PARAMETRIC CONVERGENCE ANALYSIS OF AN AGGREGATED MARKOV CHAIN

Kutluyıl Doğançay

School of Electrical and Information Engineering, University of South Australia Mawson Lakes, SA 5095, Australia phone: + (61) 883023984, fax: + (61) 883023384, email: Kutluyil.Dogancay@unisa.edu.au

web: http://people.unisa.edu.au/Kutluyil.Dogancay@unisa.edu.au

ABSTRACT

Markov chains are commonly used in system identification, modelling and statistical signal processing. In particular they provide powerful analysis tools for digital communications, computer networks and flexible manufacturing systems. For most practical systems the underlying Markov chain possesses a prohibitively large number of states. This necessitates state aggregation in an effort to maintain the computational complexity at manageable levels. In this paper we consider the aggregation of an underlying Markov chain for a parallel synchronized structure in a closed network. Such Markov chains are encountered in the modelling of computer networks and manufacturing systems, and do not have closed-form solutions, requiring numerical computation. Based on an asymptotic convergence result we provide a parametric convergence analysis of the transition rates of the aggregated Markov chain and develop reduced complexity solutions.

1. INTRODUCTION

Markov chains are commonly used to model various processes and systems in signal and speech processing, digital communications, radar, computer networks and flexible manufacturing, to name just a few. A Markov chain is a finite random process obeying the memoryless property [1]. A major challenge with using Markov chains in system modelling or analysis is that the number of states of a Markov chain can easily become prohibitive. A common solution to such "state-space explosion" is to aggregate the original Markov chain by grouping subsets of the states into aggregated states, thereby reducing the size (the number of states) of the Markov chain [2]. A reduction in the number of states translates into complexity reduction.

In parallel synchronized systems, where the parallel systems have a common input, but their output can be passed to the next system only if all systems produce an output, the presence of synchronization creates considerable difficulty with the performance analysis [3]. In this paper we consider the aggregation of such a parallel synchronized structure placed in a closed queueing network. Such structures are also referred to as closed fork-join nets. Assuming that the transition rates of packets (or tokens) are Markovian with negative exponential distribution, the closed fork-join structure becomes a generalized stochastic Petri net (GSPN) [4], which is shown at left in Fig. 1. The timed transitions T0, T1 and T2 are Markovian with transition rates λ_0 , λ_1 and λ_2 , respectively. The traditional approach to analysis of the closed fork-join GSPN involves is to solve the underlying continuous-time Markov chain (CTMC) for its stationary distribution. However for a large number of packets (tokens) in initial marking, N, the state-space of the underlying CTMC becomes excessively large, rendering the analysis too complicated if not prohibitive. Therefore the preferred method of analysis is to aggregate the CTMC as shown at right in Fig. 1.

The objective of aggregation is to reduce the state-space of the original CTMC while preserving the stationary token distribution. For the structure at right in Fig. 1 this is achieved by having marking-dependent transitions rates for Tm, denoted λ_m^k , $k = 1, \ldots, N$. The synchronized transition in the fork-join structure precludes a product-form solution, which in turn implies no closed-form solution to the underlying CTMC. In [5] stochastic complementation [2] was used to aggregate the original CTMC and to derive an important convergence property for the λ_m^k , providing an approximate numerical solution for the aggregated Markov chain using a much smaller Markov chain than the original CTMC.

In this paper we present a parametric analysis with the aim of deriving parsimonious Markov chain aggregation that provides significant complexity reduction while preserving the stationary distribution of the original Markov chain. The underlying key concept is the asymptotic convergence of the transition rates, which was proven in [5], but remains to be scrutinized further to determine the relationship between original transition rates λ_0 , λ_1 and λ_2 , and the convergence behaviour of the aggregated chain's transition rates λ_m^k , k = $1, \ldots, N$. The work presented in this paper elucidates this important relationship enabling an informed choice for the number of tokens to be used in the aggregated chain for a given N.



Figure 1: Aggregation of closed fork-join GSPN—an example for aggregated Markov chain.

The paper is organized as follows. Section 2 presents the original and aggregated Markov chains under consideration. Section 3 reviews the asymptotic convergence result for marking-dependent transition rates of the aggregated Markov chain. The parametric convergence analysis of the marking-dependent transition rates is presented and demon-



Figure 2: State transition diagrams of the underlying CTMCs for N = 1, 2 and 3.

strated in Section 4. Conclusions are drawn in Section 5.

2. THE ORIGINAL AND AGGREGATED MARKOV CHAINS

The parallel synchronized system at left in Fig. 1 has an underlying CTMC with transition diagrams shown in Fig. 2 for N = 1, 2 and 3. For a given number of tokens in initial marking, N, the number of states of the underlying CTMC is $L = (N + 1)^2$ states.

Let the $L \times L$ matrix $\mathbf{Q} = [q_{ij}]$ denote the infinitesimal generator matrix for the CTMC and let $x_t \in \{1, \ldots, L\}$ be the *L*-state CTMC where $t \geq 0$ is a real number. Suppose that the CTMC is homogeneous and irreducible. Then the stationary distribution for the CTMC, π , satisfies

$$\boldsymbol{\pi}^{T}\boldsymbol{Q} = \boldsymbol{0}^{T}, \quad \sum_{i=1}^{L} \pi_{i} = 1, \quad \pi_{i} > 0$$
 (1)

where

$$\boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_L]^T \tag{2}$$

$$\pi_i = \Pr\{x_t = i\}, \quad \text{as } t \to \infty \tag{3}$$

and ${\bf 0}$ is a column vector of zeros.

Fig. 3 shows the transition diagram of the underlying CTMC corresponding to the aggregated structure in Fig. 1. For N tokens in initial marking, the aggregated CTMC has N + 1 states compared with $L = (N + 1)^2$ states for the original CTMC. Let $\tilde{x}_t \in \{1, \ldots, N + 1\}$ be the aggregated CTMC. The states of the aggregated CTMC are given by $\tilde{x}_t = N - M(P0) + 1$. For example, if N = 3 and M(P0) = 2 (where M(P0) denotes the number of tokens in place P0), the aggregated State with marking M(Pm) to another aggregated state with M(Pm) - 1 are the marking-dependent transition rates λ_m^k , $k = 1, \ldots, N$.

The stationary distribution of the aggregated chain, $\boldsymbol{\xi} =$



Figure 3: State transition diagram of the underlying CTMC for the aggregated Markov chain.

$$[\xi_1,\xi_2,\ldots,\xi_{N+1}]^T$$
, is given by

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{N+1} \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 + \pi_3 + \pi_4 \\ \vdots \\ \pi_{N^2+1} + \dots + \pi_{(N+1)^2} \end{bmatrix}$$
(4)

where

$$\xi_i = \Pr\{\tilde{x}_t = i\} \quad \text{as } t \to \infty. \tag{5}$$

The marking-dependent transition rates are unknown and need to be calculated. Using the local balance equations

$$\xi_k \lambda_0 = \xi_{k+1} \lambda_m^k, \quad k = 1, \dots, N \tag{6}$$

the marking-dependent transition rates can be expressed as

$$\lambda_m^k = \lambda_0 \frac{\xi_k}{\xi_{k+1}} = \lambda_0 \frac{\pi_{(k-1)^2+1} + \pi_{(k-1)^2+2} + \dots + \pi_{k^2}}{\pi_{k^2+1} + \pi_{k^2+2} + \dots + \pi_{(k+1)^2}}.$$
 (7)

3. ASYMPTOTIC CONVERGENCE RESULT

In this section we review an important result for the asymptotic convergence of λ_m^k as N grows unbounded. This asymptotic convergence result is central to the ensuing parametric convergence analysis. In [5] it was shown that as $\lambda_0 \to 0$ the marking-dependent transition rates λ_m^k become independent of N. This observation is then extended to the case of $\lambda_0 > 0$ by exploiting the coupling between a given aggregated state and the new aggregated states introduced by increasing N. We formally have the following asymptotic convergence result: **Theorem 1** ([5]). As $N \to \infty$ the transition rates of the aggregated chain converge to constant values. That is,

$$\lim_{N \to \infty} \lambda_m^{k=x} = \gamma_x, \quad x = 1, 2, \dots$$

where γ_x are the converged transition rates.

An important observation in relation to the convergence rate of the marking rate transition rates as N is increased is that decoupling between a given state and new introduced states occurs faster for smaller λ_0 , all other parameters remaining constant. This leads to the natural conclusion that λ_m^k will exhibit faster convergence for smaller λ_0 while λ_1 and λ_2 are fixed. In the next section we provide a detailed analysis of transition rate convergence.

4. PARAMETRIC CONVERGENCE ANALYSIS

As N is increased, two problems arise in relation to the aggregated transition rates λ_m^k for k close to N: (1) they become difficult to calculate as a result of increased computational complexity, and (2) barring special cases (e.g., $\lambda_0 \rightarrow 0$), they cannot be approximated by resorting to the asymptotic convergence result since their converged values cannot be checked. To overcome these shortcomings of the asymptotic analysis we will make use of the following observation:

Observation 1. As the number of tokens in the merged place Pm (see Fig. 1) tends to infinity, the parallel synchronized (fork-join) structure behaves like a single delayed transition with transition rate given by the slowest branch; i.e.,

$$\lambda_m^k \to \min(\lambda_1, \lambda_2), \quad as \ N \to \infty \ and \ k \to N.$$
 (8)

This observation allows us to approximate markingdependent transition rates λ_m^k for large k and N by the smallest firing rate in the parallel branches of the fork-join structure. Thus asymptotically as $N \to \infty$ the aggregated CTMC takes the form shown in Fig. 4.



Figure 4: Aggregated CTMC as N tends to infinity where $\lambda_s = \min(\lambda_1, \lambda_2)$.

We will analyze the convergence behaviour of the marking-dependent transition rates as well as the reducibility of the aggregated Markov chain under three different conditions for λ_0 , λ_1 and λ_2 .

4.1 Case I: $\lambda_0 < \lambda_s$

As N is increased this case results in states with small M(P0) (i.e., large \tilde{x}_t where $\tilde{x}_t \in \{1, \ldots, N+1\}$) to have increasingly smaller state probabilities as a result of reduced likelihood for them to be visited. The transition diagrams characteristic of this case as $N \to \infty$ are depicted in Fig. 5. We use Observation 1 to determine the asymptotic transition rates for k close to N. Note that in this case the aggregated CTMC is no longer irreducible as a result of states close to N + 1 not being communicated asymptotically.

For large but finite N, the question of how many states M in the reduced Markov chain should be retained for a



Figure 5: Reduced aggregated CTMC with M < N+1 states as N tends to infinity for case I ($\lambda_0 < \lambda_s$).

given desired accuracy measure is answered next. Referring to the reduced aggregated Markov chain in Fig. 5 we have

$$\lambda_0 \xi_i = \gamma_i \xi_{i+1}, \quad i = 1, 2, \dots, M - 1$$
 (9a)

$$\sum_{i=1}^{M} \xi_i = 1. \tag{9b}$$

Solving the above local balance equations for ξ_M yields

$$\xi_M = \frac{1}{1 + \sum_{i=1}^{M-1} \frac{\zeta_i}{\lambda_0^i}}, \quad \zeta_i = \prod_{j=1}^i \gamma_{M-j}.$$
 (10)

The numerical approximation of the reduced aggregated Markov chain involves selection of M for given $\gamma_1, \ldots, \gamma_{M-1}$ such that ξ_M is sufficiently small, justifying the elimination of the remaining states on the grounds of vanishingly small state probabilities. If M is sufficiently large so that the remaining transition rates can be approximated by λ_s , we have

$$\xi_i = \xi_M \left(\frac{\lambda_0}{\lambda_s}\right)^{i-M}, \quad i = M+1, \dots, N+1.$$
(11)

Since $\lambda_0/\lambda_s < 1$ the states $M + 1, \dots, N + 1$ are guaranteed to have smaller state probabilities than state M:

$$\xi_M > \xi_{M+1} > \xi_{M+2} > \dots > \xi_{N+1}$$

In fact in the limit as $N \to \infty$ we have

$$\lim_{N \to \infty} \xi_N = 0.$$

Suppose that we wish to find the smallest M such that $\xi_M \leq \epsilon$ for a given threshold ϵ for the approximated reduced Markov chain. Fig. 6 shows plots of ξ_M versus M for different ratios λ_0/λ_s computed using (10) as $N \to \infty$. We observe that as λ_0/λ_s gets smaller, the aggregated Markov chain can be approximated by a smaller Markov chain by selecting a smaller M for a fixed ϵ . Conversely, larger λ_0/λ_s (subject to $\lambda_0/\lambda_s < 1$) requires larger M. For finite N the same plots can be used approximately as long as N is sufficiently large.

For given λ_0 , λ_1 , λ_2 , N and ϵ where $\lambda_0/\lambda_s < 1$, i.e., case I applies, a computational method for finding approximate aggregation can be conceived as follows: (i) compute λ_m^k , $k = 1, \ldots, N_c$, for $N_c < N$ tokens in initial marking, where N_c is the maximum number of tokens that can be



Figure 6: Plot of ξ_M versus M for different λ_0/λ_s as $N \to \infty$ ($\lambda_1 = 4, \lambda_2 = 5$). As the ratio λ_0/λ_s gets smaller, smaller M is needed to achieve $\xi_M \leq \epsilon$ for a given ϵ .

handled under the complexity constraint, (ii) approximate ξ_M using

$$\hat{\xi}_M = \frac{1}{1 + \sum_{i=1}^{M-1} \frac{\hat{\zeta}_i}{\lambda_0^i}}, \quad \hat{\zeta}_i = \prod_{j=1}^i \lambda_m^{k=M-j}$$
(12)

(iii) find smallest M satisfying $\hat{\xi}_M \leq \epsilon$. As an example, suppose N = 50, $\lambda_0 = 0.5$, $\lambda_1 = 8$, $\lambda_2 = 3$, $N_c = 20$ and $\epsilon = 10^{-4}$. Using the above approach we have obtained M = 7 and the sum of absolute differences between λ_m^k , $k = 1, \ldots, M$, for N_c tokens and the true markingdependent transition rates was 5.4×10^{-11} . The reduced Markov chain that was constructed using $N_c = 20$ tokens has only 7 states compared with $(N + 1)^2 = 2601$ states for the original Markov chain.

4.2 Case II: $\lambda_0 > \lambda_s$

Asymptotically this case results in small states to have vanishing state probabilities as illustrated in Fig. 7.



Figure 7: Reduced aggregated CTMC with N - K + 2 states as N tends to infinity for case II ($\lambda_0 > \lambda_s$).



Figure 8: Plot of ξ_K versus D = N - K for different λ_0/λ_s ($\lambda_1 = 4, \lambda_2 = 5$). As the ratio λ_0/λ_s gets smaller, larger D would be required to achieve $\xi_K \leq \epsilon$ for a given ϵ .

Referring to the reduced Markov chain in Fig. 7 we have

$$\lambda_0 \xi_i = \lambda_s \xi_{i+1}, \quad i = K, K+1, \dots, N \tag{13}$$

$$\sum_{i=K}^{N+1} \xi_i = 1.$$
 (14)

Solving the above equations for ξ_K we obtain

$$\xi_K = \frac{1}{\sum_{i=0}^{N-K+1} \left(\frac{\lambda_0}{\lambda_s}\right)^i} = \frac{\frac{\lambda_0}{\lambda_s} - 1}{\left(\frac{\lambda_0}{\lambda_s}\right)^{N-K+2} - 1}.$$
 (15)

We observe that ξ_i monotonically increases with *i* and that ξ_K only depends on the difference between *N* and *K* rather than their individual values so long as *N* is sufficiently large.

For sufficiently large N and given transition rates λ_0 , λ_1 and λ_2 complying with $\lambda_0 > \lambda_s$, a reduced aggregated CTMC can be approximately obtained by setting K to a value that returns a sufficiently small ξ_K ; i.e., $\xi_K \leq \epsilon$. It is clear from (15) that, for a fixed threshold ϵ , larger D = N - Kwould be necessary for smaller λ_0/λ_s ($\lambda_0/\lambda_s > 1$). This is illustrated in Fig. 8.

An approximate reduced aggregation can be obtained simply by determining the smallest D for given ϵ and λ_0/λ_s using (15). Thus the computational procedure for case II is much simpler than that for case I, provided that N is sufficiently large to warrant approximation of λ_m^k by λ_s . To demonstrate this, suppose N = 50, $\lambda_0 = 5$, $\lambda_1 = 8$, $\lambda_2 = 3$ and $\epsilon = 10^{-4}$, for which we have obtained K = 34 and the sum of absolute differences between λ_m^k , $k = K, \ldots, N$, (i.e., the true marking-dependent transition rates) and the asymptotic values λ_s was only 8.9×10^{-10} . The reduced Markov chain has D + 2 = 18 states, which is a significant reduction compared with the original Markov chain.

4.3 Case III: $\lambda_0 = \lambda_s$

No states can be assumed to have vanishing state probabilities in this case. Fig. 9 illustrates the transition diagram of the aggregated CTMC as N tends to infinity. Even though the aggregated Markov chain cannot be reduced, the asymptotic convergence result for the marking-dependent transition rates still holds. To see this first consider the local



Figure 9: Aggregated CTMC as N tends to infinity for case III ($\lambda_0 = \lambda_s$). No reduction is possible in this case since state probabilities do not vanish.

balance equations for a given N:

$$\lambda_0 \xi_1 = \gamma_1 \xi_2$$
$$\lambda_0 \xi_2 = \gamma_2 \xi_3$$
$$\lambda_0 \xi_3 = \gamma_3 \xi_4$$
$$\vdots$$
$$\lambda_0 \xi_{N-2} = \lambda_0 \xi_{N-1}$$
$$\lambda_0 \xi_{N-1} = \lambda_0 \xi_N$$
$$\lambda_0 \xi_N = \lambda_0 \xi_{N+1}$$

and

$$\sum_{i=1}^{N} \xi_i = 1.$$

The local balance equations imply that states, for which Observation 1 is valid, have identical state probabilities:

$$\cdots = \xi_{N-2} = \xi_{N-1} = \xi_N = \xi_{N+1}$$

We will now show that asymptotic convergence of transition rates to γ_i and validity of Observation 1 for large states can be maintained even if state probabilities do not exhibit any convergence. Consider increasing the number of tokens to N + 1. To see the effect of this on the state probabilities in conjunction with the asymptotic result (8), write

$$\lambda_0 \xi'_1 = \gamma_1 \xi'_2$$

$$\lambda_0 \xi'_2 = \gamma_2 \xi'_3$$

$$\lambda_0 \xi'_3 = \gamma_3 \xi'_4$$

$$\vdots$$

$$\lambda_0 \xi'_{N-2} = \lambda_0 \xi'_{N-1}$$

$$\lambda_0 \xi'_{N-1} = \lambda_0 \xi'_N$$

$$\lambda_0 \xi'_N = \lambda_0 \xi'_{N+1}$$

$$\lambda_0 \xi'_{N+1} = \lambda_0 \xi'_{N+2}$$

where the ξ'_i are the state probabilities after increasing the number of token in initial marking from N to N + 1. To satisfy the unit sum of state probabilities we have

$$\xi'_{N+2} = \alpha \xi_{N+1}, \quad 1 - \alpha = \xi'_{N+2}, \quad 0 < \alpha < 1$$
 (16)

where α is a scaling factor that multiplies all state probabilities ξ_i in order to allow the new state N + 2 to have a non-zero state probability that is equal to the new state probability of state N + 1. This way none of the converged transition rates are affected by the new state. From (16) we have

$$\alpha = \frac{1}{1 + \xi_{N+1}}, \quad \xi'_{N+2} = \frac{\xi_{N+1}}{1 + \xi_{N+1}}$$

For given λ_0 , λ_1 , λ_2 , N and N_c where $\lambda_0 = \lambda_s$, the following computational procedure can be used to approximate

the solution to aggregation: (i) compute λ_m^k , $k = 1, \ldots, N_c$, for $N_c < N$ tokens in initial marking, (ii) set the remaining transition rates $\lambda_m^k = \lambda_0$, $k = N_c + 1, \ldots, N$. Suppose $N = 50, \lambda_0 = 3, \lambda_1 = 8, \lambda_2 = 3$ and $N_c = 20$. Using the above procedure the sum of absolute differences between the approximated and true marking-dependent transition rates was 2.1×10^{-8} .

5. ACKNOWLEDGEMENT

This work was supported by Australian Research Council Discovery Project DP0880928.

6. CONCLUSIONS

We have presented a parametric convergence analysis for the transition rates of an aggregated Markov chain encountered in the modelling of parallel synchronized systems. The key to aggregation is to cluster the states of the underlying Markov chain into aggregated states in a systematic way that lends itself easily to mathematical analysis. Since parallel synchronized systems do not have product-form solutions [6], no closed-form solution can be found to determine their stationary distribution among other things. In the paper we have utilized an asymptotic convergence result for transition rates of the aggregated Markov chain to ease the computational burden associated with the solution. Three different cases have been identified in relation to the transition rates of the original Markov chain. For two of these cases whereby the slowest parallel branch has different transition rates to the transition rate of the initial place, we have shown that the aggregated chain can be reduced as a result of certain aggregated states having vanishing state probabilities. The third case where the slowest parallel branch and the initial place have identical transition rates does not allow any reduction. For each of these cases we have provided computational procedures and demonstrated their effectiveness by numerical examples.

REFERENCES

- [1] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. Princeton, NJ: Van Nostrand, 1960.
- [2] C. D. Meyer, "Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems," SIAM Review, vol. 31, no. 2, pp. 240–272, 1989.
- [3] T. G. Robertazzi, Computer Networks and Systems, 3rd ed. New York: Springer-Verlag, 2000.
- [4] M. A. Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis, *Modelling with Generalized Stochastic Petri Nets.* Chichester, England: Wiley, 1995.
- [5] K. Doğançay, "An asymptotic convergence result for the aggregation of closed fork-join generalized stochastic Petri nets," in *Proc. IEEE Region 10 Conference, TEN-CON 2009*, Singapore, December 2009.
- [6] E. Gelenbe and G. Pujolle, Introduction to Queueing Networks. Chichester, England: Wiley, 1987.