INVESTIGATING COUPLING PATTERNS IN
STATE-SPACE MODELS FOR SYSTEM RELIABILITY

Vitali Volovoi

Georgia Institute of Technology
Atlanta, GA 30332, USA

ABSTRACT

State-space representation in system reliability modeling provides a flexible means for modeling dynamic scenarios, but monolithic models tends to be prohibitively complex for realistic systems. Breaking models into modules that can be analyzed independently (off-line) greatly reduces this complexity, but the characterization of the coupling among these modules should provide adequate accuracy once those modules are assembled together. In this paper, the feasibility of component-based modeling of shared repair resources is investigated. The work builds upon a recently developed compact representation of competing risks scenarios that matches the relative occurrence probability ratios of multiple risks (choice) and replaces it with a single Weibull distribution.

1   INTRODUCTION

In order to model maintenance policies of a system, a state representation is required to describe the dynamics of the maintenance actions and the system states. For a small system, a custom Monte Carlo simulation maintenance model can be easily developed, but using a standardized graphical representation provides advantages when creating larger models, in particular for verification purposes. Furthermore, local representation of the state space usually scale better with the size of the problem: instead of each state representing the system as a whole (or globally, as in Markov chains), states of individual components are described along with their interactions, so that the system state can be inferred from its component states (rather then described explicitly). This local representation naturally leads to Stochastic Petri Nets (SPNs) (Natkin 1980; Molloy 1981), where individual components are denoted with small circles (called tokens) and their transitions between states (called places) are denoted with solid rectangles. Some implementations of SPNs rely on automated translation of the local-representation into the global (Markovian) model (Trivedi 2002). However, the use of Monte-Carlo simulation is also a viable alternative (Dutuit, Châtelet, Signoret, and Thomas 1997; Haas 2002) as it allows to avoid any restrictions on the type of distributions for the delays in state transitions.

In traditional SPNs tokens are indistinguishable. In contrast, tokens in so-called colored Petri nets (Jensen 1993) have unique identities (labels), so an alternative interpretation of firing facilitates the preservation of the information about the system's past states: rather than considering removing a token from the transition's input place and depositing a (different) token to the output place as two disjoint actions, one can unite these two actions into a single action of moving the same token from an input place into the output place (see also Reisig (1985)). Due to the fact that system reliability models often deal with physical components that change their states (age, fail, get repaired, etc.), considering tokens with distinct identities provides certain advantages. It must be noted that this interpretation implies a one-to-one correspondence between the inputs and outputs of a transition in order to to keep track of the token movements (effectively providing token preservation during transition firing; this firing preservation does not need to be absolute, as tokens "births" and "death" can be included as long as appropriate bookkeeping is implemented). This can be contrasted to standard constructions in traditional SPNs where the numbers of inputs and outputs

to a transitions are not related , e.g., where multiple inputs to a transition are used to model logical *AND* (resulting in a single output).

However, this token preservation upon firing does not imply any restriction on the expressive power of the SPNs with distinct tokens, as logical *AND* and similar constructions can be modeled using inhibitors and their opposites (so-called enablers), as shown in this paper. Memory (continuously changing labels) can be assigned to tokens (the result is "aging tokens" (Volovoi 2004; Volovoi 2006)). Such tokens can move freely throughout the Petri net without losing their memory. Firing delays for timed transitions can be interpreted by associating backward clocks: the clock starts when transition gets enabled, and once the clock reaches zero, the firing takes place. In standard SPNs, this clock is associated solely with the transition, and if more than one token is present in the input place, the token to be fired is selected at random. With aging tokens, a clock is associated with a token-transition pair, which allows several clocks to run simultaneously for a single transition, and often results in a more compact model.

While local nature of SPN can be referred to as component-based modeling, here we reserve the term "component-based" to hierarchical or modular construction where instead of a single model, relatively simple models for individual components are combined to construct system-level models (Dekker 1995). In this context, communication among component models is avoided during the actual simulation, and instead statistical representation of the influence of other components is taken into account.

Stochastic Petri nets can be interpreted as a special kind of discrete-event simulations, so it is useful to draw an analogy among the approaches used in the latter field to deal with modeling of complex systems (Cota and Sargent 1992). Specifically, the use of object-oriented principles (Zeigler 1990), and component-based approach (Chen and Szymanski 2001) are advocated to improve scalability of the modeling (in particular, in enable the use of parallel computing). Analogies can also be drawn with classical work in queuing theory related to product-form solutions (Nelson 1993), Norton and arrival theorems, and in particular, to Jackson's theorem (Jackson 1963), and to its extensions (and BCMP in particular (Baskett, Chandy, Muntz, and Palacios 1975)). There seems to be sufficient difference for those queuing problems and the system reliability problems, so that existing queuing results are not directly applicable (to the best of the author's knowledge).

In spirit, the present work most closely resembles (Balbo and Bruell 1988). In general, there is a compelling evidence that in both natural and engineering domains complex systems only "lightly" coupled, as modular architecture provides clear advantages in developing desirable systems properties. The evolutionary advantage of so-called nearly decomposable systems has been discussed in the context of biological systems (Simon 2002), as well as for steam engine development (Frenken and Nuvolari 2004). These concepts are also explicitly employed in the design principles of computer systems (Courtois 1977), where the importance of different time scales for interactions within and between the modules was emphasized. It is therefore logical to take full advantage of the modular structure of the systems in modeling maintenance processes. The main idea of the present paper is to pursue the hypothesis that mean-time calculations suffice for choice-free portions of the models, while matching winning ratios locally at the choice branches provides a significant improvement in the overall accuracy of the approximation.

If the units composing the system behave independent of each other then the creation of a component-based model for each type of unit is straightforward. However, if there exists coupling among some of the units then an appropriate procedure to capture the effect of the rest of the system on the component being modeled needs to be followed. One such procedure that is able to deal with both competing-risk coupling among components while retaining the compactness of the component-based representation has been proposed and validated (Volovoi and Vega 2012). In this paper the results of Ref. Volovoi and Vega (2012) are extended to show how the proposed component-based modeling procedure can be for used to capture coupling effects for single shared repairs.

2   SHARED REPAIR RESOURCE

2.1 Simple Model

Let us consider the simplest possible model that would take into account the basic structure of the competition for resources. The main idea is to consider the component to be distinct from the rest of the system, where influence of the rest of the system is represented in a as compressed form as possible. The smallest system of interest would be the system that consists of three components. First, let us consider a coupled model (see Figure 1). At this point, we don't consider specific details of the service priorities, as they can require slight modification of the model (for example, the depicted model can be considered for fixed set of priorities if immediate transitions to the repair are replaced with fast transitions that are incrementally timed as needed). Of course, if the components are indistinguishable, then in standard SPN it is possible to put all three tokens in a single place, leading to a trivial model with two places (this model would not track individual components). In the case of colored Petri nets with the clock assigned to a transition-token pair, the model is depicted in (see Figure 2). Here each color can have a matching transition distribution, so components with different properties can be treated. If specific policies (such as FIFO need to be implemented, then either additional places can be introduced for each spot in the queue, or age can be used (Volovoi 2004; Volovoi 2006)) to avoid adding additional places. In the latter scenario age-dependent fast transitions are used, so that the tokens that arrive first (and aged most) will be fired first.
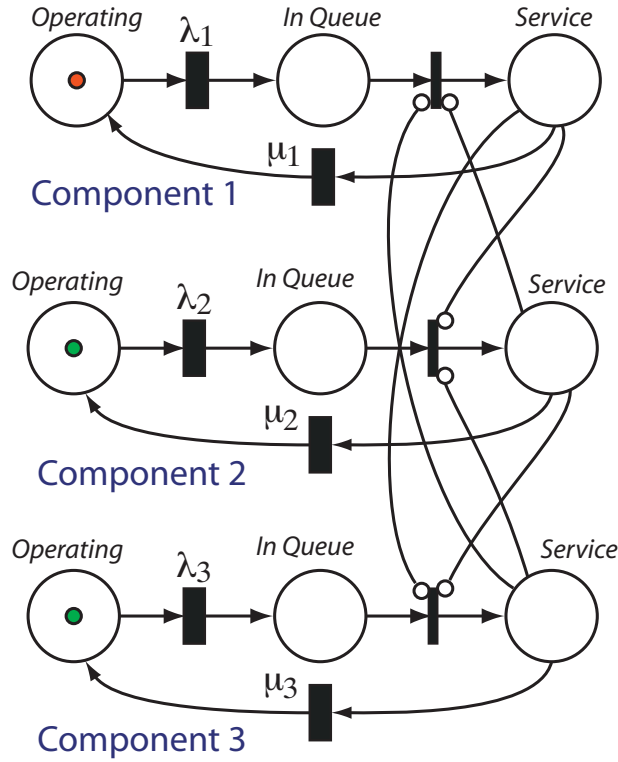

Figure 1: SPN model of a system with three components and a single repair server. Here in below a standard SPN convention of using thin bars to represent immediate transitions is used (as opposed to thick bars that represent transitions with finite delays).

The simplicity of this coupled models allows easy comparison of the quality of the component-based approximations that will be introduced next. Consider the following numerical example: three identical components with $\lambda_0 = 1$ and $\mu_0 = 2$; one million Monte-Carlo replications; and results collected at time $T = 100$. For those parameters, if both failures and repairs are exponential, the chances that the repair
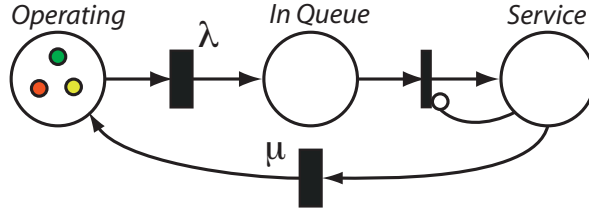
Figure 2: SPN model of a system with three components and a single repair server (colors and parallel clocks are used).

crew is busy are $P_r = 0.7896$. The answer can be also evaluated for a steady-state formula using a simple four-state Markov representation:

$$P_r = 1 - \frac{1}{1 + \frac{3\lambda_0}{\mu_0} + \frac{6\lambda_0^2}{\mu_0^2} + \frac{6\lambda_0^3}{\mu_0^3}} \approx 0.7895 \tag{1}$$

Importantly, we can clearly note that components are not independent: indeed, the availability of all three components simultaneously is $P_3 = 0.2104$, which for independent components would imply the availability of $P_{ci} = 0.5948$, which in turn would lead to the conclusion at that at least one out of three components would be available with probability $P_{1i} = 0.9335$. However, given the structure of the model (i.e., common repair source), the availability is significantly lower ($P_1 = 0.8422$). In creating component-based models we follow the principle of finding models that are as simple as possible but not simpler than that. First, let us consider a simple model depicted in Figure 3. The bottom part of the model depicts the states of the modeled component (let us call it the first component), while the top represents the combined state of the other two components (the second and third component combined, or "the rest of the system"). Note that there is only one token representing the rest of the system. Here the place "Both up" indicates that both the second and third component are operational (regardless of the state of the first component); "In Queue" implies there is a need to service at least one of the components from the rest of the system, but they are not serviced due to the fact that the first component is serviced; finally, place "Service" indicates that either the second or the third component is serviced. note that "Service" place includes the situation when both second and third components need to be serviced.

This model has two unknown parameters λ_1 and μ_1. If we fix one of the parameters (e.g., λ_1), the other can be found by iteration to ensure that service time for the rest of the system is twice as much as compared to the service time for the component (here the symmetry consideration is used). However, for some values of λ_1, converged probabilities can be actually negative. At the first glance it seems logical to assume that $\lambda_1 = 2\lambda_0$, as both components are operational. However, the true value is likely to be slightly larger to account for the fact that some failures can accumulate in queue for the rest of the system while the first component is repaired (and so the transition from "Both up" to "In Queue" is immediate in this scenario, once the repair of the first component is completed). Indeed, for $\lambda_1 = 2$ the converged value of the repair rate is $\mu = 1.352$, yielding an approximation of $P_r = 0.7793$, or about 1.3% less then the correct value). Importantly, modeling this effect explicitly can provide not only better accuracy, but also the mechanism for accounting for variable transition rates, as shown next.

2.2 Refined Component Model

The refined model is shown in Figure 4. Note that an arc terminated with a solid small circle represents an enabler that acts in the opposite manner to the inhibitor (depicted by an arc terminated with a hollow circle) (Volovoi 2006). Now, service states are divided into two places: "open service" place implies that the modeled component is operating, and therefore, any other components from the rest of the system are either serviced or joined the queue. In fact, at the modeled level of resolution this service is lumped into a single service. Under Markovian assumptions for the rest of the system, it is easy to calculate the expected
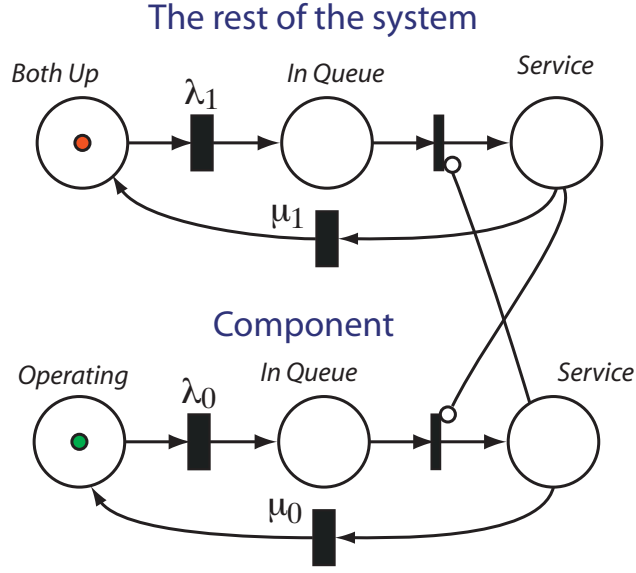
The rest of the system

Figure 3: Simple component-based model for a system with three components.

duration of that lumped service, and the appropriate rate can be calculated as $\mu_2 = \frac{\mu_0^2}{\mu_0 + \lambda_0}$. In contrast, when the modeled component fails, a transition to "closed service" mode occurs, which corresponds to the situation where (if FIFO is used) the lumped service is completed, but new failures from the rest of the system must wait until the repair of the modeled component occurs. The total mean transition rate from that place is denoted by $\hat{\mu}$ and can be calculated as follows.

First, we note that there are two possible (compressed) states when the foregoing transition occurs: q_2 - in which one component is operational, and the other needs service, and q_3, in which both need service. The ratio of probabilities for those two respective states is $\lambda_0 : \mu_0$ (this can be calculated based on the same Markov model for the rest of the system that was used to calculate μ_2). Next, we note that the mean service times for those states are $2/\mu$ and $1/\mu$, so we can evaluate the overall mean rate:

$$\hat{\mu} = \frac{2\lambda_0 + \mu_0}{\mu_0(\mu_0 + \lambda_0)} \tag{2}$$

Finally, let us introduce the winning ratio γ_e (Volovoi and Vega 2012) that represents the chance that failure occurs before the repair is completed, and use this ratio to evaluate the two remaining competing transition rates: $\nu_1 = \gamma_e \hat{\mu}$, and $\mu_1 = (1 - \gamma_e)\hat{\mu}$. Let us calculate the winning ratio for exponential service distribution:

$$\gamma_e = \int_0^\infty \lambda_0 e^{-\lambda_0 t} e^{-\mu_0 t} dt = \frac{\lambda_0}{\mu_0 + \lambda_0} \tag{3}$$

Using this value leads to $P_r = 0.7899$, or 0.04% error (within the expected accuracy of Monte Carlo simulation).

Obviously, these calculations simply confirm that no errors have been made, as all transitions follow exponential assumptions, and the solutions are Markovian. Next, let us consider fixed repair time $1/\mu_0 = 0.5$. The coupled model (Figure 2) predicts the repair time to be $P_{rf} = 0.8366$, or about 5.62% higher than the value for exponential distribution. The natural step for adjusting the refined component model (Figure 4) is to replace μ_0 and μ_2 exponential transitions with their fixed counterparts. The resulting prediction $P_{rf1} = 0.8553$ is better, with the over-prediction error of about 2.23%. While this error is not very large, it is still outside of the expected accuracy of Monte Carlo simulation. Finally, re-evaluating winning ratio

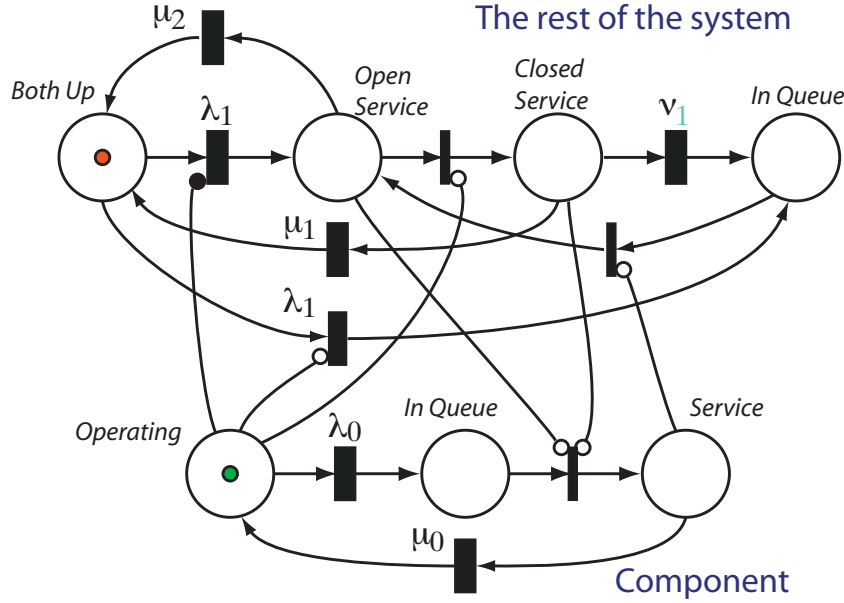$$\gamma_f = F\left(\frac{1}{\mu_0}\right) = 1 - e^{-\frac{\lambda_0}{\mu_0}} \tag{4}$$

Figure 4: Refined component-based model for a system with three components.

(here F is cumulative probability distribution for the time to failure) and using this new ratio to recalculate the (still exponential) rates $\nu_1 = \gamma_f \hat{\mu}$, and $\mu_1 = (1 - \gamma_f)\hat{\mu}$, we reduce the error by an order of magnitude: $P_{rf2} = 0.8345 \, (-0.25\%)$. This is consistent with the error improvements shown when winning ratio has been used for modeling choice "races" (**Volovoi and Vega 2012**). Parametric studies are required to ascertain the validity of this procedure.

3   CONCLUSIONS

The feasibility of a component-based modeling for shared repair resources has been explored. A simple example of three identical components has been considered. It is hoped that component-based models can be useful in understanding coupling mechanisms and building simple models of complex systems. The procedure is tested for exponential and fixed repair time. While preliminary results are encouraging, the suitability of using the winning ratio in the context of shared repairs and other coupling scenarios needs to be further investigated, including parametric investigation of the other types of distributions, *e.g.,* gamma distribution, thus establishing the ranges of applicability of the proposed approach.

REFERENCES

Balbo, G., and S. C. Bruell. 1988. "Combining Queueing Networks and Generalized Stochastic Petri Nets for the Solution of Combining Queueing Networks and Complex Models of System Behavior". *IEEE Transactions on Computers* 37 (10): 1251–1268.

Baskett, F., K. M. Chandy, R. Muntz, and F. Palacios. 1975. "Open, closed and mixed networks of queues with different classes of customers". *Journal of the ACM* 22 (2): 248–260.

Chen, G., and B. K. Szymanski. 2001. "Component-Based Simulation". In *Modeling and Simulation Conference, ESM2001*, 68–75. Delft, Netherlands: SCS Press.

Cota, B. A., and R. G. Sargent. 1992. "A modification of the process interaction world view". *ACM Trans. Model. Comput. Simul.* 2 (2): 109–129.

Courtois, P. J. 1977. *Decomposability: queueing and computer system applications*. New York: Academic Press.

Dekker, R. 1995, April. "Integrating optimisation, priority setting, planning and combining of maintenance activities". *European Journal of Operational Research* 82 (2): 225–240.

Dutuit, Y., E. Châtelet, J.-P. Signoret, and P. Thomas. 1997. "Dependability Modelling and Evaluation by Using Stochastic Petri Nets: Application to Two Test Cases". *Reliability Engineering and System Safety* 55:117–124.

Frenken, K., and A. Nuvolari. 2004. "The early development of the steam engine: an evolutionary interpretation using complexity theory". *Industrial and Corporate Change* 13 (2): 419—450.

Haas, P. J. 2002. *Stochastic Petri Nets. Modelling, Stability, Simulation*. New York: Springer.

Jackson, J. R. 1963. "Jobshop-like Queueing Systems". *Management Science* 10 (1): 131–142.

Jensen, K. 1993. *Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use*, Volume 1. Berlin: Springer.

Molloy, M. K. 1981. *On the Integration of Delay and Throughput Measures in Distributed Processing Models*. Ph. D. thesis, Department of Computer Science, University of California, Los Angelos, CA.

Natkin, S. 1980. *Les réseaux de Petri stochastiques et leur application a l'evaluation des systemes informatiques*. Ph. D. thesis, Conservatoure National des Arts et Metier, Paris.

Nelson, R. D. 1993. "The mathematics of product form queuing networks". *ACM Comput. Surv.* 25 (3): 339–369.

Reisig, W. 1985. "Petri nets with individual tokens". *Theoretical Computer Science* 41:185 – 213.

Simon, H. A. 2002. "Near decomposability and the speed of evolution". *Industrial and Corporate Change* 11 (3): 587—599.

Trivedi, S. K. 2002. *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. Second ed. John Wiley and Sons.

Volovoi, V. 2006. "Stochastic Petri Nets Modeling using SPN". In *Proceedings of Annual Reliability and Maintainability Symposium*, 75–81: IEEE.

Volovoi, V., and R. V. Vega. 2012. "On Compact Modeling of Coupling Effects in Maintenance Processes of Complex Systems". *International Journal of Engineering Science* 51:193–210.

Volovoi, V. V. 2004. "Modeling of System Reliability Using Petri Nets with Aging Tokens". *Reliability Engineering and System Safety* 84 (2): 149–161.

Zeigler, B. P. 1990. *Object Oriented Simulation wtth Hierarchical Modular Models*. New York: Academic Press.

AUTHOR BIOGRAPHY

VITALI VOLOVOI He is an Assistant Professor in the School of Aerospace Engineering, has over 80 journal and conference publications. He received a university diploma with honors in Applied Mechanics from Moscow State University in 1988, and earned his Ph.D. in Aerospace Engineering from the Georgia Institute of Technology in 1997. His principal fields of interest are system safety and reliability, risk assessment, and simulation of complex systems. His email address is vitali@gatech.edu.