoning; $\text{DGN}_r, \text{DGN}_r^{**}$ – intermediate diagnosis.

Three main stages of reasoning algorithm can be distinguished: initialization, diagnosis specifying, and final diagnosis formulation.

Initialisation of isolation procedure. The isolation algorithm starts in the time $t^*=0$ when the first symptom $s_x^1=1$ is observed (fault detection). The following steps are conducted:

• Determining the set of possible faults. The primary set of possible faults is determined based on diagnostic relation. It consists of all the faults, for which the diagnostic signal with the observed symptom is sensitive for:

$$\text{SK}(f_k) = <s_1, s_2, s_3, s_4, \ldots>.$$

where: $\text{DGN}_1$ denotes temporary diagnosis, elaborated under the condition of use of the first diagnostic signal $s_x^1=1$ but without taking into account the intervals of symptoms delays; $\text{DGN}_1^{**}$ denotes that diagnostic signal $s_x^1$ detects the fault $f_x$ according to the diagnostic relation.

• Reduction of primary set of possible faults. Let us introduce the notations $0^1_{k,x}$ and $0^2_{k,x}$ for minimal and maximal periods from $k^{th}$ fault occurring to the first, detected symptoms $s_x^1 = 1$ formulation, respectively. The faults, which occurrence should cause another symptoms to be
observed before the symptom $s_i^1$, in respect to known intervals of symptoms delays, are eliminated from the set $DGN(s_i^1)$:

$$DGN_i = \{ f_k \in DGN^*_1 : \exists j (s_j \notin S^1) \land \big( \exists k \in DGN^*_1 : \forall x (0^1_k < x) \big) \}$$  \hspace{1cm} (26)

while $DGN_i$ denotes first, temporary diagnosis elaborated while taking into account the intervals of the symptoms delays.

**Determining the set of diagnostic signals useful for further fault isolation** in the following form:

$$S^* = \{ s_j : F(s_j) \cap DGN_1 \neq \emptyset \} - s^1_i.$$  \hspace{1cm} (27)

**Defining the intervals of symptoms possible consecutive forming**. Due to the fact that the real time of fault occurring is unknown (only the time of the first symptom detection is registered) the time intervals of the appearing of the consecutive symptoms of the diagnostic signals from the set $S^*$ must be recalculated in respect to the moment of the first symptom detection. Such calculations must be conducted for the faults pointed out in the diagnosis in the following way:

$$\beta_{k,j}^1 = \begin{cases} 0 & \text{if } 0^1_k - 0^2_k = 0 \\ \beta_{k,j}^{\min} & \text{if } 0^1_k - 0^2_k < 0 \\ \beta_{k,j}^{\max} & \text{if } 0^1_k - 0^2_k > 0 \end{cases}$$  \hspace{1cm} (28)

The parameters $\beta_{k,j}^{\min}$ for the faults $f_k \in DGN_1$ and the diagnostic signals $s_j \in S^*$ are arranged in ascending order.

Iterative diagnosis specifying. The second part of the following diagnosis takes place:

- after the detection of each, successive faults symptom,
- each time when the maximal period of the symptom delay $\beta_{k,j}^{\max}$ from the ordered series of these parameters passes.

During this stage, the following steps are conducted iteratively:

- The reduction of the set of possible faults based on diagnostics relation. If the symptom $s_j^1 = 1$ ($s_j \in S^*$) was detected in the proper period of delays than the set of possible faults is reduced according to formula:

$$s_j^1 = 1 \land (s_j \in S^*) \Rightarrow DGN^*_r =$$

$$= \{ f_k \in DGN_{r-1} : q(f_k,s_j) = 1 \land (t \in [\beta_{k,j}^1,\beta_{k,j}^2]) \}$$

Such an operation is realised for all the faults from the set $f_k \in F(s_j)$.

- The reduction of the set of possible faults based on the analysis of delays interval. The faults, which occurrence should cause another symptoms $s_j^r = 1$ to be observed before the currently observed symptom, in respect to the known intervals of symptoms delays, can be eliminated from the diagnosis elaborated in previous step:

$$DGN^*_r = \{ f_k \in DGN^*_1 : \beta_{k,j}^2 < \beta_{k,j}^1 \}$$  \hspace{1cm} (31)

- The reduction of the set of possible faults after the analysis of the maximal times of the symptoms delays. The lack of symptom after predefined time period, $t > \beta_{k,j}^{\max}$, allows for the reduction of the set of possible faults due to the formula:

$$s_j = 0 \land (s_j \in S^*) \land (t > \beta_{k,j}^2) \Rightarrow$$

$$DGN_r = \{ f_k \in DGN^*_1 - f_k \}$$  \hspace{1cm} (32)

**The end of fault isolation**. The algorithm stops when all the diagnostic signals from the set $S^*$ are taken into account.

Taking into account the symptoms forming delays can increase faults distinguishability comparing with the diagnosis elaborated basing only on binary diagnostic matrix. In some cases it reduces the diagnosing time.

**Example 4.** On the base of binary diagnostic matrix from Fig.1 faults $f_1$ and $f_3$ as $f_4$ and $f_5$ are indistinguishable. Let assume the symptoms delay intervals for the first pair of undistinguishable faults as follows:

$$[0^1_{1,1},0^2_{1,1}] = [1,3], [0^1_{1,2},0^2_{1,2}] = [4,5], [0^1_{3,1},0^2_{3,1}] = [6,8], [0^1_{3,2},0^2_{3,2}] = [2,5].$$

This implicates that in case of $f_1$ fault the symptom $s_i^1 = 1$ always appears before the symptom $s_j^1 = 1$, whereas $f_3$ fault occurrence will cause reverse sequence of the symptoms. It makes their unique recognition possible.

Let us assume that delays intervals for undistinguishable faults $f_1,f_3$ are as follows:

$$[0^1_{4,1},0^2_{4,1}] = [2,3], [0^1_{4,2},0^2_{4,2}] = [4,5], [0^1_{7,1},0^2_{7,1}] = [1,3], [0^1_{4,3},0^2_{4,3}] = [7,9].$$

Both faults results with the same symptoms sequence but in spite of this they are distinguishable. Maximal $s_i^1 = 1$ symptom delay after $s_j^1 = 1$ symptom for $f_1$ fault equals 5-2=3, whereas minimum time interval between these symptoms for $f_3$ fault equals 7-3=4. Therefore if $s_i^1 = 1$ symptom is delayed relatively to $s_j^1 = 1$ symptom less than 3 seconds that indicates $f_3$ fault, if delay is bigger we are inferring about $f_1$ fault occurrence.

The algorithm has also limited ability to isolate multiple faults. In general, if the new symptom $s_j^1 = 1$ is observed, and if it does not appeared in the predefined period of symptoms delays in respect to the first observed symptom, than the new fault isolation thread is started. In that case it is assumed that this symptom is caused by another fault than the faults pointed out in the previous steps. In this case it is sometimes possible to formulate final diagnosis.
about multiple faults if the proper sets of diagnostic signals used in each isolation thread fulfill some necessary conditions. The detailed description of that problem is not in the scope of this paper. Some information about creating fault isolation threads can be found in [7].

7. FINAL REMARKS

The Section 3 presents the reasoning algorithm that takes into account the simplified information about symptoms delays and enables to elaborated proper diagnosis for dynamic systems. It was also shown, in Section 4, that one can achieve proper diagnosis without taking directly into account the information about symptom forming times, however, it leads to lower fault isolability.

The knowledge about symptoms interval delays enables, in many cases, to isolate the fault that are unisolable based on binary diagnostic matrix. However, to be able to determine the symptoms intervals we need the residual equations in the inner form or very precise expert knowledge. This is very difficult to obtain in practice. Such algorithm with detailed description was presented in Section 6.

The alternative approach was shown in Section 5. The knowledge about the sequence of symptoms generation also enables, in many cases, to isolate the fault that are unisolable based on binary diagnostic matrix. It is easier to define such a sequence than precise symptoms interval delays.

AKNOWLEDGMENTS

The authors acknowledge partial funding support of Polish Ministry of Education and Higher Education under the grant R01 012 02 DIASTER: Intelligent diagnostic system supporting control of industrial processes. Project financed from the fund supporting research in the years 2007-2009.

REFERENCES