Hypothesis Test for Unsuccessful Partial Explorations of State Spaces of Petri Nets

Eleazar JIMENEZ SERRANO, Member, IAENG

Abstract— We describe a method using statistical hypothesis test for handling the unsuccessful results of partial exploration of the state space of marked graphs PN systems. The first contribution of our method is the estimation of the probability distribution function of the state space of PN systems. The second one is to give completeness to the unsuccessful results by determining reachability with a level of confidence through the test. In this paper we focus in PN systems with the form of structurally bounded marked graphs with probability distribution of their set of reachable markings believed to be normal.

Index Terms—Petri nets, Hypothesis test, Reachability Problems, State space

I. INTRODUCTION

O UR work focuses on determining the existence of undesired states without having to explore the entire state space generated from large and bounded Petri net (PN) systems. With a heuristically guided partial exploration of the state space we visit and register a number of states believed to be smaller than the complete state space and search for the target state.

Works related to determining reachability in the state space generated from PN systems usually utilize reachability graphs exploring the entire state space but limited due to the state space explosion problem [1]. On the other hand, the list of works exploring only a portion of the state space is large but many of them face convex termination, lack of conclusiveness and their efficiency is ignored [1-4], although good results exist in the confines of specific practicability.

Lack of conclusiveness represents a wasted cost, therefore upon unsuccessful search, we will treat the results with a statistical hypothesis tests to decide with a level of confidence if the undesired state exists or not. This novel approach integrates certainty to the incompleteness of the heuristically guided partial exploration of the state space using a test for proportions.

First, from the unsuccessful results of the partial exploration, we will utilize the explored states as a sample for statistical inference. We will estimate the probability distribution function (PDF) of the state space, the proportion of states with the same cardinality as the undesired state and define a hypothesis test to determine if the undesired state exists or not in the state space. We believe we can accurately estimate the proportion of states in the state space with same cardinality as the undesired state and obtain a reliable result with the hypothesis test. With this, our partial exploration can have either a conclusive outcome if the undesired state was found or provide a level of confidence indicating if such state is likely to exist in the state space.

Second, in this paper we will analyze how appropriate is treating the results of the partial exploration of the state space with a hypothesis test for proportions in PN systems with the structure of bounded marked graphs. For this, two small examples will demonstrate that for this subclass of PN the test can give correct results.

This paper is self-contained and arranged in the following order: section two contains a small theoretical background about hypothesis test and PN systems. Section three presents the fundamentals of our method regarding the probability distribution of the state space of a PN system (a distribution which belongs to the state space of the PN without tokens), the reachability analysis with partial exploration and the form of our hypothesis test. Section four describes the algorithm for partial exploration of the state space and three sorting routines for making it heuristically guided. Section five describes the estimated normal probability distribution function of the state space. Section six contains the two theoretical examples and at the end are some conclusions.

II. HYPOTHESIS TEST AND PN SYSTEMS

A statistical hypothesis test uses observed data from an experiment for making decisions about the acceptance or not of specific characteristics in the entire population. In our research we will use them to provide a level of certainty and closure to the unsuccessful search of the partial exploration in reachability problems of PN systems for determining the existence or not of undesired states.

Assuming the size of the state space of the PN system is large and unknown, we have to calculate the number n of states we need to explore (the number of samples). The number of states we need to explore has to be enough in order not to compromise the veracity of the hypothesis test.

Then, we transform the reachability problem of the undesired state in terms of deciding with a hypothesis test for proportions if there is enough evidence indicating that the state being searched exists or not in the state space.

Let us start our treatment with a brief introduction of the PN theory. For reference and proofs on boundedness in PN systems and the theorems in the section three we address the reader to [3, 5-6].
A. Petri Nets

A Petri net is a tuple \( N = (P, T, I, O, Q) \), where \( P \) is a finite non-empty set of \( i \) places, \( T \) is a finite non-empty set of \( j \) transitions, \( I \) is the set of directed arcs connecting places to transitions, \( O \) the set of directed arcs connecting transitions to places and \( Q \) is a capacity function for the places mapping \( P \rightarrow Z^+ \). Places are graphically represented by circles, transitions by rectangles and all directed arcs by arrows. The pre-conditions of a transition \( t \) are in the set of input places \( \mathbf{\tau} \) and the post-conditions in \( \mathbf{\rho} \). The pre-events of a place \( p \) are in the set \( \mathbf{p} \) and the post-events in the set \( \mathbf{p}^* \).

A PN is called pure when there are no self-loops. A state machine is a PN such that \( |\mathbf{\tau}| = |\mathbf{\rho}| = 1 \). A marked graph is a PN such that \( |\mathbf{p}| = |\mathbf{p}^*| = 1 \).

The way to represent a state of the system is by putting tokens in the corresponding places. Tokens are black dots that exist only in the places. The function \( m \) called marking maps \( P \rightarrow Z^+ \), and \( m_0 \) is the initial marking. A PN with initial marking is called a PN system and will be denoted by \( (N, m_0) \).

We say that a place \( p \) is marked when \( m(p) > 0 \). The finite set of all possible markings (i.e. the reachability space) of \( (N, m_0) \) is denoted by \( R \). The sum of all tokens in a marking \( m \) is defined with the function \( \text{card}(m) \).

The number of states that a PN system can generate depends on the input and output arcs, the initial marking and the way how the occurrence of transitions is specified. Occurrence of single transition is carried out for the generation of the complete state space of a PN system and mainly used for analysis purposes. Occurrence of concurrent transitions is implemented for generating possibly not-complete state spaces and it is assumed in the rest of this paper except if specified differently.

The sets of arcs \( I \) and \( O \) can be represented with the pre-incident matrix and the post-incident matrix \( \mathbf{\tau}^T \) and \( \mathbf{\rho}^T \) respectively having both \( i \) rows and \( j \) columns, with values of \( |I[p_i, t_j]| \) and \( |O[p_i, t_j]| \) respectively, and a marking \( m \) as a vector \( \mathbf{m} \in Z^j \) defining the state of the system. The token game refers to the way how the dynamic behavior of the system is described with the markings evolution (the removal of existing tokens and the creation of new tokens), according with the firing of enable transitions. The enabling rule is defined as: a transition \( \tau^* \in T \) is said to be enabled at a given marking \( m \) if every of its input places has at least as many tokens as the weight of the arcs joining it and every of its output places has a number of tokens smaller than the sum of their current marking plus the weight of the arc connecting them. The set of all enabled transitions at a marking \( m \) is denoted with \( T(m)^* \).

A transition is called fireable if it is enabled. A fireable transition may fire, eliminating the marking \( m \) and creating the new marking \( m' \).

Using the initial marking vector \( \mathbf{m}_0 \), the next marking vector \( \mathbf{m} \) is mathematically calculated with the state transition function \( \mathbf{m} = \mathbf{m}_0 + (\mathbf{\tau}^T \mathbf{\rho} - 1) \mathbf{\sigma} \). In this formula, \( \mathbf{\sigma} \) is a firing count vector representing the number of times every transition has fired.

The maximal number \( \gamma \) of tokens that could exist in any PN system \( (N, m_0) \), with \( q_0 \), as the finite tokens capacity of the \( \alpha^{th} \) place, is given by

\[
\sum_{\alpha=1}^{j} (q_\alpha + 1).
\]

A PN system \( (N, m_0) \) is bounded if all places have finite tokens capacity; it is bounded at a number \( \beta \leq \gamma \) of tokens in every marking of its state space.

III. Reachability in PN Systems

State spaces are derived from a model of system behavior, like PN systems. Verification of specific properties of the complete state space is in general necessary to guarantee the correct functionality of the system and verify the compliance of design specifications. Unfortunately the usage of verification techniques of the state space is limited due to its possibly exponential growth of size, especially when concurrent occurrences of events exist in the system.

A PN \( N' \) (without tokens) with all places having finite tokens capacity has a number of markings given by

\[
\prod_{\alpha=1}^{j} (q_\alpha + 1).
\]

Let us denote \( U \) as the set of all markings that could exist in a PN \( N' \) without tokens, with size of \( \Omega \). For every \( m \in U \) let us define the variable \( X(\text{number of tokens in } m) \) as follows:

\[
X(m) = \text{card}(m).
\]

Independently of the number of places and tokens capacity in \( N' \), for \( \forall m \in U \) we create a histogram of \( X(m) \), we get a distribution of frequencies of \( X \) as seen in the fig. 1, with the appearance of a normal probability distribution function, i.e. \( X \sim \mathcal{N}(\mu, \sigma^2) \) where \( \mu = \gamma/2 \).

![Fig. 1. Histogram of X(m).](image)

A. Complete Exploration

In a PN system \( (N', m_0) \) and its set of reachable markings \( R \), for every \( m \in R \) the following two theorems belong by now to folk knowledge.

**Theorem 1.** A PN system \( (N', m_0) \) with \( m_0 \in U \) generates a state space \( R \) with markings belonging to \( U \).

**Theorem 2.** A PN system \( (N', m_0) \) generates zero markings.
when the initial marking has no tokens, i.e. \( \text{card}(m_0) = 0 \), or when \( \text{card}(m_0) = \gamma \).

Let us define the variable \( Y \) as follows for a PN system \((N', m_0)\) and every \( m \in R \)

\[
Y(m) = \text{card}(m).
\] (4)

We can create a histogram of \( Y(m) \) and identify the distribution of frequencies of \( Y \) if we explore the complete set of reachable markings \( R \).

Let us assume the PN system \((N', m_0)\) is the marked graph from the fig.2. After complete exploration of its set \( R \), the distribution of frequencies of \( Y \) is seen in the fig.3, with the appearance of a normal probability distribution function, i.e. \( Y \sim \mathcal{N}(\bar{x}, s^2) \).

![Fig. 2. Structurally bounded marked graph PN system.](image)

![Fig. 3. Histogram of \( Y(m) \).](image)

The probability distribution function turns out to be normal for the PN system in the fig.2 and the parameter easily calculated because the number \( \omega \) of states in \( R \) is small, but in very large PN systems where the complete exploration of the reachability space is not feasible, this information is not available to the analyst.

B. Partial Exploration

For the reachability problem of finding a target marking \( m_t \in R \) with cardinality of \( c \) without exploring the complete set \( R \), we can conduct a partial exploration with a heuristic algorithm targeting to visit markings with cardinality of \( c \) and to stop the search upon finding the target marking or visiting \( \omega \times p_t(m_c) \) markings with cardinality \( c \), i.e. the number of states in \( R \) multiplied by the probability of the marking \( m_c \) with cardinality \( c \) [7-8]. The disadvantage of this approach is that knowledge of the PDF is ignored. Since we ignore the number of states we should stop the exploration, complete exploration of the reachability space \( R \) is possible to happen.

For the case when we ignore the information about the set \( R \), we can use the information from the set \( U \). We ignore how the markings in \( R \) are distributed in \( U \), but since we know \( R \subseteq U \), the maximal number of markings with cardinality of \( c \) that could exist in \( R \) is \( \Omega \times p_t(m_c) \).

For the reachability problem of finding a target marking \( m_t \in R \) with cardinality of \( c \) without exploring the complete set \( R \), we can conduct a partial exploration with a heuristic algorithm targeting to visit markings with cardinality of \( c \) and to stop the search upon finding the target marking or visiting \( \Omega \times p_t(m_c) \) markings with cardinality \( c \). The disadvantage of this alternative approach is that a full exploration of the complete reachability space \( R \) is also possible to happen.

C. Probabilistic Termination

The two approaches described before seem correct in theory, but with ignored efficiency and possible convex termination, leading to a possible complete exploration. Targeting to visit specified markings is not a guarantee of reaching those markings; therefore a sample size is calculated for termination and completeness.

The heuristic algorithms could also stop the search after visiting a number of markings determined by the formula in (5) to calculate the number of samples \( n \) in a large population for a proportion \([9-10]\)

\[
n = Z^2 p(m_t)(1 - p(m_t)) / e^2.
\] (5)

The value of \( n \) is adequate whenever we know size \( \omega \) of the set of reachable markings \( R \) and the proportion \( p_t(m_c) \). For the case when we ignore those values and use \( p_t(m_c) \), the number \( n \) is just as reference to determine an approximate value.

D. Standard Hypothesis Test

The results of the unsuccessful search of the target marking \( m_t \) with partial exploration can be treated with a hypothesis test to provide certainty and completeness.

We define the pseudo-random variable \( Q \) as follows for each marking \( m \) registered in the exploration

\[
Q(m) = \begin{cases} 
1 & \text{if } \text{card}(m) = c \\
0 & \text{if } \text{card}(m) \neq c
\end{cases}.
\] (6)

From the probability mass function (PMF) \( p_q(m_c) \), we obtain \( q \) as the proportion of markings with the same cardinality as \( m \) and state the following test:

\[
H_0 : q = p_q \\
H_1 : q > p_q
\]
The test statistic computed for the hypothesis test of proportions is

$$\pi = \frac{q - \hat{p}}{\sqrt{\hat{p}(1 - \hat{p}) / n}}. \quad (7)$$

We will use a 95% level of significance in our hypothesis test. The standard normal z-table gives a value of 1.645.

The decision obtained with this hypothesis test will be: if we accept $H_0$ as true it means there might not be other markings with the same cardinality as the target marking, then the target marking is not in the set of reachable markings $R$ of the PN system. On the other hand, if we cannot reject $H_1$ then it is probable that the target marking exists.

The appropriateness of the hypothesis test relies on the sampling method, the states registered (samples) and the proportion of markings $p_t$ in the set of reachable markings $T$ with the same cardinality as the target marking. For this, in the next section we describe the former heuristic sampling method and describe two alternative methods. Later in the section five we will present how to utilize the samples for constructing the probability distribution function of $R$, how to estimate $p_T$ and will describe the statistical hypothesis test.

IV. HEURISTIC ALGORITHM OF PARTIAL EXPLORATION

The breadth-first based heuristic algorithm originally presented for the partial exploration of the state space for reaching a target marking $m_t$ is described as follows:

1: $REG = m_0 \quad \leftarrow \text{put } m_0 \text{ in register of markings}$
2: $VIS = REG \quad \leftarrow \text{set of visited markings}$
3: repeat
4: $m_c = \text{Top}(VIS) \quad \leftarrow \text{get current marking}$
5: $T(m_c)^* \leftarrow \text{set of sorted enabled transitions at } m_c$
6: if $T(m_c)^* = \emptyset$ then
7: \quad $VIS = VIS - m_c$ and goto 12
8: $\sigma = \text{Top}(T(m_c)^*)$
9: $m_n = m_c + (O - 1) \times \sigma \quad \leftarrow \text{get next marking}$
10: if $m_n \neq REG$ then
11: \quad $REG = REG + m_n$ and $VIS = VIS + m_n$
12: until $VIS = \emptyset$
13: end.

For the previous algorithm to be heuristically guided towards the target marking, the set $T(m_t)^*$ is sorted with the routine discussed next.

A. Rectilinear Distance Sorting

The sorting routine to obtain the set $T(m_t)^*$ is described as, for every firing vector $\sigma \in T(m_t)^*$ enabling a transition $t$ and creating one next marking $m_n$, the rectilinear distance $RD(\sigma)$ is the sum of absolute values of the difference $m_n(p) - m(p)) \forall p \in P$. A firing vector with zero rectilinear distance will be put at the top of the stack.

1: $T1 = T(m_t)^*$
2: repeat
3: $\sigma = \text{read } \sigma$ from $T1$
4: $RD(\sigma) = \text{Sum}(\text{Abs}(m_n(p) - m(p))) \forall p \in P$
5: put $RD(\sigma)$ in $ET$
6: delete $\sigma$ from $T1$
7: until $T1 = \emptyset$
8: $T(m_t)^* = \text{sorted } T(m_t)^* \text{ by } ET1$ in ascending order

Although with a rectilinear distance sorting routine we still obtain a convex exploration, successfully cases are registered with analogous methods in [11-13].

On the other hand, in this paper we present two additional methods taking in consideration past and future marking results. Upon accomplishing to fill the tokens quota in places of the target marking, those partial findings are discarded with the rectilinear distance sorting routine because the metric is a total scalar measurement. Our proposals consider a more specific metric based on individual scalar measurements of tokens in each place for the reached markings.

B. Forward Sorting

One proposed routine to obtain the set $T(m_t)^*$ is described as, for every firing vector $\sigma \in T(m_t)^*$ enabling a transition $t$ and creating one next marking $m_n$, if $\exists p \in I^t$ such that $m_n(p) = m(p)$, then the corresponding firing vector will be put at the top of the stack. The firing vector $\sigma$ enabling transitions all together having the largest number of pos-places fulfilling the condition will be at the top of the stack.

1: $T2 = T(m_t)^*$
2: repeat
3: $\sigma = \text{read } \sigma$ from $T2$
4: $FW(\sigma) = \{ p \in I^t \mid t \text{ is enabled at } \sigma \text{ and } m_n(p) = m(p) \}$
5: $s(\sigma) = \text{card}(FW(\sigma))$
6: put $s(\sigma)$ in $ET2$
7: delete $\sigma$ from $T2$
8: until $T2 = \emptyset$
9: $T(m_t)^* = \text{sorted } T(m_t)^* \text{ by } ET2$ in descending order

In order to fulfill the tokens quota in the places, the previous algorithm was specifically designed to foreseeing the marking in the places.

C. Backward Sorting

In our second proposal, for every firing vector $\sigma \in T(m_t)^*$ enabling a transition $t$ and creating one next marking $m_n$, if $\exists p \in I^t$ such that $m_n(p) = m(p)$, then such firing vector will not be put at the top of the stack. The firing vector $\sigma$ enabling transitions all together having the largest number of pre-places fulfilling the previous condition will be at the bottom of the stack.

1: $T3 = T(m_t)^*$
2: repeat
3: $\sigma = \text{read } \sigma$ from $T3$
4: $BW(\sigma) = \{ p \in I^t \mid t \text{ is enabled at } \sigma \text{ and } m_n(p) = m(p) \}$
5: $s(\sigma) = \text{card}(BW(\sigma))$
6: put $s(\sigma)$ in $ET3$
7: delete $\sigma$ from $T3$
8. until $T3 = \emptyset$
9. $T(m_*) = \text{sorted} \ T(m_*) \ by \ ET3 \ in \ ascending \ order$

Upon fulfilling the tokens quota in some places, the previous algorithm was designed to avoid those places to lose their token quota immediately. And a more sophisticated algorithm could include a second sorting using the index ET2 from the previous algorithm.

V. ASSUMED PDF OF THE SET OF REACHABLE MARKINGS

Progressions of the markings in a PN system are dependent of the previous marking and the inherent non-determinism in the network. Statistical methods use random samples of independent observations all having the same probability of being selected. Since we ignore the set of reachable markings $R$, it is impossible to conduct a proper random sampling and any exploration-based samplings seems inadequate in principle.

Probability samples can only be used to create mathematically sound statistical inferences about a larger target population. Non-probability sampling has no way of measuring their bias or sampling error.

Works in [14-20] discuss the estimation of state space parameters using different sampling techniques, but none of them alert about this inadequacy. Although any state space exploration algorithm, saying depth-first, breath-first, random walks, etc., might seem inadequate because they are dependent on the system’s progression, in fact the algorithms provide a systematic sampling, a type of probability sampling, however no details on the sampling procedure are given.

Our exploration algorithms and the three sorting routines give non-probability samplings and their samples will be biased towards the target marking. The empirical probability mass function (PMF) constructed with the results of the sampling will be fitted to a normal PDF since it is believed the state space has such distribution.

The results in [8] fitted the empirical PMF to a normal distribution with mean $\hat{x}$ equal to the cardinality of the target marking and standard deviation $\hat{s}$ equal to the absolute value of resting the cardinality of the initial marking to the cardinality of the target marking. The empirical PMF fitted to a normal PDF can be seen in the fig.4. Although good results were obtained for some PN systems, the method lacks of generalization.

Calculated from the sampled markings. The mean is near the cardinality of the target marking and the maximal sampling error for the mean equal to $|\bar{x} - \hat{x}|$.

Our proposed fitted normal PDF for the set $R$ is defined as

$$P_{f} [\hat{x} - 3\hat{s} \leq Y \leq \hat{x} + 3\hat{s}] = \int_{\hat{x} - 3\hat{s}}^{\hat{x} + 3\hat{s}} f(y). \quad (8)$$

The function $f(x)$ corresponds to the normal distribution and the $3\hat{s}$ assumption is adopted from the “68-95-99.7” statistical rule of the normal distribution.

A. Hypothesis Test with the Fitted Normal PDF

The form of the hypothesis test presented in the section 3-D does not change. The only difference is that the proportion of markings $p_{Y}$ used in the test corresponds to the probability $P_{f}[Y=\text{card}(m_{i})]$, i.e. the probability of the cardinality of the target marking with the fitted normal PDF of the set $R$.

VI. THEORETICAL EXAMPLE

Let us utilize the PN systems in the figure 5 and 6. In this paper the PN systems under study are restricted to be a bounded marked graph. The nets are believed to have normality in their set of reachable markings $R$, similar to the one of the PN system in the fig.2.

![Fig. 5. Example 1 - Marked graph PN system.](image)

The PN systems are rather small but the usability of the hypothesis test in larger nets is straightforward.

We conducted partial exploration of different markings using the three sorting routines: rectilinear distance (RD), forward distance (FD) and backward distance (BD). Sample size was fixed at 20 based on all sample size results. All explorations were unsuccessful, but only one sample was used in the hypothesis test.

To determine the appropriateness of the test for these examples, by first comparing the mean and variance of the set of reachable markings $R$ with the ones from the sample (Table I) we observe a significant difference exists only in one parameter.

Our second comparison is on the value of the proportion of markings with same cardinality as the target marking of the set...
Although it is not guaranteed yet that the target marking would be found with our heuristically guided partial exploration algorithm without complete exploration of the set of reachable markings \( R \), the idea of using a complementary hypothesis test to analyze reachability problems have certain level of effectiveness in this specific subclass of PN systems. Criterions about determining normality based on the PN model are in the next steps.

REFERENCES


Fig. 6. Example 2 - Marked graph PN system.

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Mean</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>3.05357</td>
<td>0.524351</td>
</tr>
<tr>
<td>Estimation</td>
<td>3.05000</td>
<td>0.786842</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 2</th>
<th>Mean</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>4.00000</td>
<td>1.609760</td>
</tr>
<tr>
<td>Estimation</td>
<td>3.50000</td>
<td>0.789474</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>card(m)</th>
<th>( p_Y(m) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.21135</td>
</tr>
<tr>
<td>3</td>
<td>0.44093</td>
</tr>
<tr>
<td>2</td>
<td>0.10799</td>
</tr>
<tr>
<td>3</td>
<td>0.38352</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>card(m)</th>
<th>Test Statistic</th>
<th>Decision</th>
<th>Eval</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.12433</td>
<td>Accept H0</td>
<td>Correct</td>
</tr>
<tr>
<td>3</td>
<td>0.60532</td>
<td>Accept H0</td>
<td>Correct</td>
</tr>
<tr>
<td>2</td>
<td>0.76578</td>
<td>Reject H0</td>
<td>Incorrect</td>
</tr>
<tr>
<td>3</td>
<td>0.10799</td>
<td>Reject H0</td>
<td>Correct</td>
</tr>
</tbody>
</table>

VII. CONCLUSION

Using a hypothesis test on the unsucceeded results of the heuristically guided partial exploration seems appropriate to determine if a target marking exists or not in the probably normal state space of a structurally bounded marked graph PN system. Comparison of the estimated values presents good result and the evaluation of the decision shows only one incorrect decision. We conclude this kind of test might be appropriate to use on larger PN systems just upon further research.