An Approximative Method for Calculating Performance Measures of Markov Processes

Juha Leino
Networking Laboratory
Helsinki University of Technology
P.O. Box 3000, FI-02015 TKK, Finland
Juha.Leino@tkk.fi

Jorma Virtamo
Networking Laboratory
Helsinki University of Technology
P.O. Box 3000, FI-02015 TKK, Finland
Jorma.Virtamo@tkk.fi

ABSTRACT

We present a new approximation method called value extrapolation for Markov processes with large or infinite state spaces. The method can be applied for calculating any performance measure that can be expressed as the expected value of a function of the system state. Traditionally, the state distribution of a system is solved in a truncated state space and then an appropriate function is summed over the states to obtain the performance measure. In our approach, the measure is obtained directly, along with the relative values of the states, by solving the Howard equations in the MDP setting. Instead of a simple state space truncation, the relative values outside the truncated state space are extrapolated using a polynomial function. The Howard equations remain linear, hence there is no significant computational penalty. The accuracy of value extrapolation, even with a heavily truncated state space, is demonstrated using processor sharing systems and data networks as examples.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Markov processes

General Terms

Performance

Keywords

Approximation, Markov processes, performance evaluation

1. INTRODUCTION

Many queueing systems, for instance those appearing in telecommunication systems, can be modelled as Markov processes. If the state space of the process is finite, the steady state distribution of the system can be solved numerically using balance equations. However, if the state space is very large or infinite, it is unfeasible to solve the state distribution (unless an analytical solution is available). In order to evaluate the performance of the system, an approximative method is needed. The traditional approach is to truncate the state space and approximate the performance measure by summing an appropriate function over the states in the truncated state space. In general, the larger the truncated state space, the more accurate results are obtained. Error bounds and rate of convergence can be analyzed in some cases [14]. There exists also more sophisticated methods for approximation such as power series algorithm [8] and structured analysis approaches [5].

In this paper, we present a novel approximation method called value extrapolation which is a generalization of the idea we first introduced in [9]. Instead of first solving the state distribution of the process using the balance equations, we consider the system in the setting of Markov decision processes (MDP's) and solve directly the performance measure from the so-called Howard equations, along with the relative values of the states (see, e.g., [13]). The method can be used to approximate any performance measure expressed as the expected value of a random variable which is a function of the system state. Instead of ignoring the states outside the truncated state space, we extrapolate their relative values using the values inside the truncated state space. This is a kind of bootstrapping idea familiar in many other fields. If the extrapolation is made using a polynomial function, the Howard equations remain a closed system of linear equations and there is no significant computational penalty due to the extrapolation. The advantage of this method is that if the extrapolation of the relative values is accurate, an expectation type performance measure representing a sum over the whole state space is automatically accurately determined from the Howard equations without any summation and separate estimation of the contribution to the sum from the state space outside the truncated state space. There are good reasons to anticipate that polynomial extrapolation does indeed work well; in some specific simple systems the relative values are exactly polynomial functions of the state. Thus the performance measure can be accurately estimated even with a heavily truncated state space. Another advantage of our method is its simplicity; it is very straightforward to apply.

We use two types of systems to evaluate the accuracy of value extrapolation. In chapter 3, we discuss processor sharing (PS) systems, in which a single resource is shared among the customers. PS type queues can be used as models for many systems, e.g., in computer systems [2] and data communications [10]. Both one-class queues with state-dependent service rate and multi-class queues are considered

Valuetools'06 October 11-13, 2006, Pisa, Italy.

as examples. If there are multiple customer classes, the capacity allocation policy affects the system performance. We study two well-known policies, discriminatory [1] and generalized [15] processor sharing.

In chapter 4, as examples of systems with multiple resources, we evaluate the performance of simple data networks. While data networks can be dimensioned using deterministic traffic models, it is essential to use a dynamic model in order to analyze the quality of performance experienced by the end users. The traditional approach is to model the system as a Markov process and solve the equilibrium state distribution using a truncated state space. However, the applicability of this approach is limited as the number of states needed for accurate results is usually very large. When value extrapolation is applied, more accurate results are obtained with the same number of states. While value extrapolation does not solve the state space explosion, it makes it possible to analyze bigger networks with higher accuracy. The approach has already been applied to performance evaluation of a simple wireless scenario [9].

2. THEORETICAL FRAMEWORK

2.1 Multidimensional Stochastic Processes

Let $X(t)$ be a continuous-time Markov process describing a system with K customer classes. The state of the process is $\mathbf{x} = (x_1, \ldots, x_K)$, where x_k is the number of class-k customers. We assume that the state space $\mathcal{S}$ of the process is large or infinite. The transition intensity from state $\mathbf{x}$ to state $\mathbf{y}$ is denoted $q_{\mathbf{xy}}$.

Value extrapolation approach is applicable to performance measures that can be expressed as a mean value of a revenue that is a function of the system state. Let the revenue rate of the process in state $\mathbf{x}$ be $r(\mathbf{x})$. The mean revenue rate of the process is denoted r. A simple yet often informative performance measure is the mean occupancy $\mathrm{E}[|\mathbf{x}|] = \mathrm{E}[\sum_i x_i]$, which can be determined by using revenue function $r(\mathbf{x}) = |\mathbf{x}|$. Similarly, other moments of $|X|$ may be determined. If the first and second moments are known, variance may be solved using equation $\mathrm{Var}[|X|] = \mathrm{E}[|X|^2] - \mathrm{E}[|X|]^2$. Using the same approach, also the distributions of the individual customer classes can be analyzed by using revenue functions such as $r(\mathbf{x}) = x_k$ or $r(\mathbf{x}) = x_k^2$.

The steady state distribution of the process satisfies the global balance equations

$$\pi(\mathbf{x}) \sum_{\mathbf{y}} q_{\mathbf{xy}} = \sum_{\mathbf{y}} \pi(\mathbf{y}) q_{\mathbf{yx}} \quad \forall \mathbf{x}, \tag{1}$$

where $\pi(\mathbf{x})$ is the steady state probability of state $\mathbf{x}$. If the state space is large, state distribution can be solved only in some special cases. In general, the state space $\mathcal{S}$ needs to be truncated to a smaller state space $\tilde{\mathcal{S}}$ in order to solve the set of equations (1). If the truncated state space is large enough, the probabilities of the truncated state space can be used to approximate the probabilities of the whole state space. When the state distribution is known, the mean performance measure can be approximated by summing over the truncated state space

$$r \approx \sum_{\mathbf{x} \in \tilde{\mathcal{S}}} r(\mathbf{x}) \pi(\mathbf{x}).$$

The larger the truncated state space is, the more accurate the results.

2.2 Value Extrapolation

Instead of solving the state probabilities using balance equations (1), we define and solve another state metric called relative value. The approach is well-known on the theory of Markov decision processes (see, e.g., [13]). Relative value $v(\mathbf{x})$ of state $\mathbf{x}$ is the conditional expected difference in cumulative revenue over infinite time horizon when starting from state $\mathbf{x}$ rather than from equilibrium:

$$v(\mathbf{x}) = \mathrm{E}\left[\int_{t=0}^{\infty} (r(X(t)) - r)\, dt \;\middle|\; X(0) = \mathbf{x} \right].$$

Relative values $v(\mathbf{x})$ satisfy the so-called Howard equations

$$r(\mathbf{x}) - r + \sum_{\mathbf{y}} q_{\mathbf{xy}} (v(\mathbf{y}) - v(\mathbf{x})) = 0 \quad \forall \mathbf{x}. \tag{2}$$

As seen, only the differences of the relative values appear in the equations, hence we may set, e.g., $v(\mathbf{0}) = 0$ when solving the equations. From the $|S|$ equations, the $|S| - 1$ unknown relative values along with the mean revenue rate r can be solved.

Similarly to the traditional solution method utilizing state probabilities and balance equations, the state space can be truncated by setting $q_{\mathbf{xy}} = 0$, $\forall y \notin \tilde{\mathcal{S}}$. In this paper, we introduce an approximation method that does the truncation more efficiently. We generalize the value extrapolation method that was briefly introduced in [9]. Instead of setting the transition intensities to zero, the relative values outside the truncated state space are extrapolated using the values inside (note that the Howard equations remain a closed set of equations). If the relative values outside $\tilde{\mathcal{S}}$ are correctly extrapolated, the mean value solved from (2) is exact. While exact results are obtained only in some special cases, extrapolation usually improves the accuracy significantly.

The extrapolation problem can be formulated as follows. We want to fit a function $f(\mathbf{x})$ to the relative values inside the truncated state space $\tilde{\mathcal{S}}$ so that it approximates also the values outside $\tilde{\mathcal{S}}$. Function f and fitting method need to be chosen so that the approximated relative values outside $\tilde{\mathcal{S}}$ are linear functions of the values inside, so that the group of equations (2) remains linear.

One linear extrapolation method is to use a polynomial function $f(\mathbf{x}) = \sum_{i=1}^{K} \sum_{j=0}^{n_i} a_{i,j} x_i^j$ and least squares fitting. The fitting can be done either globally or locally. When global fitting is used, all the $(\mathbf{x}, v(\mathbf{x}))$-pairs in $\tilde{\mathcal{S}}$ are used. The fitting can also be done locally, i.e. using only a subset $\mathcal{S}_f(\mathbf{x})$ of the truncated state space. The choice of $\mathcal{S}_f(\mathbf{x})$ may depend on the extrapolated point $\mathbf{x}$. Parameters $a_{i,j}$ are determined so that the sum of squared errors

$$Q = \sum_{\mathbf{x} \in \mathcal{S}_f} (f(\mathbf{x}) - v(\mathbf{x}))^2$$

is minimized. Function f and set $\mathcal{S}_f(\mathbf{x})$ need to be chosen so that the parameters have unambiguous values, i.e. the number of points in $\mathcal{S}_f(\mathbf{x})$ is equal or greater than the number of parameters. If the number of parameters and points are equal, the fitting reduces to ordinary polynomial fitting. The optimal parameter values are found by solving the following group of equations:

$$\frac{\partial Q}{\partial a_{i,j}} = 0 \quad \forall i, j.$$

The parameter values and hence also the function $f(\mathbf{x})$ is a linear function of relative values $v(\mathbf{x})$ inside the truncated state space $\mathcal{S}_f(\mathbf{x})$.

Mean revenue rate r can be approximated by defining the Howard equations (2) for the truncated state space and extrapolating the relative values outside $\tilde{\mathcal{S}}$ that appear in the equations. Value extrapolation does not alter the number of equations or variables in the group of equations hence there is no significant computational penalty when value extrapolation is used.

2.2.1 Example of Extrapolation

We demonstrate the extrapolation using a one-dimensional Markov process. The truncated state space of the process is $\tilde{\mathcal{S}} = \{x \mid 0 \leq x \leq N\}$. When linear extrapolation is used, the extrapolation function is of form $f(x) = a_1 x + a_0$. We use two data points in the fitting. In this particular case, the least squares fitting corresponds to finding the straight line crossing the two points, i.e. question is of ordinary fitting. However, we use the least squared error method in order to illustrate the approach. The squared error sum is

$$Q = (a_1 N + a_0 - v(N))^2 + (a_1(N-1) + a_0 - v(N-1))^2.$$

The values of parameters are solved using equations

$$\begin{cases} \frac{\partial Q}{\partial a_0} &= 2a_1(2N-1) + 4a_0 - 2v(N) - 2v(N-1) = 0 \\ \frac{\partial Q}{\partial a_1} &= 2a_1(N^2 + (N-1)^2) + 4a_0 - 2v(N) + \\ & \quad -2v(N-1) = 0. \end{cases}$$

The solution is

$$\begin{cases} a_0 = (1-N)v(N) + Nv(N-1) \\ a_1 = v(N) - v(N-1), \end{cases}$$

hence the values outside $\tilde{\mathcal{S}}$ are extrapolated using function

$$v(x) = (v(N) - v(N-1))x + (1-N)v(N) + Nv(N-1).$$

If the process is a birth-death-process, i.e. only increments and decrements of one customer are possible, the only relative values outside the truncated state space that appear in the Howard equations (2) are the outside values closest to the truncated state space. In this one-dimensional example, it would be sufficient to extrapolate the value at point $N+1$:

$$v(N+1) = 2v(N) - v(N-1).$$

3. PROCESSOR SHARING SYSTEMS

In this chapter, we study the accuracy of value extrapolation using multi-class processor sharing (PS) systems as examples. A PS system consists of a server and customers. Customers arrive at the server as a random process and depart when they have received a sufficient amount of service. Customers of a given class concurrently in the system obtain equal share of the capacity. The share of the capacity between different classes need not be egalitarian; we will consider discriminatory and generalized processor sharing systems as examples. We assume that the arrival process is Poissonian and the service requirements are exponentially distributed, hence the process is Markovian. As only one customer arrives or departs at a time, the system is a birth-death-process. PS processes can be used in modelling of many systems, for example telecommunication networks [10] or scheduling of operating systems [12].

We use the expected total number of customers $\mathrm{E}[|X|]$ as the performance measure, which, of course, is related to the mean sojourn time by Little's result. We study also the second moment $\mathrm{E}[|X|^2]$ as it is needed if variance of the mean occupancy is determined.

3.1 One-Dimensional Examples

First, we discuss simple PS systems with one customer class and state dependent service rates. The state x of the system is the number of customers in the system. Without loss of generality, we assume unit mean service requirement. The service rate $\phi(x)$ is a function of the system state. Customer arrival process is Poissonian with intensity λ. The Howard equations of the system read

$$x - r + \lambda(v(x+1) - v(x)) + \phi(x)(v(x-1) - v(x)) = 0.$$

3.1.1 Regular Service Rate Functions

If the service rate $\phi(x)$ is constant, quadratic extrapolation gives the exact mean number of customers as demonstrated in [9], which is one of our motivations for the usefulness of the method. Similarly, cubic extrapolation yields the exact result when the second moment $\mathrm{E}[|X|^2]$ is studied. If the service rate is not constant, the accuracy of value extrapolation depends on the service rate function $\phi(x)$. In some cases, exact results are obtained similarly to the case with constant service rate. One such function is, for example,

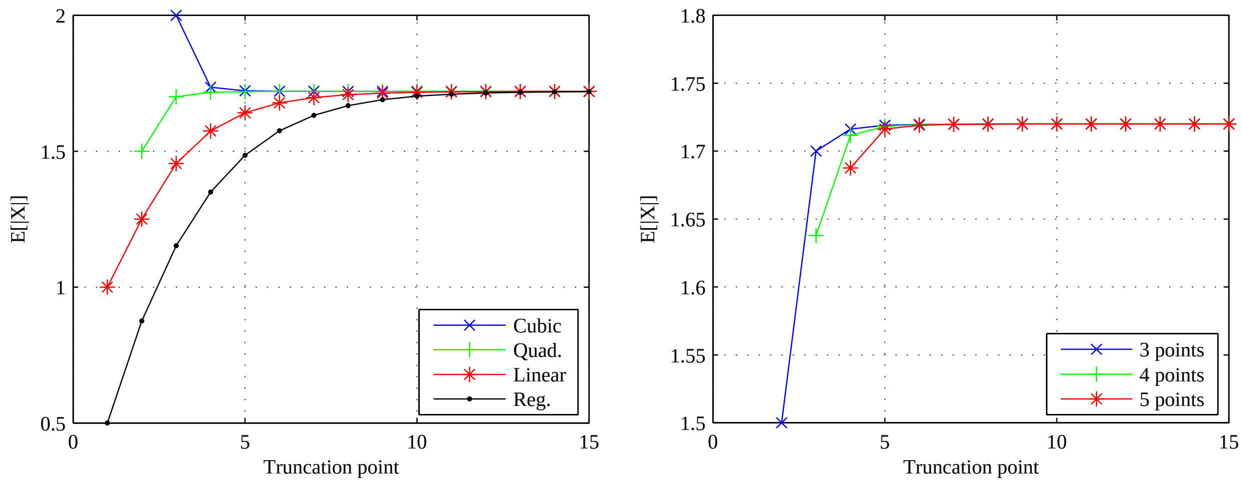
$$\phi(x) = \frac{x}{x+a}, \quad a > 0.$$

In general, value extrapolation does not yield exact results when applied to one-class processor sharing systems with arbitrary state-dependent service rates. However, it usually estimates the performance accurately even if the truncated state space is very small. We use function

$$\phi(x) = 1 - \frac{1}{2x}$$

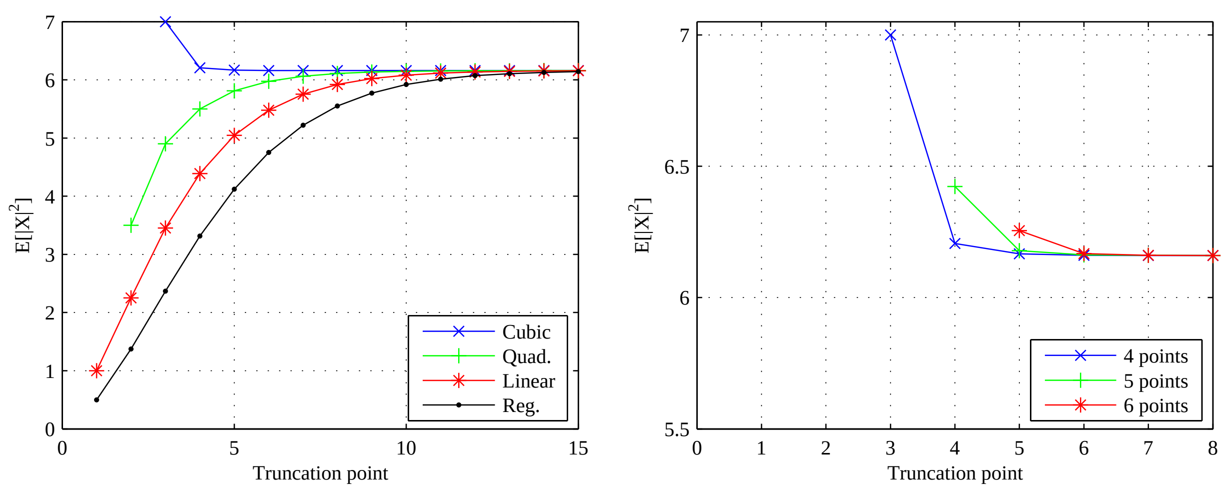
as an example and demonstrate the accuracy of value extrapolation using different extrapolation methods. The arrival intensity λ is assumed to have value $1/2$, i.e. the load of the system is 0.5.

Figure 1 illustrates the calculated mean occupancy $\mathrm{E}[|X|]$ as a function of the size of the truncated state space. In figure 1(a), the order of the extrapolation polynomial is varied between one and three. For comparison, we also include results calculated using the traditional method utilizing the state probabilities of the truncated state space. All the extrapolation methods are more accurate than the regular state space truncation when state spaces with similar sizes are considered. Quadratic extrapolation converges most rapidly hence it is used to study how the number of data points used in the fitting affects the accuracy of value extrapolation. The results are illustrated in figure 1(b). In this case, the less points are used in the extrapolation, the better results are obtained. Correspondingly, figure 2 illustrates the approximation of the second moment $\mathrm{E}[|X|^2]$. Extrapolation with a cubic polynomial converges most rapidly. In general, one may expect that for estimating $\mathrm{E}[|X|^n]$, $(n+1)$th order extrapolation is needed. Also in this case,



(a) Order of polynomial (b) Number of data points

Figure 1: $\mathrm{E}[|X|]$ **of a one-class PS queue with a regular service rate function at load** 0.5 **as a function of the truncation point. In figure (a), the order of the extrapolation polynomial is studied. In figure (b), the effect of the number of data points used in the fitting is studied using quadratic extrapolation.**



(a) Order of polynomial (b) Number of data points with cubic extrapolation

Figure 2: $\mathrm{E}[|X|^2]$ **of a one-class PS queue with load** 0.5 **with different extrapolation functions**

the straightforward polynomial fitting with four data points yields more accurate results than extrapolation with more data points and least squares fitting.

We studied also other regular service rate functions. The behaviour of value extrapolation usually resembles figures 1 and 2. Quadratic and cubic extrapolation yields the most accurate results when first and second moments are studied. Use of more data points in the fitting does not typically improve the accuracy of the approximation.

3.1.2 Irregular Service Rate Function

In the previous example, quadratic value extrapolation worked very well when only three data points were used. However, if the process is irregular, it may be better to use more data points in the fitting so that the results do not depend so much on the individual values. In order to study this issue, we construct an artificial process where the service rates are highly irregular. For each point x, the service rate $\phi(x)$ is drawn from a normal distribution with mean 1 and standard deviation $1/100$ (once drawn, the rate is fixed, i.e. there is no stochasticity in the system, just irregularity). The load of the system is 0.75. The mean occupancies with different extrapolation polynomials are illustrated in figures 3(a). Each curve is extrapolated using the minimum possible number of data points. While quadratic extrapolation yields the best results, it fluctuates strongly around the correct value. Linear extrapolation is more robust and also outperforms the traditional approach. In order to get better results, the second order polynomial is fitted using more data points and least squared error fitting. Figure 3(b) illustrates the effect of the number of points. The more points are used, the less fluctuations.

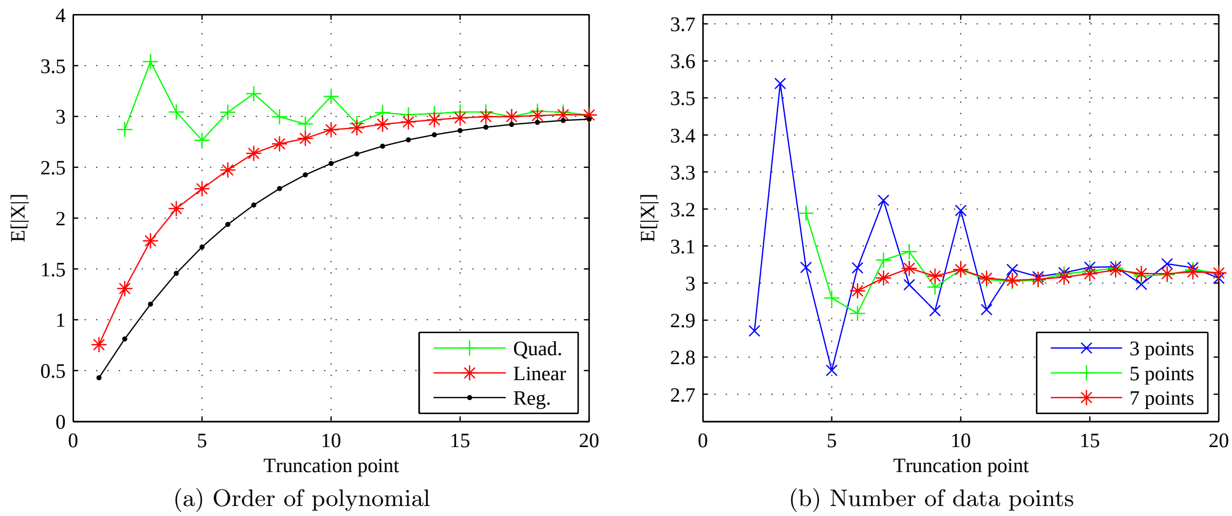


Figure 3: E$[|X|]$ of a one-class PS queue with load 0.75 and irregular service rate function as a function of the truncation point. In figure (a), the order of the extrapolation polynomial is studied. In figure (b), the effect of the number of data points used in the fitting is studied using quadratic extrapolation.

3.2 Multi-Dimensional Examples

Next, we study more general processor sharing systems with one server and K customer classes. The state of the system is $\mathbf{x} = (x_1, \ldots, x_K)$, where x_k is the number of class-k customers in the system. The arrival process of class k is Poissonian with intensity λ_k and the service requirement of class k is exponentially distributed with mean $1/\mu_k$. Without loss of generality, the server is assumed to have unit capacity. The service rate of class k is denoted $\phi_k(\mathbf{x})$.

When the first moment is studied, the Howard equations of the process are

$$|\mathbf{x}| - r + \sum_k \lambda_k (v(\mathbf{x} + \mathbf{e_k}) - v(\mathbf{x})) + \\ + \sum_k \phi_k(x)(v(\mathbf{x} - \mathbf{e_k}) - v(\mathbf{x})) = 0,$$

where $\mathbf{e_k}$ is a vector with 1 in element k and 0 elsewhere. The state space is truncated so that

$$\tilde{\mathcal{S}} = \{\mathbf{x} \mid 0 \leq x_i \leq N, \ \forall i\}\,.$$

We use one-dimensional extrapolation, i.e. the value at point $(x_1, \ldots, N+1, \ldots, x_K)$ is extrapolated using only points $(x_1, \ldots, x_i, \ldots, x_K), x_i \leq N$.

3.2.1 Discriminatory Processor Sharing

When discriminatory or weighted processor sharing (DPS) is used, class-k bandwidth is [7]

$$\phi_k(\mathbf{x}) = \frac{w_k x_k}{\sum_j w_j x_j},$$

where w_k is the weight parameter of class k. Egalitarian processor sharing is a special case of DPS with $w_k = 1$, $\forall k$.

In previous literature, many characteristics of DPS systems have been analyzed. Moments of the number of customers in a system can be exactly determined by solving a system of linear equations derived in [11]. The solution is used to validate the results of value extrapolation. It seems that the $(n+1)$th order extrapolation yields exact results when the nth moment of the occupancy is studied regardless of the parameter values. Exact results are also obtained if value extrapolation is used to approximate the mean number of customers in a specific class instead of the total number, i.e. $r(\mathbf{x}) = x_k$.

3.2.2 Generalized Processor Sharing

Another variant of processor sharing is generalized processor sharing (GPS), see, e.g., [15]. GPS capacity allocation is defined as follows:

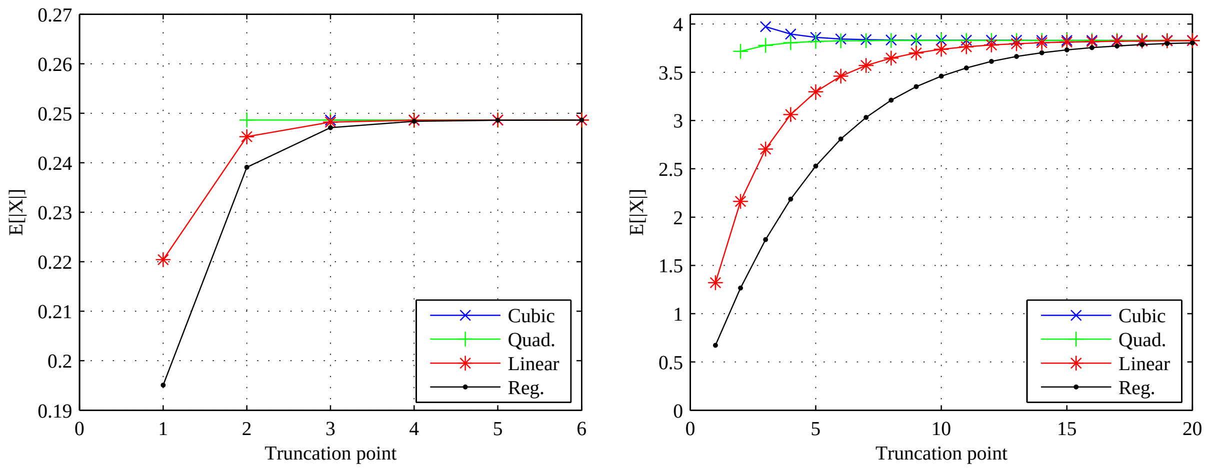
$$\phi_k(\mathbf{x}) = \frac{w_k}{\sum_{j:x_j>0} w_j}.$$

While value extrapolation yields exact results when applied to DPS, this is not the case with GPS. In order to estimate the accuracy of value extrapolation, we compare the different extrapolation methods using two customer classes. The parameter values used are $w_1 = 3/10$, $w_2 = 7/10$, $\lambda_1 = \lambda$, $\lambda_2 = 2\lambda$, $\mu_1 = 1/2$ and $\mu_2 = 2/3$. Parameter λ is varied so that the load of the system is either 0.2 or 0.8. The first moment is illustrated in figure 4. Quadratic extrapolation converges very quickly even with the higher load. Second moment E$[|X|^2]$ is illustrated in figure 5. In this case, cubic extrapolation yields the best results. In both cases, linear extrapolation clearly outperforms regular truncation.

4. APPLICATION IN DATA NETWORKS

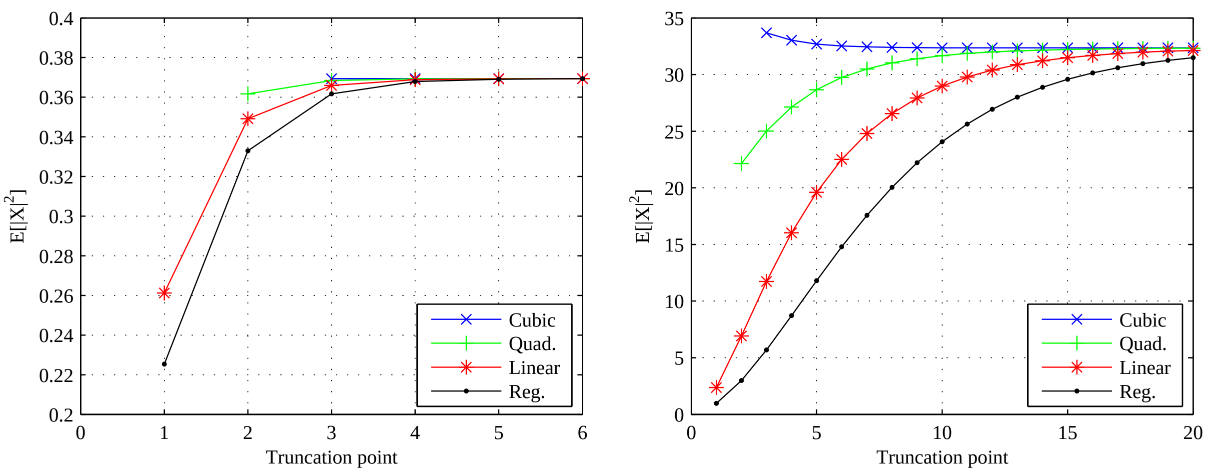
Data networks can be modelled as Markov processes [6]. We assume that the traffic of the network is elastic, i.e. the sizes of the transfered files are fixed and the transmission duration depends on the available capacity. Flows arrive randomly as a Poisson process and flow sizes are exponentially distributed.

Balanced fairness (BF) is a new resource sharing concept recently introduced by Bonald and Proutière [3] as a means to approximately evaluate the performance of fair allocations like max-min fairness and proportional fairness. Un-



(a) Load 0.2 (b) Load 0.8

Figure 4: $\mathrm{E}[|X|]$ of a GPS process with different extrapolation functions



(a) Load 0.2 (b) Load 0.8

Figure 5: $\mathrm{E}[|X|^2]$ of a GPS process with different extrapolation functions

der balanced fairness resource sharing the dynamic flow-level model becomes more tractable. While value extrapolation can be used to approximate any capacity allocation policy, the use of BF allows us to estimate the accuracy of the approximation as the exact mean occupancy can be determined in the tree networks we are analyzing [4]. BF allocation can be determined recursively starting from an empty network. When the allocation is known, the state transition intensities $q_{\mathbf{xy}}$ can be determined.

4.1 Two-Level Tree Network

The first example network is illustrated in figure 6. Assuming $C_1 < C_0$, $C_2 < C_0$ and $C_0 < C_1 + C_2$, the exact mean occupancy is [4]

$$\mathrm{E}[|X|] = \frac{(C_1 + C_2 - C_0)\rho_1\rho_2 + C_1C_2(1 - \rho_1 - \rho_2)}{(C_0 - \rho_1 - \rho_2)(C_1 - \rho_1)(C_2 - \rho_2)},$$

where $\rho_k = \lambda_k/\mu_k$ is the load of class-k traffic.

We use parameter values $\lambda_1 = 2\lambda_2$, $\mu_1 = 3$ and $\mu_2 = 2$.

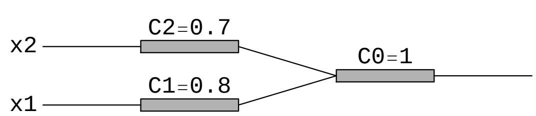
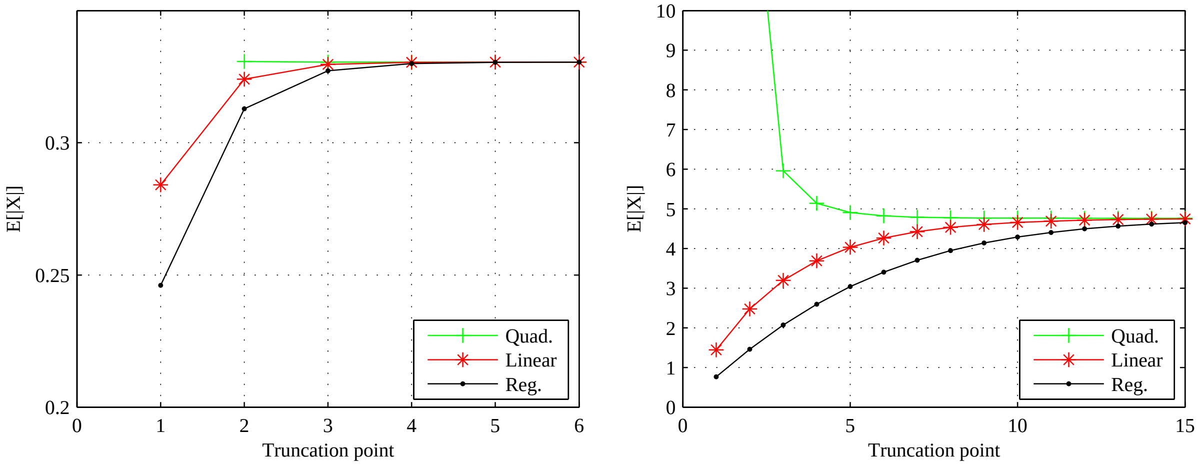


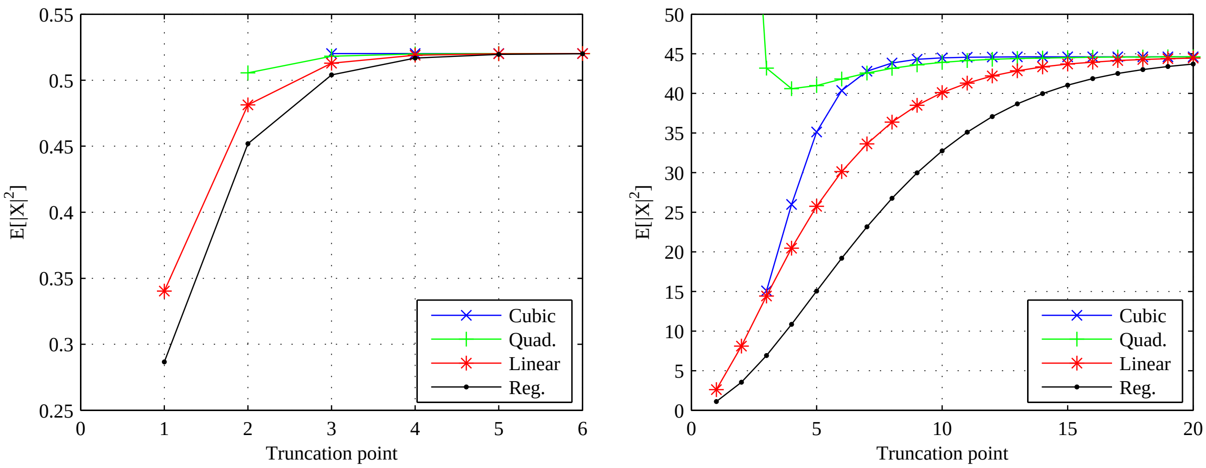
Figure 6: Two-level tree network

The amount of traffic is varied so that the total load of the system is either 0.2 or 0.8. The first moment is illustrated in figure 7. Quadratic extrapolation converges most rapidly. Also linear extrapolation outperforms regular truncation. For example, to get results within 1 percent of the exact value with load 0.8, the truncated state space needs to be at least 8×8 states when quadratic extrapolation is used, 14×14 states when extrapolation is linear and 19×19 states when the traditional method is used. The differences are even more pronounced when systems with more traffic classes are analyzed. The use of more data points in the polynomial fitting does not improve the accuracy.



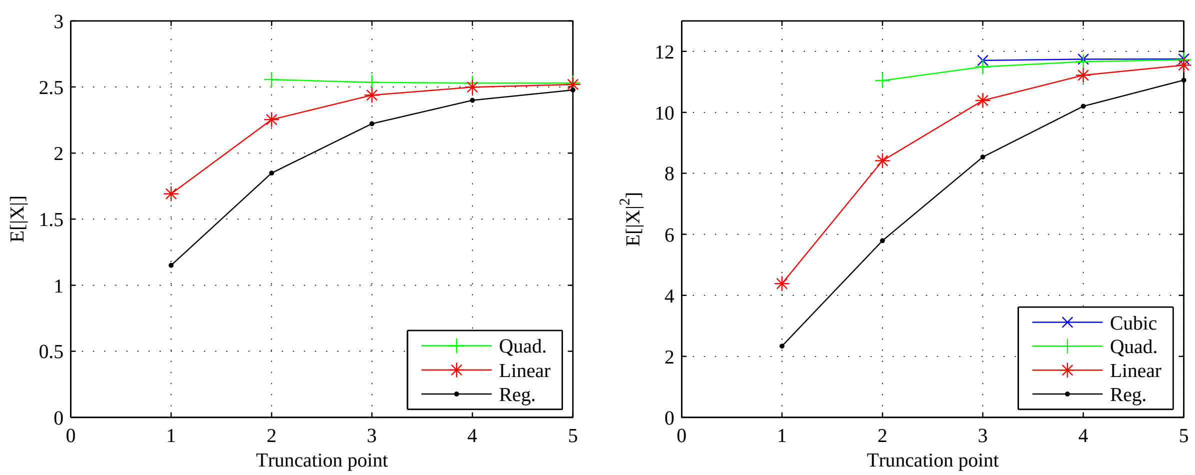
(a) Load 0.2 (b) Load 0.8

Figure 7: $\mathrm{E}[|X|]$ **of a two-level tree network with different extrapolation functions**



(a) Load 0.2 (b) Load 0.8

Figure 8: $\mathrm{E}[|X|^2]$ **of a two-level tree network with different extrapolation functions**



(a) First moment $\mathrm{E}[|X|]$ (b) Second moment $\mathrm{E}[|X|^2]$

Figure 9: Moments of the three-level tree network with load 0.6 **and different extrapolation functions**

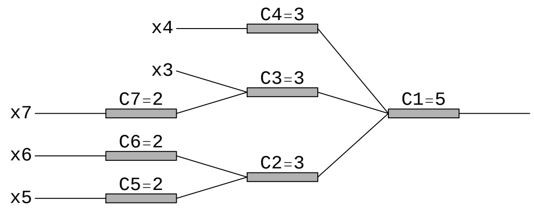


Figure 10: Three-level tree network

Similarly, second moment is illustrated in figure 8. With low network load, cubic extrapolation yields best results. When the network load is higher, quadratic extrapolation is more accurate than cubic when the size of the truncated state space is small. When the state space is bigger, cubic extrapolation is the most accurate.

4.2 Three-Level Tree Network

A more complex tree with three levels is illustrated in figure 10. Traffic parameters are $\lambda_3 = \lambda_5 = \lambda_7 = 2$, $\lambda_4 = \lambda_6 = 1$, $\mu_3 = \mu_5 = \mu_7 = 3$ and $\mu_4 = \mu_6 = 2$. With these values, the load of the system is 0.6. The accuracy results are illustrated in figure 9. Quadratic approximation estimates the first moment very accurately. In a system with multiple traffic classes, the effect on computation time is significant because the number of states needed is considerably smaller than with regular truncation. Second moment is illustrated in figure 9(b). Cubic extrapolation is the most accurate, but also the quadratic approximation converges quickly. In both cases, also linear extrapolation outperforms the regular truncation.

5. CONCLUSIONS

Markov processes with large or infinite state spaces may be used in modeling of many systems. Performance of such systems can be analyzed exactly only in some special cases, hence approximative methods are needed. The traditional approach is to truncate the state space, solve the equilibrium state distribution, and use it to approximate the performance.

In this paper, we presented a new approximative method that can be used to evaluate any performance measure expressed as the expected value of a function of the system state. Instead of solving the state probabilities using the balance equations, the performance measure is determined directly using relative values of the states and Howard equations. The advantage of this approach is that the relative values outside the truncated state space can often be well extrapolated using a polynomial function and least squared error sum fitting without computational penalty.

We demonstrated the accuracy of value extrapolation using processor sharing systems and simple data networks as examples. When the first moment of occupancy is approximated, quadratic extrapolation yielded the best results. Correspondingly, cubic extrapolation approximated the second moment most accurately. When value extrapolation is used, the size of the truncated state space needed to reach certain accuracy is only a fraction of the size needed using the traditional approach. The more customer classes in the system, the more pronounced the difference is. Linear extrapolation outperformed the regular truncation in every example studied. While the higher order polynomials usually perform better, the linear extrapolation is a robust method to get more accurate results. If the transition intensities of the process are irregular, more accurate results are obtained if more data points and least squared error sum are used in the fitting.

While value extrapolation does not solve the state space explosion, it can significantly reduce computation time when multi-class systems are studied. In addition to the use for approximation, value extrapolation also showed promise as an analytical tool in the DPS example.

6. ACKNOWLEDGMENTS

This work was financially supported by the Academy of Finland grant n:o 210275.

7. REFERENCES

[1] K. Avrachenkov, U. Ayesta, P. Brown, and R. Núñes-Queija. Discriminatory processor sharing revisited. In *Proceeding of IEEE INFOCOM 2005*, pages 784–795, 2005.

[2] G. Bolch, S. Greiner, H. de Meer, and K. Trivedi. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. John Wiley & Sons, 1998.

[3] T. Bonald and A. Proutière. Insensitive bandwidth sharing in data networks. *Queueing Systems*, 44(1):69–100, 2003.

[4] T. Bonald and J. Virtamo. Calculating the flow level performance of balanced fairness in tree networks. *Performance Evaluation*, 58(1):1–14, 2004.

[5] P. Buchholz. Structured analysis approaches for large Markov chains. *Applied Numerical Mathematics*, 31(4):375–404, 1999.

[6] P. Harrison and N. Patel. *Performance Modelling of Communication Networks and Computer Architectures*. Addison Wesley, 1993.

[7] L. Kleinrock. Time-shared systems: A theoretical treatment. *Journal of the ACM*, 14(2):242–261, 1967.

[8] G. Koole. On the use of the power series algorithm for general Markov processes, with an application to a Petri net. *Journal of Computing*, 9(1):51–56, 1997.

[9] J. Leino, A. Penttinen, and J. Virtamo. Flow level performance analysis of wireless data networks: A case study. In *Proceeding of IEEE ICC*, 2006.

[10] L. Massoulié and J. Roberts. Bandwidth sharing and admission control for elastic traffic. *Telecommunication Systems*, 15(1):185–201, 2000.

[11] K. M. Rege and B. Sengupta. Queue-length distribution for the discriminatory processor-sharing queue. *Operations Research*, 44(4):653–657, 1996.

[12] A. Silberschatz, P. Galvin, and G. Gagne. *Applied Operating System Concepts*. John Wiley & Sons, 2000.

[13] H. Tijms. *Stochastic Models: An Algorithmic Approach*. John Wiley & Sons, 1994.

[14] N. van Dijk. Truncation of markov chains with application to queueing. *Operations Research*, 39(6):1018–1026, 1991.

[15] M. van Uitert. *Generalized Processor Sharing Queues*. PhD thesis, Eindhoven University of Technology, 2003.