A Semantics for Every GSPN*

Christian Eisentraut[1], Holger Hermanns[1], Joost-Pieter Katoen[2], and Lijun Zhang[3]

[1] Saarland University — Computer Science, Germany
[2] Department of Computer Science, RWTH Aachen University, Germany
[3] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences

Abstract. Generalised Stochastic Petri Nets (*GSPN*s) are a popular modelling formalism for performance and dependability analysis. Their semantics is traditionally associated to continuous-time Markov chains (*CTMC*s), enabling the use of standard *CTMC* analysis algorithms and software tools. Due to ambiguities in the semantic interpretation of confused *GSPN*s, this analysis strand is however restricted to nets that do not exhibit non-determinism, the so-called well-defined nets. This paper defines a simple semantics for *every GSPN*. No restrictions are imposed on the presence of confusions. Immediate transitions may be weighted but are not required to be. Cycles of immediate transitions are admitted too. The semantics is defined using a non-deterministic variant of *CTMC*s, referred to as Markov automata. We prove that for well-defined bounded nets, our semantics is weak bisimulation equivalent to the existing *CTMC* semantics. Finally, we briefly indicate how every bounded *GSPN* can be quantitatively assessed.

Keywords: timed and stochastic nets, semantics, confusion, (weak) bisimulation, continuous-time Markov chains.

1 Introduction

Generalised Stochastic Petri Nets (*GSPN*s) [4,3,8] constitute a formalism to model concurrent computing systems involving stochastically governed timed behaviour. *GSPN*s are based on Petri nets, and are in wide-spread use as a modelling formalism in different engineering and scientific communities. From Petri nets they inherit the underlying bipartite graph structure, partitioned into *places* and *transitions*, but extend the formalism by distinguishing between *timed transitions* and *immediate transitions*. The latter can fire immediately and in zero time upon activation. The firing time of a timed transition is governed by a *rate*, which serves as a parameter of a negative exponential distribution. Timed transitions are usually depicted as non-solid bars, while immediate transitions are depicted as solid bars.

The precise semantics of a *GSPN* may conceptionally be considered as consisting of two stages. First, an abstract, high-level semantics describes *when* which transitions

* This work is supported by the EU FP7 Programme under grant agreement no. 295261 (MEALS) and 318490 (SENSATION), by the DFG as part of the SFB/TR 14 AVACS, and by DFG/NWO bilateral research programme ROCKS.

may fire, and *with what probability*. Speaking figuratively in terms of a token game, this semantics determines how tokens can be moved from place to place by the firing of transitions. Then second, a lower-level mathematical description of the underlying stochastic process, typically a continuous time Markov chain (*CTMC*, for short), is derived to represent the intended stochastic behaviour captured in the first stage. This Markov chain is then subject to the analysis of steady-state or transient probabilities of markings, or more advanced analysis such as stochastic model checking.

The modelling power of *GSPN*s is particularly owed to the presence of immediate transitions [12]. Unfortunately, this characteristic strength of the formalism may lead to semantically intricate situations [9,12,13,14,15,25,32]. One of the most prominent cases is *confusion* [3,8]. In confused nets, the firing order of two concurrently enabled, non-conflicting immediate transitions determines whether two subsequent transitions are in conflict or not. The net in Fig. 1 is confused, since transitions t_1 and t_2 are not in direct conflict, but firing transition t_1 first leads to a direct conflict between t_2 and t_3, which does not occur if t_2 fires first instead. Confusion is not a problem of the high-level (token game) semantics of a net, as it is entirely clear which transition may fire, and how tokens are moved in either case. It is rather a problem of the underlying stochastic process that ought to be defined by this net. Recall that the transitions t_1 through t_3 are all immediate, and thus happen without elapse of time. Thus, their firing is basically transparent to a continuous time evolution. Places p_4 and p_5 enable two distinct timed transitions with rate λ_1 and λ_2 respectively, cf. Fig. 1. Now, depending on how the confusion between the transitions (and potentially the direct conflict between t_2 and t_3) is resolved, the underlying stochastic behaviour *either* corresponds to an exponential delay with rate λ_1, *or* to a delay with rate λ_2. Which of the two delays happens is not determined by the net structure, and as such is *non-deterministic*. Figure 2 shows a graphical representation of this phenomenon as a marking graph. States correspond to markings of the net in Fig. 1, and there is an obvious graphical correspondence with respect to the representation of the firing of timed or immediate transitions by similarly shaped edges. In state $\{p_2, p_3\}$ the direct conflict between t_2 and t_3 in the net yields a non-deterministic choice.

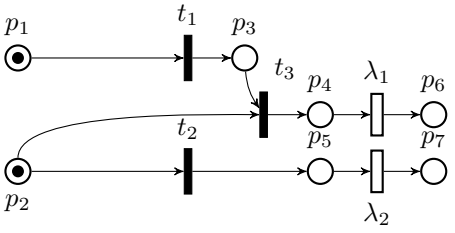

Fig. 1. Confused *GSPN*, see [3, Fig. 21]

As the resulting process is not a *CTMC*, workarounds have been developed. To resolve (or: avoid) non-determinism, *priorities* and *weights* have been introduced [1]. Intuitively, weights are assigned to immediate transitions at the net level so as to induce a probabilistic choice instead of a non-deterministic choice between (equally-prioritised) immediate transitions. Ignoring priorities, when-

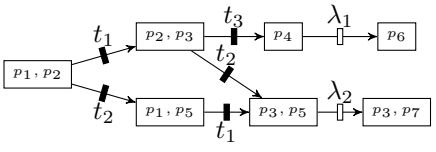

Fig. 2. Non-deterministic behaviour of the confused *GSPN* of Fig. 1

ever more than one immediate transition is enabled, the probability of selecting a certain enabled immediate transition is determined by its weight relative to the sum of the weights of *all* –including those that are independent– enabled transitions.

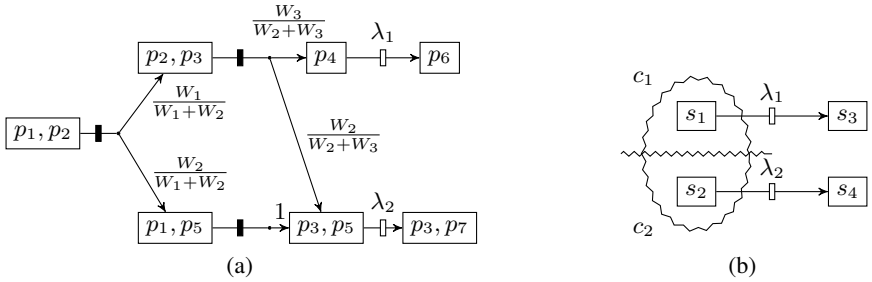

Fig. 3. (a) Probabilistic behaviour of *weighted* confused *GSPN* in Fig. 1; (b) the resulting *CTMC*

For example, for the marking depicted in Fig. 1, transition t_1 is selected with probability $\frac{W_1}{W_1+W_2}$ where W_i is the weight of transition t_i. In this way, we obtain an unambiguous stochastic process for this *GSPN*, cf. Fig. 3(a). Now, the unlabelled edges have multiple endpoints and denote probability distributions over markings. We can consider this as a semi-Markov process, which has both zero-time delay and exponentially distributed time delay edges, as worked out, for instance by Balbo [8]. In order to derive a *CTMC* from this process, sequences of zero-time delay edges are fused into probability distributions over states. For our example net, we obtain the *CTMC* in Fig. 3(b) with initial distribution μ^0 with $\mu^0(s_1) = c_1$ and $\mu^0(s_2) = c_2$ where

$$c_1 = \frac{W_1}{W_1+W_2} \cdot \frac{W_3}{W_2+W_3} \text{ and } c_2 = \frac{W_2}{W_1+W_2} + \frac{W_1}{W_1+W_2} \cdot \frac{W_2}{W_2+W_3}.$$

These quantities correspond to the reachability probability of marking $\{p_4\}$ and $\{p_3, p_5\}$, respectively from the initial marking. Unfortunately, this approach has a drawback, related to the *dependence* and *independence* of transitions, an important concept in Petri net theory. In our example net of Fig. 1, the transitions t_1 and t_2 are *independent*. Their firings happen independent from each other, as the two transitions share no places. Transitions t_2 and t_3, in contrast, are dependent, as the firing of one of them influences the firing of the other (by disabling it) via the shared input place p_2. However, the expected independence between t_1 and t_2 is not reflected in our *GSPN* above after introducing weights. Instead, the probability to reach marking p_4 (and marking p_5) under the condition that transition t_2 has fired will differ from the corresponding probability under the condition that t_1 has fired. A further conceptual drawback from a modelling perspective, is that when a new immediate transition is inserted between t_1 and t_3, then this changes these probabilities. This is irritating, since we only refine one immediate transition into a sequence of two immediate transitions. Since immediate transitions do not take time, this procedure should not result in a change of the underlying stochastic model. However, it does. We can also consider this phenomenon as a problem of locality. A local change of the net has unexpected global consequences with respect to the probabilities.

To remedy this defect, several approaches to define the stochastic process at the net level have been proposed. At the core of these approaches, immediate transitions are usually partitioned according to their conflict behaviour, based on a structural analysis of the net. The standard approach is to partition them into *extended conflict sets* (shortly, *ECS*s) [1], which is a generalisation of structural conflicts in the presence of priorities (which are not treated here). Intuitively, two transitions are in structural conflict in a marking, if both are enabled in this marking, and firing any of them will disable the other. Inside an *ECS*, weights are used to decide immediate transition firings, while no choice is resolved probabilistically across *ECS*s. For confusion-free nets, the *ECS* does provide a way of resolving conflicts probabilistically with a localised interpretation of weights. Unfortunately, for confused nets, this solution approach suffers from the same problem as our initial approach: The *ECS*s for the net in Fig. 1 are given by the partition $\{\{t_1\}, \{t_2, t_3\}\}$. As transitions t_2 and t_3 are in the same *ECS*, the decision which to fire will be resolved probabilistically according to their weights. Transitions t_1 and t_2, in contrast, are in different *ECS*. Thus, the decision will still need to be resolved non-deterministically, given that they may be enabled at the same moment. Inserting immediate transition t_4 between t_1 and t_3 as mentioned above will lead to the *ECS*s $\{\{t_1\}, \{t_4\}, \{t_2, t_3\}\}$. Thus, still only the decision between transitions t_2 and t_3 is resolved probabilistically and not influenced by t_4. So, since some decisions are forced to be non-deterministic, this approach does in general not yield a mathematically well-defined stochastic process. Moreover, it is easy to see that in our example, any partition of immediate transitions will suffer from one of the semantic problems discussed.

In summary, certain nets lead to undesirable semantic problems. Due to this fact, several researchers have identified certain classes of nets as *not well-defined* (aka. *ill-defined*) [3,14,15]. Such nets are excluded both semantically and from an analysis point of view. Several different definitions have occurred in the literature. However, ill-defined nets, with confused nets being a prominent example, are not *bad* nets *per se*. As Balbo states [7]: "*this underspecification [in confused nets] could be due either to a precise modelling choice [...] or to a modelling error*". We firmly believe that the modeller should have full freedom of modelling choices, and that such choices should not be treated as errors *by definition*.

Contribution of this paper. This paper presents a semantics for *GSPN*s that is *complete* in the sense that it gives a meaning to *every GSPN*. Our semantics is *conservative* with respect to the well-established existing semantics of *well-defined* nets. More precisely, we show that for well-defined bounded *GSPN*s, our semantics is weak bisimulation equivalent to the classical *CTMC* semantics. This entails that measures of interest, such as steady-state and transient probabilities are identical. Finally, we sketch the available analysis trajectory for our semantics, including confused bounded nets.

Outline. We first recall the definition of *GSPN*s in Section 2. In Section 3 we present the *MA* semantics for *GSPN*s based on the marking graph. The bisimulation semantics will be discussed in Section 4. In Section 5 we describe quantitative analysis approaches for arbitrary (bounded) *GSPN*s, and Section 6 concludes the paper.

2 Generalised Stochastic Petri Nets

This section introduces *GSPN*s, where, for the sake of simplicity, we do not consider transition priorities. For a set X, we use $\Sigma(X)$ to denote the set of all partitions of X. For a set of places P, a *marking* m is a multi-set over P of the form $m : P \to \mathbb{N}$. We let M denote the set of all markings over P, and use m, m_0 etc to denote its elements.

Definition 1 (Generalized stochastic Petri net). *A* generalised stochastic Petri net G *(GSPN) is a tuple* $(P, T, I, O, H, m_0, W, \mathcal{D})$ *where:*

- P *is a finite set of* places*,*
- $T = T_i \cup T_t$ *is a finite set of transitions (*$P \cap T = \emptyset$*) partitioned into the sets* T_t *and* T_i *of* timed *and* immediate *transitions,*
- $I, O, H : T \to M$ *defines the transitions'* input *places,* output *places,* inhibition *places*[4]*,*
- $m_0 \in M$ *is the initial marking,*
- $W : T \to \mathbb{R}_{>0}$ *defines the transitions'* weights*, and*
- $\mathcal{D} : M \to \Sigma(T)$ *is a marking-dependent* partition *satisfying the condition that* $T_t \in \mathcal{D}(m)$ *for all markings* $m \in M$*.*

The above definition agrees, except for the last component $\mathcal{D}$, with the classical *GSPN* definition in the literature [2,3,8]. We use the marking-dependent partition function $\mathcal{D}$ as a generalisation of the extended conflict set mentioned before. It serves to express for which immediate transitions choices are resolved probabilistically, and for which non-deterministically. This information is usually not provided in the net definition. Instead the (marking independent) *ECS* are derived based on a structural analysis of the net at hand. The reason why we include this information in an explicit form in the definition is mainly ought to formal reasons. However, it also enables (but does not enforce) a view where the choices between immediate transitions are resolved as a consequence of a conscious modelling decision, possibly decoupled from the net structure. The constraint $T_t \in \mathcal{D}(m)$ is due to the fact that all enabled timed transitions are always weighted against each other in a race. On the expense of slightly more complicated definitions in the following, we could eliminate this technicality and let $\mathcal{D} : M \to \Sigma(T_i)$.

The *input*, *output* and *inhibition* functions assign to each transition a mapping $P \to \mathbb{N}$, specifying the corresponding cardinalities. A transition has *concession* if sufficiently many tokens are available in all its input places, while the corresponding inhibition places do not contain sufficiently many tokens for an inhibitor arc to become effective. Firing a transition yields a (possibly) new marking, which is obtained by removing one or more tokens from each input place and adding tokens to the transition's output places. Immediate transitions execute immediately upon becoming enabled, whereas timed transitions are delayed by an exponentially distributed duration which is uniquely specified by a *transition rate* (i.e., a positive real number defined by the weights).

For notational convenience, we write cascaded function application with indexed notation of the first parameter. For example, we write I_t, O_t and H_t for $I(t)$, $O(t)$ and $H(t)$, respectively. The semantics of a *GSPN* is defined by its *marking graph*, which

[4] If transition t has no inhibitor places, we let $H(t) = \infty$.

is obtained by playing the "token game". Immediate transitions are fired with priority over timed transitions [2,12,3]. Accordingly, if both timed and immediate transitions have concession in a marking, only the immediate transitions become enabled. Let G be a *GSPN* with marking $m \in M$.

Definition 2 (Concession and enabled transitions).

1. *The set of transitions with* concession *in marking m is defined by:*

$$\mathit{conc}(m) = \big\{ t \in T \mid \forall p \in P.\ m(p) \geq I_t(p) \wedge m(p) < H_t(p) \big\}.$$

2. *The set of* enabled transitions *in marking m is defined by: $\mathit{en}_m = \mathit{conc}(m) \cap T_i$ if $\mathit{conc}(m) \cap T_i \neq \emptyset$, and $\mathit{en}_m = \mathit{conc}(m)$ otherwise.*

A marking m is *vanishing* whenever an immediate transition is enabled in m, otherwise it is *tangible*. Given the priority of immediate transitions over timed ones, the sojourn time in vanishing markings is zero. In a vanishing marking, none of the timed transitions which have concession is enabled. In a *tangible* marking m, only timed transitions can be enabled. The residence time in tangible marking m is determined by a negative exponential distribution with rate $\sum_{t \in \mathit{en}_m} W(t)$. The effect of executing a transition is formalised in the classical way:

Definition 3 (Transition execution). *Let the* transition execution relation $[\cdot\rangle \subseteq M \times T \times M$ *be such that for all markings $m, m' \in M$ and transitions $t \in T$ it holds:*

$$m\,[t\rangle\,m' \iff t \in \mathit{en}_m \wedge \forall p \in P.\ m'(p) = m(p) - I_t(p) + O_t(p).$$

We now recall the notion of *marking graph*, obtained from reachable markings:

Definition 4 (Reachable marking graph). *The* marking graph *of the GSPN G is the labelled digraph $\mathit{MG}(G) = (RS, E)$, where*

- *RS is the smallest set of reachable markings satisfying: $m_0 \in RS$, and $m \in RS \wedge m\,[t\rangle\,m'$ implies $m' \in RS$.*
- *The edge between m and m' is labelled by the transition t such that $m\,[t\rangle\,m'$.*

This graph describes how a net may evolve in terms of its markings. However, it fails to faithfully represent the stochastic aspects of the net. This is made more precise below.

Recall the idea that we consider certain immediate transitions probabilistically dependent from some other transitions (mainly when they are in conflict), while we consider them independent from others. Traditionally, these relations are captured by extended conflict sets (*ECS*s [1]). Here, we consider a generalisation of this concept in the form of an arbitrary immediate transitions partition $\mathcal{D}_m$. For each marking m, the partition $\mathcal{D}_m$ determines a way of resolving conflicts between immediate transitions. Each set $C \in \mathcal{D}_m$ consists of transitions whose conflicts are resolved probabilistically in m. On the other hand, transitions of different sets are considered to behave in an independent manner, i.e., we make a non-deterministic selection if several of them are enabled in m. Our semantics will be general enough that we may allow the latter even if there is a *structural* conflict between these transitions. Let us make this precise.

Assume that some transitions in the set $C \in \mathcal{D}_m$ are enabled and C is chosen to be fired. Under this condition, the probability that a specific transition fires is given as the normalised weight of the enabled transitions in C. Precisely, $\mathbf{P}_C\{t \mid m\} = 0$ if $t \notin C \cap en_m$, and otherwise:

$$\mathbf{P}_C\{t \mid m\} = \frac{W(t)}{W_C(m)} \quad \text{where} \quad W_C(m) = \sum_{t \in C \cap en_m} W(t). \tag{1}$$

If m is a vanishing marking, $W_C(m)$ denotes the cumulative weight of all enabled (i.e., immediate) transitions in C. In this case the probability $\mathbf{P}_C\{t \mid m\}$ of taking the immediate transition t in m is determined by the weight assignment W. Note that $\mathbf{P}_C\{t \mid m\}$ is 0 if t is neither enabled nor an element from C. The case that m is tangible is similar. Then only timed transitions are enabled, and recall that the set of timed transitions T_t is an element in $\mathcal{D}_m$. Thus, $C = T_t$. Accordingly,

$$W_C(m) = \sum_{t \in en_m} W(t)$$

is the exit rate from the tangible marking m. In this case, $\mathbf{P}_C\{t \mid m\}$ is the probability of taking the transition t if the tangible marking m is left.

In both cases, several distinct transition firings may lead from m to the same marking m'. These need to be accumulated. With some overload of notation we define

$$\mathbf{P}_C(m, m') = \sum_{m[t\rangle m'} \mathbf{P}_C\{t \mid m\}.$$

3 Markov Automata Semantics for *GSPN*s

Our aim is to provide a semantics to every *GSPN*. In particular, this includes nets in which multiple immediate transitions are enabled in a marking, nets with cycles of immediate transitions, as well as confused nets. Obviously, stochastic processes such as *CTMC*s do not suffice for this purpose, as they cannot express non-determinism. We therefore resort to an extension of *CTMC*s with non-determinism, *Markov automata* (*MA*s, for short) as introduced in [20]. This model permits to represent the concepts above, including a formulation in terms of a semi-Markov process with zero-timed delay and exponentially distributed time delays [8], while in addition supporting non-determinism between transition firings in vanishing markings. Figure 2 and 3(a) are in fact graphical representations of *MA*.

3.1 Markov Automata

We first introduce some preliminary notions that we shall use in the rest of the paper. A *subdistribution* μ over a set S is a function $\mu : S \mapsto [0, 1]$ such that $\sum_{s \in S} \mu(s) \leq 1$. Let $Supp(\mu) = \{s \in S \mid \mu(s) > 0\}$ denote the support of μ and $\mu(S') := \sum_{s \in S'} \mu(s)$ the probability of $S' \subseteq S$ with respect to μ. Let $|\mu| := \mu(S)$ denote the *size* of the

subdistribution μ. We say μ is a *full distribution*, or simply distribution, if $|\mu| = 1$. Let $Dist(S)$ and $Subdist(S)$ be the set of distributions and subdistributions over S, respectively. For $s \in S$, let $\delta_s \in Dist(S)$ denote the *Dirac* distribution for s, i.e., $\delta_s(s) = 1$. Let μ and μ' be two subdistributions. We define the subdistribution $\mu'' := \mu \oplus \mu'$ by $\mu''(s) = \mu(s) + \mu'(s)$, if $|\mu''| \leq 1$. Conversely, we say that μ'' can be split into μ and μ', or that (μ, μ') is a splitting of μ''. Since $\oplus$ is associative and commutative, we use the notation $\bigoplus_{i \in I}$ for arbitrary sums over a finite index set I. Moreover, if $c \cdot |\mu| \leq 1$ and $c > 0$, we let $c\mu$ denote the subdistribution defined by: $(c\mu)(s) = c \cdot \mu(s)$. For $s \in S$ and $\mu \in Subdist(S)$ let $\mu \ominus s$ denote the subdistribution μ' with $\mu'(t) = \mu(t)$ if $t \neq s$ and $\mu'(s) = 0$.

Definition 5 (Markov automaton). *A* Markov automaton A *is a quadruple* $(S, \multimap, \multimapdotinv, \mu^0)$*, where*

- S *is a non-empty countable set of* states*,*
- $\multimap \subset S \times Dist(S)$ *is a set of* immediate edges*,*
- $\multimapdotinv \subset S \times \mathbb{R}_{>0} \times Dist(S)$ *is a set of* timed edges*, and*
- $\mu^0 \in Dist(S)$ *is an* initial distribution *over the states* S*.*

It is required that every state $s \in S$ *has at most* one *outgoing* timed edge.[5]

We let s, u and their variants with indices range over S, and μ over $Dist(S)$. An immediate edge $(s, \mu) \in \multimap$ is denoted by $s \multimap \mu$. The operational interpretation of edge $s \multimap \mu$ is that from s a next state will be probabilistically determined according to distribution μ and that in s no time elapses. Similarly, a timed edge $(s, \lambda, \mu) \in \multimapdotinv$ is denoted by $s \overset{\lambda}{\multimapdotinv} \mu$. We use $\lambda, r \in \mathbb{R}_{>0}$ to denote the rate of a negative exponential distribution. An edge $(s, \mu) \in \multimap$ is said to originate from state s.

A *state* $s \in S$ is called *tangible* if no immediate edge originates from s. A *probability distribution* over states is called *tangible* if all states in its support set are tangible. We write $s \xrightarrow{\alpha} \mu$ if either (i) $\alpha = \varepsilon$ (i.e. the edge is unlabelled) and $s \multimap \mu$ or (ii) $\alpha \in \mathbb{R}_{>0}$, s is tangible and $s \overset{\alpha}{\multimapdotinv} \mu$, or (iii) $\alpha = 0$, $\mu = \delta_s$, and s has no outgoing transition. This notation combines immediate edges (i) with timed edges (ii), but timed edges are only considered from tangible states. Clause (iii) generalizes the implicit tangibility check of clause (ii) to states without outgoing edges. The inclusion of a tangibility check inside the above clauses (ii) and (iii) of $\xrightarrow{\alpha}$ will have an interesting effect, discussed in Section 4.2. We stipulate that *non-determinism* occurs in an *MA* whenever multiple immediate edges originate from a state. In that case, it is deliberately left unspecified with which probability a particular immediate edge is taken. This represents a non-deterministic choice. Obviously, *CTMC*s can be considered as special cases of *MA*s: A *CTMC* is a *MA* with $\multimap = \emptyset$.

3.2 Basic semantics of *GSPN*s

We are now in the position to define the semantics of every *GSPN*—including the *non* well-defined ones—by means of a *MA*. The intuition is rather simple. Basically

[5] This is not a restriction since the effect of two timed edges $s \overset{r}{\multimapdotinv} \mu$ and $s \overset{r'}{\multimapdotinv} \mu'$ can be combined into a single timed edge $s \overset{r+r'}{\multimapdotinv} \mu''$.

the semantics of a *GSPN* corresponds to its reachable marking graph, cf. Def. 4. States correspond to markings, taking an immediate edge in the *MA* is the counterpart to firing an immediate transition in the net, and likewise for timed edges and timed transitions. The marking graph can therefore directly be interpreted as a Markov automaton.

Definition 6 (Basic *MA* semantics for *GSPN*s). *The MA* semantics *of the GSPN* $G = (P, T, I, O, H, m_0, W, \mathcal{D})$ *is the MA* $A_G = (S, \rightarrow\!\!\blacksquare, \rightarrow\!\!\square, \mu^0)$*, where*

- $S = RS$ *is the reachable set of markings in the marking graph,*
- $\mu^0 = \delta_{m_0}$*,*
- *for every* $m \in RS$*, and each equivalence* $C \in \mathcal{D}_m$*,*
  1. *there is an edge* $m \overset{r}{\rightarrow\!\!\square} \mu$ *if and only if* m *is a tangible marking,* $r = W_C(m)$ *and* $\mu(m') = \mathbf{P}_C(m, m')$ *for all* $m' \in RS$*,*
  2. *there is an edge* $m \rightarrow\!\!\blacksquare \mu$ *if and only if* m *is a vanishing marking and* $\mu(m') = \mathbf{P}_C(m, m')$ *for all* $m' \in RS$*.*

So, the basic *MA* semantics is the marking graph of a *GSPN*. Every marking of the *GSPN* that is reachable by a sequence of (net) transitions from the initial marking corresponds to a state in the *MA*. As discussed before, in marking m of the net all enabled timed transitions t induce an exponentially distributed stochastic delay with a rate r that is the sum of all weights of enabled transitions. In this case, the probability to reach a marking m', say, by edge t is given as the edge's relative weight. This is reflected in clause 1 of the above *MA* semantics. If no timed transition is enabled in marking m, then no timed edge originates from state m.

In contrast, the enabled immediate transitions in a marking need to be represented by more than one immediate edge in the *MA*. Recall that each equivalence class $C \in \mathcal{D}_m$ corresponds to an *ECS* in *GSPN* terminology. For every such set C, the enabled transitions in C fire with a probability that is equal to their weight in relation to the sum of the weights of all enabled transitions in C. However, transitions that are in different sets in $\mathcal{D}_m$ are entirely independent. More precisely, transitions from different sets in $\mathcal{D}_m$ compete in a non-deterministic way. This is reflected in clause 2 of the above definition. The non-deterministic choice between transitions across different sets of $\mathcal{D}_m$ is represented by introducing an immediate edge for *every* set in the partition $\mathcal{D}_m$. The probabilistic decision among transitions *within a single* set, in turn, is reflected by the distribution over markings the corresponding immediate edge leads to.

3.3 Well-defined *GSPN*s

The aim of this section is to formalise and generalise well-defined *GSPN*s in terms of our new semantics. A central notion for this purpose is the concept of weak edges.

Labelled trees. The notion of weak edge is defined using labelled trees. For $\sigma, \sigma' \in \mathbb{N}^*_{>0}$, let $\sigma \leq \sigma'$ if there exists a (possibly empty) $\phi \in \mathbb{N}^*_{>0}$ such that $\sigma\phi = \sigma'$. We write $\sigma < \sigma'$ whenever $\sigma \leq \sigma'$ and $\sigma \neq \sigma'$. Let L be a set of labels. An *(infinite) L-labelled tree* is a partial function $\mathcal{T} : \mathbb{N}^*_{>0} \to L$ satisfying

- if $\sigma \leq \sigma'$ and $\sigma' \in \mathit{dom}(\mathcal{T})$, then $\sigma \in \mathit{dom}(\mathcal{T})$,

- if $\sigma i \in \mathit{dom}(\mathcal{T})$ for $i \in \mathbb{N}_{>1}$, then $\sigma(i-1) \in \mathit{dom}(\mathcal{T})$, and
- $\varepsilon \in \mathit{dom}(\mathcal{T})$.

The empty word ε is called the root of $\mathcal{T}$ and $\sigma \in \mathit{dom}(\mathcal{T})$ is a node of $\mathcal{T}$. For node σ of tree $\mathcal{T}$, let $\mathit{Children}(\sigma) = \{\sigma i \mid \sigma i \in \mathit{dom}(\mathcal{T})\}$. Node σ is a leaf of tree $\mathcal{T}$ if there is no $\sigma' \in \mathit{dom}(\mathcal{T})$ with $\sigma < \sigma'$; then $\mathit{Children}(\sigma) = \emptyset$. We denote the set of all leaves of $\mathcal{T}$ by $\mathit{Leaf}_{\mathcal{T}}$ and the set of all inner nodes of $\mathcal{T}$ by $\mathit{Inner}_{\mathcal{T}}$. If the tree only consists of the root, then $\mathit{Inner}_{\mathcal{T}} = \mathit{Leaf}_{\mathcal{T}} = \{\varepsilon\}$. In any other case the two sets are disjoint. We consider L-labelled trees with finite branching, i.e., $|\mathit{Children}(\sigma)| < \infty$ for all nodes σ.

Weak edges. Weak edges for probabilistic systems have been defined in the literature via probabilistic executions in [31], trees [17], or infinite sums [16]. We adopt the tree notation here. The material presented below concerning weak edges provides no innovation over the classical treatment, it is included for the benefit of the reader. Let $L = S \times \mathbb{R}_{>0}$. A node in an L-labelled tree is labelled by a state and the (by definition non-zero) probability of reaching this node from the root of the tree. For a node σ we write $\mathit{Sta}_{\mathcal{T}}(\sigma)$ for the first component of $\mathcal{T}(\sigma)$ and $\mathit{Prob}_{\mathcal{T}}(\sigma)$ for the second component of $\mathcal{T}(\sigma)$. If $\mathcal{T}$ is clear from the context we omit the subscripts.

Definition 7 (Weak edge tree). *Let $(S, \rightarrow\!\!\!\!\blacksquare, \rightarrow\!\!\!\!\square, \mu^0)$ be an MA. A* weak edge tree $\mathcal{T}$ *is a $S \times \mathbb{R}_{>0}$-labelled tree satisfying the following conditions*

1. $\mathit{Prob}(\varepsilon) = 1$,
2. $\forall \sigma \in \mathit{Inner}_{\mathcal{T}} \setminus \mathit{Leaf}_{\mathcal{T}} : \exists \mu : \mathit{Sta}(\sigma) \longrightarrow \mu$ *and* $\mathit{Prob}(\sigma) \cdot \mu = \xi$ *where* $\xi(\mathit{Sta}(\sigma')) = \mathit{Prob}(\sigma')$ *for all* $\sigma' \in \mathit{Children}(\sigma)$,
3. $\sum_{\sigma \in \mathit{Leaf}_{\mathcal{T}}} \mathit{Prob}(\sigma) = 1$.

A weak edge tree $\mathcal{T}$ corresponds to a probabilistic execution fragment: it starts from the root's state $\mathit{Sta}(\varepsilon)$, and resolves non-deterministic choices at every inner node of the tree, which represents the state in the MA it is labelled with. The second component of σ, $\mathit{Prob}(\sigma)$, is the probability of reaching the state $\mathit{Sta}(\sigma)$ via immediate edges in the MA, starting from the state $\mathit{Sta}(\varepsilon)$. The distribution associated with edge tree $\mathcal{T}$, denoted $\mu_{\mathcal{T}}$, is defined as $\mu_{\mathcal{T}} \stackrel{\text{def}}{=} \bigoplus_{\sigma \in \mathit{Leaf}_{\mathcal{T}}} \rho_\sigma$, where $\rho_\sigma \in \mathit{Subdist}(S)$ with $\rho_\sigma(s) = \mathit{Prob}(\sigma)$ if $s = \mathit{Sta}(\sigma)$ and $\rho_\sigma(s) = 0$ otherwise. Subdistribution $\mu_{\mathcal{T}}$ is said to be *induced* by $\mathcal{T}$. We are now in a position to define *weak edges*: For $s \in S$ and $\mu \in \mathit{Dist}(S)$, let $s \Longrightarrow \mu$ if μ is induced by some internal edge tree $\mathcal{T}$ with $\mathit{Sta}(\varepsilon) = s$.

We now generalise edges to edges originating in subdistributions over states. Let $\mu \in \mathit{Dist}(S)$. If for every state $s_i \in \mathit{Supp}(\mu)$, $s_i \Longrightarrow \mu_i'$ for some μ_i', then we write $\mu \Longrightarrow \bigoplus_{s_i \in \mathit{Supp}(\mu)} \mu(s_i)\mu_i'$. We apply a similar definition for $\xrightarrow{\alpha}$ instead of $\Longrightarrow$. Finally, for $\alpha \in \mathbb{R}$, we write $s \stackrel{\alpha}{\Longrightarrow} \mu$ if there exist μ_1 and μ_2 such that $s \Longrightarrow \mu_1$, $\mu_1 \xrightarrow{\alpha} \mu_2$ and $\mu_2 \Longrightarrow \mu$.

Intuitively, the weak edges in Def. 7 (referred to as *weak transitions* in the automata literature) are used to capture all possible evolutions along immediate edges starting from s. Thus, any edge itself is a weak edge, and note that from state s, there is always a weak edge $s \Longrightarrow \delta_s$, even if s is tangible.

Well-defined GSPNs. We are now ready to define well-defined *GSPN*s.

Definition 8 (Well-defined *GSPN*). *Let $G = (P, T, I, O, H, m_0, W, \mathcal{D})$ be a GSPN with MA semantics A_G. We say G is* well-defined*, if for every state $m \in RS$, and every pair (μ, μ') of distributions over* tangible states *it holds: $m \Longrightarrow \mu$ and $m \Longrightarrow \mu'$ implies $\mu = \mu'$.*

Different to [32], we are only interested in the probability to reach a marking, and whether it is uniquely specified, but not in the sequences of edges leading to tangible markings. Phrased differently, we are only interested in tangible state to tangible state probabilities [14,8].

It is not surprising that a well-defined *GSPN* induces a unique *CTMC*: states will correspond to those tangible markings, edge $\stackrel{r}{\multimap\!\!\square\!\!\rightarrow}$ is obtained by extending the weak edge until tangible states are reached. The uniqueness is guaranteed by the definition of well-defined *GSPN*s. This is summarised in the following definition:

Definition 9 (*CTMC* induced by a well-defined *GSPN*). *The well-defined GSPN G induces the CTMC $C_G = (S, \multimap\!\!\blacksquare\!\!\rightarrow, \multimap\!\!\square\!\!\rightarrow, \mu^0)$, where*

- *S is the set of reachable tangible markings of G,*
- *$m \stackrel{r}{\multimap\!\!\square\!\!\rightarrow} \mu$ iff μ is the unique distribution over tangible markings such that a distribution μ' exists with $m \stackrel{r}{\multimap\!\!\square\!\!\rightarrow} \mu'$ and $\mu' \Longrightarrow \mu$ in the basic MA semantics of G,*
- *μ_0 is the unique distribution over tangible markings such that $m_0 \Longrightarrow \mu_0$.*

Lemma 1. *The induced CTMC of a well-defined GSPN is unique (up to isomorphism).*

4 Bisimulation Semantics

The basic *MA* semantics we have introduced already has several advantages. It is complete, i.e. it provides semantics for every net, and it is amenable to several analysis techniques that are being established (see Sec. 5 for further details). Nevertheless, we want to address more desirable properties the current proposal does not have: (i) the semantics should be conservative with respect to the existing standard semantics for well-defined nets, (ii) immediate edges should be disregarded as much as possible, and exponential delays should be only distinguished up to lumpability. This ensures that the actual formal semantics agrees with the intuitive behaviour of a net and semantic redundancies are avoided as much as possible. For instance, the introduction of a new immediate transition between t_1 and t_3 in Fig. 1, which should be independent of every other concurrently enabled transition, should not affect the underlying semantics.

We now will implement the above requirements by defining the semantics of a *bounded GSPN* as its basic *MA* semantics modulo a behavioural equivalence, weak bisimilarity [20]. The basic *MA* semantics modulo weak bisimilarity will exactly represent the behavioural kernel of the *GSPN*. (The setting of unbounded *GSPN*s is left for further study.)

We first need the notion of a convex combination of weak edges. Let $\mu \stackrel{\alpha}{\Longrightarrow}_C \gamma$ if there exists a finite index set I, and weak edges $\mu \stackrel{\alpha}{\Longrightarrow} \gamma_i$ and a factor $c_i \in (0,1]$ for every $i \in I$, with $\sum_{i \in I} c_i = 1$ and $\gamma = \bigoplus_{i \in I} c_i \gamma_i$. This notion is standard for probabilistic automata, and inherited here for *MA*; see [31] for more details. Let the set of all splittings of immediate successor subdistributions be defined as

$$\mathit{split}(\mu) = \{(\mu_1, \mu_2) \mid \exists \mu' : \mu \Longrightarrow_C \mu' \wedge \mu' = \mu_1 \oplus \mu_2\}\,.$$

Definition 10 (Weak bisimulation [20]). *A symmetric relation $\mathcal{R}$ on subdistributions over S is called a* weak bisimulation *if and only if whenever $\mu_1 \mathcal{R} \mu_2$ then for all $\alpha \in \mathbb{R} \cup \{\varepsilon\}$: $|\mu_1| = |\mu_2|$ and for all $s \in \mathit{Supp}(\mu_1)$ there exist ${\mu_2}^{\rightarrow}, {\mu_2}^{\Delta}$: $({\mu_2}^{\rightarrow}, {\mu_2}^{\Delta}) \in \mathit{split}(\mu_2)$ and*

(i) $\mu_1(s)\delta_s \;\mathcal{R}\; {\mu_2}^{\rightarrow}$ and $(\mu_1 \ominus s) \;\mathcal{R}\; {\mu_2}^{\Delta}$
(ii) whenever $s \xrightarrow{\alpha} \mu_1'$ for some μ_1' then ${\mu_2}^{\rightarrow} \stackrel{\alpha}{\Longrightarrow}_C \mu''$ and $(\mu_1(s) \cdot \mu_1') \;\mathcal{R}\; \mu''$

Two subdistributions μ and γ are weak bisimilar, denoted by $\mu \approx \gamma$, if the pair (μ, γ) is contained in some weak bisimulation.

Note that weak bisimilarity is a relation over *distributions*, which is a natural choice for stochastic processes. Its basic idea is that two distributions μ and γ are bisimilar, if the edge of every state in the support of μ can by matched by a weak edge of a subdistribution of γ (Condition (ii)) in the usual sense of (probabilistic) bisimulation, however, enhanced by the idea that before γ is to be split into suitable subdistributions, it may perform an arbitrary sequence of weak immediate edges (Condition (i)). As it has been shown in [19], Condition (i) is the essential difference that distinguishes weak bisimulation for *MA*s from weak bisimulation for Probabilistic Automata [31]. Furthermore, although not obvious from the definition, it is exactly this condition that allows to *fuse* sequences of immediate edges into their unique final goal distribution, if existing.

Bisimulation can be lifted to a relation between *MA*s with disjoint state space. Two *MA*s A, A' are bisimilar, denoted $A \approx A'$, if their initial distributions are bisimilar in the direct sum, which is the *MA* obtained by considering the disjoint union of states and edges respectively. This shall be used in the next section to compare the semantics of models.

4.1 Revisiting well-definition

To illustrate why we consider weak *MA* bisimilarity a semantic equivalence especially well-suited for *GSPN* semantics, let us recall the standard procedure applied to derive a *CTMC* from the basic *MA* semantics underlying a well-defined *GSPN*. We illustrate this process with the *MA* from Fig. 3(a) as an example. For convenience, we repeat it in Fig. 4(a) below. This figure shows the basic *MA* semantics of the *GSPN* in Fig. 1 in the case that every immediate edge is weighted, and choices among immediate edges are always resolved probabilistically. For a shorter notation, we now denote edge probabilities by x_1, x_2 and so on. When we want to transform this *MA* into a *CTMC*, we successively remove every immediate edge by replacing a state with an outgoing immediate edge by the distribution that this immediate edge leads to. The result of this

replacement is shown in Figs. 4(b) and 4(c). Finally, when no such states remain, we obtain the *CTMC* in Fig. 4(d), where $c_1 = x_1x_3$ and $c_2 = x_2 + x_1x_4$. The effect of this iterative process of fusing transitions can also be formulated via matrix operations [3].

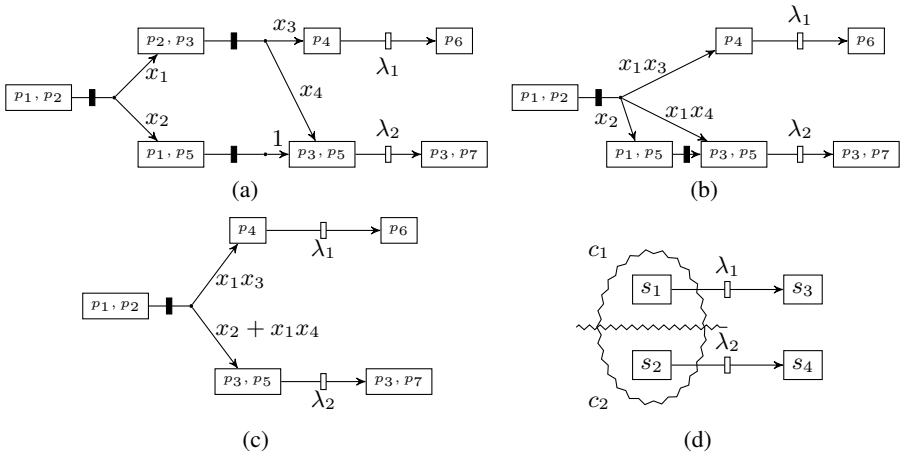

Fig. 4. From the *MA* semantics (a) a *CTMC* is obtained (d) by step-wise fusing immediate edges in (b) and (c).

In this example, this leads to a unique result, as every state has at most one outgoing immediate edge. In general, this leads to unique results whenever the net is well-defined. For nets with non-determinism, however, this approach does not lead to mathematically well-defined results.

For this purpose consider now the net in Fig. 5. Assume now that we do not resolve every choice of immediate transitions probabilistically, but only the conflict between t_2 and t_3. Hence let $\mathcal{D}_m = \{\{t_2, t_3\}, \{t_1\}, \{t_4\}\}$. Note that these are exactly the *ECS*s of the net. We then obtain the non-deterministic basic *MA* semantics in Fig. 6(a). Applying the fusing procedure as before is clearly not possible, since already in the initial state of the *MA*, the marking $\{p_1, p_2\}$, we have two outgoing immediate edges, which will finally lead to two different distributions over tangible markings.

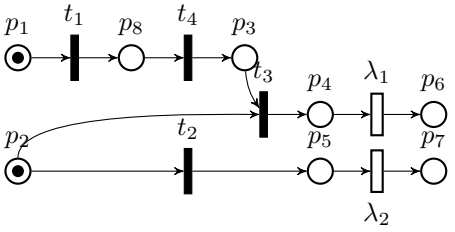

Fig. 5. Confused *GSPN* with additional transition

Although it is thus not possible to fully remove immediate edges here – as they are a necessary semantic component to express non-deterministic choice – we want to remove immediate edges whenever they can be fused. In our example, this would lead to the *MA* in Fig. 6(b). Only in the first state two immediate edges remain. They fully capture the non-deterministic behaviour of this *GSPN*.

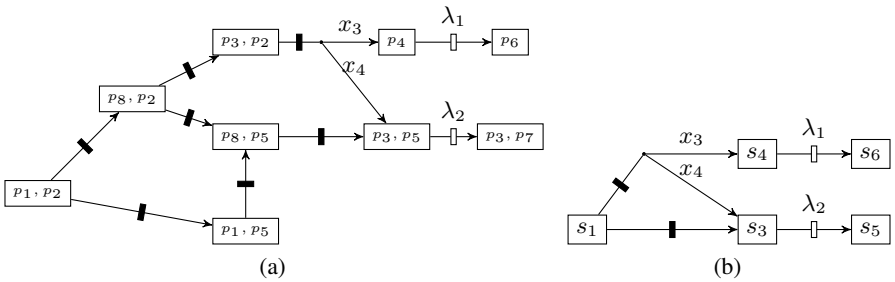

Fig. 6. A basic *MA* (a) with non-determinism and the smallest *MA* weakly bisimilar to it (b). In (b), state s_1 subsumes markings $\{p_1, p_2\}$ and $\{p_8, p_2\}$ from (a). All other markings with immediate behaviour are removed as a result of fusing them.

Weak *MA* bisimilarity has been designed to exactly perform the task of removing immediate edges by fusion when the result is uniquely defined. In fact, the *MA* in Fig. 6(b) is the (state- and transition-wise) minimal *MA* that is weakly bisimilar to the *MA* in Fig. 6(a).

Speaking more generally, weak bisimilarity gives us a powerful means to conservatively generalise the notion of tangible and vanishing markings. Formally, a tangible marking has been defined as a marking that has no outgoing immediate transitions. Markings that are not tangible are called vanishing. More intuitively speaking, as the words *tangible* and *vanishing* suggest, vanishing markings are semantically insignificant, while tangible markings constitute the semantic essence of a net's behaviour. Now, in the context of non-deterministic behaviour, besides of those states without immediate transitions, also those states with a non-deterministic choice between immediate transitions are semantically *tangible* in the literal sense (as long as the choice makes a behaviour difference in the end).

To make this precise, we will define the notion of *significant* markings as a conservative extension of tangible markings, and show that for well-defined nets, they coincide with tangible markings and vice versa.

Definition 11 (Significant marking). *Given a GSPN G and its basic MA semantics A_G, we call a marking m* insignificant *if it is vanishing* and *– in A_G – m is a state that has at least one outgoing immediate edge $m \longrightarrow \mu$ such that $\mu \approx \delta_m$. Otherwise we call marking m* significant*.*

Whereas every tangible marking is also significant, not every vanishing marking is insignificant. Only those vanishing markings are also insignificant, which have an immediate successor distribution that is semantically equivalent to the marking itself, and could thus fully replace the marking without affecting the behaviour of the net. Only in *well-defined GSPN*s significant and tangible, and vanishing and insignificant coincide respectively, as stated in the following proposition.

Proposition 1 (Preservation). *If G is a* well-defined *GSPN, then a marking m of G is tangible if and only if it is significant.*

Furthermore, the CTMC associated with a well defined *GSPN* enjoys a strong relation to the original net in terms of the MA semantics:

Proposition 2. *The basic MA semantics A_G of a well-defined GSPN G is weakly bisimilar to the CTMC C_G induced by G.*

Before we present the proof of this proposition, an auxiliary notation and a claim is needed. Throughout, it is worthwhile to recall that the states of A_G and C_G are markings of G. If G is well-defined, for every state $m \in RS$, and every pair (μ, μ') of distributions over *tangible states* it holds: $m \Longrightarrow \mu$ and $m \Longrightarrow \mu'$ implies $\mu = \mu'$. Thus, for an arbitrary distribution γ, we may write $\gamma{\models\!\Longrightarrow}\mu$ to express that μ is the unique distribution over tangible states such that $\gamma \Longrightarrow \mu$.

Claim. Let G be a well-defined *GSPN*. Then for every distribution γ and γ' over states of the basic *MA* semantics of G, it holds that $\gamma \Longrightarrow \gamma'$ implies $\gamma{\models\!\Longrightarrow}\mu$ if and only if $\gamma'{\models\!\Longrightarrow}\mu$.

This follows immediately from the uniqueness of μ.

Proof (Proposition 2). In order to prove $A_G \approx C_G$, we will provide a bisimulation $\mathcal{R}$ and show that the pair of initial distributions of A_G and C_G is contained in $\mathcal{R}$. Let S_t be the state space of C_G, the set of all reachable tangible markings of A_G. Recall that the state space of A_G is the set RS of all reachable markings. Let $\mathcal{R}$ be the symmetric closure of the relation $\{(\gamma, \mu) \in \mathit{Dist}(RS) \times \mathit{Dist}(S_t) \mid \gamma{\models\!\Longrightarrow}\mu\}$. The pair of initial distributions of A_G and C_G is contained in $\mathcal{R}$, which follows immediately from the definition of the initial distribution of C_G.

Recall that in C_G we have an edge $m \overset{r}{\multimap} \mu$ if and only if μ is the unique distribution over tangible markings such that a distribution μ' exists with $m \overset{r}{\multimap} \mu'$ and $\mu' \Longrightarrow \mu$ in the basic *MA* semantics of G. We will refer to this fact by $(\star)$ whenever used in the sequel.

We will now check that every pair of $\mathcal{R}$ satisfies the bisimulation conditions. Consider an arbitrary pair $(\gamma, \mu) \in \mathcal{R} \cap \mathit{Dist}(RS) \times \mathit{Dist}(S_t)$. Clearly $|\gamma| = |\mu|$, as $\gamma \Longrightarrow \mu$. Now consider an arbitrary state (i.e. marking) $s \in \mathit{Supp}(\gamma)$. By the definition of hyperedges and of $\models\!\Longrightarrow$ it is easy to see that there exists a splitting $\mu^{\rightarrow} \oplus \mu^{\Delta} = \mu$, such that $\delta_s{\models\!\Longrightarrow}\mu^{\rightarrow}$ and $\mu^{\Delta}{\models\!\Longrightarrow}\mu^{\Delta}$, which immediately implies $\gamma(s)\delta_s \ \mathcal{R} \ \mu^{\rightarrow}$ and $\gamma \ominus s \ \mathcal{R} \ \mu^{\Delta}$. This satisfies Clause (i) of Definition 10. Now assume $s \longrightarrow \gamma'$. Then, by Claim 4.1, we see that $\gamma(s)\gamma'{\models\!\Longrightarrow}\mu^{\rightarrow}$ and thus immediately $\gamma(s)\gamma' \ \mathcal{R} \ \mu^{\rightarrow}$. Now assume $s \overset{r}{\longrightarrow} \gamma'$. Note that this implies that s is tangible, and thus $\mu^{\rightarrow} = \gamma(s)\delta_s$. But then by $(\star)$ the result follows. This finishes Clause (ii).

Now, for the symmetric case, consider an arbitrary pair $(\mu, \gamma) \in \mathcal{R} \cap \mathit{Dist}(S_t) \times \mathit{Dist}(RS)$, and let $t \in \mathit{Supp}(\mu)$. From the definition of $\mathcal{R}$ it follows that $\gamma{\models\!\Longrightarrow}\mu$ and thus $\gamma \Longrightarrow \mu$. Hence, $(\mu(t)\delta_t, \mu \ominus t) \in \mathit{split}(\gamma)$. We then choose $\gamma^{\rightarrow} = \mu(t)\delta_t$ and $\gamma^{\Delta} = \mu \ominus t$. Then for Clause (i) it suffices to note that $\mu(t)\delta_t \, \mathcal{R} \, \mu(t)\delta_t$ and $\mu \ominus t \, \mathcal{R} \, \mu \ominus t$, as for arbitrary distributions ξ over tangible states we have $\xi{\models\!\Longrightarrow}\xi$. For Clause (ii), consider $t \overset{r}{\longrightarrow} \mu'$ in the *CTMC* C_G. Note that this is the only possible transition of t (if any), as t is tangible. But then by $(\star)$, also $t \Longrightarrow \mu'$ in A_G, and as before $\mu' \mathcal{R} \mu'$ follows.

Proposition 2 provides us with a kind of correctness criterion for the setup we presented. The MA weak bisimulation semantics indeed conservatively extends the classical semantics. Furthermore, many traditionally ill-defined and confused nets can still be related to a CTMC modulo weakly bisimilarity. This is linked to the fact that weak bisimilarity embodies the notion of lumpability, apart from immediate transition fusing.

4.2 Timeless traps

Cycles of immediate transitions are an intricate problem in classical *GSPN* theory, their circular firing is often called a timeless trap [9], see Fig 7(a) for an example. *GSPN*s with timeless traps are traditionally excluded from the analysis, basically because the firing precedence of immediate over timed transitions makes the system diverge on the cycle without letting time progress. This is an awkward phenomenon, related to Zeno computations. In our *MA* reformulation, timeless traps are represented as cycles in the *MA*, and as such do not pose specific semantic problems. Furthermore, weak bisimilarity is sensitive to cycles of immediate transitions, but only to those that cannot be escaped by firing an alternative immediate transition. This is due to a built-in fairness notion in the weak bisimulation semantics, (rooted in the inclusion of a tangibility check inside the definition of the abbreviation $\xrightarrow{\alpha}$). As a consequence, if a timeless trap can be left by firing a (finite sequence of) immediate transitions leading to a tangible marking, this is equivalent to a single immediate transition firing. This implies that the net

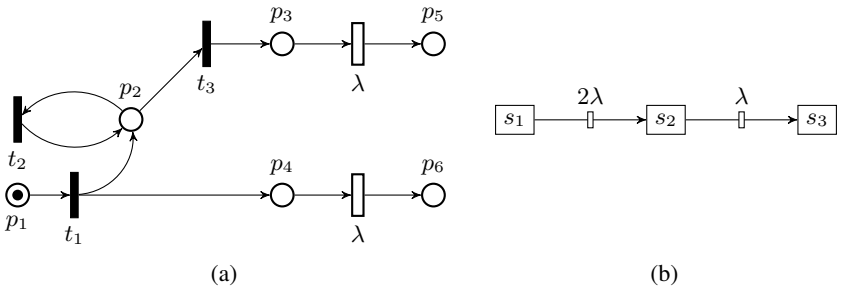

Fig. 7. A timeless trap that can be escaped by an immediate transition firing (a), and the smallest *MA* weakly bisimilar to its semantics (b). In (b), state s_1 subsumes markings $\{p_1\}$, $\{p_2, p_4\}$, and $\{p_3, p_4\}$. State s_2 subsumes markings $\{p_3, p_6\}$, and $\{p_4, p_5\}$, while state s_3 represents marking $\{p_5, p_6\}$.

in Fig. 7(a) is in fact weak bisimilar to the small chain-structured 3-state CTMC in Fig 7(b). And thus the net is analysable via the classical *CTMC* machinery. This example shows that the combination of lumping and fusing of immediate transitions as supported by weak bisimulation can have powerful effects. Variations to the definition of $\xrightarrow{\alpha}$ can induce more liberal notions of weak bisimiliarity, including the option to escape timeless traps unconditionally [27]. That option is not supported by the setup presented here, which has originally been designed to support strong compositionality properties [20]. Since compositionality is not a first-class concern in the Petri net world, this avenue seems worthwhile to be investigated further.

5 Quantitative Analysis of Markov Automata

So far, we have provided the details of a semantics of every definable *GSPN*. Thanks to Proposition 2, the steady-state and transient analysis of a well-defined *GSPN* under our semantics yields the same results as the evaluation of the induced *CTMC*. The remaining question is whether a quantitative analysis of a non well-defined *GSPN* is possible, and if so, how such analysis could be performed. Due to the possible presence of non-determinism, we can no longer consider *the* probability of a certain event. We stipulate that such probabilities depend on the resolution of non-determinism. Rather than considering, e.g., the probability to reach a state (i.e., a marking), it is common to determine the minimal and maximal reachability probabilities. These values correspond to the worst and best resolution of the non-determinism, respectively. Objectives that do not address the timing of net transitions, such as reachability, can be addressed using standard techniques for Markov decision processes (MDPs) such as linear programming, value, or policy iteration [5, Ch. 10]. Properties that involve the elapsed time are more interesting. In the following we briefly consider two such objectives: expected time and long run averages. For details we refer to [21] where Markov automata without probabilistic branching are considered. The inclusion of probabilistic branching however is rather straightforward. Long run average probabilities are the pendant to steady-state probabilities in *CTMC*s. Expected time objectives correspond to the expected time to reach a state in *CTMC*s. The counterpart to transient probabilities is a bit more involved and can be tackled using discretisation techniques advocated in [34,23].

In the following we let $A = (S, \multimap, \multimap\!\!\square, \mu^0)$ be an *MA*, $s \in S$ a state in A, and $G \subseteq S$ a set of (goal) states.

Expected time objectives. Starting from state s we are interested in the maximal, or dually, minimal, expected time to reach some state in G. Computing expected time objectives for *CTMC*s boils down to solving a linear equation system. The computation of minimal (or maximal) expected time objectives in *MA* can be reduced to a non-negative stochastic shortest path problem in MDPs [21]. Such problems can be casted as a linear programming problem [10] for which efficient algorithms and tools (such as SoPlex) exist.

Long-run average objectives. Intuitively speaking, the long-run average of being in a state in G while starting from state s is the fraction of time (on the long run) that the *MA* A will spent in states in G. We assume w.l.o.g. that G only contains tangible states, as the long-run average time spent in any vanishing state is zero. The general idea of computing the minimal long-run time spent in G is the following three-step procedure:

1. Determine the maximal end components[6] $\{A_1, \ldots, A_k\}$ of the MA at hand.
2. Determine the minimal long-run time spent in G within each end component A_j.
3. Solve a stochastic shortest path problem [10].

[6] A maximal end component is the analogue of a maximal strong component in the graph-theoretic sense, and is a standard notion for MDPs.

The first step is performed by a graph-based algorithm, whereas the last two steps boil down to solving linear programming problems. Determining the minimal expected long-run time in an end component can be reduced to a long-run ratio objective in an MDP equipped with two cost functions. Basically, it is the long-run ratio of the expected time of being in a state in G relative to the total expected time elapsed so far.

A prototypical implementation of our semantics is provided as part of the SCOOP tool, see:
`http://wwwhome.cs.utwente.nl/~timmer/scoop/webbased.html`.
This is based on translating *GSPN*s to an intermediate process-algebraic formalism [33] whose operational semantics yields Markov automata. The tool also supports expected time, timed reachability, and long-run analysis as described just above.

6 Conclusion

This paper has presented a semantics of *GSPN*s in terms of a non-deterministic variant of *CTMC*s, called Markov automata [20]. We have shown that for well-defined bounded *GSPN*s, our semantics is weak bisimulation equivalent to the *CTMC* semantics existing in the literature [8,13,4,3]. This "backward compatibility" result intuitively means that our semantics is the same as the classical *GSPN* semantics up to an equivalence that preserves all quantitative measures of interest such as transient, steady-state probabilities and CSL (without next) formulae [6]. Thus, any tool based on our *MA*-semantics yields for well-defined bounded nets the same results as popular *GSPN* tools such as GreatSPN, SMART, and MARCIE.

The main contribution of this paper is that our semantics applies to *every GSPN*. That is to say, our semantic framework is not restricted to well-specified or confusion-free nets. The key to treating confused nets is (not surprisingly) the use of non-determinism. We claim that our approach can also be applied to other stochastic net formalisms such as SANs [28,30].

The semantics closes a gap in the formal treatment of *GSPN*s, which is now no longer restricted to well-defined nets. This abandons the need for any check, either syntactically or semantically, for well-definedness. This gap was particularly disturbing because several published semantics for higher-level modelling formalisms—e.g., UML, AADL, WSDL—map onto *GSPN*s without ensuring the mapping to be free of confusion, thereby inducing ill-defined models. Our Markov automata semantics provides the basis to also cover the confused and ill-specified semantic fragments of these formalisms. Indeed, we were able to relax both notions by considering the Markov automata semantics modulo weak bisimulation. To proceed this way seemed like a natural way forward for quite some time to us, but to arrive there was an astonishingly difficult notational and technical endeavour.

Possible extensions. This paper does not consider the preservation (by the notion of weak bisimulation) of more detailed marking information such as the exact token occupancy of a place. Our notion of weak bisimulation is rather coarse and abstracts from this information. It is however straightforward to include this information by a simple extension of weak bisimulation that respects a certain state labelling, and this is fairly

routine [17,6]. The same is true for other reward structures—except rewards attached to immediate transitions, which are more involved to handle. The proof for "backward compatibility" of our semantics for unbounded (but e.g., finitely branching) *GSPN*s is left for further study.

References

1. M. Ajmone Marsan, G. Balbo, G. Chiola, and G. Conte. Generalized stochastic Petri nets revisited: Random switches and priorities. In *Petri Nets and Performance Models (PNPM)*, pages 44–53. IEEE CS Press, 1987.
2. M. Ajmone Marsan, G. Balbo, G. Chiola, G. Conte, S. Donatelli, and G. Franceschinis. An introduction to Generalized Stochastic Petri Nets. *Microel. and Rel.*, 31(4):699–725, 1991.
3. M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons, 1995.
4. M. Ajmone Marsan, G. Conte, and G. Balbo. A class of generalized stochastic Petri nets for the performance evaluation of multiprocessor systems. *ACM TOCS*, 2(2):93–122, 1984.
5. C. Baier and J.-P. Katoen. *Principles of Model Checking*. MIT Press, 2008.
6. C. Baier, J.-P. Katoen, H. Hermanns, V. Wolf, Comparative branching-time semantics for Markov chains, Inf. Comput 200 (2) (2005) 149–214.
7. G. Balbo. Introduction to stochastic Petri nets. In *Lectures on Formal Methods and Performance Analysis*, volume 2090 of *LNCS*, pages 84–155, 2001.
8. G. Balbo. Introduction to Generalized Stochastic Petri Nets. In *7th Int. School on Formal Methods for the Design of Computer, Communication, and Software Systems*, volume 4486 of *LNCS*, pages 83–131. Springer, 2007.
9. F. Bause. No way out ∞ The timeless trap. In *Petri Net Newsletter* 37:4-8, 1990.
10. D. P. Bertsekas and J. N. Tsitsiklis. An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16(3):580–595, 1991.
11. P. Buchholz. Exact and ordinary lumpability in finite Markov chains. *J. Applied Probability.*, 31:59–75, 1994.
12. G. Chiola, S. Donatelli, and G. Franceschinis. GSPNs versus SPNs: What is the actual role of immediate transitions? In *Petri Nets and Performance Models (PNPM)*, pages 20–31. IEEE CS Press, 1991.
13. G. Chiola, M. Ajmone Marsan, G. Balbo, and G. Conte. Generalized stochastic Petri nets: A definition at the net level and its implications. *IEEE TSE*, 19(2):89–107, 1993.
14. G. Ciardo, R. Zijal. Well-defined stochastic Petri nets. In *MASCOTS*, pages 278-284, 1996.
15. D. D. Deavours, W. H. Sanders. An efficient well-specified check. In *Petri Nets and Performance Models (PNPM)*, pages 124 – 133. IEEE CS Press, 1999.
16. Y. Deng, R. J. van Glabbeek, M. Hennessy, and C. Morgan. Testing finitary probabilistic processes. In *CONCUR*, pages 274–288, 2009.
17. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Weak bisimulation is sound and complete for PCTL*. *Inf. Comput.*, 208(2):203–219, 2010.
18. J. Hillston. *A Compositional Approach to Performance Modelling*. PhD thesis, University of Edinburgh, 1994.
19. C. Eisentraut, H. Hermanns, and L. Zhang. Concurrency and composition in a stochastic world. In *CONCUR*, volume 6269 of *LNCS*, pages 21 – 39. Springer, 2010.
20. C. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. In *LICS*, pages 342 – 351. IEEE, 2010.
21. D. Guck, T. Han, J.-P. Katoen, and M. R. Neuhäußer. Quantitative timed analysis of interactive Markov chains. In *NASA Formal Methods (NFM)*, volume 7226 of *LNCS*, pages 8–23. Springer, 2012.

22. H. Hermanns. *Interactive Markov Chains: The Quest for Quantified Quality*, volume 2428 of *LNCS*. Springer, 2002.
23. H. Hatefi and H. Hermanns. Model Checking Algorithms for Markov Automata. *ECEASST*, 53:1–15, 2012.
24. H. Hermanns, U. Herzog, V. Mertsiotakis, and M. Rettelbach. Exploiting stochastic process algebra achievements for Generalized Stochastic Petri Nets. In *Petri Nets and Performance Models (PNPM)*, pages 183–192, IEEE CS Press, 1997.
25. J.-P. Katoen. GSPNs revisited: Simple semantics and new analysis algorithms. In *ACSD*, pages 6–11. IEEE, 2012.
26. J. G. Kemeny, J. L. Snell, and A. W. Knapp. *Denumerable Markov Chains*. Springer, 2nd edition, 1976.
27. M. Lohrey, P. R. D'Argenio, and H. Hermanns. Axiomatising divergence. *Inf. Comput.*, 203(2):115–144, 2005.
28. J. F. Meyer, A. Movaghar, and W. H. Sanders. Stochastic activity networks: Structure, behavior, and application. In *Petri Nets and Performance Models (PNPM)*, pages 106–115. IEEE CS Press, 1985.
29. R. Milner. *Communication and Concurrency*. Prentice Hall, 1989.
30. W. H. Sanders, J. F. Meyer. Stochastic Activity Networks: Formal definitions and concepts. In *Lectures on Formal Methods and Performance Analysis (FMPA)*, volume 2090 of LNCS, pages 315 – 343. Springer, 2002.
31. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, Laboratory for Computer Science, Massachusetts Institute of Technology, 1995.
32. E. Teruel, G. Franceschinis and M. De Pierro. Well-defined Generalized Stochastic Petri Nets: A net-level method to specify priorities, *IEEE TSE*, 29(11):962–973, 2003.
33. M. Timmer, J.-P. Katoen, J. C. van de Pol, and M. I. A. Stoelinga. Efficient modelling and generation of Markov automata. In *CONCUR*, volume 7454 of *LNCS*, pages 364–379. Springer, 2012.
34. L. Zhang, M. R. Neuhäußer. Model checking interactive Markov chains. In *TACAS*, volume 6015 of *LNCS*, pages 53 – 68. Springer, 2010.