EVALUATING THE TRANSIENT BEHAVIOR OF QUEUEING SYSTEMS VIA SIMULATION AND TRANSFER FUNCTION MODELING

Jingang Liu
Feng Yang

Industrial and Management Systems Engineering Dept.
West Virginia University
Morgantown, WV, 26506, U.S.A

ABSTRACT

Characterizing the transient behavior of queueing systems is a difficult problem, which has been addressed by either simplified analytical models or simulation. We seek to capture the transient performance of systems from a new perspective: based on high-fidelity simulation experiments, we estimate a number of transfer function models (the discrete approximations of those ODEs provided by an analytical approach) which characterizes the evolution of the system's dynamic behavior.

1 INTRODUCTION

This paper is concerned about the transient behavior of queueing systems, which is graphically illustrated as in Figure 1. For a given queueing system, we are interested in the dynamic responses $\mathbf{Y}(t)$ of the system with an input process $X(t)$. We refer to the functional relationships between $\mathbf{Y}(t)$ and $X(t)$ as "transfer functions"—$X(t)$ is transferred into $\mathbf{Y}(t)$ by going through the system. Here, $X(t)$ represents the arrival process of entities into the system, and $\mathbf{Y}(t)$ can be the performance measures of a queueing system such as the moments of the number of entities in the system at time t. Although it has long been recognized that transient analysis are needed to appropriately model many complex production, service, communication and air transportation systems, characterizing the dynamic input-output relationships for queueing models remains a very difficult problem yet to be addressed adequately.

For Markov queueing models, time-dependent ordinary differential equations (ODEs) can be derived to represent this input-output dynamics. However, analytical solutions to these ODEs are rare. A few exceptions include the known solutions for the M(t)/G/∞ and M/M/1 systems presented in queueing books (Gross and Harris 1985, Kleinrock 1975). The mainstay of the transient analysis work has been the development of numerical solutions of time-dependent ODEs characterizing the transient behavior of the Markov models.
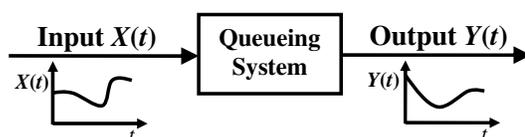
Figure 1: Time-dependent input and output in relation to a queueing system.

Ingolfsson et al. (2007) provides a fairly complete review of these methods including Rothkopf and Oren (1979), Clark (1981), Taaffe and Ong (1987), Green and Kolesar (1991), Eick, Massey and Whitt (1993a, 1993b), Jennings et al. (1996), Massey and Whitt (1997), etc. Other techniques for approximating the transient behavior of queues include fluid approximations, accurate when there is little variability, and diffusion models, which are used for heavily loaded systems. These methods can be roughly divided into two categories: those that are highly accurate but computationally intensive (comparable to detailed simulation), and those that are fast but inaccurate. Nevertheless, a common limitation of these methods is that they rely on simplistic assumptions such as Markov property and is not able to adequately capture many features of realistic systems.

Computer simulation is an alternative appealing approach to address the transient (or nonstationary) behavior of dynamic systems because of its high fidelity and flexibility and, more and more, also because of its easy of use. The shortcoming of simulation is that many replication runs are required to obtain good estimates of time-dependent performance measures, and the computational intensive nature of simulation prevents it from being used for real-time "what-if" analysis in decision making.

In light of this, we propose the simulation-based transfer function Modeling approach to analyze the transient behavior of queueing systems. The objective of this research is to estimate a *small* number of transfer function models (TFMs), which characterize the evolution of the system's dynamic

performance, based on extensive *off-line* simulation. The resulting TFMs are the discrete approximations of those ODEs provided by an analytical approach. Given the current state of the system, we can use such TSMs to predict in a timely manner the future evolution of the system under any input process, and hence allow for real-time "what-if" analysis. Furthermore, since the TFMs are estimated from detailed simulation data, they embody the high fidelity of simulation. In this paper, the dynamic responses of our TFMs are the time-dependent departures from the system and moments of the number of entities in the system. The inputs are the typically time-varying entity arrival process into the system, which may be considered as controllable variable in some decision context such as production planning.

We approach the transient analysis of queueing systems from a new perspective by integrating computer simulation, adaptive statistical methods and queueing analysis. Our methodology is three fold:

- In advance of the need to make decision, extensive off-line simulations are performed to collect sufficient amount of transient data. This can be achieved by exploiting the availability of large quantities of idle (perhaps networked) computer resources.
- Based on the simulation data collected for a queueing system, TFMs (discrete approximations of time-dependent ODEs) can be estimated representing the input-output dynamics of the system. These TSMs can be used to support real-time decision making without performing any additional simulation: Given its initial state, the system's future performances can be recursively evaluate under different scenarios for the input process to the system.
- Queueing theory, though inadequate to address the transient behavior of non-Markovian systems, may suggest an appropriate form of the TSMs representing the dynamics of a general system.

This paper describes our initial efforts to investigate the transient behavior of queueing systems via simulation and transfer function modeling.

2   STATEMENT OF THE PROBLEM

In this section, we define our research problems in more precise terms. We assume that the system is given, which implies that the distribution of service time $g(x)$ and the number of servers s does not change over the time horizon being considered. Our task is to predict over a time period $(0, T]$ the dynamic performances of a system, $m_1(t)$, $m_2(t)$, and $d(t)$, for a certain input process $a(t)$ ($t \in (0, T]$). Next, we define these inputs and outputs in mathematical terms.

Time-Varying Input Process

Let $A(t, t + \Delta t)$ be the number of jobs arrived during the interval $(t, t + \Delta t)$, and define $a(t)$ as the arrival rate $a(t) = \lim_{\Delta t \to 0} A(t, t + \Delta t)/\Delta t$ assuming that the limit exist. We assume that there is no multiple arrivals and therefore $a(t)\Delta t$ is the probability that an arrival occurs in $(t, t + \Delta t)$.

Dynamic Performance Responses

Let $N(t)$ be the number of entities in the system at time t, and define the first two moments of $N(t)$ as:

$$m_1(t) = \mathrm{E}[N(t)] \quad \text{and} \quad m_2(t) = \mathrm{E}[N^2(t)]. \qquad (1)$$

In addition to $m_1(t)$ and $m_2(t)$ we are also interested in the departure process of the system. Assuming that there is no multiple departures, we denote $D_n(t, t + \Delta t) = \mathrm{P}\{N(t) = n, \mathscr{D}(t)\}$ where $\mathscr{D}(t)$ represents the event that a departure occurs within the time interval $(t, t + \Delta t)$.

Also, denote by $D(t, t + \Delta t)$ the probability that there is a departure within $(t, t + \Delta t)$:

$$D(t, t + \Delta t) = \sum_{n=1}^{\infty} D_n(t, t + \Delta t). \qquad (2)$$

We define:

$$d_n(t) = \lim_{\Delta t \to 0} \frac{D_n(t, t + \Delta t)}{\Delta t} \quad \text{and} \quad d(t) = \lim_{\Delta t \to 0} \frac{D(t, t + \Delta t)}{\Delta t} \qquad (3)$$

These limits have the following interpretation. Given the state of the system at time t, the probability that a departure occurs during the interval $(t, t + \Delta t)$ is $d(t)\Delta t + o(\Delta t)$. Thus we may say that $d(t)$ is the departure rate from the system at time t. Similarly, $d_n(t)$ is the rate of departure at time t when there are n entities in the system.

Our objective is to establish a dynamic relationship between the the input process $a(t)$ and the outputs of interest, namely the first two moments of $N(t)$ and the departure rate $d(t)$.

3   TRANSIENT ANALYSIS OF QUEUEING SYSTEMS

In this section, we discuss our transient analysis approach based on several queueing systems.

3.1 M(t)/M/∞

For the purpose of intuition and motivation, we start with the simplest queueing model M(t)/M/∞, which is in fact analytically tractable. The analytical analysis of M(t)/G/∞ is given in Massey and Whitt (1993). Here, we use M(t)/M/∞ to illustrate what we want to obtain in the transient analysis of a queueing system.

Suppose that the arrival rate is $a(t)$ $(t \in (0,T])$, and service rate is μ. From the Kolmogorov forward equations for the state probabilities of $M(t)/M/\infty$, we have the following differential equations for the first two moments of $N(t)$.

$$m_1'(t) = a(t) - \mu \cdot m_1(t) \tag{4}$$

$$m_2'(t) = a(t) + 2a(t) \cdot m_1(t) + \mu \cdot m_1(t) - 2\mu \cdot m_2(t) \tag{5}$$

Due to the Markov property, the departure rate is simply

$$d(t) = \mu \cdot m_1(t) \tag{6}$$

Obviously, given an intial state of the system, future performances represented by $m_1(t)$, $m_2(t)$, and $d(t)$ can be evaluated based on (4), (5), and (6) in a recursive manner for a certain arrival process $a(t)$. Note that μ is considered as a fixed system parameter.

Apparently for an M(t)/M/∞ system, there is no need to incorporate computer simulation or statistical modeling into its transient analysis. Unfortunately, the situation is much more complicated for other systems. We will show next that as finite number of servers is introduced or the Markovian assumption is relaxed, we may not be able to easily obtain the solutions to those ODEs like (4), (5), and (6), or even worse, it may be impossible to analytically formulate any ODEs.

3.2 Simulation and Transfer Function Modeling

The purpose of this work is to generate for general queueing systems a very small number of TFMs characterizing their transient behavior. These TFMs are just like the ODEs provided in section 3.1. As already mentioned, our major tasks are: (i) performing preliminary queueing analysis to identify the appropriate forms of the TSMs, and (ii) running simulation experiments for data collection and estimate the TFMs, which are the discrete approximations of the time-dependent ODEs for the dynamic output performances. We next discuss these two aspects respectively.

3.2.1 Preliminary Queueing Analysis

The traditional methods of queueing theory is shown to be inadequate in terms of addressing the transient behavior of a general queueing system. Since ODEs like those for $M(t)/M/\infty$ is not available for a general queue, we seek to obtain the second best: develop TSMs, the discrete approximations of ODEs, from simulation data. Nevertheless, as already pointed out, queueing analysis can shed light on our simulation-based statistical modeling.

For the analytical analysis of a general queueing system, we make the following assumption regarding successive arrivals into the system:

Assumption 1 *The time interval between the ith entity and the $(i+1)^{th}$ entity is independent of the arrival process prior to the arrival epoch of the ith entity, for all $i = 1, 2, \ldots$. Moreover, new arriving entities do not affect the time in the system of previous entities of the same class.*

Examples of arrival processes satisfying Assumption 1 include (a) all renewal processes, and (b) a nonhomogeneous Poisson process. For a queueing system satisfying Assumption 1, the differential-difference equations of the state probabilities can be written as:

$$dp_0(t)/dt = -a(t)p_0(t) + d_1(t)$$
$$dp_n(t)/dt = -a(t)p_n(t) - d_n(t) + a(t)p_{n-1}(t)$$
$$+ d_{n+1}(t), \ n \geq 1 \tag{7}$$

where $p_n(t) = P\{N(t) = n, n \geq 0\}$ and $d_n(t)$ is defined in Section 2 as the system's departure rate at time t while having n entities in the system. By multiplying both sides of the nth equation in (7) by n and summing, the first two moment differential equations of $N(t)$ follow as:

$$m_1'(t) = dE[N(t)]/dt = a(t) - d(t) \tag{8}$$

$$m_2'(t) = dE[N^2(t)]/dt$$
$$= a(t) + 2a(t)m_1(t) + d(t) - 2M(t) \tag{9}$$

where

$$d(t) = \sum_{n=1}^{\infty} d_n(t) \quad \text{and} \quad M(t) = \sum_{n=1}^{\infty} nd_n(t) \tag{10}$$

The two expressions in (10) are finite as long as $E[N(t)]$ is finite at time t. Note that equations (8) and (9) hold for any queueing systems that satisfy Assumption 1. In addition, they also accomodate the cases where the system is temporarily overloaded (i.e., the arrival rate of entities is greater than the service rate of the system during some time intervals). Hence, we are considering a fairly general system free of the various simplistic assumptions made in a standard queueing analysis.

Unfortunately, despite the simple forms of (8) and (9), they are not ready to be used for prediction purpose. The problem lies in the fact that on the right-hand side of these ODEs, there are functions $d(t)$ and $M(t)$ involved. Except for Markovian systems, both $d(t)$ and $M(t)$ are dynamic responses themselves depending on the arrival process $a(t)$ and the initial state of the system. Hence, built upon (8) and (9), the task to predict the performances ($m_1(t)$, $m_2(t)$, and $d(t)$) boils down to describing the evolution of $d(t)$ and $M(t)$, which is difficult.

To gain some idea about the functional expressions of $d(t)$ and $M(t)$, we next provide the analytical results for a G(t)/G/1 model. Similar derivation

Proposition 1   *Consider a G(t)/G/1 with arrival process $a(t)$, and the p.d.f of the service time is $g(x)$ ($x \in [x_L, x_U]$). Under Assumption 1, we have*

$$d_n(t) = \int_{x_L}^{x_U} a(t-x)p_0(t-x)P\{A(t-x,t)=n-1\}g(x)dx$$

$$\int_{x_L}^{x_U} \sum_{\ell=2}^{n+1} P\{A(t-x,t)=n-\ell+1\}d_\ell(t-x)g(x)dx$$

Proposition 1 can be proved by recognizing that a departure will occur during the interval $(t, t+\Delta t)$ with $N(t) = n$ if one of the two following conditions holds:

1. The system was empty at time $t-x$ and one entity entered into service between $t-x$ and $t-x+\Delta t$. During its service which lasts a period of x, there were $n-1$ new arrivals
2. A departure occurred during $(t-x, t-x+\Delta t)$ with $N(t-x) = \ell$ ($\ell = 2, 3, \ldots, n+1$). During the service of the first of the remaining $\ell-1$ entities , there were $n-\ell+1$ new arrivals.

By definitions (10), we have derived the following from Proposition 1:

$$d(t) = \int_{x_L}^{x_U} a(t-x)p_0(t-x)g(x)dx$$
$$+ \int_{x_L}^{x_U} (d(t-x) - d_1(t-x))g(x)dx \quad (11)$$

$$M(t) = \int_{x_L}^{x_U} a(t-x)(E[A(t-x,t)]+1)p_0(t-x)g(x)dx$$
$$+ \int_{x_L}^{x_U} E[A(t-x,t)](d(t-x) - d_1(t-x))g(x)dx$$
$$+ \int_{x_L}^{x_U} (M(t-x) - d(t-x))g(x)dx \quad (12)$$

where $E[A(t-x,t)] = \int_{t-x}^{t} a(\tau)d\tau$ is the expected number of arrivals during $(t-x,t)$. It can be seen that $d(t)$ and $M(t)$ depends on the arrival process $a(t)$ and dynamic performances in the past. If, by appropriate approximations, we can turn the right-hand sides of (11) and (12) into expressions that only involve the input $a(\tau)$ ($\tau < t$), and the historical values of $m_1(\tau)$, $m_2(\tau)$, $d(\tau)$ and $M(\tau)$ ($\tau < t$), then from equations (8), (9), (11) and (12), we can recursively compute the dynamic responses for a certain arrival process $a(t)$ given the initial values $\{m_1(0), m_2(0), d(0), M(0)\}$.

For a G(t)/G/s system, similar derivation leads to equations—though much more complicated in form—that are of the same nature as those for a G(t)/G/1. Due to space constraint, here we only present the results of G(t)/G/1 for

the purpose of illustration. These analytical expressions provides very helpful indications as to the forms of the TFMs in our simulation-based methodology, which is discussed next.

3.2.2 Simulation-Based Transfer Function Modeling

In the proposed method, we seek to describe the transient behavior of a queueing system by four discrete TFMs:

$$\begin{aligned}
m_1(t) &\approx f_1(a(t-1), a(t-2), \ldots, \mathbf{Y}(t-1), \mathbf{Y}(t-2), \ldots) \\
m_2(t) &\approx f_2(a(t-1), a(t-2), \ldots, \mathbf{Y}(t-1), \mathbf{Y}(t-2), \ldots) \\
d(t) &\approx f_3(a(t-1), a(t-2), \ldots, \mathbf{Y}(t-1), \mathbf{Y}(t-2), \ldots) \\
M(t) &\approx f_4(a(t-1), a(t-2), \ldots, \mathbf{Y}(t-1), \mathbf{Y}(t-2), \ldots)
\end{aligned}$$
$$(13)$$

where $\mathbf{Y}(t-n) = (m_1(t-n), m_2(t-n), d(t-n), M(t-n))$ represents the multiple-output vector at time $t-n$ ($n = 1, 2, \ldots$). With these models, given the current state of the system at time 0 we will be able to predict the evolution of the system over $(0, T]$ under a certain input process $a(t)$ ($t \in (0, T]$).

The finite-difference approximations of equations like (8), (9), (11) and (12) suggest appropriate forms of equations f_1, f_2, f_3, and f_4. Take equation (11) for an example. If we assume that (i) $p_0(\tau) \approx p_0(m_1(\tau), m_2(\tau))$ and $p_1(\tau) \approx p_1(m_1(\tau), m_2(\tau))$ can be well approximated in terms of $m_1(\tau)$ and $m_2(\tau)$, and (ii) $d_1(\tau) \approx d(\tau)p_1(\tau) \approx d(\tau) \times p_1(m_1(\tau), m_2(\tau))$. Substituting these approximations into (12) and replacing the integration in (12) by finite-difference approximation, we obtain a specific function form for f_3. Notice that (8) and (9) hold for any queueing systems that satisfy Assumption 1, whereas (11) and (12) are derived for $G(t)/G/1$ only. Again, we base our discussion here on the analytical results for $G(t)/G/1$ simply for ease of illustrating our methods.

To estimate models (13) we run simulation experiments at different combinations of the input process $a(t)$ and initial condition $\mathbf{X}(0) = (m_1(0), m_2(0), d(0), M(0))$. Under a certain $a(t)$ and $\mathbf{X}(0)$, multiple replications are performed. For each replication, the state of the system is recorded over the simulation period $(0, T]$, and estimates for $\{(m_1(t), m_2(t), d(t), M(t)), t = 1, 2, 3, \ldots T\}$ can be obtained across multiple replications. Here we assume that the sampling interval is unit 1, and the simulation length T is an integer.

With the time series data collected from simulation, we can estimate the models (13) of appropriate forms suggested by preliminary queueing analysis. As mentioned earlier in Section 3.2.2, approximate functional relationships such as $p_0(\tau) \approx p_0(m_1(\tau), m_2(\tau))$ and $p_1(\tau) \approx p_1(m_1(\tau), m_2(\tau))$ are also needed for determining the appropriate forms of the TFMs (13). In some previous work (Rothkopf and

Table 1: Design of experiments for the M/G/s simulation

Settings	$a(t)$ $(t \in (0,T])$	$m_1(0)$	$m_2(0)$	$d(0)$	K(0)
1	1.8	2.3	8.47	1.83	6.08
2	1.8	4.23	25.10	2.48	12.60
3	1.8	7.84	71.28	2.84	23.60
4	2.1	2.94	13.21	2.10	8.26
5	2.1	5.16	35.80	2.65	15.74
6	2.1	8.90	90.62	2.90	27.08
7	2.4	3.69	19.99	2.34	10.84
8	2.4	6.26	0.950	2.77	19.25
9	2.4	10.09	114.80	2.92	30.75

Table 2: A system consisting of two stations in tandem

Parameters	Station 1	Station 2
Number of Servers	3	3
Mean Service Time	0.5	1
Std. of Service Time	0.1	0.2

Oren 1979, Taaffe and Ong 1987), such relationships are obtained assuming that $N(\tau)$ follows a certain distribution (e.g., negative binomial distribution). In our method, with simulation data available, we can easily fit logistic regression models to obtain the functional dependence of these probabilities upon the first two moments of $N(\tau)$, which is highly accurate in our empirical experiences.

4 EMPIRICAL EVALUATION

We applied our proposed method on a number of queueing systems. Although the systems considered here are relatively simple, we are not aware of any existing transient analysis methods of these systems that can be considered as parallel to our work.

In the remainder of this section, we will refer to a setting as a combination of the input process $a(t)$ and the initial condition $\mathbf{X}(0) = (m_1(0), m_2(0), d(0), M(0))$. For each system being investigated, simulation experiments are performed under different settings. The simulation data are divided into two sets: the estimation data set (EDS) and the validation data set (VDS). The former is used to estimate the TFMs (13), and the latter is used to evaluate the goodness of the estimated TFMs in the sense of predicting the system's dynamic evolution. The settings of the experiments included in the VDS are different from those of the experiments in the EDS.

4.1 A Single-Station System with Multiple Servers

We consider the transient behavior of a single-station system with 3 servers. The service time follows gamma distribution with a mean of 1 and standard deviation of 0.2. Simulation experiments were performed over a period of $(0,T]$ at nine different settings as given in Table 1.

Note from Table 1 that in our current experiments, the input process is Poisson with $a(t)$ held constant throughout the entire period $(0,T]$ of a simulation run, which may not be a good choice. To collect simulation data efficiently for the estimation of the TFMs, it may help to let $a(t)$ vary over

the simulation period. In addition, the levels of the arrival rate and the initial conditions were selected in a somewhat arbitrary manner in the hope of providing a good coverage of the possible settings. Developing an efficient design strategy for the simulation experiments is an important part of our on-going research.

The Experiments performed under settings 2, 4, 6, and 8 are included in the EDS, based on which the TFMs are estimated. Experiments performed under settings 1, 3, 5, 7, and 9 are included in the VDS. We compare the TFMs-predicted system evolutions with the simulated cases in the VDS. Figures 2 (a), (b), (c), and (d) show the comparison results of four dynamic responses $m_1(t)$, $m_2(t)$, $d(t)$, and $M(t)$ ($t \in (0,T]$) for Setting 3 specified in Table 1. Figure 2 displays the comparison for Setting 7.

In each figure, there are two curves and two vertical axis. The solid curve represents the evolution of the output response estimated from the simulation data in VDS, which is highly accurate and considered as "true" in our experiments. The dotted curve represents the prediction error, which is the absolute difference between the TFMs-predicted responses and the corresponding "true" values. The left vertical axis depicts the scale for the "true" responses, and the right vertical axis gives the scale for the prediction error.

For all the five settings in VDS, we generated similar plots as those in Figure 2 and 2. Obviously, the prediction provided by the TFMs tracks the dynamic performances very well. We have also applied the proposed method on single-station systems with failures and maintenances, and the resulting TFMs are able to provide very good prediction of the system's transient behavior as well.

4.2 Two Station in Tandem

We consider a system consisting of two stations in tandem. The system parameters are given in Table 2. The service time at both stations follows gamma distribution.

As in the single-station case, we simulate this two-station system under 27 different settings. Each setting represents a different combination of the input process to the whole system $a(t)$, the initial condition at the 1st station $\mathbf{X}^{[1]}(0) = (m_1^{[1]}(0), m_2^{[1]}(0), d^{[1]}(0), M^{[1]}(0))$, and the initial condition at the 2nd station $\mathbf{X}^{[2]}(0) = (m_1^{[2]}(0), m_2^{[2]}(0), d^{[2]}(0), M^{[2]}(0))$.
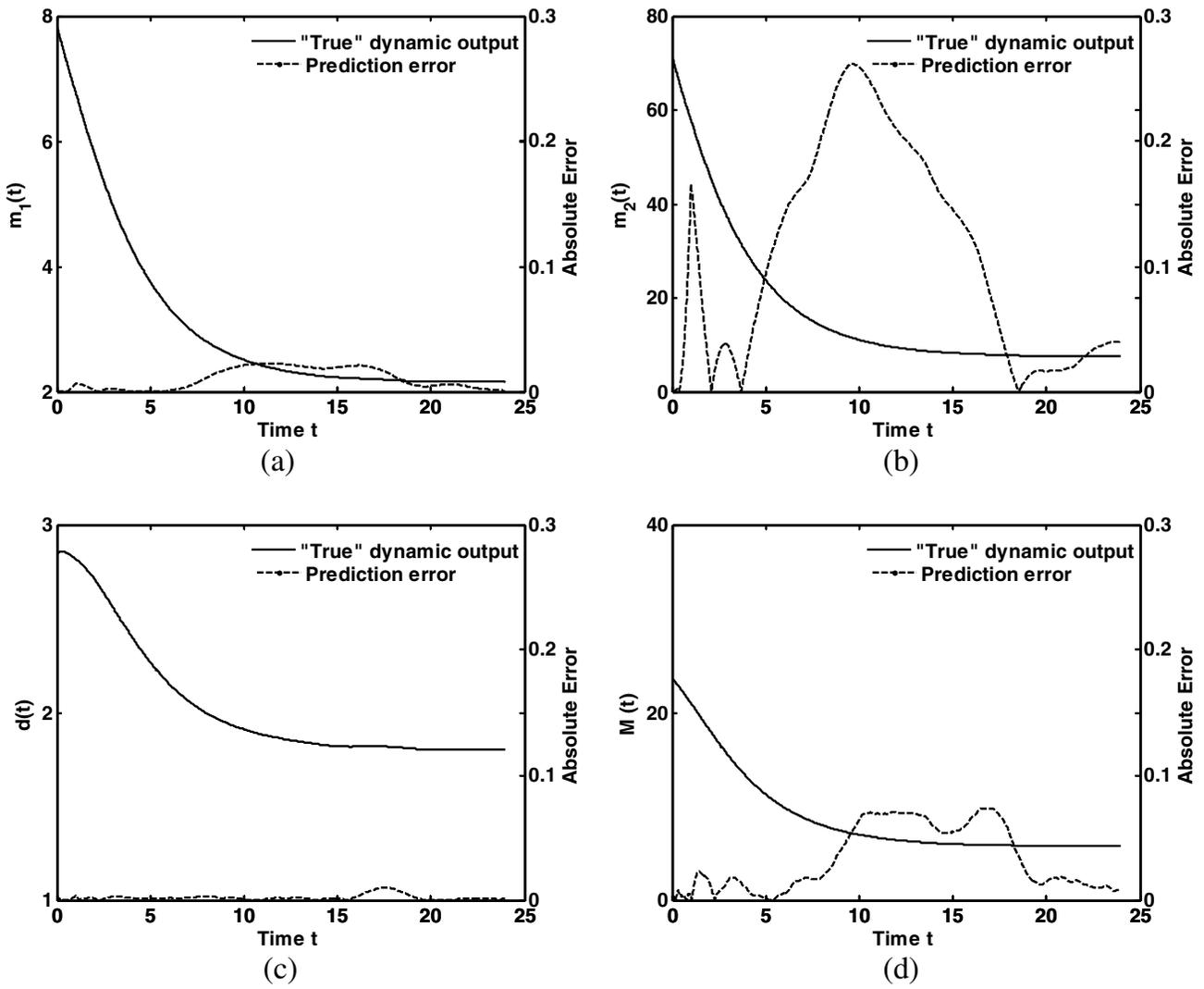
Figure 2: Comparison of the predicted dynamic responses and their "true" values (Under Setting 3 as specified in Table 1).
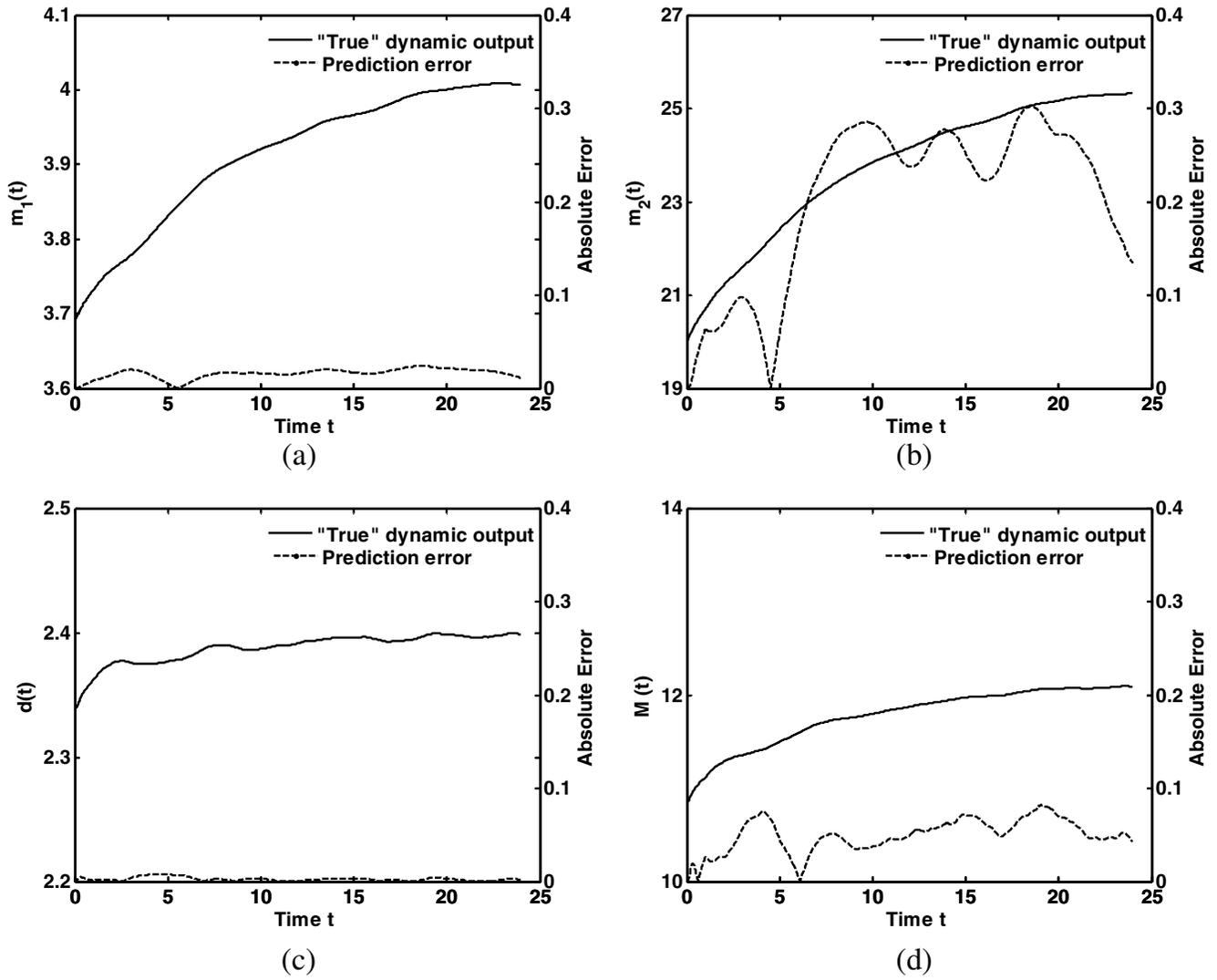
Figure 3: Comparison of the predicted dynamic responses and their "true" values (Under Setting 7 as specified in Table 1).
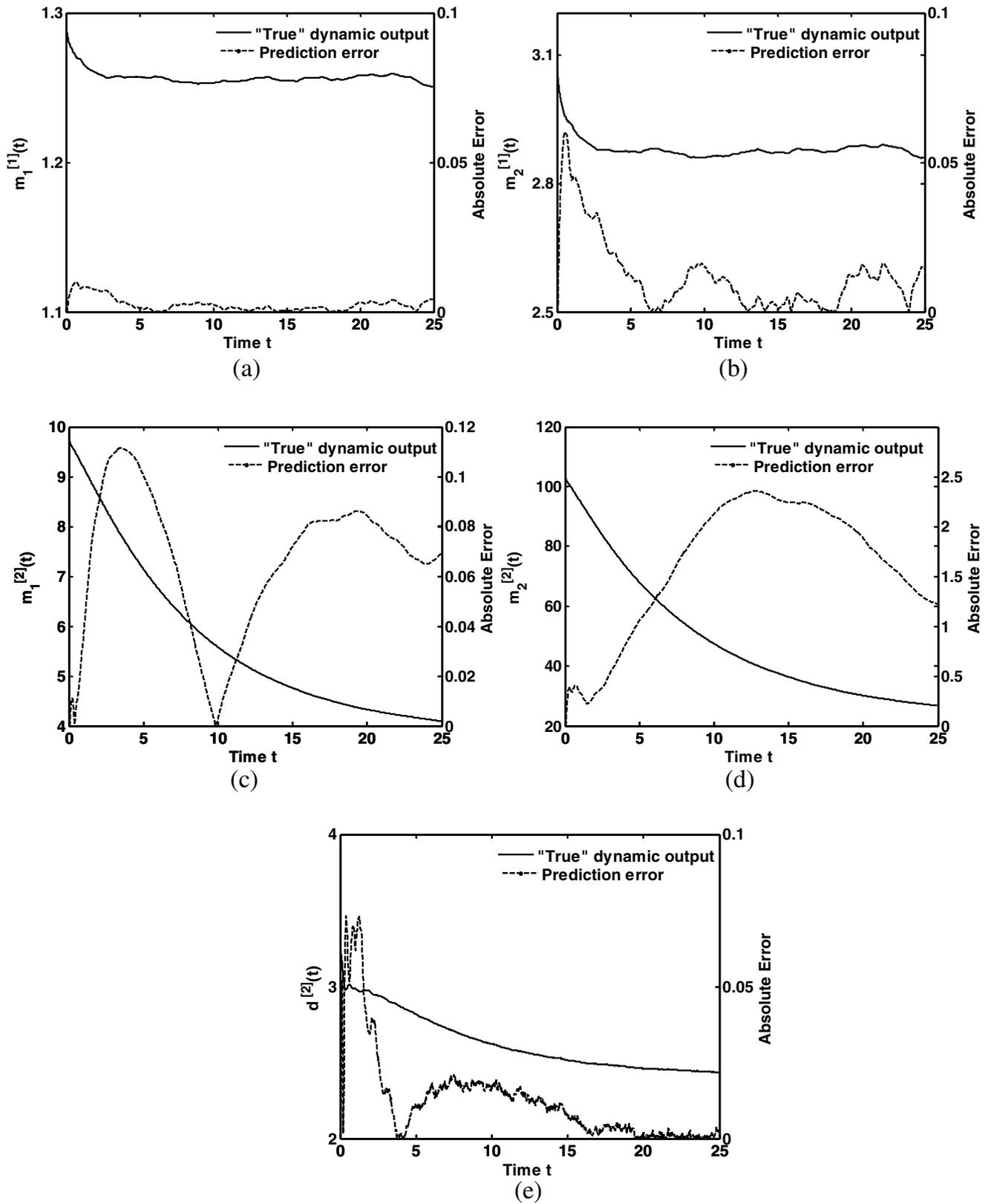
Figure 4: Comparison of the predicted dynamic responses and their "true" values for the 2-station system.

We use the superscript $[i]$ to denote station i ($i = 1, 2$). Among these 27 settings, 14 were used for the EDS, and the other 13 for the VDS.

For each station i ($i = 1, 2$), we use the simulation data to estimate a set of TFMs (denoted as TFMs[i]) characterizing the transient behavior of that station. Note that $a^{[2]}(t)$, the input process to station 2, is equal to $d^{[1]}(t)$, the departure process from station 1.

Once we have estimated TFMs[1] and TFMs[2], for given initial conditions $\mathbf{X}^{[1]}(0)$ and $\mathbf{X}^{[2]}(0)$, we can predict the transient performance of this two-station system under a certain input $a(t)$ to the whole system. Specifically, our prediction process can be divided into two steps:

1. With $a^{[1]}(t) = a(t)$ and given $\mathbf{X}^{[1]}(0)$, we use TFMs[1] to predict the transient performances for station 1, namely $(m_1^{[1]}(t), m_2^{[1]}(t), d^{[1]}(t), M^{[1]}(t))$ with $t \in (0, T]$.
2. With $a^{[2]}(t) = d^{[1]}(t)$ and given $\mathbf{X}^{[2]}(0)$, we use TFMs[2] to predict the transient performances for station 2, namely $(m_1^{[2]}(t), m_2^{[2]}(t), d^{[2]}(t), M^{[2]}(t))$ with $t \in (0, T]$.

Figure 4 provides the prediction results of TFMs[1] and TFMs[2] under one of the settings in VDS. For the case presented in Figure 4, station 1 is lightly utilized and station 2 is heavily utilized. As can be seen from Figure 4, we are able to predict with a high accuracy the dynamic performances of interest, $m_1^{[1]}(t)$, $m_2^{[1]}(t)$, $m_1^{[2]}(t)$, $m_2^{[2]}(t)$, and $d^{[2]}(t)$ (the departure rate from the 2-station system).

5 CONCLUSIONS

We proposed a new perspective to approach the transient analysis of queueing systems: simulation and transfer function modeling. Such an approach would be expected to fundamentally overcome the computational-expensive nature of simulation and the analytical intractability of realistic systems, and thus support real-time decision making. The results presented in this paper represent our initial exploration down this path, and our empirical experience based on relatively simple queueing systems is encouraging. We are currently investigating experiment design methods for simulation experiments, and model fitting strategies of TFMs. Our methods will be further adapted to realistic systems that involve multi-class entities, reentrant flows, etc.

REFERENCES

Eick, S. G., W. A. Massey, and W. Whitt. 1993a. The physics of the Mt/G/infty Queue. *Operations Research* 41: 731–742.

Eick, S. G., W. A. Massey, and W. Whitt. 1993b. Mt/G/infty queues with sinusoidal arrival rates. *Management Science* 39: 241–252.

Green, L. V., and P. J. Kolesar. 1991. The pointwise stationary approximation with nonstationary arrivals. *Management Science* 37: 84–97.

Gross, D., and C. Harris. 1985. Fundamentals of Queueing Theory. Hoboken, New Jersey: John Wiley & Sons, Inc.

Gross, D. and D. Miller.1984. The randomization technique as a modeling tool and solution procedure for transient Markov processes. *Operations Research* 32(6): 926–944.

Jennings, O. B., A. Mandelbaum, W. A. Massey, and W. Whitt. 1996. Server Staffing to Meet Time-Varying Demand. *Management Science* 42(10): 1383–1394.

Kleinrock, L. 1975. Queueing Systems. Hoboken, New Jersey: John Wiley & Sons, Inc.

Massey, W. A., and W. Whitt. 1997. Peak Congestion in Multi-Server Service Systems with Slowly Varying Arrival Rates. *Queueing Systems* 25: 157–172.

Rothkopf M. H., and S. S. Oren. 1979. A closure approximation for the nonstationary M/M/s queue. *Management Science* 25: 522–534.

Taaffe, M. R., and K. L. Ong. 1987. Approximating Nonstationary Ph(t)/M(t)/S/C Queueing Systems. *Annals of Operations Research* 8: 103–116.

AUTHOR BIOGRAPHIES

JINGANG LIU is a PhD student in the Industrial and Management Systems Engineering Department at West Virginia University. His research work has been focused on simulation and metamodeling. His e-mail address is <jliu7@mix.wvu.edu>.

FENG YANG is an assistant professor in the Industrial and Management Systems Engineering Department at West Virginia University. Her research interests include simulation and metamodeling, design of experiments, and applied statistics. Her e-mail and web addresses are <feng.yang@mail.wvu.edu> and <http://www2.cemr.wvu.edu/ yang/>.