November 1992 LIDS-P-2158

TRANSIENT ANALYSIS OF MANUFACTURING SYSTEM PERFORMANCE

Y. Narahari
N. Viswanadham
Indian Institute of Science
Department of Computer Science and Automation
Bangalore 560 012 INDIA

Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, MA 02139, USA

The research of Y.Narahari is supported by an INDO–US Science and Technology Fellowship Program.

Transient Analysis of Manufacturing Systems Performance

Y.Narahari and N.Viswanadham
Indian Institute of Science
Department of Computer Science and Automation
Bangalore – 560 012, INDIA

November 1992

ABSTRACT

Studies in performance evaluation of automated manufacturing systems, using simulation or analytical models, have always emphasized *steady-state* or *equilibrium* performance in preference to *transient* performance. In this study, we present several situations in manufacturing systems where transient analysis is very important. Manufacturing systems and models in which such situations arise include: systems with failure states and deadlocks, unstable queueing systems, and systems with fluctuating or nonstationary workloads. Even in systems where equilibrium exists, transient analysis is important in studying issues such as accumulated performance rewards over finite intervals, first passage times, sensitivity analysis, settling time computation, and deriving the behavior of queueing models as they approach equilibrium. In certain systems, convergence to steady-state is so slow that only transient analysis can throw light on the system performance. After presenting several illustrative manufacturing situations where transient analysis has significance, we discuss two applications: (1) computation of distribution of time to absorption in Markov models of manufacturing systems with deadlocks or failures, and (2) computation of distribution of manufacturing cycle time in a failure-prone manufacturing system operated over a finite shift period. We also briefly discuss computational aspects of transient analysis.

1 INTRODUCTION

Studies in performance analysis of discrete manufacturing systems and in general, discrete event dynamical systems have traditionally emphasized *steady-state* or *equilibrium* performance over *transient* or *time-dependent* performance. This paper is concerned with transient analysis of manufacturing systems performance. Transient analysis is very important in manufacturing system models that do not attain a steady state or equilibrium. Examples of such systems include, systems with failure states, unstable queueing systems, and systems with fluctuating or non-stationary workloads. Even in systems where equilibrium does exist, transient analysis is important for studying performance over finite intervals, sensitivity analysis, first passage time computation, settling time computation, and for deriving the behavior of models as they approach equilibrium.

In this paper, we view a manufacturing system as a *discrete event dynamical system* [1, 2] and consider that the evolution of a manufacturing system constitutes a discrete state space stochastic process. In particular, we focus on Markov chain models. Such a model could be generated directly or using higher level models such as queueing networks, stochastic Petri nets, or discrete event simulation [2].

1.1 STEADY-STATE ANALYSIS

Steady-state analysis has been the focus of most performance studies in the area of discrete manufacturing systems. The two recent textbooks in this area, by Viswanadham and Narahari [2], and by Buzacott and Shantikumar [3] are concerned mostly with steady-state analysis. There are also many survey articles that discuss steady-state analysis of manufacturing systems using simulation modeling [4], Markov chain models [5], queues and queueing network models [6, 7, 8], and stochastic Petri net models [9, 10].

Steady-state analysis deals mainly with customer average measures or time average measures. Performance measures such as steady-state waiting time belong to the first category whereas measures such as steady-state number of jobs in system are time average measures. In the literature, much of the analysis deals with only mean values of these performance measures. Higher moments and distributions are only occasionally computed, for special classes of systems.

There are three main reasons for the popularity of steady-state analysis:

1. There are computationally efficient and simple methods for steady-

state analysis. For example, the computation of steady-state probabilities in a Markov chain is carried out by solving a system of linear equations; the computation of performance measures in product form queueing networks is accomplished through efficient polynomial-time algorithms; and so on. Availability of a wide variety of efficient linear equation solvers, including parallelized algorithms, has made possible the solution of Markov chains with several hundred thousand states.

2. Major results in queueing theory, such as Burke's result [11], Little's law [12], Jackson's theorem [13], product form of closed queueing networks [14], the BCMP formulation [15], and the arrival theorem [16] are all concerned with steady-state analysis.

3. Developments in aggregation and decomposition methods for solving large Markov chain models or large queueing models have also focused on steady-state analysis (see, for example, the paper by Curtois [17]).

Often, manufacturing system models do not have a steady-state or do not reach a steady-state in the observation period of interest. Transient analysis becomes important in such situations. In Section 2, we will be looking at several such situations.

1.2 TRANSIENT ANALYSIS

Let us assume that a manufacturing system evolves in time as a homogeneous continuous time Markov chain (CTMC) $\{X(t) : t \geq 0\}$ with state space $S = \{0, 1, \ldots\}$ and infinitesimal generator Q. Let $i, j \in S$ and

$$p_{ij}(t) = P\{X(t) = j | X(0) = i\}$$

$$H(t) = [p_{ij}(t)]$$

The forward and backward differential equations that govern the behavior of this CTMC are respectively given by [18, 19, 2],

$$\frac{d}{dt}(H(t)) = H(t)Q \tag{1}$$

$$\frac{d}{dt}(H(t)) = QH(t) \tag{2}$$

with initial conditions $H(0) = I$ in both the cases. Note that these are first order, linear, ordinary differential equations. In terms of the individual matrix elements, the above equations become

$$\frac{d}{dt}(p_{ij}(t)) = q_{jj}p_{ij}(t) + \sum_{k \neq j} q_{kj}p_{ik}(t) \tag{3}$$

$$\frac{d}{dt}(p_{ij}(t)) = q_{ii}p_{ij}(t) + \sum_{k \neq i} q_{ik}p_{kj}(t) \tag{4}$$

The forward and backward equations have the same unique solution given by

$$H(t) = e^{Qt} \tag{5}$$

where e^{Qt} is the matrix exponential defined by the Taylor series

$$e^{Qt} = \sum_{k=0}^{\infty} \frac{(Qt)^k}{k!} \tag{6}$$

If we are interested in the state probabilities

$$\Pi(t) = [p_0(t), p_1(t), \ldots]$$

where $p_j(t) = P\{X(t) = j\}, j \in S$, then we need to solve the differential equation

$$\frac{d}{dt}(\Pi(t)) = \Pi(t)Q \tag{7}$$

The solution of the above is given by

$$\Pi(t) = \Pi(0)e^{Qt} \tag{8}$$

1.2.1 An Example

To get a feel for the equations above, let us consider a simple example [19, 2]. Consider a manufacturing system comprising a single machine that fails with failure time exponentially distributed with rate λ and gets repaired, once failed, with repair time exponentially distributed with rate μ. Assuming that the failure and repair times are independent, the system can be formulated as a CTMC with state space $S = \{0, 1\}$ where state 0 indicates, say, "machine in the up condition" and state 1 denotes "machine undergoing

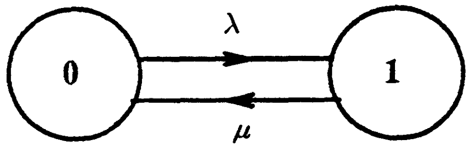

Figure 1: Markov chain model of a single machine system

repair." Figure 1 depicts the state diagram of this Markov chain. For this example, we have

$$H(t) = \begin{bmatrix} p_{00}(t) & p_{01}(t) \\ p_{10}(t) & p_{11}(t) \end{bmatrix}$$

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

The forward equations (1) in this case are given by

$$\frac{d}{dt}(p_{00}(t)) = p_{00}(t)q_{00} + p_{01}(t)q_{10}$$

$$\frac{d}{dt}(p_{01}(t)) = p_{01}(t)q_{11} + p_{00}(t)q_{01}$$

$$\frac{d}{dt}(p_{10}(t)) = p_{10}(t)q_{00} + p_{11}(t)q_{10}$$

$$\frac{d}{dt}(p_{11}(t)) = p_{11}(t)q_{11} + p_{10}(t)q_{01}$$

The backward equations are given by

$$\frac{d}{dt}(p_{00}(t)) = q_{00}p_{00}(t) + q_{01}p_{10}(t)$$

$$\frac{d}{dt}(p_{01}(t)) = q_{01}p_{11}(t) + q_{00}p_{01}(t)$$

$$\frac{d}{dt}(p_{10}(t)) = q_{10}p_{00}(t) + q_{11}p_{10}(t)$$

$$\frac{d}{dt}(p_{11}(t)) = q_{11}p_{11}(t) + q_{10}p_{01}(t)$$

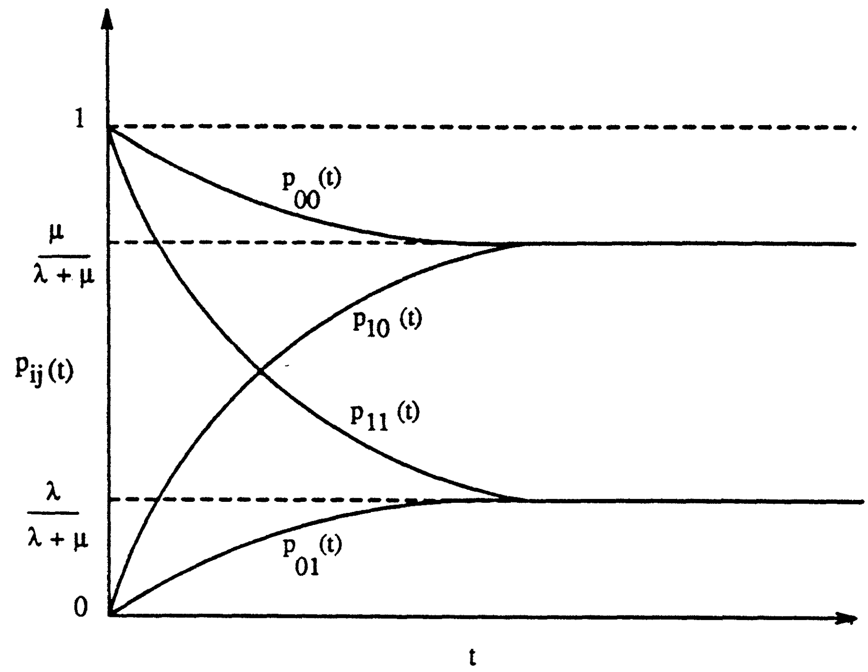

Figure 2: Evolution of transition probabilities

The solution of the coupled differential equations above is straightforward and it can be shown that the transition probabilities are given by

$$p_{00}(t) = \frac{\mu}{\lambda + \mu} + (\frac{\lambda}{\lambda + \mu})e^{-(\lambda+\mu)t} \tag{9}$$

$$p_{01}(t) = \frac{\lambda}{\lambda + \mu} - (\frac{\lambda}{\lambda + \mu})e^{-(\lambda+\mu)t} \tag{10}$$

$$p_{10}(t) = \frac{\mu}{\lambda + \mu} - (\frac{\mu}{\lambda + \mu})e^{-(\lambda+\mu)t} \tag{11}$$

$$p_{11}(t) = \frac{\lambda}{\lambda + \mu} + (\frac{\mu}{\lambda + \mu})e^{-(\lambda+\mu)t} \tag{12}$$

Figure 2 illustrates the evolution of these state probabilities. Note that

$$\lim_{t\to\infty} p_{00}(t) = \lim_{t\to\infty} p_{10}(t) = \frac{\mu}{\lambda + \mu}$$

$$\lim_{t\to\infty} p_{11}(t) = \lim_{t\to\infty} p_{01}(t) = \frac{\lambda}{\lambda + \mu}$$

The above limiting probabilities are precisely the steady-state probabilities π_0 and π_1 of the states 0 and 1, respectively. For $j = 0, 1$, the state probabilities $p_j(t)$ are given by

$$p_j(t) = p_{0j}(t)p_0(0) + p_{1j}(t)p_1(0) \tag{13}$$

1.2.2 Relevant Literature

Literature on transient analysis of Markov chain models is vast and is scattered across several inter-disciplinary areas. We shall only mention here some papers that are of direct interest.

Grassman's aticle [20] is an authentic survey on transient analysis whereas the paper by Stewart [21] discusses numerical techniques for transient analysis. More recently, Reibman and Trivedi [22, 23] have surveyed the numerical techniques while Marie *et al* [24] have discussed the transient analysis of acyclic Markov chains. Bobbio and Trivedi [25, 26] have discussed an aggregation method for transient analysis of Markov chains.

Reliability and availabilty modeling has been a major motivating factor for conducting transient analysis. For example, see the papers by Reibman *et al* [27], Bavuso *et al* [28], and de Souza de Silva and Gail [29, 30]. Analysis of fault-tolerant computer systems and performability modeling have also spurred several research efforts in transient analysis. For example, see the works by de Souza de Silva and Gail [30], Gerber [31], Meyer [32], and Trivedi *et al* [33].

Transient analysis of queueing models arising in computer and communication systems is the subject of the works by Baiocchi *et al* [34], Kotiah [35], Konstantopoulis and Baccelli [36], Tripathi and Duda [37], Upton and Tripathi [38], Massey [39], Weiss and Mitra [40], and Kobayashi [41].

In the manufacturing context, some work on transient analysis has been reported in the works of Ram [42], Manjunath [43], Viswanadham and Narahari [2], Viswanadham *et al* [44], and Malhame and Boukas [45]. Malhame and Boukas look at the statistical evolution of a manufacturing system producing a single product, under hedging point control policies. They formulate partial differential equations that describe this evolution and show that transient analysis is very important here since the convergence to steady state is very slow.

The aim of this paper is to spell out clearly the need for transient analysis of manufacturing system models and to explore the major issues of relevance.

1.3 Organization of the Paper

In this section, we have introduced the transient analysis problem in performance modeling. In the next section, we discuss several situations in manufacturing systems analysis where transient analysis is relevant. We discuss these under four categories:

1. Systems where steady state does not exist.
2. Models with absorbing states.
3. Performance computation over finite time durations.
4. Other important transient phenomena.

In Section 3, we present two illustative examples. The first is concerned with the computation of *time to absorption* in Markov models with absorbing states. This analysis can be used to study manufacturing systems with deadlocks and systems with total failure states. The second example addresses the important problem of computing, over finite intervals of time, the distribution of *manufacturing cycle time* in a failure-prone manufacturing system.

in Section 4, we briefly touch upon important computational issues in transient analysis. In Section 5, we provide a summary of the paper.

2 WHY TRANSIENT ANALYSIS ?

The aim of this section is to provide various situations in manufacturing system analysis where transient analysis assumes much significance.

2.1 SYSTEMS WITH NO STEADY STATE

It is only in special classes of Markov chain models, such as *ergodic* Markov chains, that a unique steady state or equilibrium exists. We now give some examples where a steady state does not exist.

Example 1: An Unstable Queue.
Consider an M/M/1 queue with arrival rate λ and service rate μ. The queue is stable if and only if $\lambda < \mu$ and steady-state performance measures will be meaningful only in this case. When $\lambda = \mu$, it is known that the underlying

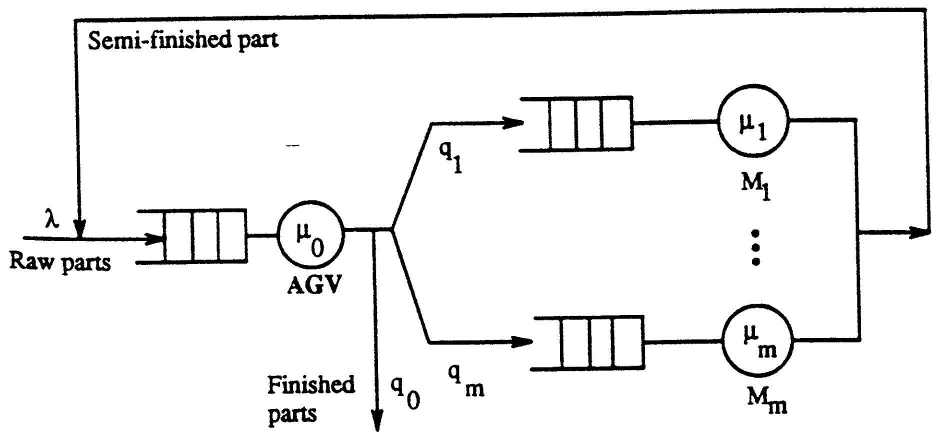

Figure 3: Open central server queueing network model

Markov chain states are all *transient* [46] and the number of customers in the system grows to infinity in the long term. If $\lambda > \mu$, all the states are *null recurrent* and the system is again unstable. Similar arguments hold for any single or multiple server queueing system.

Example 2: An Unstable Queueing Network.
Consider the *open central server queueing network* model shown in Figure 3. This is a very popular model of flexible manufacturing systems [47, 7, 3, 2]. This network is a special class of a *Jackson network* [13]. If λ is the external arrival rate of jobs and $\mu_i (i = 0, 1, \ldots, m)$ are the service rates (see Figure 3), it is known that the above network is stable if and only if $\rho_j < 1$ for all $j = 0, 1, \ldots, m$, where

$$\rho_0 = \frac{\lambda}{q_0 \mu_0}$$

$$\rho_j = \frac{\lambda q_j}{q_0 \mu_j} \qquad j = 1, \ldots, m$$

If even one of these conditions is not satisfied, the network is unstable and steady-state analysis loses significance.

Example 3: A Kanban Cell with Non-Stationary Demands.
Mitra and Mitrani [48] have studied the performance of a linear network of Kanban cells, subjected to stochastic demands. Figure 4 depicts a single Kanban cell subjected to external demands. The input to the machines is modulated by the arrival processes of demands and raw parts. Mitra and

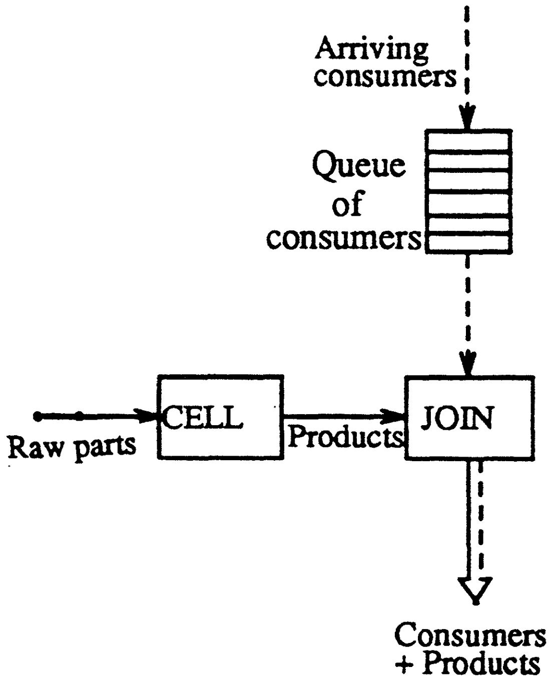

Figure 4: A Kanban cell subjected to external demands

Mitrani [48] assume that the demands for finished parts arrive according to a Poisson process. However, in the real-world context, the demands arrive in very complex fashion and the workload to the system is highly *non-stationary* . For example, during rush hours, the demands arrive rapidly and during other times, their arrival follows some stochastic pattern. The underlying queueing system belongs to the realm of non-stationary queues and the system here may be unstable or stable depending on the maximum rate of arrivals of demands and raw parts. There is a rich body of literature in the area of non-stationary queueing systems [49], where the issue of stability has been resolved for a very limited class of models.

Example 4 : Re-Entrant Lines.
Re-entrant lines [50] constitute a class of manufacturing systems models where the flows are *non-acyclic* since the parts visit the same machines several times. These are characteristic of semiconductor and thin film manufacturing. Scheduling is an important problem in these systems and several distributed policies based on buffer priorities and due dates have been formulated for these systems (see, for example, the papers by Kumar [50] and by Lu and Kumar [51]. Stability is an important issue in evaluating these scheduling policies. Not all the policies suggested in the above papers are

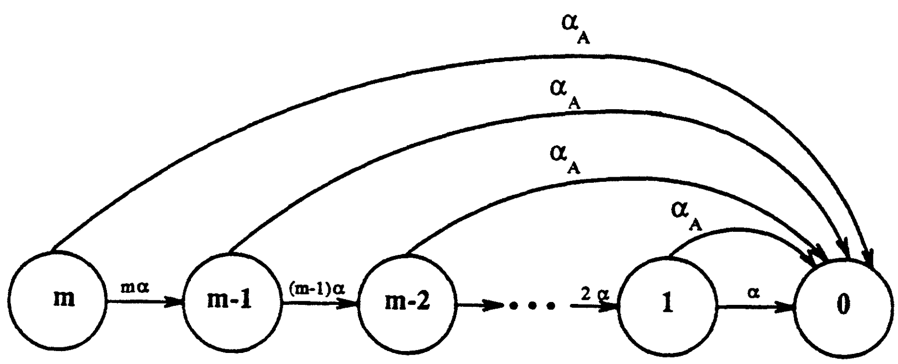

Figure 5: Markov chain model for failure-repair behavior

stable [50, 51] and performance analysis of re-entrant lines under such unstable policies can only be carried out via transient analysis.

2.2 MODELS WITH ABSORBING STATES

Markov models with absorbing states have a trivial steady-state, namely that the chain ends up in some absorbing state, remaining there forever; therefore, transient analysis alone throws some light on the system performance. We consider two examples below.

Example 5 : Reliability Analysis.
Manufacturing systems with no or limited repair of failed elements will lead to models with absorbing states. In such systems, reliability is an important performance index. Consider, for instance, a manufacturing system with m identical machines and an automated guided vehicle (AGV). Both the machines and the AGV are failure-prone and let us assume that repair is not possible. If the failure times are all independent exponential random variables, then the model that describes the failure-repair behavior of this system is a Markov chain. It is reasonable to assume that the system is *operational* only when the AGV is "up" and at least one machine is "up" (this is because the AGV is involved in the successful completion of processing of every job). In such a case, the Markov chain model has state space $S = \{0, 1, \ldots, m\}$, where state 0 corresponds to the *failed state* (all machines are down or AGV is down or both) and state i ($i = 1, \ldots, m$) indicates AGV "up" and exactly i machines "up." Figure 5 shows this Markov

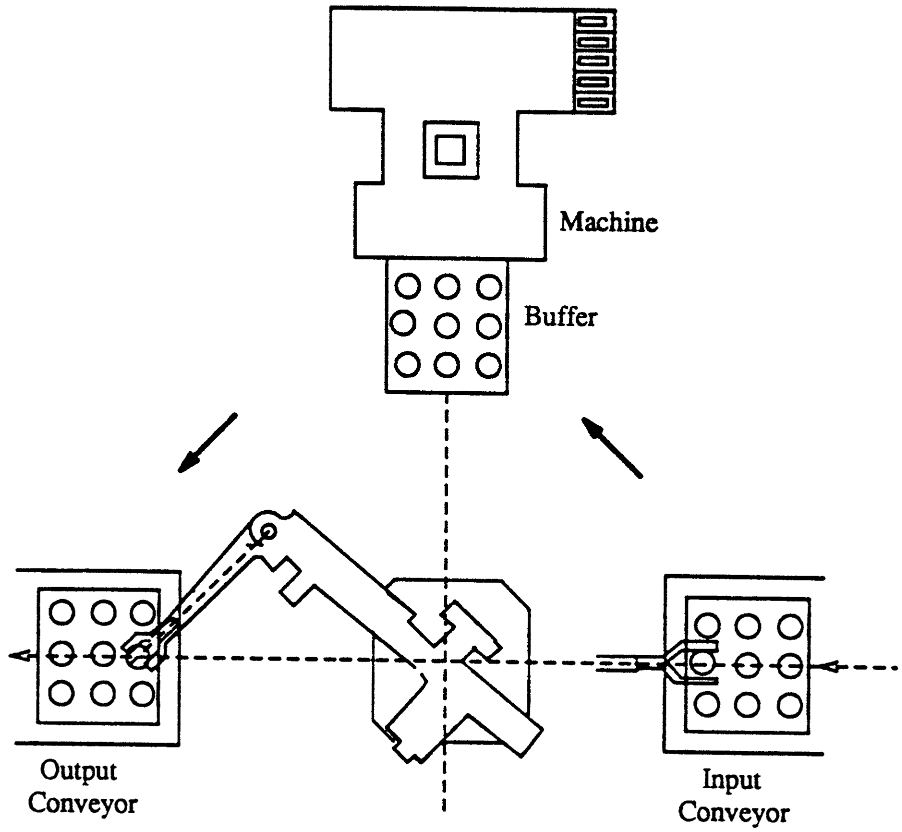

Figure 6: A robotic cell to illustrate deadlock

chain model, assuming λ_A as the AGV failure rate and λ as the failure rate of each machine. This same model is discussed in depth in [44]. State 0 is an absorbing state and the *reliability* of this system at time t is the probability that the system is not in state 0 at time t, given some initial condition. The reliability in this case can only be computed through transient analysis.

Example 6 : A Manufacturing System with Deadlocks.
This example is taken from [2]. Consider the robotic cell shown in Figure 6, where there is a single machine that produces parts, with processing time exponentially distributed with rate μ. Raw parts arrive on to an input conveyor according to a Poisson process having rate λ. A robot picks up a raw part from the input conveyor and loads it on to the machine if the machine is free or to its buffer if the machine is busy. The robot picks up the finished part and puts it on the output conveyor. Assume that arrival of raw parts into the system is inhibited whenever the machine is busy, the buffer is full, and the robot is holding a raw part. Hence, if the buffer capacity is n, the maximum number of jobs inside the system is $n + 2$. Let us assume that the robot takes negligible time to load and unload parts.

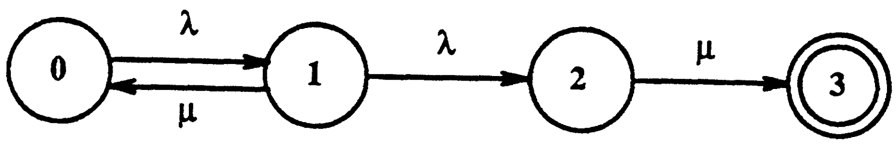

Figure 7: Markov chain model of the robotic cell

First, consider the case where there is no buffer. Here, the states of the system are $0, 1, 2, 3$, with the following interpretation:

- 0: no raw parts; machine idle.
- 1: machine processing a part, no raw parts waiting.
- 2: machine processing a part, robot holding a raw part.
- 3: machine waiting for the robot to transfer the finished part and the robot waiting for the machine to release the finished part.

The CTMC model of the above system is shown in Figure 7. In state 3, the waiting is indefinite if we assume that the robot controller and the machine controller are not programmed to react to such mutual or circular waiting. Such a state is called a *deadlock* , which stalls further activity and production in the system. In this simple example, it is easy to see how the deadlock may be prevented, but in a real-world manufacturing system having a large number of resources and concurrent interactions, deadlocks can occur commonly.

State 3 is an absorbing state in Figure 7. If we need to compute the distribution of time before the deadlock is reached or the number of parts produced before deadlock, transient analysis becomes important.

In the above example, if there is a buffer in front of the machine, the number of states will increase; in fact, if the buffer capacity is n, there will be exactly $n + 4$ states in the model and state $n + 3$ will be the absorbing state.

2.3 PERFORMANCE IN FINITE INTERVALS

In a manufacturing system, we would often be interested in computing the cumulative performance in a finite duration of time, for example in a shift period. It is not realistic to expect the system to reach a steady state during this finite observation period. We consider three examples below.

Example 7: A Wafer Fabrication Line.
In a typical semiconductor wafer fabrication line [52, 53], each *lot* of wafers goes through a large number of operations and spends several days, inside clean rooms, repeatedly visiting many workcenters. The typical cycle time and queueing time of a *lot* of wafers is much larger compared to a shift duration. Therefore, if we are interested in the production or congestion levels at the end of a shift duration, we cannot rely on steady-state performance estimates. Furthermore, some scheduling policies in such *re-entrant lines* are known to be unstable (see Example 4) and transient analysis becomes even more important.

Example 8: Interval Dependability Measures.
Fault-tolerance and flexibility are the prime attributes of advanced manufacturing systems. The degree of fault-toleance of a manufacturing system is characterized by *dependability* measures such as reliability and availability. To define these measures, we partition the system states into *operational* states (states in which the system produces useful output) and *failed* states. Given an interval $[0, t]$, the *reliability* of the system is the probability that the system never reaches a failed state during that interval. The *point availability* at time $u \in [0, t]$ is the probability that, at time u, the system is in an operational state. The *interval availability* is the fraction of time during $[0, t]$, the system is in operational states. To compute these measures, one needs to do transient analysis.

As an illustative example, we consider a manufacturing system comprising two machines M_1 and M_2 (this example is taken from [44]). Let the failure times of M_i $(i = 1, 2)$ be exponentially distributed with rate α_i and be independent. When a machine fails, assume that repair starts immediately, with repair time for machine M_i being an exponential random variable having rate β_i. The failure-repair behavior of this system is a Markov chain with four states given by

$$S = \{(11), (10), (01), (00)\}$$

where each state is a pair (x_1, x_2), with $x_i = 1$ when M_i is "up" and $x_i = 0$ when M_i is "down." Figure 8 shows this Markov chain. Obviously, the set of operational states is given by

$$S_o = \{(11), (10), (01)\}$$

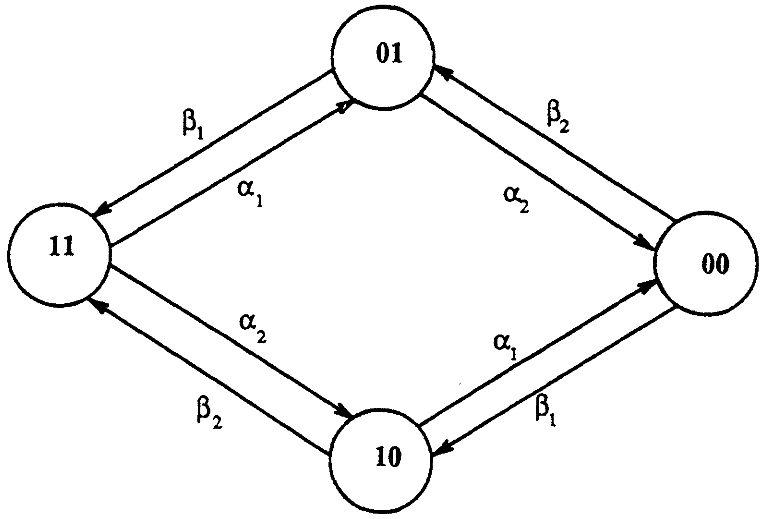

Figure 8: Failure-repair model of a two-machine system

and the set of failed states is given by

$$S_f = \{(00)\}$$

Let $\{Z(u) : u \geq 0\}$ be this Markov chain. Given an interval $[0, t]$, the reliability $R(t)$ is given by

$$R(t) = P\{Z(u) \in S_o \forall u \in [0, t]\}$$

The point availability is given by

$$PA(u) = P\{Z(u) \in S_o\}$$

The interval availability is given by

$$IA(t) = \frac{1}{t} \int_0^t P\{Z(u) \in S_o\}$$

The above failure-repair process is often referred to as the *structure state process* [44].

Example 9: Performability Measures.
Performability is a generic; composite measure of performance and dependability. There is a vast literature on performability of computer and communication systems [30]. More recently, performability has been investigated in the manufacturing systems context also [44].

We shall give a simple example, based on the system in Example 8. Assume that raw parts are always available and that parts undergo exactly one operation, either on M_1 or on M_2, and leave the system. Also, assume that machine M_i processes parts at rate μ_i. Then in state (11), the total production rate is $\mu_1 + \mu_2$. The production rates in states (10), (01), (00) are respectively, μ_1, μ_2, and zero. During the interval $[0, t]$, let $\tau_{11}, \tau_{10}, \tau_{01}, \tau_{00}$ be the total times spent the corresponding states. Note that these are random variables. The total accumulated production in the interval is then given by

$$Y(t) = \tau_{11}(\mu_1 + \mu_2) + \tau_{10}\mu_1 + \tau_{01}\mu_2$$

$Y(t)$ is called the throughput-related performability. In general, performability could be with respect to any performance measure such as throughput, lead time, queueing time, etc. To compute the distribution of $Y(t)$, one needs to do transient analysis.

2.4 OTHER TRANSIENT PHENOMENA

There are many other aspects of manufacturing system performance that can be effectively addressed only by transient analysis.

Performance under Real-Time Control Policies

When real-time control decisions are taken, for example, in the dynamic scheduling of manufacturing system operations, it is of intrinsic interest to look at the transient performance, especially if the evolution is such that it takes a long time before a steady state is reached. For instance, Malhame and Boukas [45] have considered the operation of a failure-prone, single-product manufacturing system under dynamic hedging point control policies. They characterize the transient performance using a system of coupled partial differential equations.

Settling Time of Queueing Systems

The settling time of a queueing system with a given initial number of customers in the system is the total time until the number in the system is zero. There have been a few efforts at computing the distribution of settling time of multiserver queues and open queueing networks [54, 55, 56].

The notion of settling time is analogous to the *makespan* of a manufacturing network, which is the total amount of time required to complete the processing of a given number of workpieces. Makespan computation is quite important in stochastic manufacturing systems.

Sensitivity Analysis

It is often required to determine the performance or reliability *bottleneck* of a system. In this context, it is necessary to evaluate the derivative of the desired performance measure with respect to important system parameters. The parameter with the largest derivative deserves the attention of the designers to improve the characteristics of the designed system. Such derivatives can also be used in a system optimization effort based on gradient search techniques. Sensitivity analysis often relies on transient analysis of performance.

Cut-Off Phenomenon

An interesting quantity to study in the evolution of a stochastic manufacturing system is the rate at which the steady state is approached. This depends on the time constants (eigen values) of the system [40]. There is a class of Markov chain models and queueing systems (for example, see the articles by Konstantopoulos and Baccelli [36] and Anantharam [57]) which exhibit a *cut-off phenomenon* namely, the existence of a time such that before this time, the system is far from steady state, while, after this time, the system is very close to steady state. The existence of cut-off phenomenon is a good indicator to whether a transient or a steady-state analysis is appropriate in a given setting.

3 DETAILED EXAMPLES

In this section, we illustrate transient analysis of manufacturing systems using two examples. In the first, we show the computation of distribution of *time to absorption* in a Markov model with absorbing states. In the second, we show how transient analysis may be carried out to compute distribution of cumulative measures of performance such as *manufacturing cycle time* in a failure-prone manufacturing system.

3.1 TIME TO ABSORPTION

We have observed in Section 2.2 that absorbing states occur in manufacturing system models that capure non-repairable behavior and phenomena such as deadlocks. An important quantity of interest in such systems is the time until an absorbing state is reached. Let $\{X(u) : u \geq 0\}$ be the Markov

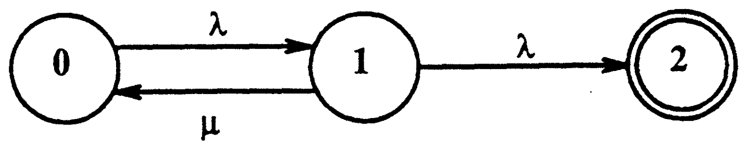

Figure 9: A Markov chain with an absorbing state

chain under consideration. Let the state space be finite and given by

$$S = \{0, 1, \ldots, m, m+1, \ldots, m+n\}$$

where $m \geq 0$, $n > 0$, the first $(m+1)$ states are transient states, and the rest of the states are absorbing states. Let 0 be the initial state and T, the time to reach any absorbing state. Define

$$p_{ij}(t) = P\{X(t) = j | X(0) = i\}$$

Then, we have, for any $t > 0$,

$$P\{T > t\} = P\{X(t) \notin \{m+1, \ldots, m+n\}\}$$

In other words, we have

$$P\{T > t\} = 1 - \sum_{j=1}^{n} p_{0,m+j}(t)$$

Hence the cumulative distribution function of T is given by

$$F_T(t) = \sum_{j=1}^{n} p_{0,m+j}(t) \tag{14}$$

The individual probabilities $p_{0,m+j}(t)$ have to be computed by solving the differential equations shown in equation (1) or equation (2).

We now show the computation of the distribution of time to absorption for a simple Markov chain. consider the Markov chain of Figure 9. There are two possible interpretations for the above model. In the first interpretation, we have a single machine system which is in state 0 when there is no part being processed, in state 1 when there is a part being processed, and in state 2 when there is a deadlock. The arrival rate of parts is λ and the service rate of each part is μ. This interpretation is similar to Example 6. The time to absorption here is the time elapsed before a deadlock is reached.

In the second interpretation, we consider a two-machine system with exponential failures and repairs. In state 0, both machines are "up" but only one of them is chosen to process parts. When this chosen machine fails, the system reaches state 1, in which the non-failed machine starts processing parts and the repair of the failed machine is in progress. If the non-failed machine now fails before completion of repair of the already failed machine, we reach state 2 and we abandon any further repair. On the other hand, if the failed machine in state 1 is repaired before the non-failed machine fails, we return to state 0. State 2 corresponds to a total failure state and the time to absorption corresponds to the time to total failure.

We know in this case that $F_T(t) = p_{02}(t)$. To compute $p_{02}(t)$, we first write down the infinitesimal generator Q of this Markov chain:

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & 0 & 0 \end{bmatrix}$$

First consider the backward equation (4) for $p_{02}(t)$:

$$\frac{d}{dt}(p_{02}(t)) = q_{00}p_{02}(t) + q_{01}p_{12}(t) + q_{02}p_{22}(t)$$

Since $q_{02} = 0$, the above becomes

$$\frac{d}{dt}(p_{02}(t)) = -\lambda p_{02}(t) + \lambda p_{12}(t)$$

The backward equation for $p_{12}(t)$ is given by

$$\frac{d}{dt}(p_{12}(t)) = q_{10}p_{02}(t) + q_{11}p_{12}(t) + q_{12}p_{22}(t)$$

Since $p_{22}(t) = 1$, the above becomes

$$\frac{d}{dt}(p_{12}(t)) = \mu p_{02}(t) - (\lambda+\mu)p_{12}(t) + \lambda$$

We shall solve for $p_{02}(t)$ by the Laplace transform method. Let $P_{ij}(s)$ denote the Laplace transform of $p_{ij}(t)$. Taking the transform on either side of the equations above, we get

$$sP_{02}(s) = -\lambda P_{02}(s) + \lambda P_{12}(s) \tag{15}$$

$$sP_{12}(s) = \mu P_{02}(s) - (\lambda+\mu)P_{12}(s) + \frac{\lambda}{s} \tag{16}$$

Simplifying using (15) and (16), we get

$$P_{02}(s) = \frac{\lambda^2}{s(s^2 + s(2\lambda + \mu) + \lambda^2)} \tag{17}$$

Now, $p_{02}(t)$ can be obtained from equation(17) by inverse Laplace transformation. It is a simple matter to show that

$$p_{02}(t) = A + Be^{-at} + Ce^{-bt}$$

where the constants are given by

$$a = \frac{2\lambda + \mu + \sqrt{\mu^2 + 4\lambda\mu}}{2}; \qquad b = \frac{2\lambda + \mu - \sqrt{\mu^2 + 4\lambda\mu}}{2}$$

$$A = \frac{\lambda}{ab}; \qquad B = \frac{\lambda(b - 2a)}{ab(b-a)}; \qquad C = \frac{\lambda}{b(b-a)}$$

In the above case, we were able to give a closed form expression for the cumulative distribution function of time to absorption, only because of the small number of states and simple structure. In general, this computation is a formidable task and in fact, is the subject of several research efforts. The problem is identical to computation of first passage times in Markov chains [58]. Marie, Reibman, and Trivedi have given a general way of obtaining such distributions efficiently for acyclic Markov chains [24]. There are several software tools that have been developed in this context and we will be briefly covering those in Section 4.

3.2 MANUFACTURING CYCLE TIME

Here, we study the performability of a two machine system. This example is adapted from the paper by Reibman [59]. The system comprises two identical machines operated as an M/M/2 queueing system with an FCFS discipline. Raw parts arrive according to a Poisson process with rate λ and the service time of each part is exponentially distributed with rate μ. The time to failure of each machine is exponentially distributed with rate α. A failed machine is repaired, the repair time being exponentially distributed with rate β.

Our aim here is to find the distribution of the manufacturing lead time (MLT) or manufacturing cycle time of a typical part. The MLT of a part is the total time elapsed between the arrival of the part into the system

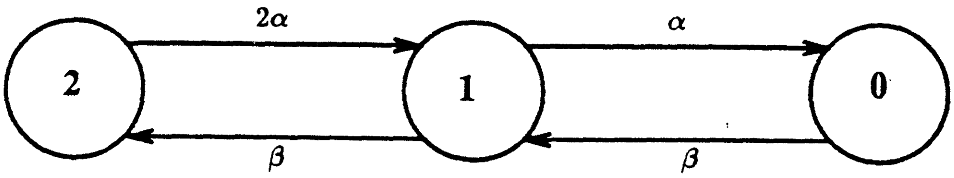

Figure 10: Failure-repair model of a two machine system

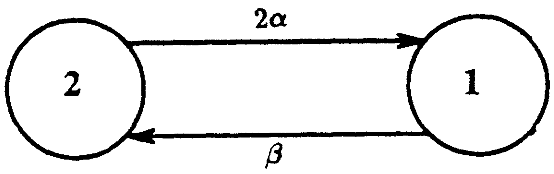

Figure 11: Simplified structure state model

and its eventual departure after finishing service. Note that an arriving raw part may find both machines "up", or exactly one machine "up", or both machines "down." Assuming that there is only one repair facility, the Markov chain model of the failure-repair behavior is shown in Figure 10. State i ($i = 0, 1, 2$) corresponds to the number of *operational* machines. Assuming α to be much less than β, the probability of having both machines down is extremely small. To simplify the example, we shall ignore state 0 and henceforth consider the structure state model shown in Figure 11. Let state 2 be the initial state and let $p_i(t)$ denote the probability of being in state i at time $t > 0$. Similar to the example in Section 1.2.1, we can write down the differential equations for computing the state probabilities. The following expressions can be easily obtained:

$$p_2(t) = \frac{\beta}{2\alpha + \beta} + \frac{2\alpha}{2\alpha + \beta} e^{-(2\alpha+\beta)t} \tag{18}$$

$$p_1(t) = \frac{2\alpha}{2\alpha + \beta} - \frac{2\alpha}{2\alpha + \beta} e^{-(2\alpha+\beta)t} \tag{19}$$

In the limit $t \to \infty$, the time-dependent probabilities $p_2(t)$ and $p_1(t)$ converge to the steady-state probabilities $\frac{\beta}{2\alpha+\beta}$ and $\frac{2\alpha}{2\alpha+\beta}$, respectively. The approach to steady state is governed by the time constant, which in this case is $(2\alpha + \beta)$.

From the above expressions, we can obtain the cumulative occupancy times in the two states, i.e., the total times during $[0, t]$ the Markov chain stays in states 1 and 2. These are obtained as

$$L_2(t) = \int_0^t p_2(u)du = \frac{\beta t}{2\alpha+\beta} + \frac{2\alpha}{(2\alpha+\beta)^2}(1 - e^{-(2\alpha+\beta)t}) \qquad (20)$$

$$L_1(t) = \int_0^t p_1(u)du = \frac{2\alpha t}{2\alpha+\beta} - \frac{2\alpha}{(2\alpha+\beta)^2}(1 - e^{-(2\alpha+\beta)t}) \qquad (21)$$

The two expressions above sum to t as can be easily verified. The state probabilities $p_1(t)$ and $p_2(t)$, and the cumulative occupancy times $L_1(t)$ and $L_2(t)$, computed above, tell us about the transient characteristics of the failure-repair model.

Let us now compute the manufacturing lead time. Let T be the MLT of a typical part in the above system, when the system is operated over an interval $[0, t]$. Let T_1 and T_2 be the MLT random variables in states 1 and 2, respectively. Then,

$$T = p_1(t)T_1 + p_2(t)T_2 \qquad (22)$$

To compute T_1 and T_2 in (22), we proceed as follows. In state 2, the system exhibits the same performance as an M/M/2 queue with parameters λ and μ, whereas, in state 1, the performance corresponds to that of an M/M/1 queue with the same parameters. If we assume that the system reaches a steady state during each visit to state 1 and state 2, then T_1 can be taken as the steady-state waiting time in system in a stable M/M/1 queue, whereas, T_2 corresponds to the steady-state waiting time in a stable M/M/2 queue. Of course, for this to happen, a necessary condition is that λ be less than μ. It is a standard result [18, 46] that T_1 is exponentially distributed with rate $\mu - \lambda$, while T_2 has been shown to have the following density function [59]:

$$f_{T_2}(t) = 1 + \frac{2\mu - \lambda}{(2\mu+\lambda)(\mu-\lambda)}\left(\frac{\lambda^2}{2\mu-\lambda}e^{-(2\mu-\lambda)t} - \mu e^{-\mu t}\right)$$

Thus we can use (22) to obtain the distribution of manufacturing cycle time over a finite interval $[0, t]$.

4 COMPUTATIONAL ISSUES

In transient analysis, we are interested in computing the transition probabilities $p_{ij}(t)$ or state probabilities $p_i(t)$ or cumulative performance measures over finite time intervals. To obtain the transition probabilities, we need to solve equations (1) or (2), and to obtain state probabilities, we need to solve equations given by (3). These are coupled, linear, first order, ordinary differential equations. The computation of cumulative measures also involves solving linear differential equations [23]. There are three basic ways in which the above differential equations may be solved:

1. Obtain a general solution by deriving and symbolically inverting Laplace transforms. Analytic Laplace transform inversion requires that the eigen values of the infinitesimal generator of the Markov model be accurately determined. If the size of the state space is N, this would have a worst-case computational complexity of $O(N^5)$.

2. Evaluate the matrix exponential series (6) directly. This approach is however beset with numerical instabilities, such as severe round off errors [22].

3. Numerically solve the differential equations using well developed techniques such as the fourth fifth order Runge-Kutta method, or the TR–BDF2 method (Trapezoid Rule– Second order Backward Difference) [22].

The above methods are not always tractable and other numerical methods have been proposed for transient analysis. Among these, *Uniformization* or *randomization* [60] has assumed prominence as an excellent numerical tool. There are also approximate techniques based on, for example, aggregation and decomposition [25, 26] and diffusion approximations [41]. In some special cases, exact closed form expressions can be obtained for transient measures, such as in acyclic Markov chains [24].

There are excellent review articles dealing with computational aspects of transient analysis. The papers by Grassman [20] and Stewart [21] are two of the earliest ones. More recently, Reibman and Trivedi [22, 23] have done a neat survey of numerical transient analysis techniques for transition probabilities, state probabilities, and cumulative measures. The article by Johnson and Malek [61] is a detailed survey on software packages for reliability and availability evaluation; many of these packages, in fact, carry out transient analysis. Much of the following discussion is based on these survey articles.

4.1 COMPUTATIONAL DIFFICULTIES

There are mainly three problems that one is confronted with in transient analysis: *largeness* , *stiffness* , and *ill-conditioning* [22].

State Space Explosion

Markov models of real-world manufacturing systems will have a large number of states, often exceeding tens of thousands. So, even an algorithm of low polynomial complexity can become intractable. Also, this will call for a large amount of storage, though, often the matrices are sparse. If the algorithms preserve the sparsity of the matrices involved, savings in storage can be obtained.

Stiffness

In a manufacturing system, the activities fall into different time scales. For example, operation times are typically much smaller compared to mean time to failure or mean time to repair. Set-up times, depending on the specific system, may be much larger or much smaller than other activity durations. The result is, the transition rates in the Markov chain model will exhibit several orders of magnitude difference. This causes the problem of stiffness. In general, we say a system of differential equations is stiff on the interval $[0, t]$ if there exists a solution component that has variation on that interval that is large compared to $\frac{1}{t}$ [22]. A component with large variation changes rapidly relative to the length of the interval.

Ill-Conditioning

Manufacturing system models often lead to transition rate matrices that are ill-conditioned. That is, small changes in the matrix elements can produce large changes in the solution. This will lead to inaccurate estimation of transient performance.

4.2 COMPUTATIONAL METHODS

We shall discuss the computational methods under various heads.

Analysis of Special Classes

Acyclic Markov chains arise frequently in reliability and performability modeling. Marie, Reibman, and Trivedi [24] have proposed a method for automatically deriving transient solutions that are symbolic in the time duration t, for acyclic chains. Their method is applicable to cumulative measures of performance and sensitivity analysis of the transient solution. Donatiello and Iyer [62] have proposed a double transform-based procedure for computing performability distributions of systems whose failure-repair behavior is described by acyclic Markov chains. Goyal and Tantawi [63] have proposed a different numerical method for the same problem. In all these cases, the acyclic structure of the Markov model plays a crucial role in the solution procedure.

Laplace Transform Inversion

This technique was illustrated in Section 3.1. This method is good for hand computation on small or special case models. It has a worst-case computational complexity of $O(N^5)$ where N is the number of states and requires that the eigen values of the transition rate matrix be accurately determined. For acyclic Markov chains, this technique is adequate, as shown in [24, 62]. Laplace transform inversion using Fourier series [64] is a promising technique but both analytic and numerical Laplace transform inversions are unstable, in general.

Computation of Matrix Exponential

For small values of t, the matrix exponential method gives accurate and efficient solutions for transient analysis. For large values of t, the exponential series has poor numerical properties even for small problems. Round-off error is a common problem with these computations [20]. There are many alternative ways of evaluating the matrix exponential [65, 66], but they are not efficient for large sized problems and for large values of t.

Numerical Solution of Differential Equations

The classical techniques for numerical solution of the differential equations (1),(2), or (6), first find the eigen values and the eigen vectors of the transition rate matrix Q. The solution is then obtained using the Lagrange-Sylvester formula [67]. This method has complexity of $O(N^4)$ when all the eigen values are distinct and $O(N^5)$ otherwise. Thus this approach is impractical for solving large models. Furthermore, for large matrices, it is difficult to accurately generate the entire eigen system.

Numerical differential equation solvers fall into two classes: explicit methods and implicit methods. Explicit methods require only function evaluations, whereas implicit methods require the solution of a linear algebraic system at each time step [27]. The Runge-Kutta method [68] is the most popular explicit method for solving differential equations. This method is widely available and is satisfactory for nonstiff problems with normal accuracy requirements. It is however not suitable for the solution of stiff equations. Popular implicit methods include, the Backwards Euler and the Trapezoid Rule [69]. These methods are very good for handling stiffness, however they are less accurate and incur substantial performance penalties on nonstiff problems.

Uniformization

Uniformization or randomization [60] is probably the most popular numerical method for transient analysis. In this method, a continuous time Markov chain is reduced to a discrete time Markov chain subordinated to a Poisson process [20, 60]. Uniformization first transforms the transition rate matrix Q to the matrix Q^* given by

$$Q^* = \frac{Q}{q} + I$$

where q is the largest magnitude of a diagonal element of Q. The solution is then given in the form of an infinite series. The series can be truncated at a desired stage and the error bounds are immediately known. Uniformization is not subject to the round-off errors encountered while directly evaluating the matrix exponential series. It is quite accurate and efficient, and allows accurate error control. It is however not very good for stiff problems.

Uniformization has now emerged as a method of choice for many typical problems in transient analysis. It is extensively used in performability evaluation [29, 30] and sensitivity analysis [70]. It has been implemented in several software packages [61, 27, 71].

Aggregation Methods

These methods are approximate and are intended to transform a stiff Markov chain into a nonstiff chain having a smaller state space. Bobbio and Trivedi [25, 26] have proposed an aggregation technique that exploits the stiffness of the chain. In thier method, the states are classified into fast and slow states. Fast states are further classified into fast recurrent subsets and a fast transient subset. A separate analysis of each of these fast subsets is done and each fast recurrent subset is replaced by a single slow state while the fast transient subset is replaced by a probabilistic switch. The resulting smaller and nonstiff chain is then analyzed using any suitable method.

Other Methods

Other methods for transient analysis include, using diffusion approximations [41], fluid approximations [40], and approximate techniques for transform inversion [35].

4.3 SOFTWARE PACKAGES

Johnson and Malek [61] have surveyed several software packages for evaluating reliability, availability, and serviceability. Several of these are useful for transient analysis.

CARE(Computer Aided Reliability Estimator program) [72] is a general purpose reliability estimation tool for large, highly reliable digital fault-tolerant avionic systems. For transient analysis, this package uses the method of convolution integral.

HARP (Hybrid Automated Reliability Predictor) [28] provides a hybrid model for evaluation of reliability and availability of large complex systems. This uses an extended stochastic Petri net model for specifying fault handling and employs the Runge-Kutta method for solving the differential equations.

METASAN (Michigan Evaluation Tool for the Analysis of Stochastic Activity Networks) [73] evaluates performability for non-repairable and repairable systems, over finite intervals of time, by analyzing or simulating a stochastic activity network model, which is an extension of stochastic Petri nets.

SHARPE (Symbolic Hierarchical Reliability and Performance Evaluator) [74] provides a hierarchical modeling framework for evaluating reliability and availability of non-repairable and repairable systems. This uses the technique of Laplace transform inversion for transient analysis.

SAVE (System Availability Estimator) [75] computes reliability and availability of all classes of systems, by doing a transient analysis using the technique of uniformization.

Marie, Reibman, and Trivedi [24] describe an algorithm called ACE (Acyclic Markov chain Evaluator) for evaluating the transition probabilities in symbolic form, for acyclic chains. Reibman, Trivedi, Sanjayakumar, and Ciardo [27] describe a software package for the specification and solution of stiff Markov chains, using the technique proposed by Bobbio and Trivedi [25, 26]. The package ESP (Evaluation Package for Stochastic Petri Nets) [76] is a stochastic Petri net-based package for transient and steady-state analysis. The tool SPNP [71] is a powerful package, developed by Ciardo, Trivedi, and Muppala, that uses stochastic Petri nets as a specification language and carries out both transient and steady-state analyses.

5 SUMMARY

In this article, we have made a case for enhancing research efforts in analyzing the transient performance of discrete manufacturing systems. There are available several computational methods and software tools for conducting transient analysis of Markov models. Application of these methods and tools can facilitate a better understanding of the manufacturing system dynamics and an improved methodology for design. In addition to the issues discussed in this paper, there are certain others that deserve attention of researchers in this area:

- Performance optimization studies using transient analysis.
- Transient analysis of semi-Markov models, M/G/1 type of models, and renewal processes.
- Improved algorithms and numerical techniques for transient analysis, including methods based on aggregation.

References

[1] Y.C. Ho. Performance evaluation and perturbation analysis of discrete event dynamical systems. *IEEE Transactions on Automatic Control*, 32:563–572, 1987.

[2] N.Viswanadham and Y. Narahari. *Performance Modeling of Automated Manufacturing Systems.* Prentice Hall, Englewood Cliffs, NJ, 1992.

[3] J.A. Buzacott and J.G. Shantikumar. *Stochastic Models of Manufacturing Systems.* Prentice Hall, Englewood Cliffs, NJ, 1993.

[4] A.M. Law. Introduction to simulation: A powerful tool for analyzing complex manufacturing systems. *Industrial Engineering*, 28(5):46–61, May 1986.

[5] Y. Dallery and S.B. Gershwin. Manufacturing flow line systems: A review of models and analytical results. Technical Report 91-002, Laboratory for Manufacturing and Productivity, MIT, April 1992.

[6] J.A. Buzacott and D.D. Yao. Flexible manufacturing systems: A review of analytical models. *Management Science*, 32:890–905, 1986.

[7] J.A. Buzacott and D.D. Yao. On queueing network models of flexible manufacturing systems. *Queueing Systems: Theory and Applications*, 1:5–27, 1986.

[8] P. Kouvelis and D. Tirupati. Approximate performance modeling and decision making for manufacturing systems: A queueing network optimization framework. *Journal of Intelligent Manufacturing*, 2:107–134, 1991.

[9] N. Viswanadham and Y. Narahari. Stochastic Petri net models for performace evaluation of automated manufacturing systems. *Information and Decision Technologies*, 14:125–142, 1988.

[10] R.Y. Al-Jaar and A.A. Desrochers. Performation evaluation of automated manufacturing systems using generalized stochastic Petri nets. *IEEE Journal on Robotics and Automation*, 6(6):621–639, December 1990.

[11] P.J. Burke. The output of a queueing system. *Operations Research*, 4:699–704, 1956.

[12] J.D.C. Little. A proof of the queueing formula $l = \lambda w$. *Operations Research*, 9:383–385, 1961.

[13] J.R. Jackson. Job shop like queueing systems. *Management Science*, 10:131–142, 1963.

[14] W.J. Gordon and G.F. Newell. Closed queueing networks with exponential servers. *Operations Research*, 15:245–260, 1967.

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

[16] M. Reiser and S. S. Lavenberg. Mean value analysis of closed multichain queueing networks. *Journal of the ACM*, 27(2):313–322, April 1980.

[17] P.J. Curtois. On time and space decomposition of complex structures. *Communications of the ACM*, 28(4):590–603, April 1985.

[18] K.S. Trivedi. *Probability and Statistics with Reliability, Queueing, and Computer Science Applications.* Prentice Hall, Englewood Cliffs, NJ, 1982.

[19] S.M. Ross. *Introduction to Probability Models.* Third Edition, Academic Press, Orlando, Florida, 1985.

[20] W.K. Grassman. Transient solution in Markovian queueing systems. *Computers and Operations Research*, 4(1):47–56, 1977.

[21] W. Stewart. A comparison of numerical techniques in Markov modeling. *Communications of the ACM*, 21(2):144–152, February 1978.

[22] A.L. Reibman and K.S. Trivedi. Numerical transient analysis of Markov models. *Computers and Operations Research*, 15(1):19–36, 1988.

[23] A.L. Reibman and K.S. Trivedi. Transient analysis of cumulative measures of Markov chain behavior. *Stochastic Models*, 5(4):683–710, 1989.

[24] R.A. Marie, A.L. Reibman, and K.S. Trivedi. Transient analysis of acyclic markov chains. *Performance Evaluation*, 7:175–194, 1987.

[25] A. Bobbio and K.S. Trivedi. An aggregation technique for the transient analysis of stiff Markov chains. *IEEE Transactions on Computers*, C-35(9):803–814, September 1986.

[26] A. Bobbio and K.S. Trivedi. Computing cumulative measures of stiff Markov chains using aggregation. *IEEE Transactions on Computers*, C-39(10):1291–1298, October 1990.

[27] A.L. Reibman, K.S. Trivedi, Sanjayakumar, and G. Ciardo. Analysis of stiff Markov chains. *ORSA Journal of Computing*, 1(2):126–133, Spring 1989.

[28] S. Bavuso, J.B. Dugan, K.S. Trivedi, E. Rothmann, and W. Smith. Analysis of some fault-tolerant architectures using HARP. *IEEE Transactions on Reliability*, R-36(2):176–185, 1987.

[29] E. de Souza e Silva and H.R. Gail. Calculating availability and performability measures of repairable computer systems using randomization. *Journal of the ACM*, 36(1):171–193, January 1989.

[30] E. de Souza e Silva and H.R. Gail. Performability analysis of computer systems: From model specification to solution. *Performance Evaluation*, 14(3):157–196, 1992.

[31] D.K. Gerber. *Performance Evaluation of Fault-Tolerant Systems using Transient Markov Models.* Masters Thesis, Department of EECS, Massachusetts Institute of Technology, 1985.

[32] J.F. Meyer. Performability: A retrospective and some pointers to the future. *Performance Evaluation*, 14:139–156, 1992.

[33] K.S. Trivedi, J.K. Muppala, S.P. Woolet, and B.R. Haverkort. Composite performance and dependability analysis. *Performance Evaluation*, 14:197–215, 1992.

[34] C. Baiocchi, A.C. Capelo, V. Comincioli, and G. Serazzi. A mathematical model for transient analysis of computer systems. *Performance Evaluation*, 3:247–264, 1983.

[35] T.C.T Kotiah. Approximate transient analysis of some queueing systems. *Operations Research*, 26(2):333–346, 1978.

[36] P. Konstantopoulos and F. Baccelli. On the cut-off phenomenon in some queueing systems. Technical Report 90-1290, INRIA, October 1990.

[37] S.K. Tripathi and A. Duda. Time dependent analysis of queueing systems. *INFOR*, 15, 1986.

[38] R.A. Upton and S.K. Tripathi. An approximate transient analysis of the M(t)/M/1 queue. *Performance Evaluation*, 2(2):118–132, 1982.

[39] W.A. Massey. Open networks of queues: Their algebraic structure and estimating their transient behavior. *Advances in Applied Probability*, 16:176–201, 1984.

[40] A.A. Weiss and D. Mitra. A transient analysis of a data network with a processor sharing switch. *AT & T Technical Journal*, 67(4):4–16, Sept-Oct 1988.

[41] H. Kobayashi. Application of the diffusion approximation to queueing networks: Part 2. Non-equilibrium distribution and computer modeling. *Journal of the ACM*, 21:459–469, 1974.

[42] R. Ram. *Performance and Performability Modeling of Automated Manufacturing Systems.* Ph D Dissertation, Department of Computer Science and Automation, Indian Institute of Science, February 1992.

[43] M.S. Manjunath. *Transient Analysis of Manufacturing Systems.* Masters Thesis, Department of Computer Science and Automation, Indian Institute of Science, Bangalore, January 1991.

[44] N. Viswanadham, Y. Narahari, and R. Ram. Performability of automated manufacturing systems. In *Control and Dynamic Systems: Volume 47*, pages 77–122. Academic Press, Inc., 1991.

[45] R.P. Malhame and El-Kebir Boukas. A renewal theoretic analysis of a class of manufacturing systems. *IEEE Transactions on Automatic Control*, AC-36(5):580–587, May 1991.

[46] R.A. Wolff. *Stochastic Processes and Queueing Theory.* Prentice Hall, Englewood Cliffs, NJ, 1989.

[47] J.G. Shantikumar and J.A. Buzacott. Open queueing network models of dynamic job shops. *International Journal of Production Research*, 19:255–266, 1981.

[48] D. Mitra and I. Mitrani. Analysis of a Kanban discipline for cell coordination in production lines, part 2: Stochastic demands. *Operations Research*, 39(5):807–823, Sept–Oct 1991.

[49] T. Rolski. Queues with non-stationary inputs. *Queueing Systems: Theory and Applications*, 5:113–130, 1989.

[50] P. R. Kumar. Re-Entrant lines. Technical report, Coordinated Science Laboratory, University of Illinois at Urabana-Champaign, 1992.

[51] S.H.Lu and P.R.Kumar. Distributed scheduling based on due dates and buffer priorities. *IEEE Transactions on Automatic Control*, 36(12):1406–1416, December 1991.

[52] L.M.Wein. Scheduling semiconductor wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 1(3):115–130, August 1988.

[53] S.X.Bai and S.B.Gershwin. A manufacturing scheduler's perspective on semiconductor fabrication. Technical Report 89–518, MIT Microsystems Research Center, March 1989.

[54] G.D. Stamoulis. *Transient Analysis of Some Open Queueing Systems.* Masters Thesis, Department of EECS, MIT, 1988.

[55] G.D. Stamoulis and J.D. Tsitsiklis. On the settling time of the G/G/1 queue. *Advances in Applied Probability*, 22:929–956, 1990.

[56] V. Anantharam. The settling time of a closed Jackson Network. Technical report, Preprint, 1988.

[57] V. Anantharam. Threshold phenomena in the transient behavior of Markovian models of communication networks and databases. *Queueing Systems: Theory and Applications*, 5:77–98, 1989.

[58] U. Sumita, J.G. Shantikumar, and Y. Masuda. Analysis of fault-tolerant computer systems. *Microelectonics and Reliability*, 27(1):65–78, 1987.

[59] A.L. Reibman. Modeling the effect of reliability on performance. *IEEE Transactions on Reliability*, 39(3):314–320, August 1990.

[60] D. Gross and D. Miller. The randomization technique as a modeling tool and solution procedure for transient Markov processes. *Operations Research*, 32:334–361, 1984.

[61] A.M. Johnson Jr and M. Malek. Survey of software tools for evaluating reliability, availability, and serviceability. *ACM Computing Surveys*, 20(4):227–269, December 1988.

[62] L. Donatiello and B.R. Iyer. Analysis of a composite performance reliability measure for fault-tolerant systems. *Journal of the ACM*, 34(1):179–199, January 1987.

[63] A. Goyal and A. Tantawi. Evaluation of performability for degradable computer systems. *IEEE Transactions on Computers*, C-36(6):738–744, June 1987.

[64] V. Kulkarni, V.F. Nicola, R.M. Smith, and K.S. Trivedi. Numerical evaluation of performability measures and job completion times in repairable fault-tolerant systems. In *Proceedings of the Sixteenth IEEE International Symposium on Fault-Tolerant Computing Systems*, July 1986.

[65] C. Moler and C.F. Van Loan. Nineteen dubious ways to compute the exponential of a matrix. *SIAM Review*, 20:801–835, 1978.

[66] G.H. Golub and C.F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, Madison, 1983.

[67] R. Bellman. *Introduction to Matrix Analysis.* Mc-Graw Hill, New York, 1969.

[68] C.W. Gear. *Numerical Initial Value Problems in Ordinary Differential Equations.* Prentice Hall, Englewood Cliffs, NJ, 1971.

[69] J.D. Lambert. *Computational Methods in Ordinary Differential Equations.* Wiley, London, 1973.

[70] P. Heidelberger and A. Goyal. Sensitivity analysis of continuous time Markov chains using Uniformization. In *Proceedings of the Second International Workshop on Applied Mathematics and Performance/Reliability Models of Computer and Communication Systems*, May 1987.

[71] G. Ciardo, J. Muppala, and K.S. Trivedi. SPNP: Stochastic Petri net package. In *Proceedings of the Third International Workshop on Petri Nets and Performance Models, Tokyo, Japan*, pages 142–151. IEEE Computer Society Press, 1989.

[72] R. Geist and K.S. Trivedi. Ultra-high reliability prediction for fault-tolerant computer systems. *IEEE Transactions on Computers*, C-32(12):1118–1127, December 1983.

[73] W.H. Sanders and J.F. Meyer. METASAN: A performability evaluation tool based on stochastic activity networks. In *Proceedings of the 1986 Fall Joint Computer Conference*, pages 807–816, November 1986.

[74] R.A. Sahner and K.S. Trivedi. Reliability modeling using SHARPE. *IEEE Transactions on Reliability*, R-36(2):186–193, June 1987.

[75] A. Goyal, S.S. Lavenberg, and K.S. Trivedi. Probabilistic modeling of computer system availability. *Annals of Operations Research*, 8:285–306, 1987.

[76] A. Cumani. ESP: A package for the evaluation of stochastic Petri nets with phase type distributed transition times. In *Proccedings of First International Workshop on Timed Petri Nets, Torino, Italy*, July 1985.