

Diagnosis of Stochastic Discrete Event Systems Based On N -gram Models With Wildcard Characters

Kunihiko HIRASHI, Miwa YOSHIMOTO, Koichi KOBAYASHI

School of Information Science,

Japan Advanced Institute of Science and Technology

Nomi-shi, Ishikawa, Japan

{hira,k-kobaya}@jaist.ac.jp

Abstract—In this paper, a new approach to the diagnosis of stochastic discrete event system is presented. We are developing a method, called *sequence profiling*, based on N -gram models. The information necessary for sequence profiling is only event logs from the target system. From event logs in the normal situation, N -gram models are constructed through a simple statistical analysis. Based on the N -gram model, the diagnoser estimates what kind of faults has occurred in the system, or may conclude that no faults occurs. When the target system is a distributed system consisting of several subsystems, event sequences from subsystems may be interleaved and the method cannot separate the event sequence from local event sequences by subsystems. To improve this situation, we introduce the wildcard characters in the short sequences used in the N -grams. This contributes to removing the effect by subsystems which may not be related to faults. Effectiveness of the proposed approach is demonstrated by application to fault diagnosis of a multi-processor system.

Index Terms—Discrete event systems, diagnosis, N -gram model, wildcard characters

I. INTRODUCTION

The problem of failure diagnosis has been studied from 1970's, and there are various schemes for solving the problem. Model-based diagnosis (MBD) is a method using system models that describe causal relationship on events occurring. There are many literature on MBD for discrete event systems (DES) (e.g., [2]). Diagnosis for stochastic DES and decentralized DES is also studied [3], [4], [5], [6]. In MBD, faults are detected by comparing event logs observed in the actual system with those given by the system model. This process is performed by finite-state automata called *diagnosers*. One of advantages of MBD is that it does not require any empirical rules depending on the target system. Moreover, MBD can be applied at low computational cost once the diagnosers are constructed. This is a desirable feature for on-line autonomous failure detection. However, MBD has the following weaknesses:

- The size of the diagnoser can be doubly exponential in the size of the system description in the worst case. This makes the diagnoser practically impossible to generate. To overcome the problem on computational cost, SAT-based diagnosis was proposed [7]. Diagnosis based on behavior reconstruction, which is an intensional representation of all behaviors that are consistent with the system observation, was also proposed [8].
- MBD requires an accurate and detailed model of the target system. However, when the configuration of the target

system changes dynamically, such as systems operated on virtual environments like in cloud computing systems, it is difficult to obtain accurate models for the entire system. Updating the system model is hard to be done in online environment.

To overcome these weaknesses, we proposed a new scheme for the diagnosis of stochastic discrete event systems [1]. The method is called *sequence profiling* (SP). The advantage of SP is that it does not require any system models and any system-specific knowledge. Only the information necessary for the diagnosis is event logs from the target system. From the event logs, N -gram models that approximate the behavior of the target system are extracted. The N -gram model was firstly introduced by Shannon [9], and are mainly used for natural language processing [10], and recently applied to text mining and web search algorithms. In the proposed method, N -gram models are used for discovering discrepancy between the observed event logs and the behavior of the system in the normal situation. The discrepancy is measured by the *specificity* of short event sequences in the event logs, where the short sequences correspond to local activities of the system. This is the main idea of the proposed approach.

One of the problem in the method of [1] is that the number of the short sequences of fixed length becomes large. Moreover, when the target system is a distributed system consisting of several subsystems, event sequences from subsystems may be interleaved. However, the method cannot separate the event sequence from local event sequences by subsystems. In this paper, we introduce *wildcard characters* in the short sequences. This will contributes to (i) reducing the number of short sequences to be considered, and (ii) removing the effect by subsystems which may not be related to faults.

The paper is organized as follows. In Section II, notations on formal languages and automata are given. In Section III, N -gram models used in the proposed method are defined. In Section IV, the proposed diagnosis algorithm is presented. In Section V, the algorithm is evaluated by using a diagnosis problem of a multi-processor system. Comparison with MBD and the effect of introducing wildcard characters are described in Section VI. Section VII is the conclusion.

II. PRELIMINARIES

Let Σ be a finite set of symbols and let Σ^* denote the set of all finite sequences over Σ . For a positive integer k , let $\Sigma^k = \{s \in \Sigma^* \mid |s| = k\}$ denote the set of all sequences of length k . Let s be a sequence over Σ . We write $|s|$ to denote the length of s . The i -th symbol of s is denoted by $s_{[i]}$ and the subsequence from the i -th position to the j -th position is denoted by $s_{[i,j]}$. Let s and w be sequences over Σ , where $|s| < |w|$. The number of occurrences of s as a subsequence of w is denoted by $O_s(w)$.

In this paper, we assume that the target system is modeled by a nondeterministic probabilistic automaton $G = (X, \Sigma \cup \{\varepsilon\}, \delta, P, x_0, F)$, where

- $X = \{x_1, \dots, x_n\}$ is the set of states,
- Σ is the set of symbols and ε is the empty sequence,
- $\delta \subseteq X \times \Sigma \times X$ is the transition relation,
- $P : X \times \Sigma \times X \rightarrow [0, 1]$ is the function defining probability of each state transition, where $\sum_{\sigma \in \Sigma, x_j \in X} P(x_i, \sigma, x_j) = 1$ holds for each state x_i ,
- $x_0 \in X$ is the initial state,
- $F \subseteq X$ is the set of final states.

We write P_{ij}^σ to denote the probability $P(x_i, \sigma, x_j)$ for the state transition (x_i, σ, x_j) . Moreover, let $P_{i\bullet}^\sigma := \sum_j P_{ij}^\sigma$. As usual, the transition relation is extended to the multi-step transition $\delta \subseteq X \times \Sigma^* \times X$ by (i) $(x_i, \varepsilon, x_i) \in \delta$ and (ii) $(x_i, s\sigma, x_k) \in \delta$ if $(x_i, s, x_j) \in \delta$ and $(x_j, \sigma, x_k) \in \delta$.

The underlying Markov chain of G is given by the set X of states, and the probability for each pair of states $P_{ij} = \sum_{\sigma \in \Sigma} P_{ij}^\sigma$.

III. N-GRAM MODELS

The N -gram model of a sequence w over Σ is the collection of following probabilities for all $y \in \Sigma^{N-1}$ and all $\sigma \in \Sigma$:

$$Pr(\sigma|y) := \frac{O_{y\sigma}(w)}{\sum_{\sigma' \in \Sigma} O_{y\sigma'}(w)} \quad (1)$$

Let $G = (X, \Sigma \cup \{\varepsilon\}, \delta, P, x_0, F)$ be a probabilistic automaton. Suppose that a finite sequence s has occurred from some unknown state of G . We consider the problem of estimating the next event to occur. Let X_s denote the set of states such that sequence s has occurred just before reaching it, i.e.,

$$X_s = \{x_j \mid \exists x_i \in X : (x_i, s, x_j) \in \delta\}. \quad (2)$$

We define an equivalence \sim on X by $x_i \sim x_j \Leftrightarrow \forall \sigma \in \Sigma : P_{i\bullet}^\sigma = P_{j\bullet}^\sigma$. We say that G is *predictable* for sequence s if all states in X_s are equivalent w.r.t. \sim . Note that if $|X_s| = 1$ then the automaton is obviously predictable for s . Moreover, G is called k -*predictable* if G is predictable for all sequence s of length k .

Whether G is k -predictable or not is determined by its structure. The automaton in Fig. 1 is 1-predictable. However, the automaton in Fig. 2 is not k -predictable for any k because G is not predictable for sequences b^*a of arbitrary length ($X_{b^*a} = \{x_1, x_2\}$).

Suppose that a probabilistic automaton G is $(N - 1)$ -predictable. Then we can obtain the following N -gram model

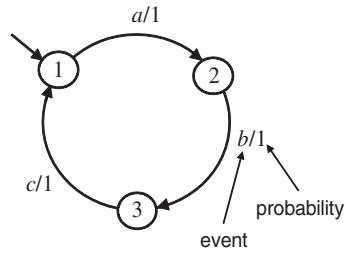


Fig. 1. Probabilistic automaton: example 1.

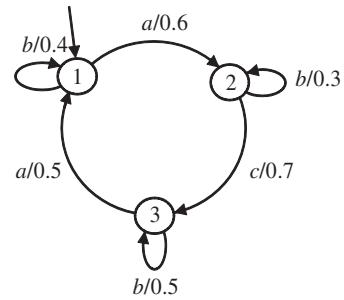


Fig. 2. Probabilistic automaton: example 2.

that correctly gives the probability of the next occurrence of events. For each $s \in \Sigma^{N-1}$ and $\sigma \in \Sigma$,

$$Pr(\sigma|s) = P_{i\bullet}^\sigma \quad (3)$$

where $x_i \in X_s$.

Suppose that G is not $(N - 1)$ -predictable and $|X_s| > 1$ for some $s \in \Sigma^{N-1}$. If the automaton has the steady state and the stationary probability $\pi = (\pi_1, \dots, \pi_n)$ is known, where π_i is the probability that the system is in state x_i , then we can build the following N -gram model that approximates the behavior of G : for each $s \in \Sigma^{N-1}$ and $\sigma \in \Sigma$,

$$Pr(\sigma|s) = \sum_{x_i \in X_s} \pi_i \cdot P_{i\bullet}^\sigma \quad (4)$$

The following theorem is a direct consequence of the theory of Markov chains.

Theorem 3.1: For a probabilistic automaton whose underlying Markov chain is ergodic, then the N -gram model in (4) is obtained by (1) for a sufficiently long sequence w that has occurred in G .

IV. FAULT IDENTIFICATION ALGORITHM

A. Fault Identification Problem

The faults considered in this paper are changes of transition probabilities in the probabilistic automaton. Let $G_i, i = 0, 1, \dots, m$ be probabilistic automata, where G_0 is an automaton that models the fault free situation. Other automata $G_i, i = 1, \dots, m$ are models for faulty situations, where G_i 's are the same automaton, except that probabilities of some state transitions are different. We call G_0 the *normal* model, and call other G_i 's *faulty models*.

In this paper, fault detection of stochastic discrete event systems is considered as the following problem:

Definition 4.1 (Fault Identification Problem): Given a sequence w_{test} of events over Σ , choose a system model from $G_i, i = 0, 1, \dots, m$ that most likely generates w_{test} .

If the normal/faulty system models $G_i, i = 0, 1, \dots, m$, are given, then the problem is solvable by model-based diagnosis. We put the following assumption on the input data:

- System models $G_i, i = 0, 1, \dots, m$, are unknown or hard to obtain.
- Event logs from the normal model G_0 are available.
- Event logs from the faulty models $G_i, i = 1, \dots, m$, are available.

B. Specificity of Sequences

We define the *specificity* of sequences based on N -gram models. We assume that the k -gram model of sequence w is obtained for all $1 \leq k \leq N$. We call the collection of these models the N^{\leq} -gram model. Based on this model, we compute the conditional probability $Pr(s|y)$ for sequence y of arbitrary length by

$$Pr(s|y) := Pr(s|y_{[j,|y|]}) \quad (5)$$

where $j = \max \{|y| - N + 2, 1\}$.

Suppose that we already have an N^{\leq} -gram model for G_0 . Let s be a sequence of length r . Then the conditional probability that sequence s appears after sequence y is given by the following recursion:

$$Pr(s | y) := Pr(s_{[1]} | y) \cdot Pr(s_{[2,|s|]} | ys_{[1]}) \quad (6)$$

We show an example. Using the 3-gram model, $Pr(abx|xyz)$ is given by

$$\begin{aligned} & Pr(abc|xyz) \\ &= Pr(a|xyz) \cdot Pr(bc|xyza) \\ &= Pr(a|xyz) \cdot Pr(b|xyza) \cdot Pr(c|xyzab) \\ &= Pr(a|xyz) \cdot Pr(b|yza) \cdot Pr(c|zab) \end{aligned}$$

where $Pr(b|xyza) = Pr(b|yza)$ and $Pr(c|xyzab) = Pr(c|zab)$ as we have assumed in (5).

Next we define the expected number of times s occurs in w based on the N^{\leq} -gram models. We use here the following number:

$$E_s^N(w) = \sum_{y \in \Sigma^{N-1}} O_y(w) \cdot Pr(s|y) \quad (7)$$

Then the specificity of sequence $s \in \Sigma^*$ in w w.r.t. the N^{\leq} -gram model as follows:

$$d_s^N(w) := \log \frac{O_s(w)}{E_s^N(w)} \quad (8)$$

If $d_s^N(w)$ is positive (negative), then s appears more (less) often than the average.

Let $\mathcal{D}^{r,N}(w)$ denote the $|\Sigma^r|$ -dimensional vector each component of which is $d_s^N(w)$ for $s \in \Sigma^r$. We can expect $\mathcal{D}^{r,N}(w)$ captures some characteristics of sequence w , and therefore, we call it the *profile* of w w.r.t. the two parameters r and N . Each

sequence w has many profiles according to the selection of r and N . We note that the dimension of $\mathcal{D}^{r,N}(w)$ is $|\Sigma^r| = |\Sigma|^r$. This is an exponential function of r , but is *independent* of the length of w .

C. Specificity of Sequences Including Wildcard Characters

A *wildcard character* is the symbol which can be replaced by any symbol in a given alphabet Σ . Let “ n ” denote the wildcard character, where $n \notin \Sigma$, and we consider a sequence $\hat{s} \in (\Sigma \cup \{n\})^*$ including n . We will call such a sequence \hat{s} a *motif*. Note that any sequence in Σ^* is a motif by the definition. Given a motif \hat{s} , let $\mathcal{C}(\hat{s})$ denote the set of all sequences in Σ^* obtained by substituting every occurrence of n in \hat{s} by a symbols in Σ . We call $\mathcal{C}(\hat{s})$ a *concretization* of \hat{s} .

We show an example of a motif and its concretization. Let $\Sigma = \{a, b\}$. Then “ $anbn$ ” is a motif of $\Sigma^{4,2}$, and its concretization is $\mathcal{C}(\hat{s}) = \{aaba, aabb, abba, abbb\}$.

The motif $anbn$ corresponds to an interleaving of sequence ab and another sequence of length 2.

Given two parameters r and h , let $\Sigma^{r,h}$ denote the set of all sequences in $(\Sigma \cup \{N\})^r$ such that the total number of “ n ”s in each sequence is exactly h .

There are ${}_4C_2 = 6$ motifs for the interleaving: $nnab, nanb, nabn, annb, anbn, abnn$. We can observe that $nnab, nabn$ and $abnn$ are identical by shifting them; $nanb$ and $anbn$ are also identical. By this reason, it suffices to consider motifs begin with a symbols in Σ . Let $\Sigma_*^{r,h}$ denote the set of such motifs in $\Sigma^{r,h}$. The number of motifs in $\Sigma_*^{r,h}$ is ${}_{r-1}C_h |\Sigma|^{r-h}$.

For a given motif \hat{s} and a sequence $y \in \Sigma^*$, we define $O_{\hat{s}}(w)$ and $E_{\hat{s}}^N(w)$ by

$$O_{\hat{s}}(w) := \sum_{s \in \mathcal{C}(\hat{s})} O_s(w) \quad (9)$$

$$E_{\hat{s}}^N(w) := \sum_{s \in \mathcal{C}(\hat{s})} E_s^N(w) \quad (10)$$

The specificity of motif \hat{s} in w w.r.t. the N^{\leq} -gram model is defined as follows:

$$d_{\hat{s}}^N(w) = \log \frac{O_{\hat{s}}(w)}{E_{\hat{s}}^N(w)} \quad (11)$$

If $d_{\hat{s}}^N(w)$ is positive (negative), then \hat{s} appears more (less) often than the average.

We can defined profiles based on motifs as follows. Let $\mathcal{D}^{r,h,N}(w)$ denote the $|\Sigma_*^{r,h}|$ -dimensional vector each component of which is $d_{\hat{s}}^N(w)$ for motif $\hat{s} \in \Sigma_*^{r,h}$.

D. Diagnosis Algorithm

Using event logs from the system in the normal situation, i.e., the situation in which no fault occurs, an N^{\leq} -gram model is constructed (Learning Phase). Once an event log is given from the system, the diagnoser estimates what kind of faults has occurred in the system, or may conclude that no faults occurs (Diagnosis Phase) (Fig. 3).

The detailed procedure of the proposed method is described as follows:

[Sequence Profiling]

- 1) Give the three parameters $r, h, N \in \mathbb{N}$, where $h < r$.
- 2) Let w_{ref} be an event sequence generated by G_0 , i.e., the system in the normal situation. Compute the N^{\leq} -gram model of w_{ref} .
- 3) Based on the N^{\leq} -gram model, compute the profile $\mathcal{D}^{r,h,N}(w_{ref})$.
- 4) Let w_i be an event sequence generated by model $G_i, i = 0, 1, \dots, m$, where $|w_0|, |w_1|, \dots, |w_m| \ll |w_{ref}|$, where there w_i 's are relatively shorter than w_{ref} . Based on the N^{\leq} -gram model, compute the profile $\mathcal{D}^{r,h,N}(w_i)$ for each $w_i, i = 0, 1, \dots, m$.
- 5) Find the vector $\mathcal{D}^{r,h,N}(w_i)$ that shows the highest correlation with $\mathcal{D}^{r,h,N}(w_{ref})$, and output the model G_i . This can be done by clustering of the specificity vectors, or by computing corelation coefficients between $\mathcal{D}^{r,h,N}(w_{test})$ and $\mathcal{D}^{r,h,N}(w_i), i = 0, 1, \dots, m$.

E. Diagnosis Under Partial Observation

Let $G = (X, \Sigma \cup \{\varepsilon\}, \delta, P, x_0, F)$ be a probabilistic automaton. An observation map is a function $\mathcal{O} : \Sigma \rightarrow \Pi \cup \{\varepsilon\}$, where Π is a different set of symbols. The domain of \mathcal{O} is naturally extended to Σ^* . We consider the fault identification problem for the case that $\mathcal{O}(w)$ is observed instead of event sequence w occurring in the models.

We can apply the same method to the probabilistic automaton defined over Π , i.e., the probabilistic automaton $G' = (X, \Pi \cup \{\varepsilon\}, \delta', P', x_0, F)$ obtained by $(x_i, \mathcal{O}(\sigma), x_j) \in \delta'$ if $(x_i, \sigma, x_j) \in \delta$, and $P'(x_i, \mathcal{O}(\sigma), x_j) = P(x_i, \sigma, x_j)$.

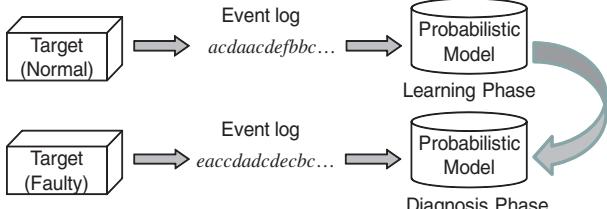


Fig. 3. Overview of the proposed approach.

V. EXPERIMENTAL RESULTS

A. A Multi-processor System

We apply the proposed method to a multi-processor system described in [7]. The system consists of 16 Processing Units (PUs) arranged on 4*4 grid layout. Fig. 4 shows the state transition diagram of a single PU, and Fig. 5 shows the system configuration. When a fault occurs in some PU, the PU will reboot and simultaneously send a message *reboot!* to 4-neighbors. If a neighbor PU receives a reboot message *reboot?*, then the PU will reboot (*IReboot*), too. After the rebooting process *rebooting*, the PU returns to the normal operation (*IAmBack*).

We model the system by a stochastic Petri net (SPN) tool StnPlay [11]. During the simulation, all event occurrences are recorded in a log. In faulty models, we give smaller values for

stochastic exponential delays of event *reboot!*. Note that PUs may reboot even in the normal model, but the frequency of rebooting is lower than that in the faulty models.

Each event log is converted into an event sequence of symbols, according to the following observation map:

- *IReboot* of PU1,2,5,6,9,10,13,14: a
- *IReboot* of PU3,4,7,8,11,12,15,16: b
- *IAmBack* of PU1,2,5,6,9,10,13,14: x
- *IAmBack* of PU3,4,7,8,11,12,15,16: y
- Other events: ε (unobservable)

Note that faulty events are unobservable. Therefore, we need to detect the fault by observation of other observable events.

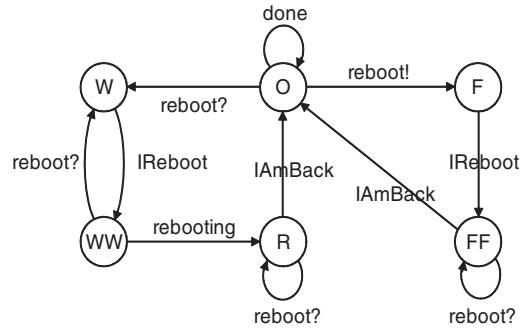


Fig. 4. A single processing unit.

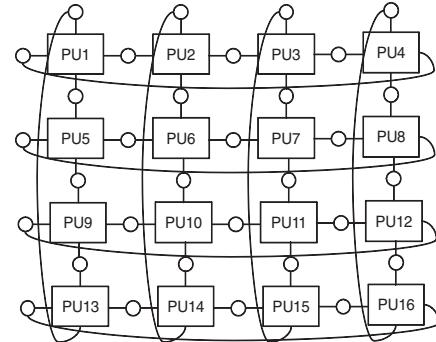


Fig. 5. System configuration.

B. Test Data

We prepare the following three models:

- Normal Model: The mean delay of *reboot!* is 10 for all PUs.
- Faulty Model 1: The mean delay of *reboot!* is 2 for PU1, and 10 for other PUs.
- Faulty Model 2: The mean delay of *reboot!* is 4 for PU1 and PU7, and 10 for other PUs.

Event logs w_{ref} and w_0 are from the normal model. Event log w_1 is from the faulty model 1, and event log w_2 and w_{test} are from the faulty model 2.

Before applying the observation map, $|w_{ref}| = 1,000,000$ and $|w_0| = |w_1| = |w_2| = 100,000$. After applying the

observation map, $|\mathcal{O}(w_{ref})| = 432,770$, $|\mathcal{O}(w_0)| = 43,360$, $|\mathcal{O}(w_1)| = 44,235$, $|\mathcal{O}(w_2)| = 44,203$, and $|\mathcal{O}(w_{test})| = 44,205$.

C. Results

Table I shows the number of event occurrences for each observable events. Since the correlation coefficients between columns for w_{test} and other event logs are more than 0.98 in all combinations, it is hard to identify the faulty models based only on the number of event occurrences.

Using vectors $\mathcal{D}^{3,0,N}(\cdot)$, it is possible to identify the correct model. Correlation coefficients between $\mathcal{D}^{3,0,N}(w_{test})$ and other $\mathcal{D}^{3,0,N}(w_i)$'s are shown in Fig. 6. From the result from $N = 3$, we can identify model 2 as the correct model.

However, this approach is not valid for profiles w.r.t. longer sequences. Fig. 7 shows the results for $\mathcal{D}^{6,0,N}(\cdot)$ and we cannot distinguish model 1 and model 2 clearly. Fig. 8 shows the results of $\mathcal{D}^{6,3,N}(\cdot)$. It is possible to identify the correct model by N -gram models for $N \geq 3$. Fig. 9 shows the results of $\mathcal{D}^{6,h,5}(\cdot)$ for different h . It is observed $h = 3$ gives the best result for the model identification.

Table II-IV shows the top 10 sequences with high specificities in the three models: Normal mode, Faulty model 1 and Faulty model 2. In the normal model, specificities are low. In the faulty model, sequences including *a* (*IReboot* of PUs in the left side) shows high specificities. This implies that the fault might occur in the left side. In the faulty model 2, however, sequences including *a* and *b* (*IReboot* of PUs in the right side) are mixed in the top 10 sequences. This implies that that faults might occur in both sides.

TABLE I
STATISTICS ON THE NUMBER OF EVENT OCCURRENCES.

	w_0	w_1	w_2	w_{test}
<i>a</i>	12,980	14,634	13,468	13,381
<i>x</i>	8,562	8,506	8,619	8,663
<i>y</i>	8,581	8,177	8,623	8,656
<i>b</i>	13,237	12,918	13,493	13,506

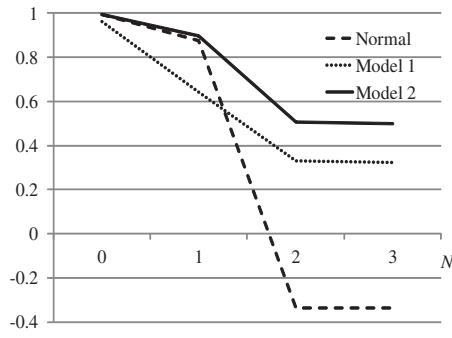


Fig. 6. Correlation coefficients between $\mathcal{D}^{3,0}(w_{test})$ and other $\mathcal{D}^{3,0,N}(w_i)$'s for $N = 0, \dots, 3$.

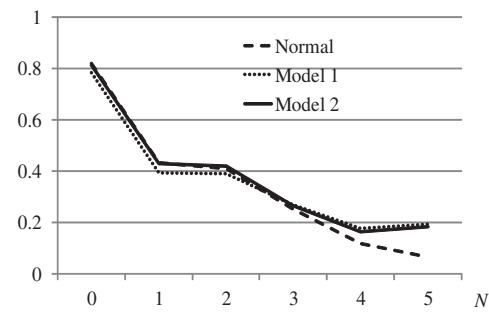


Fig. 7. Correlation coefficients between $\mathcal{D}^{6,0}(w_{test})$ and other $\mathcal{D}^{6,0,N}(w_i)$'s for $N = 0, \dots, 5$.

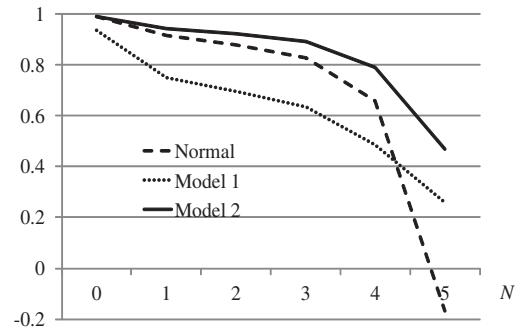


Fig. 8. Correlation coefficients between $\mathcal{D}^{6,3}(w_{test})$ and other $\mathcal{D}^{6,3,N}(w_i)$'s for $N = 0, \dots, 5$.

VI. DISCUSSION

The objective of MBD is to identify the fault that has occurred in the system by observing event sequences from the system. The objective of the sequence profiling (SP) is different from that of MBD. Faults are defined as faulty events in MBD, while they are detected as changes of transition probabilities in SP. In other words, MBD traces occurrence of events in the system model, while SP focuses on frequency of event occurrence. Changes of transition probabilities may correspond to some unobservable system errors. SP can detect whether

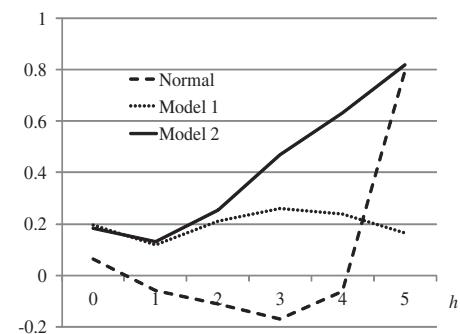


Fig. 9. Correlation coefficients between $\mathcal{D}^{6,h,5}(w_{test})$ and other $\mathcal{D}^{6,h,5}(w_i)$'s for $h = 0, \dots, 5$.

TABLE II
TOP 10 SEQUENCES IN NORMAL MODEL.

Sequence	Specificity
<i>bnnnax</i>	0.082636665
<i>ynxbnn</i>	0.078198404
<i>ynynxn</i>	0.073061752
<i>bnbbnn</i>	0.071736607
<i>xnnnxz</i>	0.070783565
<i>yxnnxn</i>	0.069698734
<i>xbnnbn</i>	0.069627383
<i>ynnbna</i>	0.06945012
<i>ynnnxx</i>	0.067944391
<i>ynnnzb</i>	0.067072112

TABLE III
TOP 10 SEQUENCES IN FAULTY MODEL 1.

Sequence	Specificity
<i>yaxnnn</i>	0.446642379
<i>xnaxnn</i>	0.385756871
<i>yaynnn</i>	0.3695851
<i>ayannn</i>	0.363505804
<i>xaynnn</i>	0.360871346
<i>ananna</i>	0.330605425
<i>axannn</i>	0.321329947
<i>xnaynn</i>	0.31243332
<i>axnann</i>	0.308730344
<i>yanynn</i>	0.304838413

some faults occur or not, and can identify event patterns related to the faults. However, additional investigation is necessary for identifying what has actually occurred in the system. By this reason, an actual use case of SP will be as follows:

- 1) Apply SP to detect occurrence of faults, and to estimate in which part of the system faults occur. This step should be done in on-line manner.
- 2) Apply the detailed analysis for the faulty subsystems. MBD or other diagnosis methods are used.

To apply SP to on-line diagnosis, it is necessary to develop a method to detect the faults by shorter w_{test} . This remains as future work.

The number of states in the multiprocessor system shown in Section V will be 6^{16} since all the subsystems are mutually independent. MBD cannot be applied to such a large system. This is the major advantage of SP comparing with MBD. Another feature of SP is that it is robust for data noise, where the noise means some of the observable events are lost, the

TABLE IV
TOP 10 SEQUENCES IN FAULTY MODEL 2.

Sequence	Specificity
<i>xbannn</i>	0.239204665
<i>yaxnnn</i>	0.178624349
<i>annnya</i>	0.167122029
<i>yaynnn</i>	0.155783797
<i>xbnynn</i>	0.144907917
<i>ayannn</i>	0.142148264
<i>yanxnn</i>	0.141006404
<i>ynaxnn</i>	0.132237506
<i>axnnny</i>	0.12804503
<i>anybnn</i>	0.127656013

order of some events are sometimes reversed, etc.

Comparing with the result in [1], what's new in this paper is introduction of the wildcard character in the N -grams. The effects of introducing wild characters are analyzed as follows:

- Introduction of the wildcard character contributes to removing the effect by subsystems which may not be related to faults.
- $|\Sigma^{r,h}|$ is relatively smaller than $|\Sigma^r|$. This works good for reducing the effect by randomness in event occurrences. Consider the situation in which $O_s(w)$ is small. Then the effect by randomness is greater than that by the faults.

The optimal number for h , the number of wildcard characters, is determined by the system mode, but is not known in advance. Therefore, it is necessary to try various values for h . Once the optimal value is known, we can use it for future diagnosis.

VII. CONCLUSION

Since order of events is taken into consideration in the N -gram model, the proposed approach is different from simple statistical analysis such as counting occurrences of each event. Short sequences used for the specificity correspond to local behavior of the system. When the target system is a distributed system like one used for the experiments, the effect of a faulty event is limited to the subsystem around the event, and other parts of the system runs normally. Introduction of wildcard characters can mask such behavior not related to the faults. In the future, we will extend the proposed approach to timed systems. Introducing events counting inter-event duration may be used.

REFERENCES

- [1] M. Yoshimoto, K. Hiraishi, Model-less Diagnosis of Discrete Event Systems Based on Sequence Analysis, ITC-CSCC2012, CD-ROM, C-W1-03 (2012)
- [2] M. Sampath, R. Sengupta, and S. Lafortune, Diagnosability of Discrete-Event Systems, IEEE Trans. Automatic Control, Vol. 40, No. 9, pp.1555–1575 (1995)
- [3] Xi Wang, I. Chattopadhyay, and A. Ray, Probabilistic Fault Diagnosis in Discrete Event Systems, 43rd IEEE Conference on Decision and Control, pp.14–17 (2004)
- [4] D. Thorsley and D. Teneketzis, Diagnosability of Stochastic Discrete-Event Systems, IEEE Trans. Automatic Control, Vol. 50, No. 4, pp.476–492 (2005)
- [5] W. Qiu and R. Kumar, Decentralized Failure Diagnosis of Discrete Event Systems, IEEE Trans. Systems, Man and Cybernetics, PART A: Systems and Humans, Vol.36, No.2, pp.384–395 (2005)
- [6] F. Liu, D. Qiu, H. Xing, and Z. Fan, Decentralized Diagnosis of Stochastic Discrete Event Systems, IEEE Tans. Automatic Control, Vol. 53, No. 2, pp.535–546 (2008)
- [7] A. Grastein, Anbulagan, J. Rintanen, and E. Kelareva, Diagnosis of Discrete-Event Systems Using Satisfiability Algorithms Proc. AAAI'07, Vol. 1, pp.305–310 (2007)
- [8] G. Lamperti, M. Zanella and P. Pogliano, Diagnosis of Active Systems by Automata-Based Reasoning Techniques Applied Intelligence, Vol. 12, No. 3 pp.217–237 (2000)
- [9] C. E. Shannon, A Mathematical Theory of Communication, Bell System Tech. J., Vol. 27, pp.379–423,623–656 (1948)
- [10] M. Nagao and S. Mori, A New Method of N-gram Statistics for large Number of n and Automatoc Extraction of Words and Phrases from Large Text Data of Japanese, Proc. COLING'94, Vol.1, pp.611–615 (1994)
- [11] <http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/db/stpnplay.html>