Faster SPDL Model Checking Through Property-Driven State Space Generation

Matthias Kuntz, Boudewijn R. Haverkort

University of Twente,
Faculty for Electrical Engineering, Mathematics and Computer Science

Abstract. In this paper we describe how both, memory and time requirements for stochastic model checking of SPDL (stochastic propositional dynamic logic) formulae can significantly be reduced. SPDL is the stochastic extension of the multi-modal program logic PDL. SPDL provides means to specify path-based properties with or without timing restrictions. Paths can be characterised by so-called programs, essentially regular expressions, where the executability can be made dependent on the validity of test formulae. For model-checking SPDL path formulae it is necessary to build a product transition system (PTS) between the system model and the program automaton belonging to the path formula that is to be verified. In many cases, this PTS can be drastically reduced during the model checking procedure, as the program restricts the number of potentially satisfying paths. Therefore, we propose an approach that directly generates the reduced PTS from a given SPA specification and an SPDL path formula. The feasibility of this approach is shown through a selection of case studies, which show enormous state space reductions, at no increase in generation time.

1 Introduction

It is extremely important to develop techniques that allow the construction and analysis of distributed computer and communication systems. These systems must work correctly and meet high performance and dependability requirements. Using stochastic model checking it is possible to perform a combined analysis of both qualitative (correctness) and quantitative (performance and dependability) aspects of a system model. Models that incorporate both qualitative and quantitative aspects of system behaviour can be modelled by various high-level formalisms, such as stochastic process algebras [12, 11], stochastic Petri nets [1], stochastic activity networks [17] (SANs), etc.

In order to do model checking of stochastic systems, over the last years a number of stochastic extensions of the logic CTL [8] have been devised. The most notable extension is the logic CSL [4] (continuous stochastic logic). More recently, in [14, 3], action-based extensions of CSL were introduced. These logics allow for the specification of desired system behaviour by means of action sequences. This makes them very well suited for modelling formalisms in which the actual system behaviour is specified as a sequence of actions or transitions, as is the case for SPAs, SPNs and SANs.

The applicability of stochastic model checking is limited by the complexity, i.e., the size of system models that are to be verified. At the heart of stochastic model checking lies the solution process of huge sparse sets of linear (differential) equations. This limits the size of systems that are practically analysable to some 10^8 states.

To overcome these limitations we can think of several approaches. One standard approach is the use of some notion of Markovian bisimulation. This approach has the following drawbacks. Computing the bisimulation quotient of a system is computationally expensive, and before reduction takes place the entire system has to be generated. Furthermore, depending on the system, the reduction in size may not be very large, and finally, due to reasons that are related to numerical analysis, the verification of the reduced system may be slower than that of the original system (cf. [13]).

We propose a different approach, which reduces the system size in many cases already during the state space generation, by exploiting the SPDL path formula that is to be verified.

Related Work For stochastic model checking we are not aware of any approach that generates the state space in a way which depends on the formula that is to be verified. For CSL model checking, in [4] an approach is described that makes states absorbing that do not functionally satisfy a given until-formula, but this state space reduction is performed only after the state space was generated. Following this proposal, in [14] model checking algorithms for SPDL path formulae were implemented. For CSL model checking this was done in [13]. For CTL model checking in [2] an approach is reported, where for interacting finite state machines equivalence relations are computed, depending on the CTL formula that is to be verified.

The paper is further organised as follows. In Section 2 we briefly introduce the syntax and semantics of SPDL; we will explain in an informal style the traditional approach to the model checking of SPDL path formulae. In Section 3 we then describe the stochastic process algebra $\mathcal{YAMPA}$, on which our property-driven state space generation approach relies. Section 4 is devoted to a denotational, symbolic property-driven semantics of $\mathcal{YAMPA}$. In Section 5 we will show the feasibility of our approach via some experimental results. Finally, Section 6 concludes the paper with a short summary and some pointers to future work.

2 SPDL - Syntax, Semantics and Model Checking

The logic SPDL is the stochastic extension of the logic PDL [9], a multi-modal program logic. PDL enriches the standard modal operator $\diamond$ ("possibly") with so-called programs, which are essentially regular expressions and tests (cf. Def. 1). In PDL, the formula $< \rho > \Phi$ means, that it is possible to execute program ρ and end in a state that satisfies Φ. SPDL adds the following extensions to PDL: The operator $< \rho >$ is replaced by the time-bounded path operator $[\rho]^I$, a probability operator $\mathcal{P}_{\bowtie p}$ to reason about the transient system behaviour, and a steady state operator $\mathcal{S}_{\bowtie p}$ to reason about system behaviour, once stationarity

has been reached. In what follows, we discuss the syntax, semantics, and a model checking procedure for SPDL.

2.1 Syntax of SPDL

Definition 1 (Syntax of SPDL). *Let p be a probability value in $[0,1]$, $q \in \mathsf{AP}$ an atomic proposition, where AP is the set of atomic propositions, and $\bowtie \in \{\leq, <, \geq, >\}$ a comparison operator. The state formulae Φ of SPDL are defined as:*

$$\Phi := q \big| \Phi \vee \Phi \big| \neg\Phi \big| \mathcal{P}_{\bowtie p}(\phi) \big| \mathcal{S}_{\bowtie p}(\Phi) \big| (\Phi)$$

Path formulae are defined as:

$$\phi := \Phi[\rho]^I \Phi,$$

where I is the closed interval $[t, t']$, Φ is assumed not to possess sub-formulae containing the steady state operator $\mathcal{S}_{\bowtie p}$.[1] *Programs ρ are described by the grammar given in Def. 2.*

Definition 2 (Programs). *Let Act be a set of actions, which are also called atomic programs, and TEST be a set of SPDL state formulae, again not containing the steady state operator $\mathcal{S}_{\bowtie p}$. A program ρ is defined by the following grammar:*

$$\rho := \epsilon \big| \Phi?; a \big| \rho; \rho \big| \rho \cup \rho \big| \rho^* \big| \Phi?; \rho \big| (\rho)$$

where $\epsilon \notin \mathsf{Act}$ is the empty program, $a \in \mathsf{Act}$ and $\Phi \in \mathsf{TEST}$.

Sequence ($;$), choice ($\cup$), and Kleene-star ($*$) have their usual meaning as known from the theory of regular expressions. The operator $\Phi?$ is the so-called test operator. Informally speaking, it tests whether Φ holds in the current state of the model. If this is the case, then execute program ρ, otherwise ρ is not executable. Following language theory, we can derive words from a program ρ (here also called program instances) according to the rules of regular expressions. The set of all these program instances is called a language.

Example 1. Throughout this paper, we use the example of a fault-tolerant packet collector, which has the following repeating behaviour. Arrivals can either be error-free (upper transition *arr*, rate λ) or erroneous (lower transition *error*, rate μ). If a data packet contains an error, this error can be correctable (*co*) non-correctable (*nco*). In case of a correctable error, the error is corrected (transition *co*) and more data packets can be received. If the error is non-correctable, the data packet has to be retransmitted (transition *rt*). In Fig. 1, the SLTS $\mathcal{M}$ for the packet collector is shown, where we assume that the number n of data packets that are to be processed is equal to four. The system has the following state labels:

[1] Mixing formulae that express transient behaviour ($\mathcal{P}_{\bowtie p}$) with formulae expressing steady state behaviour ($\mathcal{S}_{\bowtie p}$) is considered less meaningful.

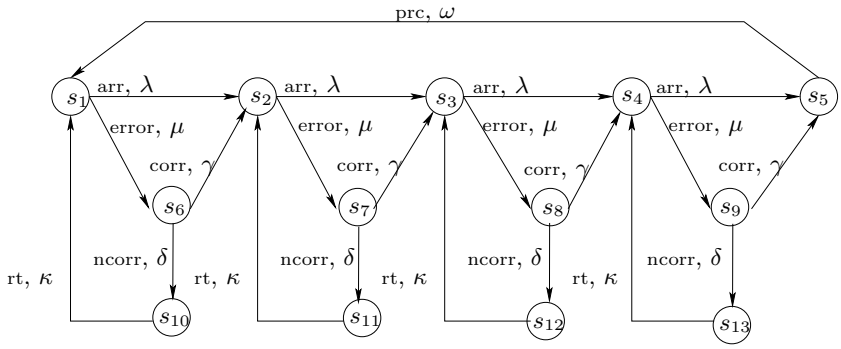

Fig. 1. Fault tolerant packet collector for $n = 4$ packets

$L(s_5) = \{\mathsf{full}\}, \quad L(s_6) = ... = L(s_9) = \{\mathsf{error}\},$
$L(s_{10}) = ... = L(s_{13}) = \{\mathsf{waitrt}\}, \quad L(s_{14}) = ... = L(s_{17}) = \{\mathsf{waitcor}\}$

The set of actions is given as follows:

$$\mathsf{Act} := \{arr, error, rt, corr, ncorr, prc\}$$

Using SPDL, we can easily express the following properties:

- $\Phi_1 := \mathcal{P}_{\bowtie p}((\neg\mathsf{full})[arr^*]^{[0,t]}(\mathsf{full}))$: Is the probability to receive N data packets without error within t time units greater or less than p?
- $\Phi_2 := \mathcal{P}_{\bowtie p}(\neg\mathsf{full}[arr; \mathsf{TEST1}?; error; rt; arr^* \cup arr^*]^{[0,t]}\mathsf{full})$: Is the probability to receive N data packets without error or with at most one non-correctable error within t time units greater or less than p, given that this non-correctable error appears in the first data packet? The test formula $\mathsf{TEST1}$ defines those states, in which it holds that 1 packet has arrived.
- $\Phi_3 := \mathcal{P}_{\bowtie p}(\mathsf{true}[arr^*; \mathsf{TEST2}?; arr; corr]^{[0,t]}\mathsf{full})$: Is the probability that the buffer is full after at most t time units and that the Nth packet contains a correctable error, given that all preceeding packets were error free, within the probability bounds given by $\bowtie p$? The test formula $\mathsf{TEST2}$ describes those states, in which it holds that $N - 1$ packets have arrived.

2.2 Semantics of SPDL

We will now show, both the model over which SPDL formulae are interpreted and the semantics of SPDL formulae.[2]

The semantic model of SPDL is a so-called stochastic labelled transition system, defined as follows.

Definition 3 (Stochastic labelled transition system (SLTS)). *An SLTS $\mathcal{M}$ is a six-tuple $(s, S, \mathsf{Act}, L, R, \mathsf{AP})$, where*

[2] The stochastic process algebra from Sections 3 and 4 and SPDL share the same semantic model.

– *s is the unique initial state,*
– *S is a finite set of states,*
– *Act is a finite set of action names,*
– *L is the state labelling function: $S \to 2^{AP}$,*
– *R is the state transition relation : $R \subseteq S \times (\mathsf{Act} \times \mathbb{R}_{>0}) \times S$,*
– *AP is the set of atomic propositions.*

Definition 4 (Semantics of SPDL).

– *The semantics of propositional logic formulae $\neg\Phi$ and $\Phi \vee \Psi$ is defined the usual way.*
– *$\mathcal{S}_{\bowtie p}(\Phi)$ asserts that the steady state probability of the Φ-states, i.e., the probability to reside in a Φ-state once the system has reached stationarity satisfies the probability bounds as given by $\bowtie p$.*
– *$\mathcal{P}_{\bowtie p}(\phi)$ asserts that the probability measure of all paths that satisfy ϕ lies within the bounds as imposed by $\bowtie p$.*
– *$\Phi[\rho]^I\Psi$ asserts that a path that satisfies this formula reaches a Ψ-state within at least t time units, but after at most t' time units. All preceeding states must satisfy Φ. Alternatively, a $\Phi \wedge \Psi$-state can be reached before the passage of t time units, but not left before at least t time units have passed. Additionally, the action sequence on the path to the Ψ-state must correspond to the action sequence of a word from the language induced by program ρ. All test formulae that are part of ρ must be satisfied by corresponding states of the path.*

2.3 Model Checking SPDL

The overall model checking algorithm of SPDL is similar to that of CTL, in the sense that it starts with the verification of atomic properties and then proceeds with the checking of ever more complex sub-formulae until the overall formula has been checked.

Model Checking SPDL

– Propositional formulae $\neg\Phi$ and $\Phi \vee \Psi$ are checked as in the CTL case.
– Steady state formulae $\mathcal{S}_{\bowtie p}(\Phi)$ can be checked as for CSL [4].
– Model checking formulae with a leading $\mathcal{P}_{\bowtie p}$ operator is more involved. We assume, we want to check whether in an SLTS $\mathcal{M}$ a state s satisfies $\mathcal{P}_{\bowtie p}(\phi)$, with $\phi = \Phi[\rho]^I\Psi$. The basic idea is to reduce the model checking problem of SPDL to one of CSL, which consists of deciding whether a continuous time Markov chain (CTMC) $\mathcal{M}^\times$ (to be constructed) and a state $s^\times$ in $\mathcal{M}^\times$ satisfies the CSL formula $\mathcal{P}_{\bowtie p}(\mathsf{F}^I\mathsf{succ})$. A path satisfies $\mathsf{F}^I\mathsf{succ}$, if within time interval I a state is reached that satisfies the atomic property succ. To reach this goal, we proceed as follows:
  1. From the program ρ we derive a deterministic program automaton A_ρ, which is a variant of deterministic finite automata.[3]

[3] For the derivation of A_ρ from program ρ we refer to [14] for a thorough discussion of this issue. As such this issue does not play a crucial role in understanding this paper.

2. Using the given SLTS $\mathcal{M}$ and the program automaton A_ρ we build a product Markov chain. $\mathcal{M}^\times$. The state space of $\mathcal{M}^\times$ is the product of $\mathcal{M}$ and A_ρ, i.e., its states are of the form (s_i, z_i), where s_i is a state of $\mathcal{M}$ and z_i a state of A_ρ. Additionally, $\mathcal{M}^\times$ possesses one new, absorbing state: the state $FAIL$.
   In $\mathcal{M}^\times$ a transition $(s_i, z_i) \xrightarrow{\lambda} (s_j, z_j)$ is kept, where λ is the rate of the transition from s_i to s_j, iff the following two constraints are satisfied:
   - (s_i, z_i) must satisfy Φ, this is the case iff s_i satisfies Φ.
   - Both s_i and z_i must be capable to perform the same action, and if the current action is associated with a test, then s_i must also satisfy this test.

   If one of these two constraints is violated, we have to introduce a transition $(s_i, z_i) \xrightarrow{\lambda} FAIL$ and delete transition $(s_i, z_i) \xrightarrow{\lambda} (s_j, z_j)$.
3. Finally, to compute the probability measure of the paths that satisfy ϕ we proceed as follows. All states (s_j, z_j) of $\mathcal{M}^\times$ for which s_j is a Ψ-state and z_j is an accepting state of A_ρ are replaced by the newly introduced absorbing success state $SUCC$, labelled with the special, newly introduced atomic state formula succ, thereby redirecting all incoming transitions from the old states to the new $SUCC$ state.
4. At this point, it is possible to check, whether $\mathcal{P}_{\bowtie p}(\Phi[\rho]^{[t,t']}\Psi)$ is functionally satisfiable: If in $\mathcal{M}^\times$ a path to a succ state exists, then $\mathcal{P}_{\bowtie p}(\Phi[\rho]^{[t,t']}\Psi)$ can be satisfied at least on the functional level.
5. On $\mathcal{M}^\times$ (which was transformed as described in step 3) we can compute the probability measure of all paths satisfying the CSL formula $\mathcal{P}_{\bowtie p}(\mathsf{F}^{[t,t']}\mathsf{succ})$, which is equal to the probability measure of the paths satisfying the original formula $\mathcal{P}_{\bowtie p}(\Phi[\rho]^{[t,t']}\Psi)$ in the original model $\mathcal{M}$.

3 Stochastic Process Algebras

In the past 15 years, a number of stochastic process algebras have been devised, such as PEPA [12] and TIPP [11]. Here, we use the stochastic process algebra $\mathcal{YAMPA}$ (yet another Markovian process algebra), that is used in the tool CASPA [16], which we use for our empirical studies. Instead of giving a formal account of $\mathcal{YAMPA}$, we will introduce its most important operators by means of a small example.

Example 2. In fig. 2 we list the $\mathcal{YAMPA}$ specification of the fault tolerant packet collector of example 1. In line (1) we can specify the maximum number of packets that must arrive, before processing starts. We see in this specification some "syntactic sugar" that eases the concise specification of complex systems, e.g., guarded choice in line (3). In lines (2) and (3) we find that process `Arr` is parameterised with parameter `i`, that can take the maximum value `max`. This parameter records the number of packets that arrived. In line (2) we see that `Arr` is initialised with `i` $= 0$, i.e., zero packets arrived in the beginning.

The overall system consists of the processes `Arr` and `Errorhandler` that are composed in parallel and that have to synchronise over the actions `error,`

```
(1)    int max = 15000;
(2)    System := Arr(0)|[error, corr,ncorr]| Errorhandler
(3)    Arr(i [max]) := [i=0] -> (arr, lambda);Arr(i+1) +
(4)                                    (error, mu);((corr, 1);Arr(i+1) + (ncorr,1);(rt,kappa);Arr(0))
(5)                    [i<max, i > 0] -> (arr, lambda);Arr(i+1) + (error, mu);((corr, 1);Arr(i+1) +
(6)                                    (ncorr,1);(rt,kappa);Arr(i-1))
(7)                    [i=max] -> (prc, omega);Arr(0)
(8)    Errorhandler := (error, 1);((corr, gamma);Errorhandler + (ncorr, delta);Errorhandler)
```

Fig. 2. Example $\mathcal{YAMPA}$ specification

`corr, ncorr`, i.e., these actions must be performed by both processes at the same time. For all other actions, the processes can evolve independently. `(arr, lambda);Arr(i+1)` (line (3)) is an example of prefix: After an exponentially distributed delay time, which is governed by rate `lambda`, action `arr` can be taken. In line (4) we find an example of choice: This process can either behave as `(arr, lambda);Arr(i+1)` or `(error, mu);((corr, 1);Arr(i+1) + (ncorr,1);(rt,kappa);Arr(0))`. In line (3) to (7) we see examples of guarded choice: Depending on the actual value of `i` different branches of the specification in lines (3) to (7) can be taken. In line (3), this branch of the specification can only be taken, if the value of parameter `i` is equal to zero. Process `Arr(i [max])` possesses cyclic (recursive) behaviour, as, after `arr` it can again behave as `Arr`.

4 A Property-Driven Symbolic Semantics for $\mathcal{YAMPA}$

In this section we introduce the new property-driven semantics for $\mathcal{YAMPA}$. In Sec. 4.1 we will give the general idea of this semantics. Sec. 4.2 introduces multi-terminal binary decision diagrams (MTBDDs) as data structure to represent SLTSs. In Sec. 4.3 the semantics rules is introduced by means of a small example, and in Sec 4.4 their formal definition is given.

4.1 General Idea

In Section 2.3 we have presented a straight-forward model checking procedure for SPDL path formulae. The size of the product CTMC, before it is reduced is the product of the sizes of the original model $\mathcal{M}$ and the program automaton A_ρ. During the model checking procedure, many states are merged into the states $FAIL$ resp. $SUCC$. This means, we needlessly generate a state space that is much larger than actually required, which is both a waste of memory space and time.

To overcome this weakness in the usual model checking procedure we propose an approach that generates only those states that are actually needed to verify the property at hand. In order to reach this goal, we introduce a property-driven semantics for the stochastic process algebra $\mathcal{YAMPA}$, that uses the path formula that is to be verified to direct the state space generation process. This new semantics cuts off state space generation as soon as it becomes clear a path

is either not satisfying, i.e., it leads to a $FAIL$ state, or satisfying, i.e., leads to a $SUCC$ state. This significantly reduces the number of states and transitions that are generated.

We will use the symbolic semantics of [15] as a basis for our new SPA semantics. Like in [15], the property-driven semantics maps the SPA specification directly to the MTBDD representation of its underlying SLTS. The semantics proceeds in a compositional manner, according to the syntactic structure of the process term at hand. Additionally to [15], the new semantics takes, as already said, during generation of the SLTS the SPDL property that is to be verified into account. We chose MTBDDs as data structures for the SLTS representation as it was shown convincingly [18] that MTBDDs allow a compact representation of even huge state spaces.

4.2 Multi-Terminal Binary Decision Diagrams Encode SLTSs

MTBDDs [10] are an extension of BDDs [6] for the graph-based representation of pseudo-Boolean functions, i.e., functions of type $\mathbb{B}^n \mapsto \mathbb{R}$. Informally spoken, MTBDDs are collapsed binary decision trees, i.e., each non-terminal nodes has exactly two outgoing edges.

They are collapsed in the sense that structural properties of the binary trees are used to reduce the size of the graph.

MTBDDs are very well suited for the representation of the semantic model of SPAs. We will demonstrate that by means of a small example.

Example 3. Consider Fig. 1 from Example 1. To represent this SLTS as an MTBDD we have to find ways to represent its "ingredients" in an appropriate way. That means we have to find representations for: the actions, the states, and the transition relation. All these can be encoded binarily resp. by means of pseudo-Boolean functions:

- Actions: The system has six actions: $arr, error, corr, ncorr, rt$, and prc, therefore, we need three variables a_1, a_2, a_3 to encode them:

$$\begin{aligned} &Enc_{Act}(arr) = \neg a_3 \wedge \neg a_2 \wedge \neg a_1 = 000, \\ &Enc_{Act}(error) = 001 \quad Enc_{Act}(corr) = 010 \quad Enc_{Act}(ncorr) = 011 \\ &Enc_{Act}(rt) = 100 \quad Enc_{Act}(prc) = 101 \end{aligned}$$

- States: The system has 13 states, i.e., we need 4 Boolean variables z_1 to z_4 to encode them:

$$\begin{aligned} &Enc_s(s_1) = \neg z_4 \wedge \neg z_3 \wedge \neg z_2 \wedge \neg z_1 = 0000, \\ &Enc_s(s_1) = 0001 \quad \cdots \quad Enc_s(s_{13}) = 1011 \end{aligned}$$

- Transition relation: A single transition $s \xrightarrow{a,\lambda} s'$ can be encoded as pseudo-Boolean function: $TR(s, a, \lambda, s')$. $TR(s, a, \lambda, s')$ is the conjunction of the binary variables that encode the source state s, target state s' and action a.

For source and target states we need two disjoint sets of Boolean variables, respectively denoted z_i and t_i. The pseudo-Boolean function obtained so, has as function value rate λ. Transition relation R is than the disjunction over all possible $TR(s, a, \lambda, s')$. For example $TR(s_1, arr, \lambda, s_2)$ can be encoded as follows:

$$TR(s_1, arr, \lambda, s_2) =$$
$$\underbrace{\neg z_4 \wedge \neg z_3 \wedge \neg z_2 \wedge \neg z_1}_{s_1} \wedge \underbrace{\neg a_3 \wedge \neg a_2 \wedge \neg a_1}_{arr} \wedge \underbrace{\neg t_4 \wedge \neg t_3 \wedge \neg t_2 \wedge t_1}_{s_2}$$

In terms of MTBDDs, the variables that encode states and actions are the non-terminal nodes and the transition rates are the values of the leaf nodes. In Fig. 3 we show the MTBDD representation of two transitions of the SLTS: $s_1 \xrightarrow{arr,\lambda} s_2$ and $s_1 \xrightarrow{error,\mu} s_6$. Note, that we put the action variables on top of the MTBDD, as this yields smaller MTBDDs.[4]

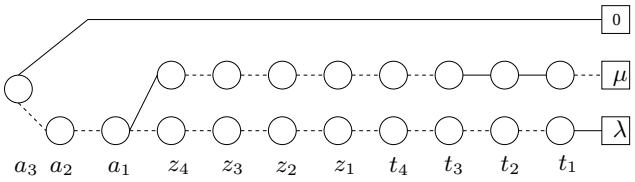

Fig. 3. MTBDD encoding transitions $TR(s_1, arr, \lambda, s_2)$ and $TR(s_1, error, \mu, s_6)$

4.3 Property-Driven Symbolic Semantics - Introduction and Example

Here, we present the general idea behind our semantics and introduce in greater detail the semantic rules for the operators of $\mathcal{YAMPA}$. Due to limited space we will not give the formal description of semantic rules for all operators. Generally, we want to encode the transitions of a given process algebraic description P by an MTBDD. The symbolic representation $[\![P]\!]$ is built from P's parse tree and the transition relation of the deterministic program automaton $\mathcal{A}_\rho$ that is attached to the path formula we want to verify. The parse tree is traversed in a depth-first manner, thereby constructing $[\![P]\!]$ inductively from smaller portions of the overall specification. Finally, we obtain the MTBDD representation of P's transitional behaviour, taking the restrictions imposed by the path formula at hand into account.

Definition 5. *The symbolic representation $[\![P]\!]$ of a process algebra term P consists of the following parts:*

- *The MTBDD $B(P)$, encoding the transition relation,*
- *a list of encodings of process variables X, that appear in P, denoted $Enc_S(X)$,*
- *the encoding of the initial state of P, denoted $Enc_S(s_P^{DS})$,*

[4] In practice further optimisations are possible, but not important here.

– *the transition relation* δ_{A_ρ} *for* A_ρ,
– *the current state of* A_ρ.

Before we list the formal rules for the property-driven semantics, we will give another example.

Example 4. We want to generate the SLTS for the specification from Example 2, with `max = 2`. and SPDL formula $\Phi_1 := \mathcal{P}_{\bowtie p}((\neg\mathsf{full})[arr^*]^{[0,t]}(\mathsf{full}))$ from Example 1. We assume, that the actions and their encodings are globally known, i.e., we know the number of Boolean variables required for their encoding, which is three (like in Example 3). As we derive the MTBDD representation of the SLTS directly from the given specification we do not know in advance the size of the state space and therefore the number of Boolean variables to encode the states and the transition relation. Therefore, we take in the beginning as small a number as possible, and extend the number of variables, if required. The initial state of the specification `Arr(0)|[error, corr,ncorr]| Errorhandler` can be encoded by one Boolean variable $Enc_S(s_1) = \neg z_1 = 0$. Given Φ_1, we check if $\neg\mathsf{full}$ is satisfied, which is the case, then we check whether a transition labelled with arr is possible, which is the case, i.e., we add $Enc_S(s_2) = z_1 = 1$. As for s_2 the condition full is not satisfied, $s_2 \neq SUCC$. The MTBDD encodes at this point the transition relation R consisting of $TR(s_0, arr, \lambda, s_1)$. In s_1 a second transition, labelled by *error* is possible, we see from Φ_1 that *err* does not belong to the actions that yield a satisfying path, i.e., we have to introduce a transition to the failure state $FAIL$, which has no encoding up to now. To do so, we have to extend the number of Boolean variables that encode states, i.e., the states s_1 and s_2 are re-encoded:

$$Enc_S(s_1) = \neg z_2 \wedge z_1 = 00 \;\; Enc_S(s_2) = \neg z_2 \wedge z_1 = 01$$
$$Enc_S(FAIL) = z_2 \wedge \neg z_1 = 10$$

Now, we can introduce a new transition encoding: $TR(s_1, error, \mu, FAIL)$. The overall transition relation R is now the disjunction of $TR(s_0, arr, \lambda, s_1)$ and $TR(s_1, error, \mu, FAIL)$

The state s_2 corresponds to `Arr(1)|[error, corr,ncorr]| Errorhandler`, i.e., $\neg\mathsf{full}$ is satisfied, and again *arr* and *error* transitions are possible, due to the restrictions imposed by the path formula, *error* leads to the $FAIL$ state, i.e., we introduce a new transition: $TR(s_2, error, \mu, FAIL)$. For *arr* we add a new transition from s_2 to s_3, as s_3 satisfies full, $s_3 = SUCC$, and $TR(s_2, arr, \lambda, SUCC)$, where $Enc_S(s_3) = 11$. In Fig. 4 we find the MTBDD encoding the transition relation of this SLTS.

4.4 Property-Driven Symbolic Semantics - Formal Definition

Process Variables A process variable X specifies a reference state within a surrounding rec$X$ operator. Therefore, process variables are encoded in a similar fashion as states. Within each sequential component[5] process variables having

[5] A sequential component is a process term which does not include the parallel composition operator.

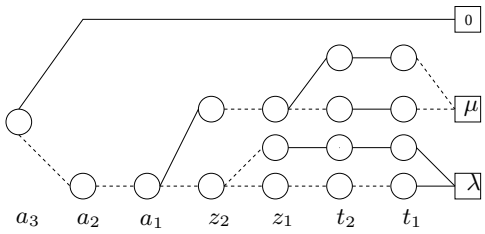

Fig. 4. MTBDD representation of the fault-tolerant packet collector's SLTS for `max` = 2 and Φ_1

(1) **if** not first appearance of X within present seq. component **then**
(2) **skip** /* do nothing */
(3) **if** no free encodings available **then** /* need to extend the set of possible encodings */
(4) Extend the number of Boolean variables
(5) Extend all existing encodings
(6) $B(X) := 0$ /*In case of stop: $B(\mathsf{stop}) := ...$

Fig. 5. Algorithm for process variable X and stop

the same name get the same encoding. Upon first appearance of a process variable X, the MTBDD associated with X is the 0-MTBDD (cf. Fig. 5). The stop process is a special case of a process variable (a process constant).

Prefix $P := (a, \lambda); Q$ For a given formula $\Psi := \mathcal{P}_{\bowtie p}(\Phi_1[\rho]^I\Phi_2)$, we want to generate the symbolic representation of P, $[\![P]\!]$. To construct $B(P)$ we have to distinguish the following cases:

1. If the current state s_P^{DS} satisfies Φ_1 and in A_ρ's current state z an a-labelled transition to a state z' is possible, s_P^{DS} satisfies the test formula Ξ, possibly attached to A_ρ's a-transition, then, we can introduce a transition from s_P^{DS} to the encoding of Q's initial state.
2. If, additionally to case 1, the target state of A_ρ is an accepting state and s_Q^{DS} satisfies Φ_2, then a transition from the encoding of s_P^{DS} to the encoding of state $SUCC$[6] is introduced.
3. If the state s_P^{DS} satisfies Φ_1, but no transition labelling in A_ρ's current state matches a, then we have to introduce a transition from the encoding of P to the encoding of the error state $FAIL$.
4. If state s_P^{DS} does not satisfy Φ_1, then we have to introduce a transition from the encoding of P to the encoding of the error state $FAIL$.
5. If state s_P^{DS} does not satisfy the test formula, attached to A_ρ's a transition, then we have to introduce a transition from the encoding of P to the encoding of the error state $FAIL$.

In Fig. 6 we give the formal description of the prefix algorithm.[7] We only give the first, second, and third case from above, the remaining cases can be treated similarly.

[6] $SUCC$ can be handled like stop.
[7] In this and the following algorithm we omit details on choosing fresh Boolean variables and possibly extending encodings.

Case 1:
(1) **if** $((s_P^{DS} \models \Phi_1) \wedge (z \xrightarrow{a}_{A_\rho} z' \wedge s_P^{DS} \models \Xi))$
(2) $\quad B(P) := TR(s_P^{DS}, a, \lambda, s_Q^{DS})$
Case 2:
(3) **if** $((s_P^{DS} \models \Phi_1) \wedge (s_Q^{DS} \models \Phi_2) \wedge (z \xrightarrow{a}_{A_\rho} z' \wedge s_P^{DS} \models \Xi))$
(4) $\quad B(P) := TR(s_P^{DS}, a, \lambda, SUCC)$
Case 3:
(5) **if** $((s_P^{DS} \models \Phi_1) \wedge (z \not\xrightarrow{a}_{A_\rho} z' \wedge s_P^{DS} \models \Xi))$
(6) $\quad B(P) := TR(s_P^{DS}, a, \lambda, FAIL)$

Fig. 6. Algorithm for prefix $P := (a, \lambda); Q$

Choice $P := Q + R$ Here, we can assume that $[\![Q]\!]$ and $[\![R]\!]$ are already available. To derive $[\![P]\!]$ from $[\![Q]\!]$ and $[\![R]\!]$ we have to proceed as follows: A new initial state is introduced for $Q + R$. All transitions emanating from the initial states of the subprocesses Q and R have to be copied, as they may also take place in the initial state of the overall process.

Recursion $P = \mathrm{rec}X : Q$ When constructing $[\![P]\!] = [\![\mathrm{rec}X : Q]\!]$ from $[\![Q]\!]$ we can distinguish the following cases:

1. X does not appear (unbound) in Q: In this case we simply identify the symbolic representation of $\mathrm{rec}X : Q$ with that of Q.
2. X appears in Q and s_Q^{DS} satisfies Φ_2 and the current state in A_ρ is an accepting state, then the process variable X is identified/replaced by the process constant $SUCC$.
3. X appears in Q and either s_Q^{DS} does not satisfy Φ_2 or the current state in A_ρ is not an accepting state: In this case the process variable X is identified with the encoding of the initial state of Q.

Parallel Composition $P := Q|[L]|R$ To derive $[\![P]\!]$ from P and $\Phi = \mathcal{P}_{\bowtie p}(\Phi_1[\rho]^I\Phi_2)$, we must not assume that $[\![Q]\!]$ resp. $[\![R]\!]$ are already available. Instead, we have to derive $[\![P]\!]$ from Q and R step by step, by respecting the same conditions as for the prefix operator, i.e., depending on the current state of s_P^{DS} of P, and z of A_ρ, we add transitions either to a "regular" successor of s_P^{DS} or to $FAIL$, resp. $SUCC$.

In Fig. 7 the algorithm for the derivation of $[\![P]\!]$ from Q and R for a single transition is given. We list only a few of the possible cases. This procedure has to be repeated, until all potential transitions that are possible are generated. This can be done using standard depth- or breadth-first search applied to P's parse tree.

5 Empirical Results

For our case studies we have employed the symbolic stochastic model checker CASPA. All results have been computed on a standard PC with Pentium IV 3.2 GHz processor, 1 GB RAM, running the operating system SuSe Linux 10.0.

Case 1:
(1) **if** $(a \notin L \wedge (s_Q^{DS} \models \Phi_1) \wedge (z \xrightarrow{a}_{A_\rho} z' \wedge s_Q^{DS} \models \Xi))$
/*with Ξ being a test formula attached to the current transition of A_ρ. */
(2) $B(Q') := TR(Enc_S(Q) \circ Enc_S(R), a, \lambda, Enc_S(Q') \circ Enc_S(R))$
(3) $B(P) := B(P) + B(Q')$
Case 2:
(4) **if** $(a \notin L \wedge (s_Q^{DS} \models \Phi_1) \wedge (z \not\xrightarrow{a}_{A_\rho} z' \wedge s_Q^{DS} \models \Xi))$
(5) $B(Q') := TR(Enc_S(Q) \circ Enc_S(R), a, \lambda, FAIL)$
(6) $B(P) := B(P) + B(Q')$
Case 3:
(7) **if** $(a \in L \wedge (s_Q^{DS} \models \Phi_1) \wedge (z \xrightarrow{a}_{A_\rho} z' \wedge s_Q^{DS} \models \Xi) \wedge (s_R^{DS} \models \Phi_1) \wedge (s_R^{DS} \models \Xi))$
(8) $B(Q') := TR(Enc_S(Q) \circ Enc_S(R), a, \lambda, Enc_S(Q') \circ Enc_S(R'))$
(9) $B(P) := B(P) + B(Q')$
Case 4:
(10) **if** $(a \in L \wedge (s_Q^{DS} \models \Phi_1) \wedge (s_{Q'}^{DS} \models \Phi_2) \wedge (z \xrightarrow{a}_{A_\rho} z' \wedge s_Q^{DS} \models \Xi) \wedge (s_R^{DS} \models \Phi_1) \wedge (s_{R'}^{DS} \models \Phi_2) \wedge (s_R^{DS} \models \Xi))$
(11) $B(Q') := TR(Enc_S(Q) \circ Enc_S(R), a, \lambda, SUCC)$
(12) $B(P) := B(P) + B(Q')$

Fig. 7. Algorithm for parallel composition $P := Q|[L]|R$

5.1 Fault-Tolerant Packet Collector

Let us consider the system from Example 1. We will check the SPDL path formulae presented there. In Table 1 we find the model sizes for these formulae. In columns three to five, we list the maximum size of the product CTMC that is generated for model checking SPDL without property-driven state space generation, which is the product of the size of the automaton and the system model. In columns six to eight we list the state space sizes as they are generated when using the property-driven approach proposed in this paper, and on which model checking is actually carried out. We see, that we can avoid the generation of many states, thereby reducing the memory requirements for SPDL model checking. We see in Table 2 that for both formulae the property-driven state space generation also requires less time than the traditional approach.

`max`	State space size	Not Property-driven			Property-driven		
		Φ_1	Φ_2	Φ_3	Φ_1	Φ_2	Φ_3
5,000	15,001	15,001	60,004	45,003	5,002	20,003	10,003
15,000	45,001	45,001	180,004	135,003	15,002	60,003	30,003
30,000	90,001	90,001	360,004	270,003	30,002	120,003	60,003
50,000	150,001	150,001	600,004	450,003	50,002	200,003	100,003

Table 1. State space sizes for Φ_1 to Φ_3 (Packet collector)

5.2 Kanban System

The Kanban manufacturing system was first described as a stochastic Petri net in [7]. We consider a Kanban system with four cells, a single type of Kanban cards and the possibility that some workpieces may need to be reworked. We will check the following properties:

max	Not Property-driven			Property-driven		
	Φ_1	Φ_2	Φ_3	Φ_1	Φ_2	Φ_3
5,000	2.9 sec.	3.3 sec.	3.1 sec.	2.0 sec.	2.8 sec.	2.9 sec.
15,000	10.00 sec.	10.8 sec.	11.2 sec.	6.9 sec.	9.0 sec.	9.0 sec.
30,000	21.4 sec.	22.7 sec.	22.5 sec.	17.8 sec.	18.9 sec.	19.6 sec.
50,000	37.9 sec.	45.3 sec.	44.4 sec.	33.6 sec.	40.4 sec.	32.8 sec.

Table 2. State space generation times for Φ_1 to Φ_3 (Packet collector)

n	State space size	Not Property-driven			Property-driven		
		Φ_1	Φ_2	Φ_3	Φ_1	Φ_2	Φ_3
5	2,546,432	22,917,888	33,103,616	43,289,344	83	13	159
8	133,865,325	1,204,787,925	1,740,249,225	2275710525	189	13	240
10	1,005,927,208	9,053,344,872	13,077,053,704	17,100,762,536	276	13	294
12	5,519,907,575	49,679,168,175	71,758,798,475	93,838,428,775	364	13	348
15	46,998,779,904	-	-	-	496	13	411

Table 3. State space sizes for Φ_1 and Φ_2 (Kanban)

n	Not Property-driven			Property-driven		
	Φ_1	Φ_2	Φ_3	Φ_1	Φ_2	Φ_3
5	0.8 sec.	0.7 sec.	0.7 sec.	0.1 sec.	0.1 sec.	0.1 sec.
8	4.7 sec.	4.2 sec.	4.5 sec.	0.2 sec.	0.2 sec.	0.2 sec.
10	11.4 sec.	10.8 sec.	11.0 sec.	0.5 sec.	0.5 sec.	0.5 sec.
12	21.7 sec.	21.5 sec.	22.1 sec.	0.8 sec.	0.7 sec.	0.7 sec.
15	-	-	-	1.6 sec.	1.5 sec.	1.5 sec.

Table 4. State space generation times for Φ_1 and Φ_2 (Kanban)

- Φ_1: Is the requirement, that within t time units exactly three reworks are required in station 1 satisfied with a probability that is at most p?
- Φ_2 :: Is the probability that a single job needs at most t time units to go through all 4 stations greater than p percent?
- Φ_3: Is the probability to reach station 4, within t time units, given in station 1 are no reworks required and in stations 2 and 3 in total exactly 2 reworks are necessary within $\bowtie p$?

From Table 3 we observe that for the formulae Φ_1 to Φ_3 the state space of the product CTMC is dramatically smaller than that of the original system, which stems from the fact that for all three formulae only very specific paths in the system are of interest. We can observe that for Φ_2 the size of the product CTMC is independent of the number of Kanban cards, which is not surprising, as we consider a specific card that goes through the system. In the second column we find the size of the original state space, in columns three to five we show the maximum size of the state space for the traditional approach, and in columns six through eight we list the final state space on which model checking actual is performed. We see in Table 4 for all three formulae that property-driven state space generation requires less time than the traditional approach. This is not surprising, as billions of states and even more important, billions of transitions of the original model do not to be explored in the property-driven approach.

Conf	State space size	Not Property-driven	Property-driven
		Φ_1	Φ_1
C1	753,664	2,260,992	53,306
C2	123,760	371,280	1,475
C3	381,681,664	1,145,044,992	6,554,329

Table 5. State space generation times for Φ_1 (fault-tolerant multi-processor)

5.3 Fault-Tolerant Multiprocessor System

This example is based on [17]. The original model consists of N computers each of which has the following components: Memory modules, CPUs, I/O ports, and error handlers. Each of these computer components consists of several subcomponents, that can fail, leading to the failure of one computer. The overall system is operational if at least one computer is operational.

We have generated the CTMC for three different configurations: C1 is the configuration consisting of two computers with three memory modules each; C1 has about 750,000 reachable states. C2 consists of 3 computers, with one memory module each. C3 comprises 3 computers and 3 memory modules each.

We will check the following formula Φ_1: Does the probability that computer failures and subsequently a system failure is only due to memory failures lie within the bounds as given by $\bowtie p$, given that the maximum time to reach a system failure state is at most t?

In Table 5 we show the model sizes for the above formulae. In column three, we list the maximum size of the product CTMC that is generated for model checking SPDL without property-driven state space generation, which is the product of the size of the automaton and the system model. In column 4 we give the model size, when applying the property-driven state space generator. We do not list the model generation times here, which are below 0.1 sec. for all configurations, in both the property-driven and the non-property-driven case.

6 Conclusions

In this paper we have introduced a property-driven symbolic semantics for the stochastic process algebra $\mathcal{YAMPA}$. We have shown its usage of a property-driven semantics for model checking probabilistic SPDL path formulae reduces both time and memory requirements. These savings can be considerable, as shown for the Kanban system, where an overhead of several billion states could be avoided. The numerical algorithms for stochastic model checking have a time complexity at least linear in state space size, so that an enormous overall time gain can be expected.

Generally, when doing numerical analysis of CTMCs with a huge state space some caution is required. As reported in [5], the accuracy of the numerical analysis depends on many factors, e.g. state space ordering, the actual iterative solution method, etc. But it must be stressed, that this is a problem that applies to all approaches that rely on numerical analysis. In fact, the probability masses

on both the model, generated using property-driven state space generation, and the model using the "traditional" approach are identical. The experiments we conducted, on both the reduced and non-reduced model did not yield any differences.

In the future we plan to combine this property-driven semantics with some notion of bisimulation reduction in order to obtain further state-space reductions and to investigate the possibilities to transfer the results from [2] to the stochastic case.

References

1. M. Ajmone Marsan, G. Balbo, and G. Conte. A Class of Generalized Stochastic Petri Nets for the Performance Evaluation of Multiprocessor Systems. *ACM Transactions on Computer Systems*, 2(2):93–122, May 1984.
2. A. Aziz, T. Shiple, V. Singhal, R. Brayton, and A. Sangiovanni-Vincentelli. Formula-dependent equivalence for compositional CTL model checking. *Form. Methods Syst. Des.*, 21(2):193–224, 2002.
3. Ch. Baier, L. Cloth, B. R. Haverkort, M. Kuntz, and M. Siegle. Model checking markov chains with actions and state labels. *IEEE Transactions on Software Engineering*, 33(4):209–224, 2007.
4. Ch. Baier, B. Haverkort, H. Hermanns, and J.P. Katoen. Model-Checking Algorithms for Continuous-Time Markov Chains. *IEEE Trans. Software Eng.*, 29(7):1–18, July 2003.
5. A. Bell. *Distributed Evaluation of Stochastic Petri Nets.* PhD thesis, RWTH Aachen, Fakultät für Mathematik, Informatik und Naturwissenschaften, 2003.
6. R.E. Bryant. Graph-based Algorithms for Boolean Function Manipulation. *IEEE Transactions on Computers*, C-35(8):677–691, August 1986.
7. G. Ciardo and M. Tilgner. On the use of Kronecker operators for the solution of generalized stochastic Petri nets. Technical Report 96-35, ICASE, 1996.
8. E.M. Clarke, E.A. Emerson, and A. Sistla. Automatic verification of finite state concurrent systems using temporal logic specifications: A practical approach. In *10th ACM Annual Symp. on Principles of Programming Languages*, pages 117–126, 1983.
9. M. Fischer and R. Ladner. Propositional dynamic logic of regular programs. *J. Comput. System Sci.*, 18:194–211, 1979.
10. M. Fujita, P. McGeer, and J.C.-Y. Yang. Multi-terminal Binary Decision Diagrams: An efficient data structure for matrix representation. *Formal Methods in System Design*, 10(2/3):149–169, April/May 1997.
11. H. Hermanns, U. Herzog, and J.-P. Katoen. Process algebra for performance evaluation. *Theoretical Computer Science*, 274(1-2):43–87, 2002.
12. J. Hillston. *A Compositional Approach to Performance Modelling.* Cambridge University Press, 1996.
13. J.-P. Katoen, T. Kemna, I. Zapreev, and D. Jansen. Bisimulation minimisation mostly speeds up probabilistic model checking. In *Proc. 13th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'07)*, LNCS 4424, pages 76–92. Springer, 2007.
14. M. Kuntz. *Symbolic Semantics and Verification of Stochastic Process Algebras.* PhD thesis, Universität Erlangen-Nürnberg, Institut für Informatik 7, 2006.

15. M. Kuntz and M. Siegle. Deriving symbolic representations from stochastic process algebras. In *Process Algebra and Probabilistic Methods, Proc. PAPM-PROBMIV'02*, pages 188–206. Springer, LNCS 2399, 2002.
16. M. Kuntz, M. Siegle, and E. Werner. CASPA - A Tool for Symbolic Performance and Dependability Evaluation. In *Proceedings of EPEW'04 (FORTE co-located workshop*, pages 293 – 307. Springer, LNCS 3236, 2004.
17. W. H. Sanders and L. M. Malhis. Dependability evaluation using composed SAN-based reward models. *Journal of Parallel and Distributed Computing*, 15(3):238–254, 1992.
18. M. Siegle. Advances in model representation. In L. de Alfaro and S. Gilmore, editors, *Process Algebra and Probabilistic Methods, Joint Int. Workshop PAPM-PROBMIV 2001*, pages 1–22. Springer, LNCS 2165, September 2001.