KYBERNETIKA — VOLUME 37 (2001), NUMBER 6, PAGES 625–646

PROGNOSIS AND OPTIMIZATION OF HOMOGENEOUS MARKOV MESSAGE HANDLING NETWORKS[1]

PAVEL BOČEK, TOMÁŠ FEGLAR, MARTIN JANŽURA AND IGOR VAJDA

Message handling systems with finitely many servers are mathematically described as homogeneous Markov networks. For hierarchic networks is found a recursive algorithm evaluating after finitely many steps all steady state parameters. Applications to optimization of the system design and management are discussed, as well as a program product 5P (Program for Prognosis of Performance Parameters and Problems) based on the presented theoretical conclusions. The theoretic achievements as well as the practical applicability of the program are illustrated on a hypermarket network with 34 servers at different locations of the Czech Republic.

1. INTRODUCTION AND PROBLEM STATEMENT

By message handling we understand transmission of digital messages between objects called servers. Typical messages are data files, computer programs or electronic mail. Typical servers are computers. A *message handling network* (briefly, MHN) is a system defined by a set of servers $\mathcal{S} = \{S_1, \dots, S_m\}$ where $m > 1$, and by two sets of rules $\mathcal{R}_G$ and $\mathcal{R}_T$. The rules $\mathcal{R}_G$ specify how the messages are generated and how they enter and exit the servers from $\mathcal{S}$. The rules $\mathcal{R}_T$ specify how the messages are transmitted between the servers of $\mathcal{S}$.

The rules $\mathcal{R}_G$ can be reduced to the *convention* that $\mathcal{S}$ is extended by a virtual server S_0 representing the collection of all the network users who produce and/or consume the transmitted messages, and to the *assumption* that all the users altogether produce X messages per *time unit* (in symbols, TU), of an average size β [bit]. The number X may be random, with the expectation

$$EX = \lambda \ [1/\mathrm{TU}].$$

The parameters β and λ are related to the average rate R [bit/TU] of information generated by the collection of all users by the formula

$$\beta\lambda = R.$$

[1]Supported by the Grant Agency of the Czech Republic under Grant 102/99/1137.

If $\mathcal{S} = \{S_0, S_1, \ldots, S_m\}$ then the rules of transmission of messages from S_0 to the remaining servers $S_1, \ldots, S_m$, and vice versa, are contained in the set $\mathcal{R}_T$.

Thus the rules $\mathcal{R}_T$ must be able to specify for all $1 \le j \le m$ the numbers

$$n_j(t) \in Z_+, \quad Z_+ = \{0, 1, \ldots\},$$

of messages in the servers S_j at a time $t \ge 0$. In addition, we may set $n_0(t) = +\infty$, which means that there is an infinite number of messages in the virtual (external) server S_0. State of the MHN at a time $t \ge 0$ is thus described by a vector

$$n(t) = (n_1(t), \ldots, n_m(t))$$

from the state space Z_+^m, and the rules $\mathcal{R}_T$ must be able to specify evolution $n(t)$, $t \ge 0$, of the state of MHN. We assume that $n(0) = (0, \ldots, 0) \in Z_+^m$ and that the states $n(t)$ for $t > 0$ are random vectors with values in Z_+^m. More precisely, we assume that $n(t)$, $t \ge 0$, is a homogeneous Markov process. A Z_+^m-valued homogeneous Markov process is specified by a semigroup of stochastic matrices

$$P(s) \equiv (P_{n,\tilde{n}}(s))_{n,\tilde{n} \in Z_+^m}, \quad s \ge 0$$

where

$$P_{n,\tilde{n}}(s) = \Pr\left(n(t+s) = \tilde{n} \,|\, n(t) = n\right), \quad t \ge 0,$$

and for all s, $\tilde{s} \ge 0$

$$P(s)\, P(\tilde{s}) = P(s + \tilde{s}), \quad P(0) = I,$$

with I being the identity matrix.

Our rules $\mathcal{R}_T$ are thus reduced to the parameters β, λ and to the semigroup $P(s)$, $s \ge 0$. Under standard technical assumptions (see, e.g., Norris [10] for a detailed treatment), the semigroup is uniquely determined by a matrix

$$Q = (Q_{n,\tilde{n}})_{n,\tilde{n} \in Z_+^m}$$

of intensities of transitions from the states n to $\tilde{n}$ satisfying for every $s > 0$ the equations

$$\frac{\mathrm{d}}{\mathrm{d}s} P(s) = P(s)\, Q \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}s} P(s) = Q P(s).$$

These equations together with the above considered relations imply in particular that

$$\left(\frac{\mathrm{d}}{\mathrm{d}s} P(s) \right)_{s=0} \triangleq \lim_{s \to 0} \frac{P(s) - P(0)}{s} = Q,$$

i.e.

$$P(s) = I + s \cdot Q + o(s) \quad \text{as } s \to 0. \tag{1}$$

Therefore if a state $\tilde{n} \in Z_+^m$ differs from $n \in Z_+^m$ then for every $t \ge 0$

$$\Pr\left(n(t+s) = \tilde{n} \,|\, n(t) = n\right) = s \cdot Q_{n,\tilde{n}} + o(s) \quad \text{as } s \to 0,$$

and for every $n \in Z_+^m$ and $t \ge 0$

$$\Pr\left(n(t+s) = n \,|\, n(t) = n\right) = 1 + s \cdot Q_{n,n} + o(s) \quad \text{as } s \to 0.$$

It follows from here that $Q_{n,\tilde{n}} \geq 0$ for $n \neq \tilde{n}$ and

$$Q_{n,n} = -\sum_{\tilde{n} \in Z_+^m,\, \tilde{n} \neq n} Q_{n,\tilde{n}} \quad \text{for every } n \in Z_+^m. \tag{2}$$

Assuming that Q is irreducible and non-explosive (see again Norris [10], section 3.5), our problem is to find, whenever it exists, a stationary (steady-state) probability distribution

$$\pi = (\pi_n : n \in Z_+^m) \tag{3}$$

on the state space Z_+^m for a given MHN under consideration. In other words, the problem is to prove that

$$\lim_{n \to \infty} P(t) = \Pi,$$

where all rows of the matrix Π are identical, equal π. Indeed, then

$$\pi P(t) = \pi \quad \text{for all } t \geq 0,$$

i. e. then π satisfies the standard Markov stationarity condition.

We shall see that a stationary distribution π enables to evaluate very easily the steady-state expectation of the corresponding Markov process

$$\nu = (\nu_1, \ldots, \nu_m) \triangleq E\, n = \sum_{n \in Z_+^m} n\, \pi_n, \tag{4}$$

which is sufficient for prognosis of performance of the corresponding MHN and for prognosis of eventual problems under extremal payloads. In this sense a reasonably fast evaluation of the distribution π can be used to evaluate various designs of MHN's and to choose among them the optimal one.

It follows from (1) that if there exists a row vector $\pi = (\pi_n : n \in Z_+^m) \in (0,1)^{Z_+^m}$ solving for a matrix of transition intensities Q the system of equations

$$\pi Q = 0 \quad \text{and} \quad \sum_{n \in Z_+^m} \pi_n = 1 \tag{5}$$

then π is a stationary distribution of the corresponding Markov process. Thus for all MHN's observing the transmission rules $\mathcal{R}_T$ under consideration we reduced the problem of estimation of performance parameters and prognosis to the statistical estimation of b, λ and Q and to the solution of equations (5).

Z_+^m-valued time-homogeneous Markov processes $n(t) = (n_1(t), \ldots, n_m(t))$, $t \geq 0$, considered above are traditional mathematical models of queuing networks with m servers $S_1, \ldots, S_m$ where $n_j(t)$ is a random size of queue of customers in (or in front of) the server S_j at time $t \geq 0$. A systematic theory of such networks has been presented, e. g., by Walrand [14]. A more recent treatment can be found in van Dijk [13]; see also corresponding chapters in Walrand [14], Pattavina [12] and Higginbottom [3]. In this paper we extend the theory presented in Walrand [14] and van Dijk [13] to the situation where the network customers are messages and the servicing of messages follows rules prescribed by message handling protocols. We also propose special statistical procedures for estimation of parameters β, λ and Q, and for testing hypotheses about these parameters. Further, for an important class of hierarchic networks we derive a recursive algorithm solving analytically the equations (5). It is able to find in a reasonable time exact solutions π for quite large MHN's (large m). Finally, we report about our program 5P (Program for Prognosis of Performance Parameters and Problems) evaluating on the basis of this algorithm the solutions π and related parameters serving for prognosis of performances of MHN's in the steady-state, and for prognosis of eventual problems in these networks.

2. SIMPLIFICATION OF MATRIX Q

Let us consider an arbitrary MHN specified in Section 1, defined by a set of servers $\mathcal{S} = \{S_0, S_1, \ldots, S_m\}$, positive parameters β, λ, and a matrix of intensities of transitions Q with zero sums of rows and nonnegative non-diagonal elements.

By definition, the matrix Q is infinitely dimensional and thus at the first sight practically intractable. In this section we discuss conditions which essentially simplify its structure.

Consider special values of the state vector n, namely

$$e_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0) \in Z_+^m \quad \text{for } 1 \le i \le m$$

where 1 is at the place i, and

$$e_0 = (0, 0, \ldots, 0) \in Z_+^m.$$

We have seen in Section 1 that the elements $Q_{n,\tilde{n}}$ of the matrix Q characterize probabilities of transitions $n \mapsto \tilde{n}$ of the network states $n(t) = n$ in time intervals $(t, t+s)$ of a very short duration s. Consider the following three events in an interval of a very short duration s.

E_1: No message is transmitted between the servers of $\mathcal{S}$ (this implies $\tilde{n} = n$);

E_2: One message is transmitted between the servers of $\mathcal{S}$ (this implies $\tilde{n} = n - e_j + e_k$ for some $0 \le j,\ k \le m,\ j + k \ne 0$);

E_3: More than one message is transmitted between the servers of $\mathcal{S}$.

We assume that the transmission of messages from servers of $\mathcal{S}$ is autonomous (independent) in the sense that the probability of E_3 is negligible (like $o(s^2)$) with respect to the probability of $E_1 \cup E_2$. As easy to see, $E_1 \cup E_2$ implies $\tilde{n} = n - e_j + e_k$ for some $0 \le j,\ k \le m$. Therefore

$$Q_{n,\tilde{n}} = 0 \quad \text{unless} \quad \tilde{n} = n - e_j + e_k \quad \text{for some } 0 \le j,\ k \le m.$$

This means that the matrix Q is sparse in the sense that majority of its elements is zero. The potentially nonzero elements $Q_{n,\tilde{n}}$ are situated on or near the diagonal.

Moreover, the potentially nonzero elements of the matrix Q are assumed to be of the form

$$Q_{n,n-e_j+e_k} = \lambda_{j,k} \frac{\phi_{n-e_j}}{\phi_n} \quad \text{for all } 0 \le j,\ k \le m,\ j \ne k, \tag{6}$$

where $\phi_n > 0$ for $n \in Z_+^m$ characterizes a potential of the network in the state n to transmit messages (we can put formally $\phi_n = 0$ for $n \notin Z_+^m$). The matrix

$$\Lambda = (\lambda_{j,k})_{j,k=0}^m \quad \text{with } \lambda_{j,j} = 0 \quad \text{for } 0 \le j \le m \tag{7}$$

and the sequence ϕ_n, $n \in Z_+^m$, characterize uniquely the whole matrix Q. Indeed, by (2), the diagonal elements of Q are given by the formula

$$Q_{n,n} = -\sum_{j,k=0}^{m} \lambda_{j,k} \frac{\phi_{n-e_j}}{\phi_n} \tag{8}$$

and the elements $Q_{n,\tilde{n}}$ appearing neither in (6) nor in (8) are zero.

In the most common case, where the potential of the network to transmit messages is not influenced by the state n, we put $\Phi_n = 1$ for all $n \in Z_+^m$. In this case the matrix Q is determined by $m(m+1)$ nonnegative parameters – the nondiagonal elements of Λ.

Let us point out that the number $m(m+1)$ of unknown parameters of the matrix Λ can be reduced by 1. Indeed, the overall intensity λ of transmission of messages from the server S_0 to the set of servers $\{S_1, \dots, S_m\}$ must satisfy for every $n \in Z_+^m$ the relation

$$\lambda = \sum_{k=1}^{m} Q_{n,n+e_k}.$$

Therefore, by (6),

$$\lambda = \sum_{k=1}^{m} \lambda_{0,k} \tag{9}$$

where λ is assumed to be given.

Note that the intensities $Q_{n,\tilde{n}}$ as well as $\lambda_{j,k}$ are assumed to be measured in the same units as the intensity λ, i.e. in [1/TU].

3. SOLUTION FOR GENERAL MHN'S

In this section we consider an arbitrary MHN specified in Sections 1 and 2, defined by a set of servers $\mathcal{S} = \{S_0, S_1, \dots, S_m\}$, positive parameters β, λ, an $(m+1) \times (m+1)$ matrix Λ considered in (7), and a sequence of potentials ϕ_n considered in (6). The problem is to solve the system of equations (5) for the matrix Q with the nonzero elements given by (6) and (8).

It is easy to see that, under our assumptions, (5) is equivalent to

$$\sum_{j,k=0}^{m} \left(\pi_{n-e_j+e_k} Q_{n-e_j+e_k,n} - \pi_n Q_{n,n-e_j+e_k}\right) = 0 \quad \text{for all } n \in Z_+^m \tag{10}$$

and

$$\sum_{n \in Z_+^m} \pi_n = 1. \tag{11}$$

The infinite system of equations (10) can be considerably simplified by seeking for a solution $\pi = (\pi_n : n \in Z_+^m)$ in the form

$$\pi_n = c\,\phi_n \prod_{\ell=1}^{m} w_\ell^{n_\ell} \quad \text{for } n = (n_1, \ldots, n_m) \in Z_+^m, \tag{12}$$

where $w = (w_1, \ldots, w_m)$ is a vector of positive constants not depending on n and $c > 0$ is a normalization constant which is to be specified from equation (11). After substituting (12) in (10) and defining $w_0 = 1$, we obtain

$$\sum_{j,k=0}^{m} \left(\phi_{n-e_j+e_k} \prod_{\ell=1}^{m} w_\ell^{n_\ell} \frac{w_k}{w_j} \lambda_{k,j} \frac{\phi_{n-e_j}}{\phi_{n-e_j+e_k}} - c\,\phi_n \prod_{\ell=1}^{m} w_\ell^{n_\ell} \lambda_{j,k} \frac{\phi_{n-e_j}}{\phi_n} \right) = 0$$

and, after an obvious simplification,

$$\prod_{\ell=1}^{m} w_\ell^{n_\ell} \sum_{j=0}^{m} \phi_{n-e_j}\, w_j \left(\sum_{k=0}^{m} (w_k \lambda_{k,j} - \lambda_{j,k} w_j) \right) = 0.$$

From here we obtain the following result.

Theorem 1. If the system of equations

$$\sum_{k=0}^{m} (w_k \lambda_{k,j} - \lambda_{j,k} w_j) = 0, \quad 1 \le j \le m \text{ with } w_0 = 1 \tag{13}$$

has a positive solution $w_1, \ldots, w_m$ then (12) with c satisfying (11) is a stationary distribution of the MHN under consideration. If $\phi_n = 1$ for all $n \in Z_+^m$ then c satisfying (11) exists if and only if $w_\ell < 1$ for all $1 \le \ell \le m$. In this case the stationary distribution is

$$\pi_n = \prod_{\ell=1}^{m} (1 - w_\ell)\, w_\ell^{n_\ell} \quad \text{for } n = (n_1, \ldots, n_m) \in Z_+^m. \tag{14}$$

Proof. The only assertion which remains to be proved is that the reduced system of equations for $1 \le j \le m$ figuring in (13) is equivalent to the full system for $0 \le j \le m$ which is required in the last equality preceding Theorem 1. This follows from the fact that the rank of the full system of $m+1$ equations is at most m because the sum of these equations is zero,

$$\sum_{j=0}^{m} \sum_{k=0}^{m} (w_k \lambda_{k,j} - \lambda_{j,k} w_j) = 0.$$

Therefore any one of these equations can be deleted. □

Remark 1. In fact, we have proved that the distribution π satisfies the stronger partial balance equations

$$\sum_{k=0}^{m} \left(\pi_{n-e_j+e_k} Q_{n-e_j+e_k,n} - \pi_n Q_{n,n-e_j+e_k}\right) = 0$$
$$\text{for every } j = 0, \ldots, m \text{ and all } n \in Z_+^m.$$

Remark 2. The solution $(w_1, \ldots, w_m)$ of (13) is unique providing the matrix Q given by (6) is irreducible (cf. Theorem 3.2 in Whittle [15]).

Theorem 1 is not an entirely new result. As already mentioned in Section 1, it can be obtained by adapting similar results of Walrand [14] or van Dijk [13] to the networks specified in Sections 1 and 2.

4. SOLUTION FOR HIERARCHIC MHN'S

Let us consider the same MHN as in the previous section with $\phi_n = 1$ for all $n \in Z_+^m$. Such an MHN is defined by a set of servers $\mathcal{S} = \{S_0, S_1, \ldots, S_m\}$, positive parameters β, λ, and an $(m+1) \times (m+1)$ matrix $\Lambda = (\lambda_{j,k})$ with $\lambda_{j,k} \geq 0$ and zeros on the diagonal. Obviously,

$$\lambda_{j,k} + \lambda_{k,j} = 0$$

is equivalent to the assumption that there is no message transmission link between the servers S_j and S_k. The matrix Λ thus defines a symmetric binary relation of "being connected by a link" on $\mathcal{S}$: Two servers S_j and S_k are connected by a link if at least one of the intensities $\lambda_{j,k}$ and $\lambda_{k,j}$ is positive. This relation is not reflexive ($\lambda_{jj} = 0$ for all $0 \leq j \leq m$) and it needs not to be transitive. It defines a graph $G(\tilde{\mathcal{S}})$ on every nonvoid subset of servers $\tilde{\mathcal{S}} \subset \mathcal{S}$.

We shall suppose that the graph $G(S_0, S_1, \ldots, S_m)$ is connected (which implies the irreducibility of the matrix Q) and its subgraph $G(S_1, \ldots, S_m)$ is a tree. This means that the virtual server representing the collection of network users, is connected to the servers $S_1, \ldots, S_m$ and that there is a hierarchy between the servers $S_1, \ldots, S_m$. To describe this hierarchy, define subsets of servers

$$S_K = \{S_k : k \in K\} \quad \text{for} \quad K \subset \{1, \ldots, m\} \triangleq K_0.$$

For every $k \in K_0$ let us denote by $C(k) \subset K_0 \setminus \{k\}$ the set of servers connected to k.

(1) We fix a unique *root server* $S_r \in \{S_1, \cdots, S_m\}$. This is a first order server, superordinated to all those from $S_{K_0 \setminus K_1} \neq \emptyset$ where $K_1 = \{r\}$.

(2) For every $i = 2, 3, \ldots$ we set

$$K_i = \bigcup_{j \in K_{i-1}} P(j)$$

where $P(j) = C(j) \setminus K_{i-2}$ for every $j \in K_{i-1}$. Thus $S_{P(j)} \subset S_{K_i}$ is the set of servers subordinated to the server $S_j \in S_{K_i-1}$, and $S_{n(j)} \in S_{K_{i-2}}$, where $\{n(j)\} = C(j) \cap K_{i-2}$, is the unique server superordinated to $S_j \in S_{K_{i-1}}$. The set of the ith order servers disjointly decomposes as follows

$$S_{K_i} = S_{B_i} + S_{L_i}$$

where S_{L_i} with $L_i = \{j \in K_i; P(j) = \emptyset\}$ is the (possibly empty) set of *leaf servers of the second order* and its relative complement S_{B_2} is the set of *branch servers of the ith order*.

(3) After finitely many steps, say ℓ, we obtain $S_{K_\ell} = S_{B_\ell} + S_{L_\ell}$ where

$$S_{B_\ell} = \emptyset \quad \text{and} \quad L_\ell \neq \emptyset,$$

i. e. ℓ is the maximal order of the tree. Then

$$S_L = \bigcup_{i=2}^{\ell} S_{L_i} \quad \left(\text{i.e. } L = \bigcup_{i=2}^{\ell} L_i \neq \emptyset\right)$$

is the set of *leaf servers* of the network and

$$S_B = \bigcup_{i=2}^{\ell-1} S_{B_i} \quad \left(\text{i.e. } B = \bigcup_{i=2}^{\ell-1} B_i \text{ where } B_i \neq \emptyset\right)$$

is the set of *branch servers* of the network. Obviously,

$$S_{K_0} = S_{\{r\}} + S_B + S_L \quad (\text{i.e. } \{1, \dots, m\} = \{r\} + B + L), \tag{15}$$

where S_B contains branch servers of all orders $1 < i < \ell$ while S_L may contain leaf servers of only some of the orders $1 < i \leq \ell$ (e. g. all servers in S_L may be of order ℓ).

Note that hierarchic networks of the described type (with S_0 connected by a link to all the leaf servers) are quite common in the practice.

Next follows a recursive algorithm which solves equations (13) for these networks. In this algorithm, $S_{n(j)}$ again denotes the (unique) superordinated server connected by a link with $S_j \in S_{K_0-\{r\}}$, and $S_{P(j)}$ denoted the (nonvoid) set of subordinated servers connected by a link with $S_j \in S_{K_0-L}$.

Step 1: Put $c_j = b_j = 0$ for all $j \in L$.

Step 2: Put

$$c_j = \sum_{k \in P(j)} \frac{\lambda_{k,j}(c_k + \lambda_{0,k})}{\lambda_{k,j} + b_k + \lambda_{k,0}} \quad \text{and} \quad b_j = \sum_{k \in P(j)} \frac{\lambda_{j,k}(b_k + \lambda_{k,0})}{\lambda_{k,j} + b_k + \lambda_{k,0}} \quad \text{for } j \in B_i$$

and $i = \ell - 1, \ell - 2, \dots, 1$ where $B_1 = \{r\}$.

Step 3: Put

$$w_r = \frac{c_r + \lambda_{0,r}}{b_r + \lambda_{r,0}}.$$

Step 4: Put

$$w_j = \frac{w_{n(j)}\lambda_{n(j),j} + c_j + \lambda_{0,j}}{\lambda_{j,n(j)} + b_j + \lambda_{j,0}} \quad \text{for } j \in B_i.$$

Theorem 2. For hierarchic networks under consideration, the above defined algorithm solves after finitely many operations the system of equations (13). The complexity of this algorithm is proportional to the network size m.

Proof. For $j \in L$ we get from Steps 1 and 4

$$w_j = \frac{w_{n(j)}\, \lambda_{n(j),j} + \lambda_{0,j}}{\lambda_{j,n(j)} + \lambda_{j,0}}$$

which verifies the jth equation of (13). Let us now consider $\ell - 1 < i < 1$ and suppose that for all $j \in B_{i+1} + L_{i+1}$ the equations of (13) with w_j given in Step 4 have already been verified. We are interested in the equation of (13) for $j \in B_i$. Since $P(j) \subset B_{i+1} + L_{i+1}$ and $n(k) = j$ for $k \in P(j)$, we obtain

$$w_{n(j)}\, \lambda_{n(j),j} - w_j\, \lambda_{j,n(j)} + \sum_{k \in P(j)} \left(\frac{w_j \lambda_{j,k} + c_k - \lambda_{0,k}}{\lambda_{k,j} + b_k + \lambda_{k,0}} \lambda_{k,j} - w_j\, \lambda_{j,k} \right) + \lambda_{0,j} - w_j\, \lambda_{0,j} = 0.$$

After standard calculations we obtain from here the solution

$$\begin{aligned} w_j &= \frac{w_{n(j)}\lambda_{n(j),j} + \sum_{k\in P(j)} \frac{\lambda_{k,j}(c_k+\lambda_{0,k})}{\lambda_{k,j}+b_k+\lambda_{k,0}} + \lambda_{j,0}}{\lambda_{n(j),j} + \sum_{k\in P(j)} \frac{\lambda_{j,k}(b_k+\lambda_{k,0})}{\lambda_{k,j}+b_k+\lambda_{k,0}} + \lambda_{j,0}} \\ &= \frac{w_{n(j)}\lambda_{n(j),j} + c_j + \lambda_{0,j}}{\lambda_{j,n(j)} + b_j + \lambda_{j,0}}, \end{aligned}$$

i.e. the solution prescribed by Step 4. If $j = r$ then the equation (13) differs from the previous one only by putting

$$\lambda_{n(j),j} = \lambda_{j,n(j)} = 0.$$

Therefore we obtain from the previous calculations the solution

$$w_r = \frac{c_r + \lambda_{0,r}}{b_r + \lambda_{r,0}}$$

which coincides with that given in Step 3. By taking into account the disjoint decomposition (15), we see that the values $w_1, w_2, \ldots, w_m$ defined by the algorithm solve the equations (13) for all $1 \le j \le m$. The proportionality of the complexity to m is easily seen from Steps $1-4$ if one takes into account that the number of elements in L (and thus the number of substitutions in Step 1) is bounded by m, and also that the maximal hierarchic order ℓ (and thus the number of computations in Steps 2 or 4) is also bounded by m. □

5. SPECIFICATION OF PARAMETERS β, λ AND Λ

In this section we consider the MHN's introduced in Sections 1, 2 and studied in Section 3 and 4. It was assumed there that the parameters of the networks are given, namely that there are given the average message size β [bit], average intensity of arrivals of messages into the network λ [1/TU], the nondiagonal elements of the matrix Λ, i. e. the intensities λ_{jk} [1/TU] for $0 \le j,\ k \le m$, and the sequence of potentials $\phi_,$, $n \in Z_+^m$. It was mentioned that if these parameters are at the disposal then it is possible to compute variables $w_1, \dots, w_m$ enabling an easy prognosis of performances of MHN's and prognosis of problems such as buffer overflows, congestions, unacceptable message delays (for more about this see the next section).

In this section we study methods for specification and verification of these parameters. For simplicity we restrict ourselves to the case $\phi_n = 1$ for $n \in Z_+^m$ considered in Sections 3 and 4, so that the attention is focused on the parameters β, λ and Λ. We propose methods for statistical estimation of these parameters, and for testing hypotheses about them.

Performance of any MHN is sensitive to the flow of information from the network users to the network, i. e. it depends on intensities and sizes of messages produced by the collection of users (virtual server S_0). The intensities, and possibly also the sizes, depend on hours of a day, days of a week and weeks of a year. We are usually interested in performances of MHN's during the periods of peak activities of users, when the flow of information from them (in average) culminates and the network is under maximal pressure. Therefore the statistical inference discussed below should be done under the extremal circumstances when the activity of users achieves a global maximum. However, one might be interested in the network performances under different circumstances, e. g. under various locally maximal activities of users. However, in every case assumptions about the users should be clarified as precisely as possible, and it should not be forgotten that conclusions drawn from the mathematical MHN model can be taken seriously only in situations where these assumptions are fulfilled. When we discuss in the sequel the inference about parameters β, λ and Λ, we have in mind the situation of globally maximal activity of users, or another well defined situation, and we assume that this situation remains unchanged during collecting empirical data.

5.1. Inference about β and λ

Suppose that users (or potential users) of an MHN produced empirical data (β_1, t_1), $\dots, (\beta_N, t_N)$, where β_i denotes the size of the ith message (in [bit]), and t_i the daytime (in [TU]) when this message was sent to the MHN (in the case of potential users of a planned MHN, t_i is the time when the message was ready for sending). We suppose that $t_1 < t_2 < \cdots < t_N$. Then one can deduce from the empirical data that the total amount of information

$$\sum_{i=1}^{N-1} \beta_i$$

was prepared for the network in the time interval (t_1, t_N) so that

$$\hat{R} = \frac{\sum_{i=1}^{N-1} \beta_i}{t_N - t_1} \quad \text{[bit/TU]} \tag{16}$$

is an estimate of the average rate of information R produced by the collection of users.

If we assume that the arrivals of the sampled messages are realizations of a homogeneous Poisson process then the interarrival times $t_{i+1} - t_i$, $1 \leq i \leq N-1$, are independent realizations of a random variable Y with the exponential density

$$f(y) = \lambda e^{-\lambda y} \quad \text{for } y \geq 0 \tag{17}$$

where λ is the intensity parameter which we are interested in. This parameter can be estimated from the available data $t_1, t_2, \ldots, t_N$ by several different methods. The maximum likelihood method leads to the estimate

$$\hat{\lambda} = \frac{N-1}{t_N - t_1} \quad \text{[1/TU]}. \tag{18}$$

An alternative class of minimum disparity methods can be found in Menéndez et al [6].

We see that the maximum likelihood estimate (18) relies on the assumption that the differences $t_2 - t_1, \ldots, t_{N-1}$ are distributed by the density (17) to the degree that it ignores the data $t_2, \ldots, t_{N-1}$. This means that this estimate is very sensitive to the deviations of the true distribution densities from (17) (and thus to violations of the assumption that the messages arrive to the network as a homogeneous Poisson process). Alternatives to (18), which are much more robust with respect to violations of the above mentioned assumptions, are the minimum disparity estimators systematically studied in Menéndez et al [6].

An alternative to the statistical estimation of R and λ, based on the empirical knowledge contained in data $(\beta_1, t_1), \ldots, (\beta_N, t_N)$, is an expert estimation based on a theoretical a priori collected knowledge. Expert estimates $\hat{R}$ and $\hat{\lambda}$ can be tested with the help of empirical data by using the disparity or entropy tests studied in Menéndez et al [7], Darbellay and Vajda [1] and Esteban et al [2], or by the special tests proposed by Menéndez et al [5], Morales et al [8] or Pardo et al [11].

For any estimates $\hat{\lambda}$ and $\hat{R}$, an estimate $\hat{\beta}$ of the average message size β follows from the formula $\beta\lambda = R$ of Section 1, namely

$$\hat{\beta} = \frac{\hat{R}}{\hat{\lambda}} \quad \text{[bit]}.$$

For example, for the statistical estimates $\hat{R}$ and $\hat{\lambda}$ given by (16) and (18) we get the intuitively appealing sample mean

$$\hat{\beta} = \frac{1}{N-1} \sum_{i=1}^{N-1} \beta_i \quad \text{[bit]}. \tag{19}$$

If we use a robust estimator $\hat{\lambda}$ which differs from (18) then we obtain in this manner a robust formula for $\hat{\beta}$, which will be different from the sample mean (19).

5.2. Inference about Λ

In this subsection we assume that the parameters β and λ are already specified and the problem which remains is specification of $\Lambda = (\lambda_{j,k})_{j,k=0}^{m}$ with $\lambda_{j,j} = 0$ for $0 \le j \le m$. To simplify the notation, we drop from double subscripts the comma, e. g. we put

$$\lambda_{jk} = \lambda_{j,k}.$$

The intensities λ_{jk} cannot be statistically estimated as easily as the overall input intensity λ in 5.1. From this, and also from the interpretational point of view, it is convenient to decompose these intensities $0 \le j,\ k \le m$ by the formula

$$\lambda_{jk} = \mu_j p_{jk}, \tag{20}$$

where μ_j is an average intensity of service in the server S_j and p_{jk} is a probability of routing a message from S_j to S_k and where we put $\mu_0 = \lambda$ (cf. below). The separate estimation of the intensities μ_j, $0 \le j \le m$, and of the stochastic matrices of routing probabilities

$$P = (p_{ij})_{i,j=0}^{m} \tag{21}$$

looks more hopefully than the direct estimation of the matrix Λ.

It follows from (9) and (20) that

$$\mu_0 = \sum_{k=1}^{m} \lambda_{0k} = \lambda.$$

For the remaining μ_j, $1 \le j \le m$, we get from the definition that

$$\mu_j = \sum_{k=0}^{m} \mu_{jk}\, p_{jk}, \tag{22}$$

where $\mu_{jk} > 0$ is the intensity with which S_j serves the messages routed to S_k. We can put for every $j \ne k$

$$\mu_{jk} = \left(\tau_{jk} + \frac{\beta}{\alpha_{jk} B_{jk} R_{jk}}\right)^{-1} \quad [1/\mathrm{TU}], \tag{23}$$

where β was introduced above and

- τ_{jk} [TU] is an average time needed to prepare a message for transmission and to activate the transmission link from S_j to S_k,
- $\alpha_{jk} \in (0,1)$ is coefficient of effectivity of the link from S_j to S_k (ratio of the effective transmission rate and the transmission rate); depends on the data link protocol,
- R_{jk} [bit/TU] is the transmission rate of the link from S_j to S_k.
- B_{jk} [bit/TU] is the number of links from S_j to S_k.

The data transmission rates R_{jk} are standard technical parameters of data links, and the coefficients of effectivity α_{jk} are standard technical parameters of data link protocols. The time delays τ_{jk} can be estimated by using an expert knowledge, or by using randomly sampled empirical data and employing the statistical estimators cited in Subsection 5.1.

The only open problem which remains is estimation of the routing probability matrices P considered in (21) which are needed in (20) as well as in (22). This problem is studied separately in the next subsection.

5.3. Inference about P

From (20) and the assumption that the diagonal elements of Λ are zero we see that the diagonal elements of $P = (p_{jk})_{j,k=0}^{m}$ are zero. In this subsection we study estimates

$$\hat{P} = (\hat{p}_{jk})_{j,k=0}$$

of the matrix P under assumption that there are at the disposal estimates $(\hat{f}_1, \dots, \hat{f}_m)$ and $(\hat{f}^1, \dots, \hat{f}^m)$ with the following interpretation:

$\hat{f}_j$ is the probability that a message enters the network through the server S_j

$\hat{f}^j$ is the probability that a message leaves the network from the server S_j

for every $j \in K_0$. One can use expert estimates or relative frequency estimates based on the empirical data. (If the MHN is not yet realized then only the first option is applicable.) In this subsection we skip the symbol ˆ in all estimates, i. e. we denote all estimates simply by f_j, f^j and p_{jk}.

We propose three different heuristic methods for estimation of the matrix P based on the evidence $(f_1, \dots, f_m)$ and $(f^1, \dots, f^m)$. The attention will be restricted to the hierarchic networks of Section 4. We shall need the notation $n(j)$ and $P(j)$ introduced there, and also the disjoint decomposition (with + denoting the disjoint union of sets)

$$K_0 = \{1, \dots, m\} = L + B + \{r\} \quad \text{(cf. (15))}.$$

Note that $n(j)$ is undefined for $j = r$ and $P(j) = \emptyset$ for $j \in L$. By $\Pi(j)$ we denote for every $1 \le j \le m$ the set of indices of the server S_j and of all servers directly or indirectly subordinated to S_j. Formally $\Pi(j)$ can be defined by induction as follows

$$\Pi(j) = \begin{cases} \{j\} & \text{if } \ j \in L \\ \{j\} + \sum_{k \in P(j)} \Pi(k) & \text{if } \ j \in B \\ \{1, \dots, m\} & \text{if } \ j = r. \end{cases}$$

For the sake of brevity we define

$$f_A = \sum_{j \in A} f_j$$

for $A \subset K_0$, and analogously f^A.

Method A.
Put $p_{0k} = f_k$ for $k \in K_0$.
For every $1 \le j \le m$ put

$$p_{jk} = \begin{cases} f^j & \text{if } k = 0 \\ f^{\Pi(k)} & \text{if } k \in P(j) \\ 1 - f^{\Pi(j)} & \text{if } k = n(j) \\ 0 & \text{otherwise.} \end{cases}$$

Method B.
Put $p_{0k} = f_k$ for $k \in K_0$.
For every $1 \le j \le m,\ j \ne r$, put

$$p_{jk} = \begin{cases} \frac{f^j}{f_{\Pi(j)}+f^{\Pi(j)}} & \text{if } k = 0 \\ \frac{f^{\Pi(k)}}{f_{\Pi(j)}+f^{\Pi(j)}} & \text{if } k \in P(j) \\ \frac{f_{\Pi(j)}}{f_{\Pi(j)}+f^{\Pi(j)}} & \text{if } k = n(j) \\ 0 & \text{otherwise} \end{cases}$$

and

$$p_{rk} = \begin{cases} f^r & \text{if } k = 0 \\ f^{\Pi(k)} & \text{if } k \in P(r) \\ 0 & \text{otherwise.} \end{cases}$$

Both these methods are based on rational assumptions. But Method A does not take sufficiently into account where the messages entered the network. Method B takes into account where the message entered the network but assumes a well organized centralistic protocol under which all messages are passing through the root server S_r. Both types of assumptions are questionable and they represent extremes of some kind.

Certain compromise between these extremes might be the following combination of both methods. It uses probabilities π_j of a *local communication* in the subnetworks of servers $S_{\Pi(j)}$, $1 \le j \le m$. The local communication probabilities are defined by

$$\pi_j = \frac{F_{\Pi(j)}^{\Pi(j)}}{f_{\Pi(j)}}, \quad 1 \le j \le m,$$

where $F_{\Pi(j)}^{\Pi(j)}$ is the probability that a message enters the network through the set of servers $\Pi(j)$ and leaves the network from the same set of servers. Under our assumptions we have $\pi_r = 1$. For $j \in B$ we can say only that in typical situations $0 < \pi_j < 1$.

Method C. For every $0 \le j,\ k \le m$ put

$$p_{jk} = \pi_j \, p^A_{jk} + (1 - \pi_j) \, p^B_{jk},$$

where p^B_{jk} is the solution by the Method B and p^A_{jk} is the solution by the Method A for the subnetwork of servers $S_{\Pi(j)}$, i.e.

$$p^A_{jk} = \begin{cases} \frac{f^j}{f^{\Pi(j)}} & \text{if } k = 0 \\ \frac{f^{\Pi(k)}}{f^{\Pi(j)}} & \text{if } k \in P(j) \\ 0 & \text{otherwise.} \end{cases}$$

A disadvantage of the Method C is that it requires an additional inference about the vector $(\pi_j : j \in B)$. Expert estimates are in this situation probably difficult and statistical inference requires many random samplings.

Thus neither of these methods is completely satisfactory. The exact rigorous solution, however, can be obtained only under additional information.

Method D (general). Similarly as above, let F^k_j for every $j,\ k \in K_0$ denote the probability that a message enters the network through the server S_j and then leaves the network from the server S_k. Suppose that the probabilities

$$F = \{F^k_j\}_{j,k \in K_0}$$

are given. We have $\sum_{j,k \in K_0} F^k_j = 1$, and let us again write $F^D_A = \sum_{j \in A} \sum_{k \in D} F^k_j$.

Due to the tree structure of the network, the path of every message is linear and uniquely given. Thus, we may observe

$$\begin{aligned} p^D_{jk} &= \frac{\Pr(\text{message enters } S_j \text{ and then passes to } S_k)}{\Pr(\text{message enters } S_j)} \\ &= \begin{cases} F^{K_0}_k & \text{for } j = 0,\ k \in K_0, \\ \frac{1}{\gamma(j)} F^j_{K_0} & \text{for } j \in K_0,\ k = 0, \\ \frac{1}{\gamma(j)} F^{\Pi(j)^c}_{\Pi(j)} & \text{for } j \in K_0,\ k = n(j), \\ \frac{1}{\gamma(j)} F^{\Pi(j)}_{\Pi(k)^c} & \text{for } j \in K_0,\ k \in P(j), \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where $\gamma(j) = F^j_{K_0} + F^{\Pi(j)^c}_{\Pi(j)} + \sum_{k \in P(j)} F^{\Pi(k)}_{\Pi(k)^c}$ is for every $j \in K_0$ an appropriate normalizing constant.

Method D' (with independence). Providing we have only the "entrance" and "exit" probabilities $(f_1, \ldots, f_m)$ and $(f^1, \ldots, f^m)$ as above, we may add the independence assumption in order to obtain the probabilities F, namely

$$F^k_j = f_j \cdot f^k.$$

Then the above formulas simplify, and we have

$$p_{jk}^{D'} = \begin{cases} f_k & \text{for } j=0,\ k \in K_0, \\ \frac{1}{\gamma(j)}\, f^j & \text{for } j \in K_0,\ k=0, \\ \frac{1}{\gamma(j)}\, f_{\Pi(j)}(1-f^{\Pi(j)}) & \text{for } j \in K_0,\ k=n(j), \\ \frac{1}{\gamma(j)}(1-f_{\Pi(k)})\, f^{\Pi(k)} & \text{for } j \in K_0,\ k \in P(j), \\ 0 & \text{otherwise,} \end{cases}$$

where now $\gamma(j) = f^j + f_{\Pi(j)}(1-f^{\Pi(j)}) + \sum_{k\in P(j)}(1-f_{\Pi(k)}) \cdot f^{\Pi(k)}$.

The estimate $P^{D'}$, in spite of being derived from the exact solution P^D, cannot be considered universally better than those obtained under A, B, and C. The independence assumption is rather strong and can be easily violated in practical situations. It depends on the number and type of users connected to the particular servers. (E.g., one should expect $F_j^j = f_j \cdot f^j$ for messages entering and leaving the same server S_j.)

6. PROGNOSIS OF PERFORMANCE AND OPTIMIZATION

In this section we use solutions $w_1, \ldots, w_m$ of the equations (13). Before going further notice that if the intensities from the matrix Λ can be decomposed as assumed in (20) then these equations can be transformed into the form

$$\sum_{k=0}^{m} y_k\, p_{kj} = y_j, \quad 1 \le j \le m \quad \text{with} \quad y_0 = \mu_0 = \lambda. \tag{24}$$

The desired $w_1, \ldots, w_m$ are then obtained from solution $y_1, \ldots, y_m$ of this system by formula

$$w_j = \frac{y_j}{\mu_j} \quad \text{for } 1 \le j \le m.$$

From $w_1, \ldots, w_m$ and the basic network parameters β, $\lambda, \mu_1, \ldots, \mu_n$ and P can be done conclusions about the performance of the network and prognosis of values of many performance parameters. The most important conclusion is that if $w_j \ge 1$ for at least one j then the number n_j of messages in the server S_j will increase to infinity and the whole network will collapse. When the remaining basic network parameters remain then the initial condition y_0 in (24), and therefore also solutions $y_1, \ldots, y_m$ and $w_1, \ldots, w_m$, are increasing functions of λ. Therefore

$$C = \beta \sup \left\{ \lambda > 0 : \sup_{1 \le j \le m} w_j < 1 \right\} \quad [\text{bit/TU}] \tag{25}$$

is a capacity of the network. It is a sharp upper bound on the amount of information which can be transmitted by the network with a finite delay.

In the rest of this section we assume that $0 < w_n < 1$ for all $1 \le j \le m$ and the conclusions are valid in the steady state of the network, achieved for $t \to \infty$. Then, applying the above results and the well-known Little's formula (cf., e. g. Nelson [9], Section 7.1), we arrive at the following conclusions.

(i) Average number of messages in the server is

$$\nu_j = \frac{w_j}{1 - w_j}$$

and the average number of all messages in the network is

$$N = \sum_{j=1}^{m} \nu_j.$$

(ii) A message spends in the server S_j

$$\frac{\nu_j}{\lambda} \quad \text{time units (TU),}$$

namely

$$\frac{w_j \nu_j}{\lambda} \quad \text{TU}$$

by waiting in a queue and

$$\frac{w_j}{\lambda} \quad \text{TU}$$

by the processing and transmission.

(iii) The average number of messages passing through the server S_j per one TU is $\mu_j w_j$, of them $\mu_j w_j p_{jk}$ are routed into the server S_k.

(iv) Average delay of a message in the network is

$$\frac{N}{\lambda} \quad \text{TU.}$$

Of this time

$$\frac{1}{\lambda} \sum_{j=1}^{m} w_j \nu_j \quad \text{TU}$$

is spent by waiting in the server queues and

$$\frac{1}{\lambda} \sum_{j=1}^{m} w_j \quad \text{TU}$$

by processing in the servers and transmission in the communication links.

Prognosis of the network capacity by means of (25), and of the performance parameters by means of the formulas in (i) – (ii), enables to detect eventual problems in existing MHN's or in their projects. By combining the prognosed parameters with cost functions and network management rules, one can optimize projects of planned MHN's, or innovations of existing MHN's, like admission of new users or modernization of hardware and software.

7. PROGRAM 5P AND AN ILLUSTRATIVE EXAMPLE

We prepared a Program for Prognosis of Performance Parameters and Problems (5P). Its basic unit is a subprogram for evaluation of solutions of Section 4. Inputs and outputs are provisional and simple, in order to replace them easily when 5P will be incorporated into professional program products of potential users. The input data structure describing servers $S_1, \ldots, S_m$ is proposed so that it is independent on m and on the structure of links between the servers. Therefore 5P imposes no a priori limitations on the network size m, and the only limitation is the computing time. Solutions for $m \approx 10^3$ are very fast on most modern PC's. Input data of the subprogram concerning each server S_j are as follows.

— Index $n(j)$ of the superordinated server $S_{n(j)}$ (not for $n = r$).

— Transmission rate $R_{jn(j)}$ [bit/sec] of one link to $S_{n(j)}$ (not for $j = r$).

— Number of links to $S_{n(j)}$ (not for $j = r$).

— Time $t_{jn(j)}$ [sec] for activation of the links to $S_{n(j)}$.

— Probabilities f_j and f^j.

We assume the symmetry of the transmission rates, $R_{jk} = R_{kj}$ for $1 \leq j,\ k \leq m$, and $R_{j0} = \infty$ for all $1 \leq j \leq m$.
Input data of the subprogram concerning the network are as follows.

— Coefficient of effectivity of links α (assumed to be the same for all links).

— Average message size β [bit].

— Input intensity λ [1/hour].

Output data of the subprogram concerning each server S_j are as follows.

— Intensity of service μ_j [1/hour].

— Intensities $\lambda_{n(j)\,j}$ and $\lambda_{jn(j)}$ [1/hour] (not for $j = r$).

— Intensities λ_{j0} and λ_{0j} [1/hour].

— Solution w_j.

The program is implemented in Borland Pascal for Windows, version 7. It allows graphical realizations of all performance parameters (except the MHN capacity) as functions of variable input intensity λ. More details about it can be found in Janžura and Boček [4].

Next follow an example analyzing a hypermarket MHN physically covering the Czech Republic and consisting of 34 servers. The input data for the subprogram are in Table 1, where the root server $S_r = S_1$ and the number of links between all servers is 1. The network data were as follows

$$\alpha = 0.8333, \quad \beta = 30\,000 \text{ [bit]}, \quad \lambda = 100 \text{ [1/hour]}$$

and the Method A of Section 5.3 was used to evaluate the routing probabilities p_{jk} for $j \neq 0$. Output data of the subprogram and some of the output data of the program are in Table 2.

Table 1. Input subprogram data for servers $S_1, \ldots, S_{34}$.

S_j	$S_{n(j)}$	$R_{jn(j)}$	$t_{jn(j)}$	f_j	f^j
S_1	–	–	–	0.61181	0.39898
S_2	S_1	4800	6	0.001	0.001
S_3	S_1	4800	6	0.001	0.013
S_4	S_1	4800	6	0.001	0.013
S_5	S_1	64000	4	0.03986	0.11997
S_6	S_5	4800	6	0.001	0.013
S_7	S_5	4800	6	0.001	0.001
S_8	S_5	9600	6	0.001	0.001
S_9	S_5	9600	6	0.001	0.001
S_{10}	S_5	56000	6	0.001	0.001
S_{11}	S_5	56000	6	0.001	0.001
S_{12}	S_1	256000	2	0.14791	0.11999
S_{13}	S_{12}	800	6	0.001	0.001
S_{14}	S_{12}	4800	6	0.001	0.001
S_{15}	S_{12}	4800	6	0.001	0.013
S_{16}	S_{12}	4800	6	0.001	0.001
S_{17}	S_{12}	19200	6	0.001	0.001
S_{18}	S_{12}	19200	6	0.001	0.013
S_{19}	S_1	256000	2	0.00299	0.005
S_{20}	S_{19}	64000	6	0.00299	0.005
S_{21}	S_{20}	4800	6	0.001	0.001
S_{22}	S_{20}	4800	6	0.001	0.001
S_{23}	S_{20}	4800	6	0.001	0.001
S_{24}	S_{20}	4800	6	0.001	0.001
S_{25}	S_{20}	4800	6	0.001	0.001
S_{26}	S_{20}	48000	6	0.001	0.001
S_{27}	S_{20}	4800	6	0.001	0.001
S_{28}	S_{20}	48000	6	0.001	0.001
S_{29}	S_{20}	48000	6	0.001	0.013
S_{30}	S_{20}	48000	6	0.001	0.0131
S_{31}	S_{20}	9600	6	0.001	0.001
S_{32}	S_{20}	9600	6	0.001	0.001
S_{33}	S_{20}	100000000	1	0.08372	0.11998
S_{34}	S_{19}	100000000	1	0.08372	0.11998

Table 2. Output subprogram data and some performance characteristics.

S_j	μ_j	$\lambda_{n(j)j}$	$\lambda_{jn(j)}$	λ_{0j}	λ_{j0}	w_j	ν_j	ν_j/λ	$\mu_j w_j$
S_1	1302.38754	0	0	39.898	796.81372	0.10709	0.1199337	0.001199	139.4726817
S_2	172.45665	0.17246	172.28419	0.1	0.17246	0.00127	0.001271615	0.000013	0.219019946
S_3	172.45665	0.17246	172.28419	1.3	0.17246	0.01517	0.015403674	0.000154	2.616167381
S_4	172.45665	0.17246	172.28419	1.3	0.17246	0.01517	0.015403674	0.000154	2.616167381
S_5	719.92081	33.01557	686.90525	11.997	28.69604	0.04181	0.043634352	0.000436	30.09988907
S_6	172.45665	0.17246	172.28419	1.3	0.17246	0.0151	0.015331506	0.000153	2.604095415
S_7	172.45665	0.17246	172.28419	0.1	0.17246	0.0012	0.001201442	0.000012	0.20694798
S_8	267.5994	0.2676	267.3318	0.1	0.2676	0.00079	0.000790625	0.000008	0.211403526
S_9	267.5994	0.2676	267.3318	0.1	0.2676	0.00079	0.000790625	0.000008	0.211403526
S_{10}	494.82318	0.49482	494.32835	0.1	0.49482	0.00045	0.000450203	0.000005	0.222670431
S_{11}	494.82318	0.49482	494.32835	0.1	0.49482	0.00045	0.000450203	0.000005	0.222670431
S_{12}	1513.98044	233.01673	1280.96371	11.999	223.93285	0.03178	0.032823119	0.000328	48.11429838
S_{13}	38.66235	0.03866	38.62369	0.1	0.03866	0.0052	0.005227181	0.000052	0.20104422
S_{14}	172.45665	0.17246	172.28419	0.1	0.17246	0.00119	0.001191418	0.000012	0.205223414
S_{15}	172.45665	0.17246	172.28419	1.3	0.17246	0.01509	0.015321197	0.000153	2.602370849
S_{16}	172.45665	0.17246	172.28419	0.1	0.17246	0.00119	0.001191418	0.000012	0.205223414
S_{17}	370.06097	0.37006	369.69091	0.1	0.37006	0.00057	0.000570325	0.000006	0.210934753
S_{18}	370.06097	0.37006	369.69091	1.3	0.37006	0.00705	0.007100055	0.000071	2.608929839
S_{19}	1638.17352	303.75013	1334.42339	0.5	4.89814	0.05775	0.061289467	0.000613	94.60452078
S_{20}	761.69571	75.18698	686.50873	0.5	2.27747	0.04469	0.046780626	0.000468	34.04018128
S_{21}	172.45665	0.17246	172.28419	0.1	0.17246	0.0012	0.001201442	0.000012	0.20694798
S_{22}	172.45665	0.17246	172.28419	0.1	0.17246	0.0012	0.001201442	0.000012	0.20694798
S_{23}	172.45665	0.17246	172.28419	0.1	0.17246	0.0012	0.001201442	0.000012	0.20694798
S_{24}	172.45665	0.17246	172.28419	0.1	0.17246	0.0012	0.001201442	0.000012	0.20694798
S_{25}	172.45665	0.17246	172.28419	0.1	0.17246	0.0012	0.001201442	0.000012	0.20694798
S_{26}	480.71961	0.48072	480.23889	0.1	0.48072	0.00046	0.000460212	0.000005	0.221131021
S_{27}	172.45665	0.17246	172.28419	0.1	0.17246	0.0012	0.001201442	0.000012	0.20694798
S_{28}	480.71961	0.48072	480.23889	0.1	0.48072	0.00046	0.000460212	0.000005	0.221131021
S_{29}	480.71961	0.48072	480.23889	1.3	0.48072	0.00545	0.005479865	0.000055	2.619921875
S_{30}	480.71961	0.48072	480.23889	1.31	0.48072	0.00549	0.005520306	0.000055	2.639150659
S_{31}	267.5994	0.2676	267.3318	0.1	0.2676	0.00079	0.000790625	0.000008	0.211403526
S_{32}	267.5994	0.2676	267.3318	0.1	0.2676	0.00079	0.000790625	0.000008	0.211403526
S_{33}	3375.4913	282.59613	3092.89517	11.998	282.59613	0.01001	0.010111213	0.000101	33.78866791
S_{34}	3375.4913	282.59613	3092.89517	11.998	282.59613	0.01102	0.011142794	0.000111	37.19791413

Moreover, the program calculated the average delay of messages in the network

$$\frac{1}{\lambda}\sum_{i=1}^{34}\nu_i = 0.04281 \text{ hours } \doteq \text{ 2 minutes and and 34 seconds,}$$

of which time the messages spend in average

$$\frac{1}{\lambda}\sum_{i=1}^{34} w_j\,\nu_j = 0.00226 \text{ hours } \doteq \text{8 seconds}$$

in the server queues (buffers) and

$$\frac{1}{\lambda}\sum_{i=1}^{34} w_j = 0.04055 \text{ hours } \doteq \text{ 2 minutes and 26 seconds}$$

in the servers and transmission lines. Further, the program calculated that for

$$\lambda_{\max} = 933 \text{ messages per hour}$$

the maximal solution $w_{\max}$ of equations (13) achieves 1. Therefore the capacity of the network is

$$C = \beta \cdot \lambda_{\max} = 30\,000 \times 933 = 27990000 \text{ [bit/hour]} \doteq \text{ 2.8 Mbits per hour.}$$

If the message flow into the network approaches the critical value 2.8 Mbits/hour then the congestion of the network is unavoidable.

(Received January 26, 2001.)

REFERENCES

[1] G. A. Darbellay and I. Vajda: Entropy expressions for multivariate continuous distributions. IEEE Trans. Inform. Theory *46* (2000), 709–712.

[2] M. Esteban, M. Castellanos, D. Morales, and I. Vajda: A comparative study of the normality tests based on sample entropies. Comm. Statist. Simulation Comput., to appear.

[3] G. N. Higginbottom: Performance Evaluation of Communication Networks. Artech House, Boston 1998.

[4] M. Janžura and P. Boček: Stochastic Methods of Prognosis of Parameters in Data Networks (in Czech). Research Report No. 1981, Institute of Information Theory and Automation, Prague 2000.

[5] M. Menéndez, D. Morales, L. Pardo, and I. Vajda: Inference about stationary distributions of Markov chains based on divergences with observed frequencies. Kybernetika *35* (1999), 265–280.

[6] M. Menéndez, D. Morales, L. Pardo, and I. Vajda: Minimum disparity estimators for discrete and continuous models. Appl. Math. *46* (2001), 401–420.

[7] M. Menéndez, D. Morales, L. Pardo, and I. Vajda: Approximations to powers of ϕ-disparity goodness of fit tests. Comm. Statist. Theory Methods *30* (2001), 105–134.

[8] D. Morales, L. Pardo, M. C. Pardo, and I. Vajda: Extension of the Wald statistics to models with dependent observations. Metrika *52* (2000), 97–113.

[9] R. Nelson: Probability, Stochastic Processes, and Queueing Theory. Springer, New York 1995.
[10] J. R. Norris: Markov Chains. Cambridge University Press, Cambridge 1997.
[11] M. C. Pardo, L. Pardo, and I. Vajda: Consistent tests of homogeneity for independent samples from arbitrary models, submitted.
[12] A. Pattavina: Switching Theory: Architecture and Performance in Broadbard ATM Networks. Wiley, New York 1998.
[13] N. M. van Dijk: Queueing Networks and Product Forms. A System Approach. Wiley, New York 1993.
[14] J. Walrand: Introduction to Queueing Networks. Prentice–Hall, Englewood Cliffs, N.J. 1988.
[15] P. Whittle: Systems in Stochastic Equilibrium. Wiley, Chichester 1986.

Mgr. Pavel Boček, RNDr. Martin Janžura, CSc. and Ing. Igor Vajda, DrSc. Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mails: boček,janžura,vajda@utia.cas.cz

Ing. Tomáš Feglar, CSc., I2S a. s., Sněmovní 9, 118 00 Praha 1. Czech Republic.
e-mail: feglar@i2s.cz