Performance Evaluation of Client-Server Distributed Information Systems

Leonid Boguslavsky, Kenneth Sevcik
Computer Systems Research Institute,
University of Toronto,
Toronto, M5S 1A1, Canada
 {bogu,kcs}@csri.toronto.edu

Andrey I. Lyakhov
Institute of Control Sciences,
Russian Academy of Science,
Moscow, 117806, Russia
 bogu@ivs.msk.su

Abstract

For performance evaluation of client-server distributed information systems, a closed exponential queueing network with several classes of requests is proposed and analyzed. Exact methods are developed to calculate the model's stationary characteristics for use as estimates of performance measures. In the large scale case, when the number of processes being executed in parallel in the system is large, asymptotic approximations based on the Laplace method are obtained. These approximations permit rapid computation of performance estimates with a guaranteed precision.

Keywords

Modelling, distributed systems, client-server, asymptotic analysis.

1 INTRODUCTION

To provide high speed access to large common information collections for many users, modern distributed information systems (DIS) have client-server architectures (as shown in Figure 1). Large mainframes as well as small personal computers can be clients of the DIS. In this type of architecture, each client node (CN) has its own local memory units (LMUs) which are storage devices with small access delay and large transfer rate, such as solid-state memory devices or high-speed magnetic disks. The server controls a hierarchy of common storage devices and provides all necessary file management functions. This hierarchy usually consists of several levels of devices with different access rates including disks and an automated tape library (ATL).

Fig.1. Client-server distributed information system

Fig.2. Client node model.

2 MODEL DESCRIPTION

Let us consider an arbitrary CN with number $i \in \overline{1, N}$ where N is the number of CNs in the DIS. The number of statistically identical processes which are executed concurrently by this CN is equal to J_i , and this CN includes M_i LMUs. The behavior of processes is illustrated in Figure 2. Each process computes on the Central Processing Unit (CPU) for an average time μ_{0i}^{-1} and then accesses LMU number $j \in \overline{1, M_i}$ with probability p_{ji} , or the remote storage via the network interface unit (NIU) of CN i with probability q_i , or (with probability p_{0i}) completes and enters the thinking phase (TH) with the average duration λ_i^{-1} . Here we assume exponential distributions for all access times and durations of thinking and computing phases. The mean access times to LMU number $j \in \overline{1, M_i}$ and to the NIU of CN i are equal to μ_{ji}^{-1} and κ_i^{-1} , respectively. The thinking phase is modelled by an IS-station, while the CPU, NIU and LMUs of each CN are modelled as single servers (SSs) where the service discipline is PS for the CPU and FCFS for other units.

The server consists of the server NIU (SNIU), the control processor (CP), a collection of S disks and the ATL (as shown in Figure 3). After entering the server, a remote access request from CN i is served by the SNIU for an average time κ_0^{-1} and then either begins its service by the ATL for an average time γ_0^{-1} (i.e. the required file is read from a tape) with probability ζ_i (which in general depends on the number i of the CN), or is directed to the CP straight away. The required file movement to the highest storage hierarchy level and its transformation (if this file is archived) can be represented as a sequence of CP work phases (with the mean time γ_i^{-1}) followed the access to disks. The probability of an access to disk $j \in \overline{1, S}$ is equal to h_{ji} and this access has average duration (without the waiting time) of $(\mu_j^*)^{-1}$. A request finishes its service (i.e. the required file or block is ready to transfer through the network) after a given CN work phase with probability $h_{0i} = 1 - h_{1i} - \dots - h_{Si}$.

A user of a DIS can either copy the required file to its local site or access the file by a distributed file system (such as the Network File System (Sandberg *et al.*, 1985)). All CNs and the server are connected by two types of communication networks: 1) a channel-to-channel (CTC) network used to carry a very fast file transfer traffic; 2) a high-speed, multi-access local area network (MLAN) such as Ethernet or Token Ring, used to carry remote access request traffic and to transmit back the file blocks when the file is not copied as a whole.

Denote by $1 - \xi$ the probability of copying the file back to the CN. Then, after finishing a remote access request service, the server is sending a block from the required file through the MLAN with probability ξ and is sending the file as a whole by the CTC network with probability $1 - \xi$. The CTC network is constructed by connecting the I/O channels of all CNs and of the server to multiport channel-to-channel adapter. When copying a block of the required file from the server to CN i , the block is transferred, at first, by the server channel (SC) and then by the channel of CN i (CCN i). Thus, we can model this file transfer by two SSs (one for SC and other for CCN i) with rates μ_c and μ_{ci} , respectively, i.e. μ_c^{-1} and μ_{ci}^{-1} are average transfer times for SC and CCN i , respectively. So we model the CTC network by $N + 1$ SS centers (see Figure 4) (as did Drakopoulos and Merges (1992)).

Fig.3. Server model.

Fig.4. Queueing model of a client-server distributed information system

As a model of the MLAN, we use two IS-stations (MLANI and MLANR) to model passage through the MLAN of requests from CNs to the server and of the required file blocks (with the distributed access) in the reverse direction. The mean values of the request and block passage times for CN i at the IS-stations are the inverse values of parameters ϑ_i^I and ϑ_i^R and can be calculated by the iterative method of Drakopoulos and Merges (1992), depending on the type of MLAN and the rates of request and block transfers from and to all CNs. On the other hand, these rates are estimated by solving the suggested model with known values of ϑ_i^I and ϑ_i^R . Thus, we can obtain the values of these parameters iteratively by solving our model and applying the method of Drakopoulos and Merges (1992).

The service at all NIU's is required twice for any remote access: first, with request passage, and second, with response passage. Taking into account that our goal is to estimate throughputs of DIS's nodes, we can assume that the NIU service is only required with the request entering any NIU and SNIU, but this service time includes also the real NIU's service time with a response passage. It allows us to describe the processes/requests/responses concerning with CN $i \in \overline{1, N}$ by threads of the stable class i . If a thread of class i is thinking or at the CPU then the process of CN i is in the thinking or computing phase; otherwise, the process is waiting for (or getting) access to storage. If it is in any node LMU j , $j \in \overline{1, M_i}$, then the process requires access to this LMU. If it is in NIU, MLANI, SNIU, ATL, CP or D j (disks), $j \in \overline{1, S}$, then a request from CN i is performed by the corresponding units. Finally, if the thread occupies any of nodes MLANR or SC and CCN i , then either a response block or the required file itself is transferred by the MLAN or by the CTC network, respectively, to CN i . After finishing the storage access, a thread returns to CPU node (i.e. to a computing phase). Thus we can model the client-server DIS by the closed queueing network shown in Figures 2 to 4 with N classes of threads.

3 PERFORMANCE MEASURES AND EXACT EVALUATION METHODS

The main performance measures for the DIS are the following: the mean time T_i^{ep} for executing a process in CN i , utilizations and throughputs of the DIS nodes with respect to requests from CN i and the mean number of "thinking" processes T_i in CN i for all $i \in \overline{1, N}$. Table 1 presents all notations for these measures.

We can see from Figures 2 to 4 that all measures are related by the following balance equations for $i \in \overline{1, N}$:

$$\begin{aligned}
 \lambda_i T_i &= p_{0i} \mu_{0i} U_i^p, & q_i \mu_{0i} U_i^p &= \kappa_i U_i^I = \Lambda_i^I = \Lambda_i^S; \\
 p_{ji} \mu_{0i} U_i^p &= \mu_{ji} U_{ji}^I, & j &\in \overline{1, M_i}; & \kappa_0 U_{0i}^I &= \Lambda_i^I, \\
 \zeta_i \kappa_0 U_{0i}^I &= \Lambda_i^M; & \mu_j^* U_{ji}^d &= h_{ji} \gamma_i U_i^u, & j &\in \overline{1, S}; \\
 (1 - \xi) \Lambda_i^S &= \mu_c U_{0i}^k = \mu_{ci} U_i^k, \\
 h_{0i} \gamma_i U_i^u &= \Lambda_i^S, & \xi \Lambda_i^S &= \Lambda_i^R.
 \end{aligned} \tag{1}$$

Utilizations for CN i				
Nodes	CPU	NIU	LMU j	CCN
Notation	U_i^p	U_i^l	U_{ji}^l	U_i^k
Utilizations for requests from CN i				
Nodes	CP	SNIU	D j	SC
Notation	U_i^u	U_{0i}^l	U_{ji}^d	U_{0i}^k
Throughputs for requests from CN i				
Nodes	Server	ATL	MLANI	MLANR
Notation	Λ_i^s	Λ_i^M	Λ_i^l	Λ_i^R

Table 1: Notations for performance measures.

Finally, using Little's formula, we obtain

$$T_i^{ep} = \frac{J_i - T_i}{\lambda_i T_i}.$$

So we need to calculate only one measure for each class i (we choose T_i as this measure) and compute all other measures by using these balance equations.

For real DIS, the number n of CNs providing the parallel execution of a large number of processes and so consisting of large number of LMUs is quite small. The other $N - n$ CNs are usually personal computers (PCs), and there is only one executing process at a time in each of them. Let us group the $N - n$ PCs (small CNs) into R kinds so that for any PC $i \in \overline{n+1, N}$ of kind $r \in \overline{1, R}$:

$$\begin{aligned}
\hat{\lambda}_i &= \lambda_{r+n}^*, & \frac{q_i}{p_{0i}\kappa_0} &= \bar{\rho}_{0,r+n}^l, & \frac{\zeta_i q_i}{p_{0i}\gamma_0} &= \bar{\rho}_{r+n}^M, \\
\frac{q_i}{p_{0i}h_{0i}\gamma_i} &= \bar{\rho}_{r+n}^u, & \frac{q_i h_{ji}}{p_{0i}h_{0i}\mu_j^*} &= \hat{\rho}_{j,r+n}^d, & j &= \overline{1, S}, \\
\frac{(1-\xi)q_i}{p_{0i}\mu_c} &= \bar{\rho}_{0,r+n}^k
\end{aligned} \tag{2}$$

where

$$\hat{\lambda}_i^{-1} = \lambda_i + \frac{1}{p_{0i}\mu_{0i}} + \sum_{j=1}^{M_i} \frac{p_{ji}}{p_{0i}\mu_{ji}} + \frac{q_i}{p_{0i}\kappa_i} + \frac{q_i}{p_{0i}\vartheta_i^l} + \frac{\xi q_i}{p_{0i}\vartheta_i^R} + \frac{(1-\xi)q_i}{p_{0i}\mu_{ci}}.$$

We denote by N_r the number of PC's of kind r . In addition, we divide M_i LMU, $i \in \overline{1, n}$, for each of the large CN's into s_i groups and S server disks into s_0 groups so that for any unit k belonging to group j the following equations are true: for LMU group $j \in \overline{1, s_i}$ of CN i

$$\frac{p_{ki}}{p_{0i}\mu_{ki}} = \bar{\rho}_{ji},$$

and for disk group $j \in \overline{1, s_0}$

$$\frac{q_i h_{ki}}{p_{0i}h_{0i}\mu_k^*} = \bar{\rho}_{ji}^d \text{ with all } i \in \overline{1, n}, \text{ and } \hat{\rho}_{k,r+n}^d = \bar{\rho}_{j,r+n}^d \text{ with all } r \in \overline{1, R}.$$

Nodes	CPU	NIU	LMU j	CCN	CP	SNIU
Notation	l_i^p	l_i^l	l_{ji}^l	l_i^k	l_i^u	l_{0i}^l
Nodes	D j	SC	ATL	MLANI	MLANR	TH
Notation	l_{ji}^d	l_{0i}^k	l_i^M	l_i^l	l_i^R	l_i^t

Table 2: Notations for numbers of threads in DIS nodes.

There are m_{ji} uniformly loaded units (with the same utilization) in group j , $j \in \overline{1, s_i}$, $i = \overline{0, n}$.

Now we can identify a state of the queueing network in Figures 2 to 4 by the vector $\vec{l} = (\vec{l}_i, i = \overline{1, n+R})$, where $\vec{l}_i = [l_i^0, l_i^u, l_{0i}^l, (l_{ji}^d, j = \overline{1, s_0}), l_i^M, l_{0i}^k]$, with $i \in \overline{1, n}$ $\vec{l}_i^0 = [l_i^p, (l_{ji}^l, j = \overline{1, s_i}), l_i^l, l_i^k]$ and with $i > n$, this component is absent. Here components of vector $\vec{l}_i$ are numbers of threads of CN i for $i \in \overline{1, n}$ or of PCs of kind $i - n$ for $i > n$ in the corresponding nodes (see Table 2). In addition we denote by l_i^t the number of "thinking" processes in CN i , which is equal to the difference between the whole number of processes J_i (with $i \in \overline{1, n}$) or N_{i-n} (with $i > n$) and the sum of all components of vector $\vec{l}_i$.

Thus, we obtain the following expression for the stationary probability (Baskett *et al.*, 1975, Boguslavsky and Lyakhov, 1992):

$$\pi(\vec{l}) = \pi(\vec{0}) K_0(\vec{l}) \prod_{i=1}^n \hat{K}_i(\vec{l}_i) \prod_{i=1}^{n+R} K_i^*(\vec{l}_i), \quad (3)$$

where $K_0(\vec{l}) = \bar{l}_0^u! \bar{l}_0^l! \bar{l}_0^M! \bar{l}_0^k! \prod_{j=1}^{s_0} (m_{j0} - 1 + \bar{l}_j^d)!$; $\bar{l}_0^u$, $\bar{l}_0^l$, $\bar{l}_0^M$, $\bar{l}_0^k$ and $\bar{l}_j^d$ are sums of the appropriate components of vector $\vec{l}$ for all i ;

$$K_i^*(\vec{l}_i) = \frac{(\lambda_i^*)^{-l_i^t}}{l_i^t!} \frac{(\bar{\rho}_{0i}^l)^{l_{0i}^l}}{l_{0i}^l!} \frac{(\bar{\rho}_{ci}^M)^{l_i^M}}{l_i^M!} \frac{(\bar{\rho}_{ci}^u)^{l_i^u}}{l_i^u!} \frac{(\bar{\rho}_{0i}^k)^{l_{0i}^k}}{l_{0i}^k!} \prod_{j=1}^{s_0} \frac{(\bar{\rho}_{ji}^d)^{l_{ji}^d}}{l_{ji}^d!}; \quad (4)$$

further, for $i = \overline{1, n}$:

values of $\bar{\rho}$ with different subscripts in (4) are also defined by (2) with the replacement of $r + n$ by i ,

$$\frac{1}{\lambda_i^*} = \left(\frac{1}{\vartheta_i^l} + \frac{\xi}{\vartheta_i^R} \right) \frac{q_i}{p_{0i}} + \lambda_i^{-1};$$

finally,

$$\hat{K}_i(\vec{l}_i^0) = (\bar{\rho}_{ci}^p)^{l_i^p} (\bar{\rho}_{ci}^l)^{l_i^l} \bar{\rho}_{ci}^{l_{0i}^k} \prod_{j=1}^{s_i} \bar{\rho}_{ji}^{l_{ji}^l} \frac{(m_{ji} - 1 + l_{ji}^l)!}{l_{ji}^l!},$$

$$\bar{\rho}_{ci}^p = \frac{1}{p_{0i} \mu_{0i}}, \quad \bar{\rho}_{ci}^l = \frac{q_i}{p_{0i} \kappa_i}, \quad \bar{\rho}_{ci} = \frac{(1 - \xi) q_i}{p_{0i} \mu_{ci}}.$$

We refer to values of $\bar{\rho}$ with different subscripts as load factors for the corresponding nodes.

So we have for measure T_i :

$$T_i = (\lambda_k^* / \lambda_i) \tau(k) T_k^*, \quad i \in \overline{1, N}, \quad (5)$$

where $k = i$ and $\tau(k) = 1$ with $i \leq n$ and with $i > n$ PC i belongs to kind $k - n$ and $\tau(k) = 1/N_{k-n}$;

$$T_k^* = \sum_{\vec{l} \in \mathcal{L}} l_k^t \pi(\vec{l}), \quad (6)$$

$\mathcal{L}$ is the set of vectors $\vec{l}$ such that for any $i \in \overline{1, n + R}$ all components of vector $\vec{l}_i$ are non-negative integers, and their sum is not larger than J_i (with $i \leq n$) or N_{i-n} (with $i > n$). Thus, to estimate the DIS performance, we should, firstly, substitute the real values of the DIS parameters to equations (2) to obtain load factors; secondly, use formulae (3), (4) and (6) to calculate T_i^* for all $i = \overline{1, n + R}$ and then (by (5)) T_i for all $i = \overline{1, N}$; and finally, apply the balance equations (1) to obtain the values of other performance measures.

When the number of processes $J = \sum_{i=1}^n J_i + \sum_{r=1}^R N_r$ is large, the computational requirements of the exact methods are high. These requirements can be decreased by using convolution or exact MVA algorithms, but they still increase rapidly with the growth of J even for these algorithms. Thus we must use approximate methods (such as approximate MVA) the computational requirements of which are small and independent of J . Unfortunately, using such approximations, we cannot predict the error and be confident of the derived results. So, for this large scale case, we'll use the asymptotic Laplace method to develop approximations for T_i^* and, thus, for all performance measures. Together with the scale independence and small computational requirements of these approximations, they have the property that the order of their errors decreases with J ; for example, the asymptotic approximation $T = [1 + O(\delta)]T^0$ means that the order of approximation $T \approx T^0$ is equal to δ . In addition, these asymptotic methods allow us to find directly the DIS bottlenecks with respect to requests of some specific client as well as overall. We assume that some DIS node is a bottleneck if its utilization is near one.

4 ASYMPTOTIC APPROXIMATIONS

For the large scale case, we enumerate groups of LMUs and disks as followed:

$$m_{ji} = \text{const}, \quad j = \overline{1, v_i}; \quad \frac{m_{ji}}{J} = \alpha_{ji} = \text{const}, \quad j = \overline{v_i + 1, s_i}; \quad i = \overline{0, n}. \quad (7)$$

In this section, we use the notation of asymptotic analysis (Boguslavsky and Lyakhov, 1992, Olver, 1974): J is the large parameter ($J \rightarrow \infty$) defining the queueing network scale. Values of other queueing network parameters are defined by referring to J : $m \sim J^{-\gamma}$, $\gamma > 0$, means that the order of some parameter m is equal to the order of $J^{-\gamma}$; $m = \text{const}$ means that for any large J the parameter m is a positive constant; $m/J = \text{const}$ means that the order of m is equal to the order of J ; finally, for any small positive x , the relation $m = O(x)$ means that the order of $|m|$ is not larger than the order of x . We assume that

$$J \rightarrow \infty, \quad n = \text{const}, \quad R = \text{const}, \quad s_i = \text{const}, \quad i = \overline{0, n};$$

$$\beta_i = \frac{J_i}{J} = \text{const}, \quad i = \overline{1, n}; \quad \beta_{n+r} = \frac{N_r}{J} = \text{const}, \quad r = \overline{1, R}. \quad (8)$$

Using the approach of McKenna and Mitra (1982) based on the application to (3) and (6) of the Euler factorial formula and the multinomial one, we represent the formula for T_i^* as a multiple Laplace integral:

$$T_i^* = G^{-1} A_i \int_0^\infty \dots \int_0^\infty \Psi_i(\vec{z}) H(\vec{z}) \exp[-Jg(\vec{z})] d\vec{z}, \quad i = \overline{1, n+R}, \quad (9)$$

where $\vec{z} = (z_{jh}, \quad j = \overline{1, b_h}, \quad h = \overline{0, n})$; $b_h = s_h + d_h$, $d_h = 3$ with $h = \overline{1, n}$ and $d_0 = 4$; $A_i = \beta_i / (\lambda_i^* W_i)$;

$$W_i = \max \{W_i^0, \bar{\rho}_i^p, \bar{\rho}_i^l, \bar{\rho}_{ci}, (\bar{\rho}_{ji}, j = \overline{1, s_i})\}, \quad i = \overline{1, n},$$

and for $i = \overline{n+1, n+R}$ $W_i = W_i^0$ where

$$\begin{aligned} W_i^0 &= \max \left\{ \frac{1}{J\lambda_i^*}, \bar{\rho}_{0i}, \bar{\rho}_i^M, \bar{\rho}_i^u, \bar{\rho}_{0i}^k, (\bar{\rho}_{ji}^d, j = \overline{1, s_0}) \right\}, \quad i = \overline{1, n+R}; \\ d\vec{z} &= \prod_{h=0}^n \prod_{j=1}^{b_h} dz_{jh}; \quad H(\vec{z}) = \prod_{h=0}^n \prod_{j=1}^{v_h} z_{j+d_h, h}^{m_{jh}-1}; \quad \Psi_i(\vec{z}) = 1/Q_i(\vec{z}); \\ g(\vec{z}) &= \sum_{h=0}^n \left(\sum_{j=1}^{b_h} z_{jh} - \sum_{j=v_h+d_h}^{b_h} (\alpha_{j-d_h, h} - J^{-1}) \ln z_{jh} \right) - \sum_{i=1}^{n+R} \beta_i \ln Q_i(\vec{z}); \\ Q_i(\vec{z}) &= B_i + S_i(\vec{z}) + \sum_{j=1}^{b_0} \varphi_{ji} z_{j0}, \quad B_i = (N\lambda_i^* W_i)^{-1}; \\ S_i(\vec{z}) &\equiv 0 \text{ with } i > n \text{ and } S_i(\vec{z}) = \sum_{j=1}^{b_i} \rho_{ji} z_{ji} \text{ otherwise.} \end{aligned} \quad (10)$$

Here ρ_{ji} with $j = 1, 2, 3$ are equal to $\bar{\rho}_i^p$, $\bar{\rho}_i^l$, $\bar{\rho}_{ci}$ divided by W_i and $\rho_{ji} = \bar{\rho}_{j-3, i}/W_i$ with $j = \overline{4, b_i}$; $\varphi_{ji} = \bar{\rho}_{j-4, i}^d/W_i$ with $j = \overline{5, b_0}$ and for $j = \overline{1, 4}$ values of φ_{ji} are equal $\bar{\rho}_{0i}^l$, $\bar{\rho}_i^M$, $\bar{\rho}_i^u$, $\bar{\rho}_{0i}^k$ divided by W_i . Thus, all B_i , ρ_{ji} and φ_{ji} are $O(1)$. For $i = \overline{1, n}$ denote by j_i the indice from the subset $\{1, \dots, v_i + 3\}$ with the maximal value of ρ_{ji} . The partition function G is represented by the same integral (9) where the coefficient A_i and function $\Psi_i(\vec{z})$ are omitted.

The function $g(\vec{z})$ defined in (10) has the absolute minimum on $Z = \{\vec{z} : z_{jh} \geq 0, \quad j = \overline{1, b_h}, \quad h = \overline{0, n}\}$ in the point $\vec{z}^* = (z_{jh}^*, \quad j = \overline{1, b_h}, \quad h = \overline{0, n})$ which is the solution of the system A, which consists of the following equations and inequalities. Firstly, for $h = \overline{1, n}$

$$\begin{aligned} 1 - \frac{\alpha_{j-3, h} - J^{-1}}{z_{jh}^*} - \frac{\beta_h \rho_{jh}}{Q_h(\vec{z}^*)} &= 0, \quad j = \overline{v_h + 4, b_h}, \\ a_{jh} &\geq 0, \quad z_{jh}^* = 0, \quad j = \overline{1, v_h + 3}, \quad j \neq j_h, \\ (a_{jh} > 0, \quad z_{jh}^* = 0) &\text{ or } (a_{jh} = 0, \quad z_{jh}^* \geq 0), \end{aligned}$$

where $a_{jh} = 1 - \beta_h \rho_{jh} / Q_h(\vec{z}^*)$. Secondly,

$$1 - \frac{\alpha_{j-4, 0} - J^{-1}}{z_{j0}^*} - \sum_{i=1}^{n+R} \frac{\beta_i \varphi_{ji}}{Q_i(\vec{z}^*)} = 0, \quad j = \overline{v_0 + 5, b_0},$$

$$(a_{j0} > 0, \quad z_{j0}^* = 0) \text{ or } (a_{j0} = 0, \quad z_{j0}^* \geq 0), \quad j = \overline{1, v_0 + 4},$$

where

$$a_{j0} = 1 - \sum_{i=1}^{n+R} \frac{\beta_i \varphi_{ji}}{Q_i(\bar{z}^*)}.$$

Obviously, $z_{jh}^* \leq \text{const}$ for all $j = \overline{1, b_h}$ and $h = \overline{0, n}$.

Now we formulate the main statement defining the form of the asymptotic approximation for T_i^* .

Statement 1. Let the large scale conditions (7) and (8) be satisfied. Then the following approximation is valid

$$T_i^* = [1 + O(J^{-\delta})] A_i / Q_i^*, \quad i = \overline{1, n + R}, \quad (11)$$

where $\delta = 1/2$ and $Q_i^* = Q_i(\bar{z}^*)$. Moreover, if $z_{ji}^* > 0$ with some $i \in \overline{1, n}$ then this approximation for this i has the following form:

$$T_i^* = [1 + O(J^{-\delta})] (\lambda_i^* \rho_{ji} W_i)^{-1}. \quad (12)$$

To prove this Statement, we use the Taylor formula for $\Delta g(\bar{z}) = g(\bar{z}) - g(\bar{z}^*)$ and show that for points $\bar{z}$ which do not belong to

$$\{\bar{z}: |Q_i(\bar{z}) - Q_i^*| = O(J^{-1/2}), \quad i = \overline{1, n + R}\}$$

the relation $\exp[-J\Delta g(\bar{z})] = O(e^{-\omega J})$ with $\omega > 0$ is valid.

It turns out that in some sufficiently general case C one can prove that the error of approximations (11) and (12) is $O(J^{-1})$ instead of $O(J^{-1/2})$ (as in Statement 1). This case C is described by the following conditions. Firstly, for any $h \in \overline{1, n}$ either $a_{jh} = \text{const} > 0$ or

$$z_{jh}^* = \text{const} > 0 \text{ and } 1 - \rho_{jh} / \rho_{jh} = \text{const} > 0, \quad j = \overline{1, v_h + 3}, \quad j \neq j_h.$$

Secondly,

$$a_{j0} = \text{const} > 0 \text{ or } z_{j0}^* = \text{const} > 0, \quad j = \overline{1, v_0 + 4}.$$

Let us introduce the following sets of indices i and j :

$$I_1^0 = \{i_1, \dots, i_X\} = \{i: i \in \overline{1, n}, \quad z_{ji}^* = \text{const} > 0\}$$

and

$$I_0 = \{\bar{j}_1, \dots, \bar{j}_Y\} = \{j: j \in \overline{1, v_0 + 4}, \quad z_{j0}^* = \text{const} > 0\}.$$

X and Y are the numbers of indice values in these sets. Then the final condition of the case C is the following: either $Y = 0$ or $n + R \geq X + Y$ and there exists such subset $I_1 = \{i_{X+1}, \dots, i_{X+Y}\} \subset \{1, \dots, n + R\} - I_1^0$ that

$$|\det \|F_{kl}\|_{k,l=\overline{1, X+Y}}| = \text{const} > 0$$

where

$$F_{kl} = \begin{cases} \rho_{ji} & \text{with } i = i_l, \quad j = j_i & \text{for } k = l \in \overline{1, X}, \\ 0 & & \text{for } k, l \in \overline{1, X}, \quad k \neq l, \\ \varphi_{ji} & \text{with } i = i_l, \quad j = \bar{j}_{k-X} & \text{otherwise.} \end{cases}$$

Statement 2. Let the large scale conditions (7) and (8) and conditions of the case C be satisfied. Then the approximations (11) and (12) are valid with $\delta = 1$.

The validity of this Statement follows from the use of the Lemma and Statement 1 of Lyakhov (1994) for evaluation of the integral in (9).

To simplify the application of the approximation (11), one can use the solution $\bar{z}^0$ of the system of equations

$$1 - \frac{m_{jh}}{Jz_{jh}^0} - \frac{\beta_h \rho_{jh}}{Q_h(\bar{z}^0)} = 0, \quad j = \overline{1, b_h}, \quad h = \overline{1, n},$$

and

$$1 - \frac{m_{j0}}{Jz_{j0}^0} - \sum_{i=1}^{n+R} \frac{\beta_i \varphi_{ji}}{Q_i(\bar{z}^0)} = 0, \quad j = \overline{1, b_0},$$

instead of $\bar{z}^*$. This substitution is possible because it is easy to prove (using the form of the system A defining the vector $\bar{z}^*$) that $z_{jh}^0 = z_{jh}^* + O(J^{-\delta})$ for all $j = \overline{1, b_h}$ and $h = \overline{0, n}$ where the value of δ has been defined in Statements 1 and 2.

Due to these Statements, the balance equations (1) and the form of the system A, we have the following

Corollary. Let $z_{j,i}^0 > 0$ with some $i \in \overline{1, n}$. Then the utilization $U_{j,i}$ of the node of CN i which corresponds to the index j_i (i.e. $U_{j,i}$ is equal to U_i^p , U_i^l , U_i^k or U_{ji}^l for some $j \in \overline{1, M_i}$) is defined by the expression $U_{j,i} = 1 - O(J^{-\delta})$; that is, this node is a bottleneck of CN i . Let $z_{j0}^0 > 0$ with some $j \in \overline{1, v_0 + 4}$. Then we have the similar approximation $U_j = 1 - O(J^{-\delta})$ for the utilization U_j of the server node which corresponds to the index j (i.e. U_j is equal to the sum of utilizations U_i^u , U_{0i}^l , U_{0i}^k or U_{ji}^d for some $j \in \overline{1, S}$ for all $i = \overline{1, n + R}$); that is, this node is a bottleneck of the DIS in a whole.

5 NUMERICAL RESULTS

Here we give some results which illustrate the accuracy and efficiency of our asymptotic approximations. Let us consider the DIS consisting of the only large client node (LC) representing a large mainframe with M_1 identical LMUs providing the parallel execution of J_1 processes, and N_1 identical personal computers (PCs) where only one process is executing at a time and only one LMU is present in each of them. The hierarchical server storage consists of two disk levels (with numbers of disks equal to m_{10} and m_{20}) and ATL. For all disks belonging to the same level $j = 1, 2$, the mean access time and the probability of access are equal to $(\mu_j^*)^{-1}$ and h_j/m_{j0} for any CN, respectively, where h_j is the probability of access to level j . Thus, in terms of section 3, we have $n = 1$, $R = 1$, $s_0 = 2$, $s_1 = 1$.

To illustrate the accuracy, we compare the values of performance measures derived by the approximations and by exact methods. Note that asymptotics are effective only with values of J_1 and N_1 being larger than ten, but several dozens of hours are required in the common case to calculate the performance measures by exact methods (with the use of IBM PC 486) even with $J_1 = N_1 = 10$. Let us consider the concrete case of values of system parameters which are given in Table 3. It is easy to see that this is the case of

Numbers of processes and devices						
Parameters	J_1	N_1	M_1	m_{10}	m_{20}	
Values	80	40	23	23	46	
Probabilities of access						
Parameters	q_1	q_2	$p_{01} = p_{02}$	ξ		
Values	0.04	0.05	0.001	0.2		
Parameters	$\zeta_1 = \zeta_2$		h_1	h_2		
Values	0.05		0.792	0.198		
Mean access times (msec)						
Parameters	μ_{01}^{-1}	μ_{02}^{-1}	μ_{11}^{-1}	μ_{12}^{-1}	κ_1^{-1}	κ_2^{-1}
Values	680	80	11.98	16.67	12.5	100
Parameters	γ_0^{-1}	κ_0^{-1}	$\gamma_1^{-1} = \gamma_2^{-1}$	$(\mu_1^*)^{-1}$	$(\mu_2^*)^{-1}$	μ_c^{-1}
Values	200	10	0.099	2.88	23	12.5
Parameters	μ_{c1}^{-1}	μ_{c2}^{-1}	$(\vartheta_1^I)^{-1}$	$(\vartheta_2^I)^{-1}$	$(\vartheta_1^R)^{-1}$	$(\vartheta_2^R)^{-1}$
Values	15.63	50	14.29	20	33.33	25

Table 3: Values of system parameters.

evenly distributed load when load factors for all server nodes and for all LC nodes, except of the processor, satisfy the equations

$$\bar{\rho}_{0i}^I = \bar{\rho}_i^M = \bar{\rho}_i^u = \bar{\rho}_{ji}^d = \bar{\rho}_{0i}^k = \bar{\rho}_i^s \quad \text{for } i = 1, 2 \text{ and } j = 1, 2 \quad (13)$$

and

$$\bar{\rho}_{11} = \bar{\rho}_1^I = \bar{\rho}_{c1} = \bar{\rho}_1^0. \quad (14)$$

According to the balance equations (1), these equations lead to the following equations of utilizations: $U_1^I = U_{11}^I = U_1^k$, $U_{0i}^I = U_{ji}^d = U_i^u = U_i^M = U_{0i}^k$ with $i = 1, 2$ and $j = 1, 2$ where $U_i^M = \Lambda_i^M / \gamma_0$ is the ATL utilization.

Equations (13) and (14) allow us to simplify significantly the exact formulae (3) - (6) by gathering all $\bar{S} = S + 4$ server units into the first group and all $\bar{M}_1 = M_1 + 2$ LC units (except of the CPU node) into the second group, which significantly decreases the computational requirements. Accordingly, the form of the Laplace integral in (9) is also simplified so that $v_1 = 1$, $b_1 = 2$, $j_1 = 1$, $v_0 = 0$, $b_0 = 1$, $\alpha_{11} = \bar{M}_1 / J$, $\alpha_{10} = \bar{S} / J$, $\rho_{11} = J \bar{\rho}_1^p$, $\rho_{21} = J \bar{\rho}_1^0$, $\varphi_{1i} = J \bar{\rho}_i^s$, $i = 1, 2$.

Table 4 shows the numerical results obtained for the values of parameters indicated in Table 3 and for the mean thinking time $T_i = \lambda_1^{-1} = \lambda_2^{-1} = 50$ sec by exact formulae (columns with title E) and by asymptotics (columns with title A). Here: $R_1^I = M_1 U_{11}^I$ and $R_2^I = N_1 U_{12}^I$ are the average numbers of busy LMUs at the LC and at all PCs; R_{ki}^d , $k = 1, 2$, is the average number of busy disks of level k which are accessed by requests from the LC ($i = 1$) or from PCs ($i = 2$). These results show that the processor of the LC is a bottleneck, i.e. $U_1^p \approx 1$.

As follows from the form of the asymptotic approximation, the approximate values of all utilizations and throughputs normalized by J are not changed with the scale (i.e. sum of J_1 and N_1) if values of J_1 , N_1 , $\bar{M}_1 - 1$ and $\bar{S} - 1$ as well as service rates of all single servers in the models of the LC, CTC network and the server also increase proportionally.

Measures	LC ($i = 1$)		PCs ($i = 2$)	
	E	A	E	A
T_i	31.26	33.97	13.72	13.57
U_i^p	0.920	1.000	0.0858	0.0849
R_i^l	7.19	7.81	4.35	4.30
$U_i^l = U_i^k$	0.313	0.340	0.0343	0.0339
$\Lambda_i^S = \Lambda_i^l \text{ (sec}^{-1}\text{)}$	25.01	27.17	13.72	13.57
$\Lambda_i^M \text{ (sec}^{-1}\text{)}$	1.250	1.359	0.686	0.679
R_{1i}^d	5.75	6.25	3.16	3.12
R_{2i}^d	11.50	13.50	6.31	6.24
$U_i^u = U_{0i}^l = U_{0i}^k$	0.250	0.272	0.00343	0.00339
$\Lambda_i^R \text{ (sec}^{-1}\text{)}$	5.00	5.44	2.74	2.72

Table 4: Values of performance measures.

So we can introduce a scale parameter H proportional to J (let $H = 4$ in the above example when $J = 120$) and investigate its effect on the accuracy of the approximation. Decreasing the scale, we have $J_1 = 40$, $N_1 = 20$, $M_1 = m_{01} = 11$ and $m_{02} = 22$ with $H = 2$, and $J_1 = 20$, $N_1 = 10$, $M_1 = m_{01} = 5$ and $m_{02} = 10$ with $H = 1$.

The curves in Figure 5 show the value of the server throughput Λ^S normalized by H with respect to the LC and PC's requests for various values of the mean thinking time T_i (that is, a load parameter). The solid curves are obtained by exact formulae with different values of the scale parameter H , while the plotted curves obtained by asymptotics are independent of H . In particular, we can see that for $0 \leq T_i < 50$ the server throughput remains unchanged (in asymptotics), because $z_{11}^* > 0$ in this case, and so the processor of the LC is a bottleneck, i.e. $U_i^p \approx 1$. For $T_i > 50$, the server throughput is monotonically decreasing with the growth of T_i . Note that due to the balance equations (1), the curves for other performance measures versus T_i have the same form.

Figure 5 shows a very small relative approximation error (it is within 2%) with evaluation of the server throughput with respect to PC's requests (see curves with title PC) even for small scale ($H = 1$) and for any load. Let us consider this error d with respect to the LC requests (curves with title LC in Figure 5) for various T_i and H (see Figure 6). According to Statements 1 and 2, the maximum of d is reached with $T_i = 50$ (thus, Table 4 shows the maximal difference between performance estimates obtained by exact formulae and by asymptotics) where $z_{11}^* = 0$ and $a_{11} = 0$, while d decreases quickly when T_i passes this value. When the scale parameter H increases above 1 (up to 4 in Figure 6), the maximal value of d decreases proportionally to $H^{1/2}$; for example, the maximum of d is equal to 19% for $H = 1$ and to 8.7% for $H = 4$. Moreover, the interval of the load parameter values where d is larger than some fixed level (7% in Figure 6), becomes narrower with the increase of H : $d > 7\%$ for $18 < T_i < 156$ if $H = 1$, for $36 < T_i < 90$ if $H = 2$, and for $47 < T_i < 60$ if $H = 4$.

Although these numerical results has been obtained only in the case of evenly distributed load, the derived conclusions remain true also for the general case where the computational requirements of our asymptotics are also very small (fractions of a second) and practically independent of the structure, scale and parameters of the DIS.

Fig. 5 Server throughput with respect to LC and PS's requests versus the mean thinking time for various scale factors.

Fig. 6 Relative approximation error versus the mean thinking time for various scale factors.

6 CONCLUSION

In this paper, we have suggested a closed exponential queueing network with several classes of requests as a model of a client-server distributed information systems. Its analysis has allowed us to evaluate performance measures by developing the exact methods for calculating the main stationary characteristics. In the case of a large number of processes executing in parallel in the system, asymptotic approximations based on the Laplace method have been obtained which provide a high calculation speed with a guaranteed precision and which identify the system bottlenecks. Its accuracy and efficiency have been illustrated by numerical results.

7 REFERENCES

- Baskett, F., Chandy, K.M., Muntz, R.R. and Palacios, F.G. (1975) Open, closed and mixed networks of queues with different classes of customers. *J.ACM*, **22**, 248-60.
- Boguslavsky, L. and Lyakhov, A.I. (1992) *Performance Evaluation Methods for Multiprocessors*. Nauka, Moscow [in Russian].
- Drakopoulos, E. and Merges, M.J. (1992) Performance analysis of client-server storage systems. *IEEE Trans. Comput.*, **41**, 1442-52.
- Lyakhov, A. (1994) Asymptotic analysis of nonuniform queueing network model of multiprocessor and multiterminal systems. *Automation and Remote Control*, **55**, 281-9.
- McKenna, J. and Mitra, D. (1982) Integral representations and asymptotic expansions for closed Markovian queueing networks: normal usage. *Bell Syst. Techn. J.*, **61**, 661-83.
- Olver, F.J.W. (1974) *Introduction to Asymptotics and Special Functions*. Academic Press, New York and London.
- Sandberg, R. *et al.* (1985) Design and implementation of the SUN network file system. *Proc. 10th Usenix Conf.*, Portland, Oregon, 119-30.

8 BIOGRAPHIES

Leonid B. Boguslavsky is a visiting professor in the Department of Computer Science at the University of Toronto. He also is the Head of the Laboratory of the Institute of Control Sciences, Russian Academy of Science in Moscow. His current research interests are in performance evaluation and modelling of computer systems, with focus on multiprocessor architecture, computer networks, and parallel processing. Boguslavsky received M.S. and Ph.D. degrees in computer science from the Moscow Institute of Transport Engineering. He is an author of three books, on traffic control in computer networks, the design of computer networks, and the performance evaluation of multiprocessors, respectively. He is a member of the American Mathematical Society.

Andrey I. Lyakhov is a Senior Research Scientist at the Institute of Control Sciences, Russian Academy of Science in Moscow, where he received his Ph.D in computer science. He holds also an M.S. degree in computer science from the Moscow Engineering and Physics Institute. His current research interests include modelling and analysis of large scale computer systems and networks. The focus of his work is the asymptotic analysis of large scale queueing network models to obtain highly effective measure approximations with corresponding error estimations. Lyakhov has co-authored a book, *Performance Evaluation Methods for Multiprocessors*.

Kenneth C. Sevcik is a Professor of Computer Science with a cross-appointment in Electrical and Computer Engineering at the University of Toronto. He is Director of the Computer Systems Research Institute, and past Chairman of the Department of Computer Science. He holds degrees from Stanford (B.S., Mathematics, 1966) and the University of Chicago (Ph.D., Information Science, 1971). His primary area of research interest is in developing techniques and tools for performance evaluation, and applying them in such contexts as distributed systems, database systems, local area networks, and parallel computer architectures.