Exploiting Stochastic Process Algebra Achievements for Generalized Stochastic Petri Nets*

Holger Hermanns† Ulrich Herzog Vassilis Mertsiotakis Michael Rettelbach
Universität Erlangen-Nürnberg, IMMD VII, Martensstr. 3, 91058 Erlangen, Germany

Abstract

Constructing large Generalized Stochastic Petri Nets (GSPN) by hierarchical composition of smaller components is a promising way to cope with the complexity of the design process for models of real hardware and software systems. The composition of nets is inspired by process algebraic operators. A solid theoretical framework of such operators relies on equivalences that are substitutive *with respect to the operators. Practically important, such equivalences allow* compositional reduction *techniques, where components may be replaced by smaller, but equivalent nets without affecting significant properties of the whole model. However, substitutive equivalence notions for GSPN have not been published. In this paper we adopt operators and equivalences originally developed in the context of Stochastic Process Algebras to GSPN. The equivalences are indeed substitutive with respect to two composition operators, parallel composition and hiding. This bears the potential to exploit hierarchies in the model definition to obtain performance indices of truly large composite GSPN by stepwise compositional reduction. We illustrate the effect of composition as well as compositional reduction by means of a running example. A case study of a workstation cluster highlights the potential of compositional reduction.*

1 Introduction

Designing realistic Petri Net models of complex hardware and software systems is usually difficult and error prone. The nice visual representation typical of Petri Nets is not really helpful for complex nets. Apart from the size of the net itself, bulks of synchronisations between distinct parts typically obstruct the design process. Research on process algebras [20] has inspired efforts to introduce *compositionality* into Petri nets, a methodology where complex models are constructed in a stepwise fashion out of smaller building blocks [2, 21]. Such building blocks, nets themselves, are hierarchically superposed by means of composition operators, for example parallel composition. Concerning stochastic extensions of Petri Nets, some suggestions in this direction have recently been proposed. [5] considers composition in the context of Stochastic Petri Nets. Superposition of GSPNs has been considered in [9]. [25] investigates compositional construction of Stochastic well-formed nets (SWN) [7]. In [3, 10] composition operators on GSPN have been introduced as main constituents of a methodology to construct complex GSPN models.

The question arises whether it is possible to exploit hierarchies of such GSPN at the level of solution, i.e., performance analysis. We answer this question positively. A solid framework for exploiting hierarchies relies on *equivalences* that are *substitutive* with respect to the composition operators. Substitutivity implies that the behaviour of nets can be compared and that equivalently behaving nets are interchangeable when composed with other nets inside a complex model. Equivalence relations for non-stochastic Petri Nets that are substitutive with respect to certain composition operators have already become a valuable research topic [21]. These equivalence notions are usually inspired by process algebraic experiences and mainly based on bisimulations [20].

Recently, a notion of equivalence for Stochastic Petri Nets has been defined by Buchholz [5]. This equivalence is based on a bisimulation that was originally developed in the context of stochastic process algebras [4, 17, 12]. It is interesting to note that this relation coincides with lumpability [19] on the underlying Markov Chain (MC). Hence, partitioning the reachability graph with respect to this equivalence induces lumping of the MC. Moreover, since it is substitutive with respect to parallel composition, it can equally be applied to components of a hierarchical model in order to reduce their reachability graphs. This implies that the size of the state space of the hierarchical model is reduced as well, but with the additional gain that the original state space of the hierarchical model needs not be constructed. In the context of process algebras, this general technique is known as *compositional reduction* and has succesfully been applied to a variety of non-stochastic concurrent systems in order to perform verification of functional properties, see e.g. [6]. The only requirement is that the equivalence is substitutive with respect to composition. The development of compositional construction and reduction techniques for performance analysis needs is one of the main issues of Stochastic Process Algebras [11].

*This research is supported in part by the German National Research Council *DFG* under *SFB 182*

† Corresponding author, hrherman@informatik.uni-erlangen.de

This paper addresses the issue of composition and substitutive equivalences in the framework of GSPN. We introduce two composition operators, hiding and parallel composition, that are borrowed from the ISO specification language LOTOS [18] . Based on our previous work in the context of Stochastic Process Algebras [14, 22] we develop an equivalence relation for composite GSPN. After showing its substitutivity we discuss and formally prove how this equivalence supports compositional construction and reduction of the MC underlying a composite GSPN.

In the usual definition of GSPN [1], the association of a MC with the reachability graph is not always straightforward, because immediate transitions do not have a counterpart on the MC level. In some cases an unambiguous association is even impossible, due to so-called confusion [8]. The usual way for confusion free nets is to distinguish between *tangible* and *vanishing* markings and to calculate firing rates using weights or priorities associated to immediate transitions. Our approach to remove immediate transitions is different, nevertheless leading to the same resulting MC. We propose a compositional technique based on a substitutive equivalence. Immediate transitions can be eliminated from the net if they do not have an observable impact on functional as well as stochastic properties. For this purpose, the notion of observability [20], central for (Stochastic) Process Algebras, is extended to GSPN. Since nets may appear as components of other nets, the elimination of immediate transitions can be applied to components, as well. In this case, immediate transitions do not contribute to the reachability graph of the whole net, if they have been eliminated out of components. This is the main advantage of the approach. The reachability graph can be generated along the (parallel) structure of the model. When applying compositional reduction during this generation, immediate transitions are eliminated as soon as possible, eventually leading to a MC. This is particularly useful if the reachability graph of the whole net is far too large to be stored in memory, due to excessive occurrence of vanishing markings.

The paper is organised as follows. In Section 2 we give the definition of compositional GSPN and define composition operators on them. Section 3 deals with our substitutive equivalence, Markovian observational equivalence. In Section 4 we explain compositional generation of the MC. Section 5 highlights the usefulness of Markovian observational equivalence by means of a small case study. Section 6 concludes the paper.

2 Compositional GSPN

This section introduces *compositional generalized stochastic Petri Nets* (CN), and defines composition operators on them. A CN is composed out of GSPNs by superposition of immediate transitions. For this purpose, immediate transitions are labelled with names drawn from a set of (potentially) synchronizing names $Sync$. When composed in parallel, immediate transitions with the same name have to synchronize if their names appear in a list of (actually) synchronizing names $s_1, \ldots, s_n$. Only internal immediate transitions named τ will happen in total independence of the environment. The distinct name τ does not appear in $Sync$. In order to allow internalization of transitions we introduce a hiding operator that turns specific names into τ's. Immediate transitions serve a variety of purposes when modelling systems with GSPN. For the sake of simplicity, we focus on their role to model synchronization aspects. In particular we neither consider priorities among immediate transitions nor weights. Weights of immediate transitions usually resolve nondeterminism between immediate transitions. Nondeterminism in our approach can be resolved by appropriate synchronization due to composition with other nets.

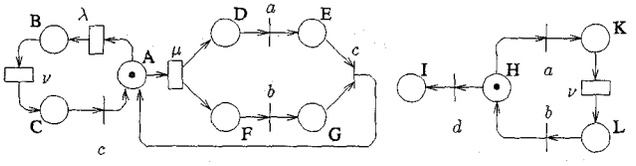


Figure 1: Two CN, N_1 and N_2

Definition 2.1 *A CN is a 7-tuple* $(P, M, I, F, \Lambda, L, q)$, *where*

- P *is a non empty set of places,*
- M *is a set of Markovian transitions,*
- I *is a set of immediate transitions,*
- $F \subseteq \big(P \times (M \cup I)\big) \cup \big((M \cup I) \times P\big)$ *is the flow relation.*
- $\Lambda : M \mapsto]0, \infty[$, *where* $\Lambda(m)$ *is the rate of the exponential probability distribution associated with the Markovian transition* m,
- $L : I \mapsto Sync \cup \{\tau\}$, *where* $L(i)$ *is the name of the immediate transition* i.
- $q : P \mapsto I\!N$ *is the initial marking.*

Figure 1 presents two examples of CN. We have labelled each transition t with its respective rate $\Lambda(t)$ (if $t \in M$) or name $L(t)$ (if $t \in I$) instead of its identifier t. Uppercase letters are used to range over places $p \in P$ in our pictorial representations of CN.

If $I = \emptyset$ the definition above leads to the class of ordinary Stochastic Petri Nets (SPN). We now recapitulate the usual definitions of reachability set (RS) and reachability graph (RG).

Definition 2.2 *The reachability set RS_N of a CN N is the smallest set of markings containing the initial marking and satisfying* $(q' \in RS_N \wedge \exists t \in (M_N \cup I_N) : q'\,[t\rangle\,q'') \Rightarrow q'' \in RS_N$. *In this definition $q'\,[t\rangle\,q''$ indicates that the system changes from marking q' to marking q'' due to the firing of t.*

The reachability graph RG_N of a CN N is a labelled directed multigraph whose set of nodes is RS_N and whose arcs $A \subseteq RS_N \times (M_N \cup I_N) \times RS_N$ are given by $(q', t, q'') \in A \iff q'\,[t\rangle\,q''$.

In the sequel, the subscript N is omitted if clear from the context. We shall indicate with $q\,[t\rangle$ that $\exists q' \in RS : q\,[t\rangle\,q'$. We call a marking q *vanishing*, iff $\exists i \in I : q\,[i\rangle$, otherwise we call it *tangible*. Presets (and postsets) of transitions t are defined as usual, ${}^\bullet t := \{p \in P \mid (p,t) \in F\}$ ($t^\bullet := \{p \in P \mid (t,p) \in F\}$).

Typically, the stochastic process associated to a GSPN N is obtained by only considering tangible markings. Immediate firing sequences are amalgamated with preceding Markovian delays using weights of immediate transitions, see [1] for the details. We deviate from this treatment in order to achieve compositionality. The stochastic process associated with a CN N is not determined using weights of immediate transitions. In contrast, it is regarded to be dependent on names of immediate transitions, since they govern the synchronization of N when composed with other nets. Only if synchronization is made impossible due to hiding, immediate firing sequences can be eliminated, eventually leading to a fully determined stochastic process.

The concrete difference is that only *internal* immediate transitions cause that markings which enable them vanish, whereas other immediate transitions do not have this unrestricted implication. For this purpose, we use the notion of stability. We call a marking q *stable*, iff no transition $i \in I$ exists with $q\,[i\rangle \wedge L(i) = \tau$, otherwise we call it *unstable*. The view is taken that only internal immediate transitions have priority over Markovian transitions. In the presence of synchronization, this distinction pays off.

We are now ready to introduce composition operators that allow for a hierarchical specification of CN. *Parallel composition* is a LOTOS-style operator where immediate transitions whose names appear in a list $s_1, \ldots, s_n$ have to synchronize. Transitions that are not labelled with an element of this list can fire independently of the transitions of the other net. We refer to CN that are built out of smaller nets using the composition operators defined below as *composite* GSPN (cN). Nets appearing as components of a given cN are refered to as *net components* (Nc).

Definition 2.3 *Let* $N_1 = (P_1, M_1, I_1, F_1, \Lambda_1, L_1, q_1)$ *and* $N_2 = (P_2, M_2, I_2, F_2, \Lambda_2, L_2, q_2)$ *be two CN with* $P_1 \cap P_2 = \emptyset$, $M_1 \cap M_2 = \emptyset$ *and* $I_1 \cap I_2 = \emptyset$. *Let* $S = \{s_1, \ldots, s_n\} \subseteq Sync$ *be a set of names and* $\tilde{I}_1 = \{i_1 \in I_1 \mid L_1(i_1) \in S\}$, $\tilde{I}_2 = \{i_2 \in I_2 \mid L_2(i_2) \in S\}$ *and* $\tilde{I} = \{(i_1, i_2) \in (\tilde{I}_1 \times \tilde{I}_2) \mid L_1(i_1) = L_2(i_2)\}$.

The parallel composition of N_1 and N_2 via a list of names $s_1, \ldots, s_n$, denoted $N_1 \overset{s_1,\ldots,s_n}{\|} N_2$, *is the composite GSPN* $(P, M, I, F, \Lambda, L, q)$, *where*

- $P = P_1 \cup P_2$ *and* $M = M_1 \cup M_2$,
- $I = (I_1 - \tilde{I}_1) \cup (I_2 - \tilde{I}_2) \cup \tilde{I}$,
- $F = F_1 \cap (P_1 \times (I_1 - \tilde{I}_1)) \cup F_2 \cap (P_2 \times (I_2 - \tilde{I}_2)) \cup \{(p, (i_1, i_2)) \in P \times \tilde{I} \mid p \in {}^\bullet i_1 \vee p \in {}^\bullet i_2\} \cup \{((i_1, i_2), p) \in \tilde{I} \times P \mid p \in i_1^\bullet \vee p \in i_2^\bullet\}$
- $\Lambda = \Lambda_1 \cup \Lambda_2$ *and* $q = q_1 \cup q_2$,
- $L(i) = \begin{cases} L_1(i) & \text{if } i \in I_1 - \tilde{I}_1, \\ L_2(i) & \text{if } i \in I_2 - \tilde{I}_2, \\ L_1(i_1) & \text{if } i = (i_1, i_2) \in \tilde{I}. \end{cases}$

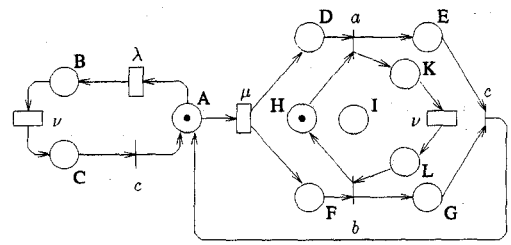


Figure 2: Composite GSPN $N_1 \overset{a,b,d}{\|} N_2$

As an example consider the net in Figure 2. It is a composite GSPN $N_1 \overset{a,b,d}{\|} N_2$ of the nets in Figure 1 synchronized over a, b and d. Note that it is possible to cut off transitions by means of synchronization. This has happened to the transition named d of N_2, since synchronization on d is forced, but no partner transition in N_1 exists.

Definition 2.4 *Let* $N = (P, M, I, F, \Lambda, L, q)$ *be a CN. Hiding of names* $h_1, \ldots, h_n$, *denoted* $N \setminus \{h_1, \ldots, h_n\}$, *results in the composite GSPN* $(P, M, I, F, \Lambda, L', q)$, *where* $L'(i) = L(i)$ *for all* $i \notin \{h_1, \ldots, h_n\}$, *and* $L'(i) = \tau$ *for all* $i \in \{h_1, \ldots, h_n\}$.

Lemma 2.1 *CN are closed under parallel composition and hiding.*

Hence, composite GSPN are compositional GSPN and may appear as components of larger composite nets. An example is $(N_1 \overset{a,b,d}{\|} N_2) \setminus \{a,b\}$, where N_1 and N_2 are the CN from above. It is depicted in Figure 3.

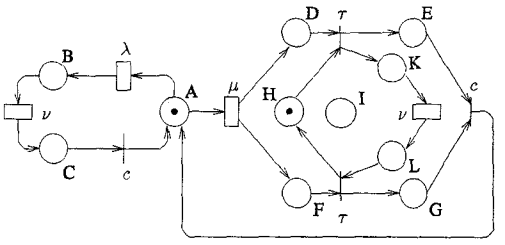
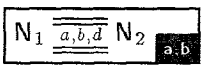


Figure 3: Composite GSPN 

3 Equivalence Notions

In this section we adopt the equivalences of [14] defined in the context of stochastic process algebra to compositional GSPN. We first define an equivalence notion for purely Stochastic Petri Nets, Markovian bisimilarity. This relation is the starting point for *Markovian observational equivalence*, a relation introduced to compare CN. We characterise it as the fixed-point of successively finer relations and show how to partition the reachability graph into classes of equivalent markings using this characterization. This partitioning will be exploited in Section 4 for compositional generation of the reachability graph.

First, we define a function that computes the sum of rates of all Markovian firings enabled in a marking q', and leading to any marking in a given set.

Definition 3.1 *For $q' \in M$ and $B \subseteq M$ let $E(q', B) := \{m \in M \mid q' \, [m\rangle \, q'' \wedge q'' \in B\}$. We define $\gamma^M : RS_N \times 2^{RS_N} \mapsto]0, \infty[$ as $\gamma^M(q', B) := \sum_{m \in E(q', B)} \Lambda(m)$.*

Definition 3.2 *Let $N = (P, M, \emptyset, F, \Lambda, L, q)$ be a SPN. An equivalence relation $\mathcal{B} \subseteq RS_N \times RS_N$ is a* Markovian bisimulation, *if $(q', q'') \in \mathcal{B}$ implies for all equivalence classes*[1] *$C \in RS_N/_{\mathcal{B}}$ that $\gamma^M(q', C) = \gamma^M(q'', C)$.*

Two markings q' and q'' are Markovian bisimilar, *written $q' \sim q''$, if $(q', q'') \in \mathcal{B}$ for some Markovian bisimulation $\mathcal{B}$.*

This notion of equivalence is the stochastic counterpart of Milners strong bisimilarity [12]. It has been adopted to SPN in [5]. On the level of the Markov Chain $\sim$ induces a *lumpable* partition [17]. The effect on the MC is comparable with symmetry exploitation in SWN. However, the general power of lumping based on Markovian bisimilarity goes beyond what is known for symmetry exploitation [23].

We now turn our attention towards CN where names of immediate transitions are used for synchronisation purposes. In order to keep track of the impact of immediate transitions on the behaviour of the system, we distinguish between observable and unobservable firings of transitions. In particular, firing of transitions named τ is not observable. The notion of an observer is due to Milner, who defined the notion of *observational equivalence* in his Calculus of Communicating Systems [20]. An observer synchronizes with a net by means of its observable firings in a specific pattern. Roughly speaking, two nets are equivalent, if no observer is able to distinguish them.

For $m \in M$ we shall indicate with $q' \, [\underline{\lambda}\rangle \, q''$ that $q' \, [m\rangle \, q'' \wedge \Lambda(m) = \lambda$. Similarly, if $i \in I$ then $q' \, [\underline{l}\rangle \, q''$ abbreviates that $q' \, [i\rangle \, q'' \wedge L(i) = l$. We extend this definition to words of $(Sync \cup \{\tau\} \cup \,]0, \infty[)^*$ in the usual way. For example, $q' \, [\underline{0.2 \; a \; \tau \; 5}\rangle \, q''$ is a shorthand for $\exists q_1, q_2, q_3 : q' \, [\underline{0.2}\rangle \, q_1 \, [\underline{a}\rangle \, q_2 \, [\underline{\tau}\rangle \, q_3 \, [\underline{5}\rangle \, q''$. Furthermore, $q' \, (\underline{\tau}\rangle \, q''$ abbreviates $q' \, [\underline{\tau^*}\rangle \, q''$ and $q' \, (\underline{t}\rangle \, q''$ abbreviates $q' \, [\underline{\tau^* \; t \; \tau^*}\rangle \, q''$ if $t \neq \tau$. Note that $q' \, (\underline{t}\rangle \, q''$ denotes that there is a firing sequence with exactly one observable transition involved, whereas $q' \, (\underline{\tau}\rangle \, q''$ includes the case that no unobservable firing has to take place at all.

With these definitions, we are ready to define a variant of observational equivalence using observable firing sequences.

Definition 3.3 *Let $N = (P, M, I, F, \Lambda, L, q)$ be a CN. An equivalence relation $\mathcal{B} \subseteq RS_N \times RS_N$ is a* weak Markovian bisimulation, *if $(q', q'') \in \mathcal{B}$ implies for all equivalence classes $C \in RS_N/_{\mathcal{B}}$:*
(i) For all immediate transitions $i \in I$ it holds that whenever $q' \, (\underline{L(i)}\rangle \, \hat{q}'$ then, for some $\hat{q}'' \in RS_N$, $q'' \, (\underline{L(i)}\rangle \, \hat{q}''$ and $(\hat{q}', \hat{q}'') \in \mathcal{B}$.
(ii) For all stable $\hat{q}' \in RS_N$, it holds that if $q' \, (\underline{\tau}\rangle \, \hat{q}'$ then, for some stable $\hat{q}'' \in RS_N$, $q'' \, (\underline{\tau}\rangle \, \hat{q}''$ and $\gamma^M(\hat{q}', C) = \gamma^M(\hat{q}'', C)$.

Two markings are Markovian observational equivalent, *written $q' \approx q''$, if $(q', q'') \in \mathcal{B}$ for some weak Markovian bisimulation $\mathcal{B}$.*

For SPN all markings are stable. Hence, only condition *(ii)* must be satisfied. Taking into account that we can choose q'' for $\hat{q}''$, the above definition obviously coincides with Markovian bisimulation equivalence, when applied to SPN. On the other hand, for nets without Markovian transitions merely condition *(i)* has to be checked. In this case, the above equivalence is exactly Milners observational equivalence as defined e.g. in [21] for labelled Petri Nets. So, Markovian observational equivalence merges observational equivalence for immediate transitions with Markovian bisimulation for Markovian transitions. Because of the priority of *internal* immediate transitions, the latter need not be checked for unstable markings (condition *(ii)*). Furthermore, if q' is stable, it is sufficient that q'' can invisibly and immediately descend to a stable marking $\hat{q}''$ that is Markovian bisimulation equivalent to q'.

[1] If $\mathcal{S}$ is an equivalence relation on $\mathcal{X}$ we denote by $\mathcal{X}/_{\mathcal{S}}$ the set of equivalence classes of $\mathcal{X}$ induced by $\mathcal{S}$.

Markovian observational equivalence is a weak Markovian bisimulation (the largest, indeed). It can be characterized as a fixed-point of successively finer relations. This characterization is the skeleton of an algorithmic computation of $\approx$ on a given reachability graph.

Proposition 3.1 *For a CN N with finite reachability set, let $\mathcal{R}_k \subseteq RS_N \times RS_N$ be a family of equivalence relations such that*

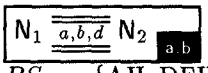
- $\mathcal{R}_0 = RS_N \times RS_N$
- $\mathcal{R}_{k+1} = \mathcal{R}_k \cap E_k$ *where* $(q', q'') \in E_k$ *if*

  (i) For every immediate transitions $i \in I$ it holds that whenever q' $(\underline{L(i)})$ $\hat{q}'$ then, for some $\hat{q}'' \in RS_N$, q'' $(\underline{L(i)})$ $\hat{q}''$ and $(\hat{q}', \hat{q}'') \in \mathcal{R}_k$.

  (ii) For every $C \in RS_N/\mathcal{R}_k$ and all stable $\hat{q}' \in RS_N$ it holds that if q' $(\underline{\tau})$ $\hat{q}'$ then for some stable $\hat{q}'' \in RS_N$, q'' $(\underline{\tau})$ $\hat{q}''$ and $\gamma^M(\hat{q}', C) = \gamma^M(\hat{q}'', C)$.

Markovian observational equivalence is the unique fixed-point of this family of relations.

Proof: We have to show (1) that a unique fixed-point exists, and (2) that $\approx$ is a fixed-point. (1) is clear, since $\mathcal{R}_0$ is finite and unique, and $|\mathcal{R}_k| > |\mathcal{R}_{k+1}|$ until $\mathcal{R}_k = \mathcal{R}_{k+1}$. To verify condition (2), observe that for each weak Markovian bisimulation $\mathcal{B}$, $\mathcal{R}_k \supseteq \mathcal{B}$ holds for arbitrary k, because *(i)* and *(ii)* are simply reformulations of Definition 3.3, whence $\mathcal{R}_k \supseteq \approx$, in particular $\mathcal{R}_n \supseteq \approx$. The opposite direction, $\mathcal{R}_n \subseteq \approx$ follows from the observation that $\mathcal{R}_n$ is a weak Markovian bisimulation itself. □

In particular there is a finite number n such that $\mathcal{R}_n = \mathcal{R}_{n+i} = \approx$ for all $i > 0$. We use this characterization to compute a partition of the RG by stepwise refinement. As an example consider the graph depicted in Figure 4. It is the reachability graph of the composite GSPN 

from Figure 3. The reachability set is $RS = \{\mathrm{AH}, \mathrm{DFH}, \mathrm{EFK}, \mathrm{EFL}, \mathrm{EGH}, \mathrm{BH}, \mathrm{CH}\}$. Deviating from Definition 2.2 we have annotated the arcs with their respective rates or names instead of using transition identifiers from $(M \cup I)$.

We compute a partition of the graph starting with $RS \times RS = \mathcal{R}_0$. To compute $\mathcal{R}_1$ we have to check for each pair of $\mathcal{R}_0$ whether or not condition *(i)* and *(ii)* are satisfied. This procedure will split $\mathcal{R}_0$ into three partitions $\{\mathrm{DFH}, \mathrm{EFK}, \mathrm{BH}\} \times \{\mathrm{DFH}, \mathrm{EFK}, \mathrm{BH}\}$, $\{\mathrm{EFL}, \mathrm{EGH}, \mathrm{CH}\} \times \{\mathrm{EFL}, \mathrm{EGH}, \mathrm{CH}\}$ and $\{(\mathrm{AH}, \mathrm{AH})\}$, indicated with dotted lines. As an example we consider a representative pair of this refinement $\mathcal{R}_1$. To check that $(\mathrm{BH}, \mathrm{DFH}) \in \mathcal{R}_1$ we verify condition *(ii)* using DFH $(\underline{\tau})$ EFK and $\gamma^M(\mathrm{BH}, RS) = \nu = \gamma^M(\mathrm{EFK}, RS)$.

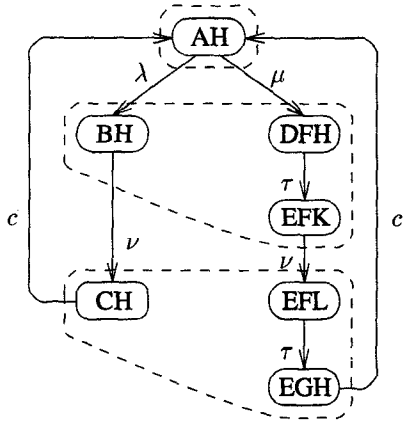
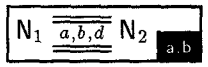


Figure 4: Reachability graph of 

The other elements of $\mathcal{R}_1$ result from similar reasoning. The next step yields $\mathcal{R}_2$, which is identical with $\mathcal{R}_1$. We have reached the fixed-point and hence computed Markovian observational equivalence on this RG.

The fact that $\approx$ extends both Markovian bisimulation and observational equivalence has important implications. First, partitioning the reachability graph induces a *lumpable* partition, since $\approx$ inherits this property from $\sim$. In addition, immediate firing sequences between equivalent markings can be eliminated from the graph if they are unobservable (see Proposition 4.2). The combination of both properties induces that BH and DFH are equivalent in the above example.

We exploit these features by factorizing the reachability graph with respect to Markovian bisimulation equivalence. An observable reachability graph (ORG) is obtained by representing each equivalence class as a macro state, together with appropriate edges between macro states for edges in RG that cross the border of equivalence classes.

Definition 3.4 *The observable reachability graph ORG_N of a CN N is a labelled directed graph whose set of nodes is $ORS_N = RS_N/\approx$ and whose set of arcs $A \subseteq ORS_N \times (Sync \cup \{\tau\} \cup \,]0,\infty[) \times ORS_N$ is*
$\{(Q', \gamma^M(q', Q''), Q'') \mid q'' \in Q' \wedge Q' \neq Q'' \wedge q''\textit{stable}\} \cup$
$\{(Q', a, Q'') \mid a \neq \tau \wedge \exists q' \in Q', q'' \in Q'' : q' \ [\underline{a}\rangle \ q''\} \cup$
$\{(Q', \tau, Q'') \mid Q' \neq Q'' \wedge \exists q' \in Q', q'' \in Q'' : q' \ [\underline{\tau}\rangle \ q''\}$.

Remark that existential and universal quantifiers are interchangable in this definition, because all markings in a partition are equivalent. The observable reachability graph corresponding to Figure 4 is depicted in Figure 5.

Markovian observational equivalence is defined on the markings of a single CN. In general we want to compare nets. So, we lift this definition to pairs of CN. Two nets are equivalent iff their reachability graphs are Markovian

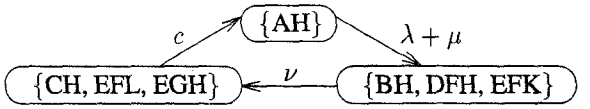


Figure 5: Observable RG of $N_1 \overset{a,b,d}{\approx} N_2$ a.b

observational equivalent. This obviously holds iff their observable reachability graphs are isomorphic.

Definition 3.5 *Two CN* $N_1 = (P_1, M_1, I_1, F_1, \Lambda_1, L_1, q_1)$ *and* $N_2 = (P_2, M_2, I_2, F_2, \Lambda_2, L_2, q_2)$ *are Markovian observational equivalent, if an isomorphism* $\phi : ORS_{N_1} \mapsto ORS_{N_2}$ *exists such that*

$$q_1 \in Q \iff q_2 \in \phi(Q) \quad and \quad \phi(ORG_{N_1}) = ORG_{N_2}$$

Following this definition, the net N_3 depicted in Figure 6 is Markovian observational equivalent to the composite net of Figure 3.

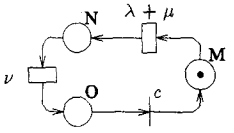


Figure 6: Net N_3**, equivalent to** $N_1 \overset{a,b,d}{\approx} N_2$ a.b

Lemma 3.1 $N_1 \approx N_2$, *iff a weak Markovian bisimulation* $\mathcal{B}$ *on* $RG_{N_1} \cup RG_{N_2}$ *exists with* $(q_1, q_2) \in \mathcal{B}$.

4 Compositional Reduction

Typical application of compositionality result in large cN that are constructed using a hierarchy of instances of parallel composition and hiding. If the reachability graph of such net models is constructed at once we typically observe state space explosion. In order to avoid this, we propose *compositional* reduction. The key idea is to generate the reachability graph in a stepwise fashion along the parallel structure of a composite GSPN. In every step, the composite net is reduced without touching the behavioural properties of the net. The reduction performs lumping and elimination of internal immediate transitions.

Substitutivity of $\approx$ with respect to the composition operators is crucial in order to perform this compositional reduction techniques. Substitutivity is also known as the *congruence* property. Intuitively, it allows to replace components by equivalent ones inside a composite GSPN without affecting its observable behaviour, specifically performance indices. The nets in Figure 3 and 6 are equivalent. However, without being assured that $\approx$ is substitutive we can not expect to get the same results when using the latter instead of the former inside a cN.

Proposition 4.1 *Let* N_1, N_2 *and* N_3 *be CN. Markovian observational equivalence is substitutive with respect to parallel composition and hiding, i.e.*
$N_1 \approx N_2$ *implies* $N_1 \overset{s_1,\ldots,s_n}{\approx} N_3 \approx N_2 \overset{s_1,\ldots,s_n}{\approx} N_3$,
$N_1 \approx N_2$ *implies* $N_3 \overset{s_1,\ldots,s_n}{\approx} N_1 \approx N_3 \overset{s_1,\ldots,s_n}{\approx} N_2$,
$N_1 \approx N_2$ *implies* N_1 $h_1,\ldots,h_n$ $\approx$ N_2 $h_1,\ldots,h_n$.

Proof *(sketch)*: In order to show the first two implications, let $s_1, \ldots, s_n$ be a list of names, N_1, N_2, N_3 be CN with initial markings q_1, q_2, q_3 such that $N_1 \approx N_2$. Lemma 3.1 implies that a weak Markovian bisimulation $\mathcal{B}$ exists on $RG_{N_1} \cup RG_{N2}$ that contains (q_1, q_2). We have to show that $N_4 \approx N_5$, where N_4 (N_5, respectively) denotes $N_1 \overset{s_1,\ldots,s_n}{\approx} N_3$ ($N_2 \overset{s_1,\ldots,s_n}{\approx} N_3$). Let $\mathcal{U}$ abbreviate $RG_{N_4} \cup RG_{N_5}$. Using Lemma 3.1 it is sufficient to show that a weak Markovian bisimulation $\mathcal{B}' \subseteq \mathcal{U} \times \mathcal{U}$ exists that includes the pair of initial markings $(q_1 \cup q_3, q_2 \cup q_3)$. The relation $\mathcal{B}' := (\{(q' \cup q''', q'' \cup q''') \mid q' \mathcal{B} q'' \wedge q''' \in RG_{N_3}\} \cap (\mathcal{U} \times \mathcal{U})) \cup Id_{\mathcal{U}}$ contains this pair, where $Id_{\mathcal{U}}$ is the identity relation on $\mathcal{U}$. $\mathcal{U}$ is an equivalence relation, because $\mathcal{B}$ is transitive and symmetric. In addition, it can be proven that $\mathcal{B}'$ satisfies *(i)* and *(ii)* of Definition 3.3. Verifying the second implication is completely analogous, since parallel composition (Definition 2.3) is symmetric. The proof of the third implication is similar. □

In order to show condition *(ii)* of Definition 3.3 it is crucial that not only tangible markings are considered, but also *stable* vanishing markings. Vanishing markings that do not enable *internal* immediate transitions potentially result in tangible markings in the presence of synchronisation. This is due to the fact that all emanating immediate transitions might be victims of failed synchronisation in a larger context that cuts them off.

Proposition 4.1 is the formal justification that we can apply compositional reduction to composite nets. Consider the following cN N, composed from two instances[2] of the composite net of Figure 2.

$$N = \left[N_1 \overset{a,b,d}{\approx} N_2 \text{ a.b} \right] \overset{c}{=} \left[N_1 \overset{a,b,d}{\approx} N_2 \text{ a.b} \right] \text{ c}$$

It consists of 20 places and 14 transitions, not shown due to space constraints. Instead of analyzing the reachability graph of the whole cN, which has 33 reachable markings,

[2] Definition 2.3 requires that the sets of places and transition are disjoint for Nc of a parallel composition. N violates this condition. However, we can assume that all place and transition identifiers of one of the Nc $N_1 \overset{a,b,d}{\approx} N_2$ of N are appropriately modified. The same is true for the Nc of $N_3 \overset{c}{=} N_3$, appearing below.

we exploit subsitutivity. From $\boxed{\mathsf{N}_1 \overset{a,b,d}{\Rightarrow} \mathsf{N}_2}_{a.b} \approx \mathsf{N}_3$ (cf. Figure 6) it follows with Proposition 4.1 that $\mathsf{N} \approx \boxed{\mathsf{N}_3 \overset{c}{\Rightarrow} \mathsf{N}_3}_{c}$. Again, the reachability graph of $\boxed{\mathsf{N}_3 \overset{c}{\Rightarrow} \mathsf{N}_3}_{c}$ (9 markings) can be partitioned following Proposition 3.1 leading to an ORG that only contains tangible markings. It gives rise to an equivalent net N_4, depicted in Figure 7.

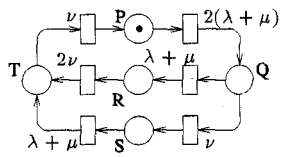


Figure 7: Net N_4, equivalent to N

The analysis of the underlying MC is straightforward. The reduction step from $\boxed{\mathsf{N}_3 \overset{c}{\Rightarrow} \mathsf{N}_3}_{c}$ to N_4 mainly exploits symmetries, similar to SWN decolorization. Note, however, that the general power of Markovian observational equivalence goes beyond what is known for symmetry exploitation. This is exemplified in the reduction step that lead to N_3, since $\boxed{\mathsf{N}_1 \overset{a,b,d}{\Rightarrow} \mathsf{N}_2}_{a.b}$ (Figure 3) bears no obvious symmetries.

The concrete interrelation between the Markov chain of N_4 and that of the whole cN N is as follows: The former is a lumped version of the latter. This is a general property of $\approx$. The MC of N can be derived from RG_{N} by means of the standard algorithm [1]. It is independent from the choice of weights or priorities among immediate transitions.

Proposition 4.2 *Let N_1 and N_2 be CN such that $I_1 \neq \emptyset$ and $I_2 = \emptyset$. If $N_1 \approx N_2$ then ORG_{N_2} is isomorphic to a lumped version of the MC associated to N_1.*

Proof *(sketch)*: From $N_1 \approx N_2$ it follows that $\forall i \in I_1 : L(i) = \tau$. We only consider the case that $I_1 = \{i', i''\}$ and that there is exactly one $q \in RS_{N_1}$ such that $q\ [i'\rangle\ q' \wedge q\ [i''\rangle\ q''$. This is the crucial situation where usually weights (or priorities) have to be taken into account. Let us assume that weights are assigned such that i' happens with probability π and i'' with $(1-\pi)$, respectively. Let Q denote the equivalence class of q, analoguous with Q' and Q''. Since $N_1 \approx N_2$ there is an isomorphism ϕ such that $\phi(ORG_{N_1}) = ORG_{N_2}$. Since $I_2 = \emptyset$ it is clear that $\phi(Q) = \phi(Q') = \phi(Q'')$, whence we have $q' \approx q''$ because ϕ is bijective. With respect to marking $q''' \in RS_{N_1}$ (contained in Q''') such that $q'''\ [m\rangle\ q$ and $\Lambda(m) = \lambda$ we have $(Q''', \lambda, Q) \in ORG_{N_1}$ (if $Q''' \neq Q$, otherwise the proof is simple).

The algorithm of [1] associates a MC to N_1 where state s''' has a pair of successor states s' and s'' that can be reached with rates of the form $\pi\lambda$ and $(1-\pi)\lambda$, respectively. The correspondence to q', q'' and q''' is clear. The *vanishing* marking q has no corresponding state on the MC level. It now follows from $q' \approx q''$ that lumping on the MC will unify s' and s'' [22] and that this lumped state is reachable from the lumped state containing s with the sum of the former rates, i.e. λ. This arc in the lumped MC can be mapped onto $(Q''', \lambda, Q) \in ORG_{N_1}$ and is the cornerstone of the construction of an isomorphism between ORG_{N_1} and the lumped MC of N_2. For arbitrary conflict situations between (sequences of) immediate transitions the proof follows similar lines. As in our special case, the lumped MC is generally independent of the actual assignment of weights to immediate transitions. □

As a whole, compositional reduction with $\approx$ may be applied to larger composite nets recursively. If finally the observable reachability graph does not contain immediate arcs, then Proposition 4.2 assures that we have obtained a lumped version of the MC of the original net. This lumped MC has been derived in a stepwise fashion out of the component's behaviour without constructing the whole reachability graph and even without constructing the whole net.

There are, however, some implicit restrictions to the method. First, all names must be hidden before an ORG without immediate arcs is attainable. This requirement is necessary but by far not sufficient to guarantee the succesful elimination of immediate firing sequences. The problem is related to the fact that we have excluded weights and priorities for immediate transitions. Consider a net containing a marking q, where two *internal* immediate transitions, say i_1 and i_2, are both enabled and in conflict. Then, it is not determined which one will fire first. This nondeterminism is problematic, if the two markings q_1 and q_2, reachable via $q\ [i_1\rangle\ q_1$ and $q\ [i_2\rangle\ q_2$ do not fall into the same equivalence class. Then, it lacks information which kind of behaviour will be observed with what probability. The stochastic process associated with the net is not fully determined. If otherwise $q_1 \approx q_2$, no problem arises. Similarly, if at least one of the transitions is observable, the situation can potentially be improved due to synchronisation in a larger cN. However, without means to associate probabilities or priorities to immediate transitions the composite net model will frequently suffer from a lack of determinacy.

5 Application example

In this section we apply the method of CN to the modelling of a network of workstations. This example has originally been investigated in the context of Stochastic Process Algebras [22]. We present it here with some minor changes, in order to meet Definition 2.1. The important point is that

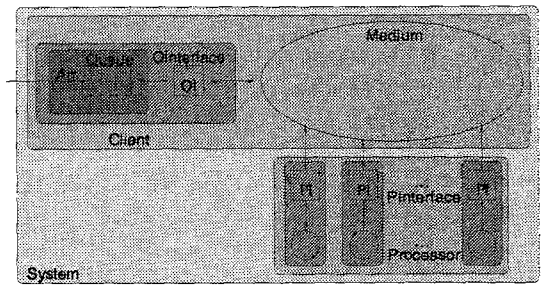


Figure 8: A cluster of workstations

compositional techniques that arise naturally in the context of SPA can be adapted succesfully to GSPN.

The model consists of a number of processors (eg. a workstation cluster) accepting jobs from a common client. The stations are connected with each other via a Token Ring. In particular, communication times caused by the protocol and by the transmission of jobs are explicitly considered. The structure of the system is depicted in Figure 8.

The interface of the client puts request on the medium (sendreq), if a job has appeared in its queue. It then waits for response. If all servers are currently busy, it receives a fullconf-signal. Then the client waits an exponentially distributed phase with rate *wait* before repeating its request. Eventually it will receive an availconf-signal, indicating that at least one server is free and *exactly one* is ready to accept a job. Now the interface transmits the job to this server via the medium (jobtrans).

The medium transfers messages between stations. This transfer causes message delays. The length of the delay is supposed to be dependent on the type of the message. Short protocol messages are transmitted with the higher rate $msgdelay$, job transfers with the lower rate $jobdelay$.

The interface of each processor is tightly connected with the respective processor. If the processor is free the interface awaits a query from the medium and returns availresp. The medium only synchronises with *one* interface, because the availresp signals of the interfaces happen independently of each other. If, on the other hand, all processors are full, they synchronise together with the medium on fullresp.

We choose this simplified model of a token ring because we do not want to blur the concept with technical details. A more detailed specification of the protocol could be modelled hierachically using the composition operators.

The composite GSPN model of the system described above is depicted in Figure 9. The parallel structure clearly resembles the block structure of Figure 8. As motivated in the preceding sections, immediate transitions are used for synchronisation purposes. If this purpose is fullfilled their names are hidden in order to prevent further synchronisation with other net components of the cN.

We investigated the system with maximal queue length 2 and with two servers. The reachability graph of this composite net consists of 882 markings and 2798 arcs. Applying the algorithm of Proposition 3.1 we compute an observable reachability graph that contains 45 markings and 113 arcs. This ORG contains no immediate state changes. Thus, Proposition 4.2 implies that this ORG is a representation of the lumped Markov Chain associated with the whole cN.

Alternatively, we can generate the state space along the hierarchical structure of the model, and exploit substitutivity of $\approx$ for compositional reduction in each step. This stepwise procedure is depicted in Figure 10. In each step we have noted the number of markings (first row) and arcs (second row) of RG (first column) and ORG (second column), i.e. before and after reduction. As an example, the RG of the arrival process has 2 markings and 2 arcs, the RG of the queue possesses 3 markings and 6 arcs. The cN of both components (shown in the upper left of Figure 9) has 6 reachable markings and 12 arcs. Applying the refinement algorithm leads to an ORG with 3 markings and 5 arcs. Due to our substitutivity result, a corresponding net with 3 places and 5 transitions can be used instead of the larger CN in subsequent construction and reduction steps.

The resulting ORG of this stepwise procedure is (by construction) isomorphic to the lumped Markov Chain of the whole cN. Note that the largest intermediate state space that has to be stored (61 markings and 156 arcs) is more than one order of magnitudes smaller than that obtained without compositional reduction. The reduction sketched in Figure 10 is partly based on lumping of symmetric components. This is especially true for the lower half of this reduction tree. The reduction obtained in the upper half does not exploit symmetries. It essentially relies on early elimination of (hidden) immediate transitions.

6 Conclusion

In this paper we have introduced composition operators for GSPN together with a substitutive notion of equivalence. We have shown that substitutivity has important effects on the construction and solution of GSPN models. Hierarchical definition of GSPN by composition of smaller nets is not only a means to effectively construct complex GSPN. Hierarchy can be exploited during the generation of the state space by stepwise compositional reduction. Substitutivity of our equivalence ensures that behavioural properties (performance properties, in particular) remain unchanged whereas the size of the state space is reduced. We have also supplied an algorithm to compute reduced state spaces by means of successive refinement.

The class of GSPN we considered is quite restrictive. We have excluded priorities as well as weight assignments to immediate transitions. Therefore, nondeterminism might

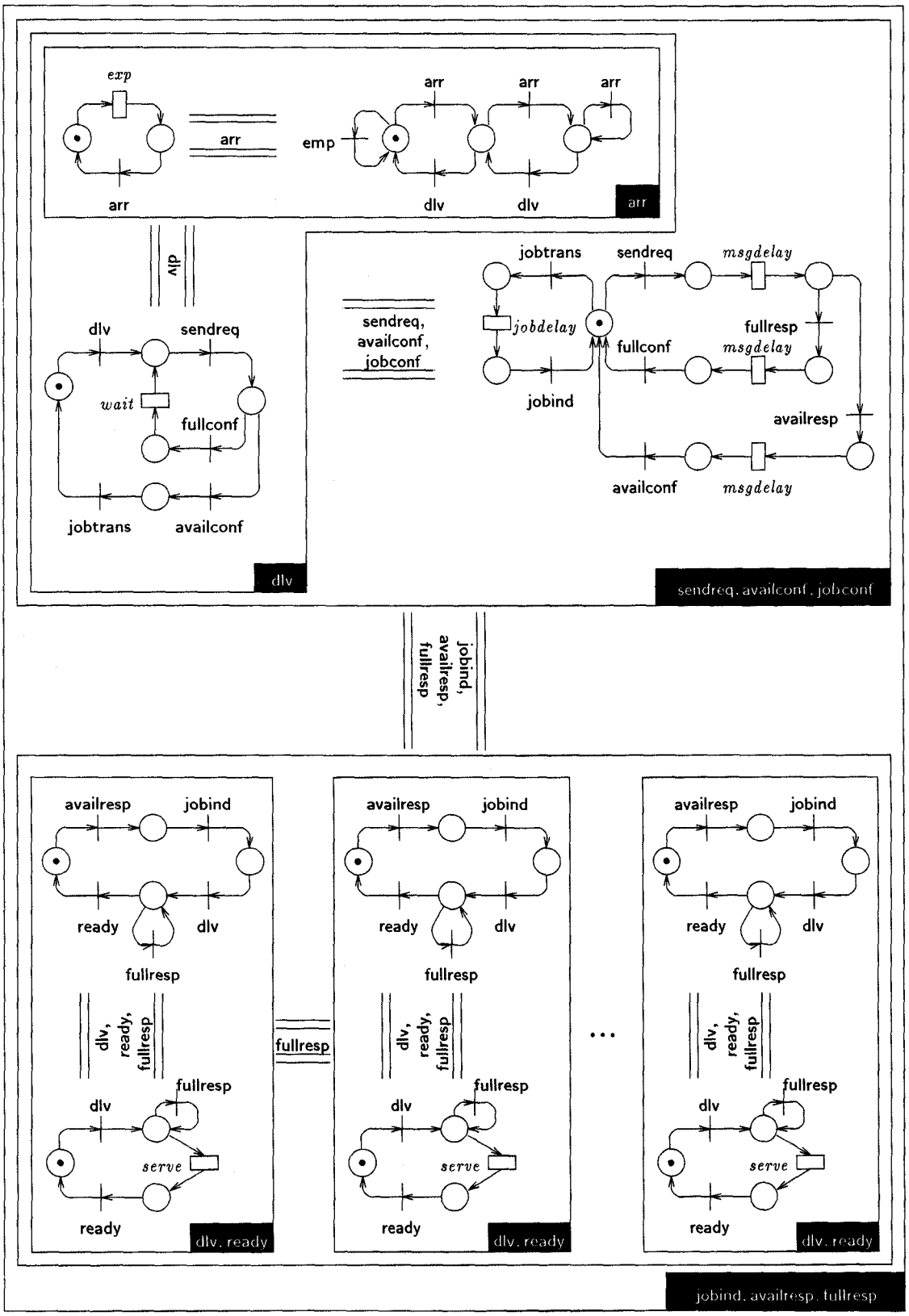


Figure 9: Composite GSPN of the workstation cluster example

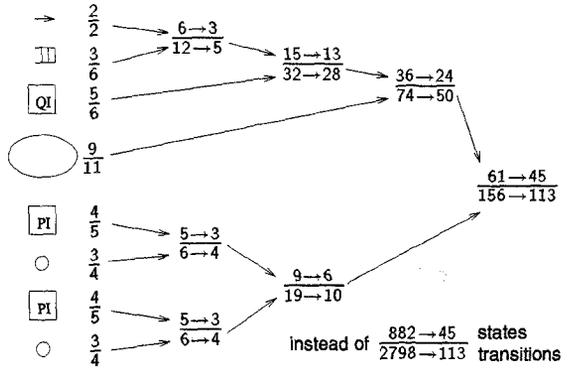


Figure 10: Stepwise compositional reduction

hamper the association of a Markov Chain to some CN, as outlined in Section 4. It is a demanding task for future work to overcome these restrictions. For Stochastic Process Algebras, this issue is discussed in [22].

One effect of our compositional reduction method is mechanised detection and exploitation of symmetries in a hierarchical specification. SWN allow explicit specifications of such symmetries. This eases their exploitation, because a reduced state space can be constructed directly. We claim that combining SWN with compositional reduction will broaden the applicability of both methods.

From a conceptual point of view, further operators should be included without sacrificing substitutivity. In particular, *sequential* composition of nets complements parallel composition in a useful way. Stochastic Process Algebras that include this (and other) operator(s), e.g. [13, 15], can serve as a starting point for their integration into GSPN.

Acknowledgements The authors would like to thank Marina Ribaudo, Joost-Pieter Katoen, Markus Siegle and the anonymous referees for many valuable suggestions.

References

[1] M. Ajmone Marsan, G. Balbo, and G. Conte. *Performance Models of Multiprocessor Systems*. MIT Press, 1986.

[2] E. Best, R. Devillers, and J. Hall. The Petri Box Calculus: A new causal algebra with multilabel communication. In *Advances in Petri Nets*. Springer LNCS 609, 1992.

[3] O. Botti and F. De Cindio. Process and Resource boxes: An Integrated PN Performance Model for Application and Architectures. In *IEEE proc. of the Int. Conference on Systems, Man and Cybernetics*, 1993.

[4] P. Buchholz. Markovian Process Algebra: Composition and Equivalence. In [16].

[5] P. Buchholz. An notion of equivalence for Stochastic Petri Nets. In *Application and Theory of Petri Nets*. Springer LNCS 935, 1995.

[6] G. Chehaivbar, H. Garavel, N. Tawbi, and F. Zulian. Specification and Verification of the Powerscale Bus Arbitration Protocol: An Industrial Experiment with LOTOS. In *Formal Description Techniques IX*. Chapmann Hall, 1996.

[7] G. Chiola, C. Dutheillet, G. Franceschinis, and S. Haddad. Stochastic well-formed coloured nets for symmetric modelling application. *IEEE Trans. on Computers*, 42(11), 1993.

[8] G. Chiola, M. Ajmone Marsan, G. Balbo, and G. Conte. Generalized Stochastic Petri Nets: A definition at the net level and its implication. *IEEE Trans. on Software Engineering*, 19(2), 1993.

[9] S. Donatelli. Superposed Generalized Stochastic Petri Nets: Definition and Efficient Solution. In *Application and Theory of Petri Nets*. Springer LNCS 815, 1994.

[10] S. Donatelli and G. Franchesinis. The PSR Methodology: Integrating Hardware and Software Models. In *Application and Theory of Petri Nets*. Springer LNCS 1102, 1996.

[11] N. Götz, H. Hermanns, U. Herzog, V. Mertsiotakis, and M. Rettelbach. Stochastic Process Algebras. Chapter 1 of *Quantitative Methods in Parallel Systems*. Springer, 1995.

[12] H. Hermanns and M. Rettelbach. Syntax, Semantics, Equivalences, and Axioms for MTIPP. In [16].

[13] H. Hermanns and M. Rettelbach. Towards a superset of Basic LOTOS for Performance Prediction. In [24].

[14] H. Hermanns, M. Rettelbach, and T. Weiß. Formal characterisation of immediate actions in SPA with nondeterministic branching. *The Computer Journal*, 38(7) 1995.

[15] U. Herzog. A Concept for Graph-Based Stochastic Process Algebras, Generally Distributed Activity Times and Hierarchical Modelling. In [24].

[16] U. Herzog and M. Rettelbach, editors. *Proc. of the 2nd PAPM Workshop*. Universität Erlangen–Nürnberg, IMMD Arbeitsbericht, 27(4) 1994.

[17] J. Hillston. *A Compositional Approach to Performance Modelling*. PhD thesis, University of Edinburgh, 1994.

[18] ISO. *LOTOS : A Formal Description Technique Based on the Temporal Ordering of Observational Behaviour*, 1989.

[19] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Springer, 1976.

[20] R. Milner. *Communication and Concurrency*. Prentice Hall, 1989.

[21] L. Pomello, G. Rozenberg, and C. Simone. A survey of equivalence notions for net based systems. In *Advances in Petri Nets*. Springer LNCS 609, 1992.

[22] M. Rettelbach. *Stochastische Prozeßalgebren mit zeitlosen Aktivitäten und probabilistischen Verzweigungen*. PhD thesis, Universität Erlangen–Nürnberg, 1996.

[23] M. Ribaudo. On the Aggregation Techniques in Stochastic Petri Nets and Stochastic Process Algebras. *The Computer Journal*, 38(7) 1995.

[24] M. Ribaudo, editor. *Proc. of the 4th PAPM Workshop*. Dpto. di Informatica, Universita di Torino, 1996.

[25] I. Rojas. Compositional construction of SWN models. *The Computer Journal*, 38(7) 1995.