Biuletyn WAT
Vol. LXIII, Nr 2, 2014

A Simulation Method For Effectiveness Evaluation of Distributed Systems

ZBIGNIEW WESOŁOWSKI

Military University of Technology, Faculty of Cybernetics,
Institute of Computer and Information Systems,
2 Kaliskiego Str., 00-908 Warsaw, Poland, zwesolowski@wat.edu.pl

Abstract. This paper presents a simulation method for evaluating the effectiveness of distributed systems based on passing messages. For mathematical modelling of these systems, two paradigms were used, i.e. the paradigm of experimental modelling and the paradigm of probabilistic modelling. The paper proposes effectiveness measures of the considered family of systems. An algorithm of stochastic simulation for effectiveness evaluation of distributed systems was presented.
Keywords: distributed system, stochastic simulation, effectiveness analysis of systems

1. Definitions of Fundamental Terms

1. A system is a separate part of the real world, whose properties and phenomena occurring in it are under investigation.
2. A structure of the system is a set of elements and interactions between them conditioned by their belonging to this system.
3. Behaviour of the system is the range of actions made by this system in conjunction with itself or its surroundings, which includes the other systems as well as physical environment. It is the response of the system to various stimuli or inputs, whether internal or external ones.
4. A computer network is a collection of computers, routers, links and any intermediate nodes which are connected so, as to enable telecommunication between the terminals.

5. A distributed system is a software system consisting of many components located on networked computers to ensure their mutual communication and coordination of actions by passing messages.
6. Computer simulation is a method of concluding about behaviour of systems based on data generating by computer programs simulating this behaviour [14, 15, 16, 45].

Let us enter markings:
$\mathbb{N}$ — the set of natural numbers,
$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathbb{R}$ — the set of real numbers,
$\mathbb{R}_+ = (0,\infty) \subset \mathbb{R}$, $\mathbb{R}_0 = \mathbb{R}_+ \cup \{0\}$, $\mathbb{R}_{[0,1]} = [0,1] \subset \mathbb{R}$, $T = \mathbb{R}$ — the set of time parameters,
$W(\alpha, \beta)$ — the Weibull distribution with the parameters α ($\alpha > 0$) and β ($\beta > 0$).

2. Introduction

One of the challenges of modern science is the study and modelling of phenomena occurring in the physical structures composed of many interacting elements [32]. Results of previous works on physical systems (e.g. turbulent flows), biological systems (e.g. living organisms, ant colonies), social systems (e.g. transport systems, economic systems, social nets), or information systems (e.g. computer networks, parallel and distributed computing systems) allow us to conclude that no one understands occurrences and processes happening in them [3, 32]. The consequence of this fact is, inter alia, that have not yet developed reliable mathematical models of these phenomena. This means that should be considered to be open, inter alia, problems such as: the problem of prediction, the problem of effectiveness evaluation, the problem of control systems with sufficiently high complexity [3].

The subject of the work is the issue of evaluation of the effectiveness of distributed systems. For solution of this problem, both an analytical and a probabilistic approach are used.

Analytical methods for effectiveness evaluation of distributed systems and related methods [5, 21, 23, 44, 48] mainly use a queuing theory [19, 42] and a graph theory [11, 47]. These methods have limited practical application, because they require, among other things, making an assumption that arrivals are determined by a Poisson process and job service times have the exponential distribution.

Probabilistic methods for effectiveness evaluation of considered family of systems and related methods [7, 8, 9, 12, 20, 25, 29, 33, 38] employ diverse branches of mathematics and computer science, including: Markov processes [41, 42], Petri nets [37], statistical estimation [6, 27, 35], randomized algorithms [30, 31], and computer simulation [14, 15, 16, 22, 45].

Using the methodology of probabilistic modelling and computer simulation gives hope to develop a reliable method for solving the problem under consideration.

3. Mathematical Model of the Distributed System

Mathematical modelling is one of the most widely used methods of survey systems [24, 40]. Constructing mathematical models and using instrumentation of mathematics for analysis of their behaving are nature of this method.

A mathematical model of a distributed system determines an ordered pair,

$$D = (N, U), \tag{1}$$

where: N is the model of the structure, and U is the model of the dynamics.

The model N, Eq. (1), determines a network,

$$N = (V, E, \mathbf{c}), \tag{2}$$

where: $V = \{H, R\} = \{v_1, v_2, \ldots, v_{m_V}\}$ is the set of elements, $H = \{h_1, h_2, \ldots, h_{m_H}\}$ is the set of computers, $R = \{r_1, r_2, \ldots, r_{m_R}\}$ is the set of routers, where $m_V = m_H + m_R$;
$E = \{e_i = (v, w)_i = (v_i, w_i) \in V \times V : v_i \neq w_i, i = 1, \ldots, m_E\} \subset V \times V$ is the set of unidirectional network connections (or network links) among pairs of different elements of the distributed system structure;
$\mathbf{c} = [c_1, c_2, \ldots, c_{m_E}]^T \in \mathbb{R}_+^{m_E \times 1}$ is the vector of permissible bandwidths of all network links.

Assumption 1. The directed graph (V, E) is connected.

□

Let $L = \{1, \ldots, m_E\}$ be the set of numbers of network links from the set E.
Assumption 1 implies that all pairs of different elements of the network N, Eq. (2), are interfacing with paths. Let,

$$P = \{p_1, p_2, \ldots, p_{m_P}\}, \tag{3}$$

be the set of acyclic paths connecting all pairs of different elements from the set V, where: $p_j = p_{(v,w)_j} = p_{(v_j, w_j)} = (v_j, r_j^{(1)}, r_j^{(2)}, \ldots, r_j^{(n_j)}, w_j)$ is the path linking a pair of elements $(v_j, w_j) \in V \times V$, $v_j \neq w_j$, which contains the nodes $v_j, r_j^{(1)}, r_j^{(2)}, \ldots, r_j^{(n_j)}, w_j$ and the edges $(v_j, r_j^{(1)})$, $(r_j^{(1)}, r_j^{(2)})$, …, $(r_j^{(n_j)}, w_j)$ joining these nodes, where: $r_j^{(l)} \in V$, for: $l = 1, \ldots, n_j$, $j = 1, \ldots, m_P$, $m_P \leq m_{max}$, $m_{max} = \sum_{i=2}^{m_V} \frac{m_V!}{(m_V - i)!}$ is the maximum number of acyclic paths in a complete directed graph composed of m_v vertices. Let $Q = \{1, ..., m_P\}$ be the set of numbers of paths from the set P.

Rules of the data transmitting in the network N, Eq. (2), determines a binary routing matrix,

$$\mathbf{A} = [a_{ij}]_{m_E \times m_P}, \tag{4}$$

where:

$$a_{ij} = \begin{cases} 1, & \text{if } e_i \in p_j, \\ 0, & \text{if } e_i \notin p_j, \end{cases}$$

where: $e_i \in E$, for $I \in L$, and $p_j \in P$, for $j \in Q$.

How it is possible to notice, the elements a_{ij} of the matrix $\mathbf{A}$ assume the value one if the link $e_i \in E$ belongs to the path $p_j \in P$, and the value zero otherwise.

Behaving of the distributed system D, Eq. (1), is associated with passing messages between pairs of different computers $(\alpha, \omega) \in H \times H$, $\alpha \neq \omega$, along paths connecting these computers, where: α is called the sender or source and ω is called the receiver or sink.

Assumption 2. The passing messages between any pair of different computers from the set H are carried out along the shortest acyclic path connecting these computers. There is only one such a path.

Justification. This assumption does not reduce the generality of deliberations, because with substantial issues of effectiveness evaluation of distributed systems is the one associated with sending many messages through shared network links.

□

Let,

$$S = \{s \in Q : p_s = p_{(\alpha_s, \omega_s)} \in P \wedge (\alpha_s, \omega_s) \in H \times H, \alpha_s \neq \omega_s\}, \tag{5}$$

be the set of numbers of the shortest acyclic paths connecting all pairs of different computers from the set H, which has been designated by one of the static network routing algorithms [28], where: $|S| = m_S$, and $m_S = m_H(m_H - 1)$. Let,

$$L_s = \{l \in L : a_{ls} = 1\}, \quad s \in S, \tag{6}$$

be the set of numbers of links belonging to the path of the number $s \in S$.
Let,

$$S_l = \{s \in S : a_{ls} = 1\}, \quad l \in L, \tag{7}$$

be the set of numbers of the shortest paths passing through the link $l \in L$.
Let $\Theta = \{\theta_s : s \in S\}$ be the set of messages, where θ_s is the message of the length $m_{\theta_s} \in \mathbb{N}$ [b]. Let,

$$X = \{x_1, x_2, \ldots, x_{m_S}\}, \tag{8}$$

be the distributed computational task executed in an environment of the distributed system. The task X is the compound of many processes x_s, $s \in S$. The process x_s is located on the computer $\alpha_s \in H$ and consists in passing the message θ_s to the computer $\omega_s \in H$, $\alpha_s \neq \omega_s$, along the path $p_s \in P$.
As it can be seen, L_s, Eq. (6), can be interpreted as the set of numbers of network links from the set E used by the process $x_s \in X$, Eq. (8), while S_l, Eq. (7) can be interpreted as the set of numbers of processes from the set X that uses the link $l \in L$.
A model of dynamics of the message θ_s, passing along the link $l \in L_s$, determines an ordered triple,

$$U_{sl} = (T, \mathbb{R}_+, \mathrm{f}_{sl}), \quad l \in L_s, s \in S, \tag{9}$$

where: $\mathrm{f}_{sl} : T \to \mathbb{R}_+$ is the map determining the bit rate of the message θ_s passing along the link $l \in L_s$.
A model of dynamics of the message θ_s passing along the path $s \in S$ determines an ordered triple,

$$U_s = (T, \mathbb{R}_+, \mathrm{f}_s), \quad s \in S, \tag{10}$$

where: $\mathrm{f}_s : T \to \mathbb{R}_+$ is the map determining the bit rate of the message θ_s passing along the path $s \in S$.
The model U, Eq. (1), determines an ordered triple,

$$U = (T, \mathbb{R}_+^{m_S}, \mathbf{f}),$$

where: $\mathbf{f} : T \to \mathbb{R}_+^{m_S}$ is the map determining the bit rate of passing messages from the set Θ in an environment of the distributed system D, Eq. (1).

4. Experiments

Let us consider the experiment $Y = \{Y_s : s \in S\}$ compound of many attempts Y_s, $s \in S$. The attempt $Y_s = \{Y_{sl} : l \in L_s\}$ is compound of many tests Y_{sl}, $l \in L_s$. Each test Y_{sl} consists in observation of the values $y_{sl}(m) \in \mathbb{R}_+$ of the bit rate of the message θ_s passing along the link $l \in L_s$ observed in the time interval $[t_m, t_{m+1}) \subset \mathbb{R}$, where:

$t_m = m\Delta t$, for $m = 0, 1, ..., m_Y - 1$, $t_0 = 0$ is the moment of beginning the experiment, $\Delta t > 0$ is the length of the time interval between successive observations, $m_Y \in \mathbb{N}$ is the number of observations. Let $M = \{0, 1, ..., m_Y - 1\}$ be the set of numbers of tests.

As it can be seen, the observations $y_{sl}(m)$, $m \in M$, are interpreted as values of the map f_{sl}, Eq. (9).

The vector,

$$\mathbf{y}_{sl} = [y_{sl}(0), y_{sl}(1), ..., y_{sl}(m_Y - 1)]^T, \quad l \in L_s, s \in S, \tag{11}$$

is called a result of the test Y_{sl} or a sample.

Let us assume that the observed value $y_s(m) \in \mathbb{R}_+$ of the bit rate of the message θ_s, passing along the path $s \in S$, satisfies a relationship of the form,

$$y_s(m) = \min_{l \in L_s}\{y_{sl}(m)\}, \quad m \in M, s \in S.$$

As it can be seen, the observations $y_s(m)$, $m \in M$, are interpreted as values of the map f_s, Eq. (10).

The vector,

$$\mathbf{y}_s = [y_s(0), y_s(1), ..., y_s(m_Y - 1)]^T, \quad s \in S,$$

is called a result of the attempt Y_s or a sample.

Let us establish, that for modelling of the distributed system behaviour, based on the samples $\mathbf{y}_{sl}$, $l \in L_s$, and $\mathbf{y}_s$, $s \in S$, will be used the probabilistic modelling paradigm [30, 34]. This paradigm is derived from the concept of uncertainty, according to which there is no general common law or accuracy of determining the development of physical phenomena. This approach is at the heart of modern theory of identification [24].

Let $(\Omega, \Xi, \mathrm{P})$ be the probabilistic space, where: $\Omega = \{\omega_m : y(m) \in \mathbb{R}_+, m \in \mathbb{N}_0\}$ is the set of elementary events, ω_m is the elementary event meaning that the observed quantity assumes the value $y(m) \in \mathbb{R}_+$; Ξ is the set of random events being a σ-field of all subsets of the set Ω; $\mathrm{P} : \Xi \to \mathbb{R}_0$ is the probabilistic measure.

The random variable,

$$\psi_{sl} : \Omega \to \mathbb{R}_+, \quad l \in L_s, s \in S, \tag{12}$$

defined on the probability space $(\Omega, \Xi, \mathrm{P})$, is called a stochastic model of the test Y_{sl}. Observations that are components of the sample $\mathbf{y}_{sl}$ are interpreted as realizations of the random variable ψ_{sl}, i.e. $\psi_{sl_{\omega_m}} = y_{sl}(m)$, for $m \in M$.

Assumption 3. The random variable ψ_{sl}, Eq. (12), has the distribution $W(\alpha_{sl}, \beta_{sl})$, for: $l \in L_s$, $s \in S$.
Justification. The Weibull distribution [13, 26] is one of the most important probability distributions used, inter alia, in system identification and in reliability theory. It is commonly used as a stochastic model of observations with positively concentrated values. Its special cases are exponential and Rayleigh distributions.

□

The random variable,

$$\psi_s : \Omega \rightarrow \mathbb{R}_+, \quad s \in S, \tag{13}$$

defined on the probability space $(\Omega, \Xi, \mathrm{P})$, is called a stochastic model of the attempt Y_s. Observations that are the components of the sample $\mathbf{y}_s$ are interpreted as realizations of the random variable ψ_s, i.e. $\psi_{s_{\omega_m}} = y_s(m)$, for $m \in M$.

Assumption 4. The random variable ψ_s, Eq. (13), has the distribution $W(\alpha_s, \beta_s)$, for $s \in S$.
Justification. Reasoning is analogous to justify assumption 3.

□

5. Stochastic Simulation of the Distributed System Behaviour

Simulation of behaviour of the system D, Eq. (1), consists in generating realizations y_{sl} of the random variables ψ_{sl}, Eq. (12), this way so that there are fulfilled inequalities: $y_{sl} > 0$ and $\sum_{s \in S_l} y_{sl} \le c_l$, for $l \in L_s$, and then on computation of the realization y_s of the random variable ψ_s, Eq. (13), from the formula $y_s = \min_{l \in L_s}\{y_{ls}\}$, for $s \in S$. For established $y_s \in \mathbb{R}_+$ a run time of the process $x_s \in X$, Eq. (8), is evaluated from a relationship of the form $z_s = m_{\theta_s} / y_s$ [s]. So, a run time of the task X is evaluated from a relationship of the form $z = \max_{s \in S}\{z_s\}$.

Let us consider the simulation experiment G consisting in repeated observation of the run times $z(i) \in \mathbb{R}_+$, $i = 0, 1, ..., m_G - 1$, of the task X. Let $I = \{0, 1, ..., m_G - 1\}$ be the set of numbers of runs of the experiment G. The number of simulation runs $m_G \in \mathbb{N}$ can be determined using the methods based on the Chebyshev's inequality [26, 30].

The vector $\mathbf{z} = [z(0), z(1), ..., z(m_G - 1)]^T$ is called a result of the simulation experiment G or sample.
The random variable,

$$\zeta : \Omega \rightarrow \mathbb{R}_+, \tag{14}$$

defined on the probability space (Ω, Ξ, P), is called a stochastic model of the simulation experiment *G*. Observations that are components of the sample **z** are interpreted as realizations of the random variable ζ, i.e. $\zeta_{\omega_i} = z(i)$, for $i \in I$.

Assumption 5. The random variable ζ, Eq. (14), has the distribution $W(\alpha, \beta)$.
Justification. Reasoning is analogous to justify assumption 3.

□

In the attachment, an algorithm for generating the sample **z** was presented.

6. Effectiveness Analysis of the Distributed System

Effectiveness analysis of the system *D*, Eq. (1), relies on the estimation, based on the sample **z**, of the following measures:

— the expected value of the random variable ζ;
— the variance of the random variable ζ;
— the distribution function of the random variable ζ;
— the survival function of the random variable ζ;
— the empirical distribution function from the sample **z**;
— the empirical survival function from the sample **z**.

The expected value $\mu = \mathrm{E}(\zeta)$ of the random variable ζ determines the average length of the run time of the task *X*.
The variance $\sigma^2 = \mathrm{E}[(\zeta\text{-}\mu)^2]$ of the random variable ζ determines the mean square deviation of a length of the run time of the task *X* from its expected value.
The distribution function $\mathrm{F}(t) = \mathrm{P}(\zeta \le t)$ of the random variable ζ determines the probability of an event that executing the task *X* will end up at least to the moment *t*.
The survival function $\mathrm{R}(t) = 1 - \mathrm{F}(t) = \mathrm{P}(\zeta > t)$ of the random variable ζ determines the probability of an event that executing the task *X* will not end up to the moment *t*.
The empirical distribution function from the sample **z** is a nonparametric maximum likelihood estimator of the distribution function of the random variable ζ.
The empirical survival function from the sample **z** is a nonparametric maximum likelihood estimator of the survival function of the random variable ζ.

Let $\hat{\alpha}(\mathbf{z})$ and $\hat{\beta}(\mathbf{z})$ are the evaluations of the parameters of the distribution $W(\alpha, \beta)$ calculated by maximum likelihood [6, 18, 26, 27] based on the sample **z**. Estimators of the aforementioned effectiveness measures take the following forms:

— the estimator $\hat{\mu}: \mathbb{R}_+^{m_G} \to \mathbb{R}_+$ of the expected value of the random variable ζ has the form $\hat{\mu}(\mathbf{z}) = \hat{\beta}(\mathbf{z})\Gamma\left(1 + \frac{1}{\hat{\alpha}(\mathbf{z})}\right)$, where Γ is the gamma function;

— the estimator $\hat{\sigma}^2 : \mathbb{R}_+^{m_G} \to \mathbb{R}_+$ of the variance of the random variable ζ has the form $\hat{\sigma}^2(\mathbf{z}) = \hat{\beta}^2(\mathbf{z})\left[\Gamma\left(1+\frac{2}{\hat{\alpha}(\mathbf{z})}\right) - \Gamma^2\left(1+\frac{1}{\hat{\alpha}(\mathbf{z})}\right)\right]$,

— the estimator $\hat{\mathrm{F}} : \mathbb{R} \times \mathbb{R}_+^{m_G} \to \mathbb{R}_{[0,1]}$ of the distribution function of the random variable ζ has the form $\hat{\mathrm{F}}(t,\mathbf{z}) = 1 - \exp\left[-\left(\frac{t}{\hat{\beta}(\mathbf{z})}\right)^{\hat{\alpha}(\mathbf{z})}\right] 1_{(0,\infty)}(t)$, where $1_{(0,\infty)} : T \to \{0, 1\}$ is the indicator function;

— the estimator $\hat{\mathrm{R}} : \mathbb{R} \times \mathbb{R}_+^{m_G} \to \mathbb{R}_{[0,1]}$ of the survival function of the random variable ζ has the form $\hat{\mathrm{R}}(t,\mathbf{z}) = 1 - \hat{\mathrm{F}}(t,\mathbf{z})$;

— the empirical distribution function $\hat{\mathrm{F}}^e : \mathbb{R} \times \mathbb{R}_+^{m_G} \to \mathbb{R}_{[0,1]}$ from the sample $\mathbf{z}$ has the form $\hat{\mathrm{F}}^e(t,\mathbf{z}) = \frac{1}{m_G}\sum_{i=1}^{m_G} 1_{(-\infty,t]}(z(i))$;

— the empirical survival function $\hat{\mathrm{R}}^e : \mathbb{R} \times \mathbb{R}_+^{m_G} \to \mathbb{R}_{[0,1]}$ from the sample $\mathbf{z}$ has the form $\hat{\mathrm{R}}^e(t,\mathbf{z}) = 1 - \hat{\mathrm{F}}^e(t,\mathbf{z})$.

Example 1. Let us consider an issue of the effectiveness analysis of the distributed system, which structure determines a network (Fig. 1) of the form:

$$N = (V, E, \mathbf{c}), \tag{15}$$

where: $V = \{H, R\} = \{v_1, v_2, \ldots, v_{m_V}\}$, $m_V = 6$, is the set of elements,
$H = \{h_1, \ldots, h_{m_H}\}$, $m_H = 4$, is the set of computers, and $R = \{r_1, \ldots, r_{m_R}\}$, $m_R = 2$, is the set of routers, where: $v_1 = h_1$, $v_2 = h_2$, $v_3 = r_1$, $v_4 = r_2$, $v_5 = \mathrm{h}_3$, $v_6 = h_4$;
$E = \{e_1, \ldots, e_{m_E}\}$, $m_E = 10$, is the set of unidirectional network links, where:
$e_1 = (v_1, v_3)$, $e_2 = (v_3, v_1)$, $e_3 = (v_2, v_3)$, $e_4 = (v_3, v_2)$, $e_5 = (v_3, v_4)$,
$e_6 = (v_4, v_3)$, $e_7 = (v_4, v_5)$, $e_8 = (v_5, v_4)$, $e_9 = (v_4, v_6)$, $e_{10} = (v_6, v_4)$;
$\mathbf{c} = [c_1, \ldots, c_{m_E}]^T \in \mathbb{R}_+^{m_E \times 1}$ is the vector of permissible bandwidths of all network links, where: $c_i = 100$ [Mbps], for $i = 1, \ldots, m_E$.

The set P, Eq. (3), has the form $P = \{p_1, \ldots, p_{m_P}\}$, $m_P = 30$, where: $p_1 = p_{(v_1,v_2)} = (v_1, v_3, v_2)$, $p_2 = p_{(v_1,v_3)} = (v_1, v_3)$, …, $p_{29} = p_{(v_6,v_4)} = (v_6, v_4)$, $p_{30} = p_{(v_6,v_5)} = (v_6, v_4, v_5)$.
The set S, Eq. (5), has the form $S = \{1, 4, 5, 6, 9, 10, 21, 22, 25, 26, 27, 30\}$. The sets L_s, $s \in S$, have the forms: $L_1 = \{1, 4\}$, $L_4 = \{1, 5, 7\}$, …, $L_{27} = \{10, 6, 4\}$, $L_{30} = \{10, 7\}$.

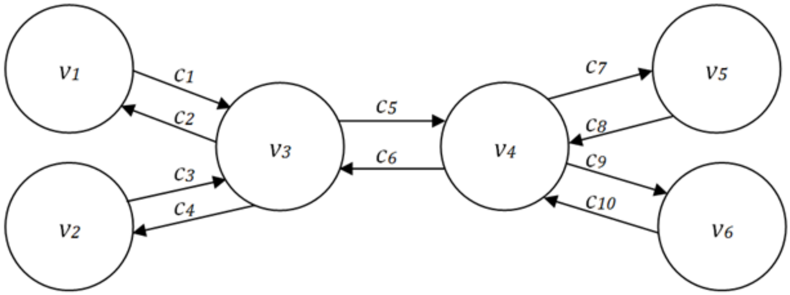

Fig. 1. Scheme of the network N, Eq. (15)

Let $\mathbf{z} = [z(0), z(1), ..., z(m_G - 1)]^T$ be the result of the simulation experiment G generated by algorithm 1 (see Attachment) based on the following input data:

— the net N, Eq. (15), represented in the form of the sets H and R and the vector $\mathbf{c}$;
— the routing matrix $\mathbf{A}_{10\times 30}$ of the net N, Eq. (15);
— the set S ;
— the sets L_s, $s \in S$;
— the samples $\mathbf{y}_{sl} = [y_{sl}(0), y_{sl}(1),..., y_{sl}(m_Y - 1)]^T$, $l \in L_s$, $s \in S$;
— the lengths $m_{\theta_s} = 100$ [Mb] of the messages $\theta_s \in \Theta$, $s \in S$;
— the number $m_G = 20$ of observation of the simulation experiment G.

Table 1 shows evaluations of the effectiveness indicators, i.e. $\hat{\mu}(\mathbf{z})$ and $\hat{\sigma}^2(\mathbf{z})$, while Fig. 2 shows evaluations of the effectiveness functions, i.e. $\hat{\mathrm{F}}(t,\mathbf{z})$, $\hat{\mathrm{R}}(t,\mathbf{z})$ and $\hat{\mathrm{F}}^e(t,\mathbf{z})$, $\hat{\mathrm{R}}^e(t,\mathbf{z})$, of the distributed system.

Table 1
Evaluations the effectiveness indicators of the distributed system

$\hat{\mu}(\mathbf{z})$	$\hat{\sigma}^2(\mathbf{z})$
10.778	4.11495

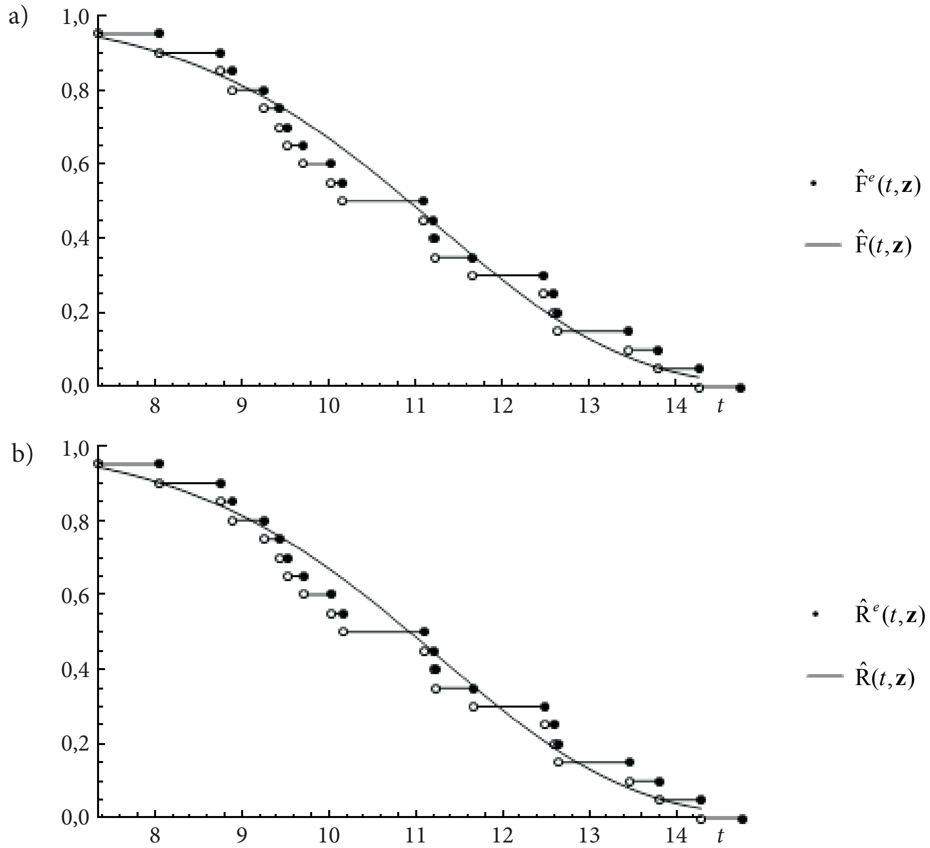

Fig. 2. Plots of evaluations of the effectiveness functions of the distributed system: (a) a plot of the empirical distribution function $\hat{F}^e(t,\mathbf{z})$ and a plot of the evaluation $\hat{F}(t,\mathbf{z})$ of the distribution function F(t) of the random variable ζ, Eq. (14); (b) a plot of the empirical survival function $\hat{R}^e(t,\mathbf{z})$ and a plot of the evaluation $\hat{R}(t,\mathbf{z})$ of the survival function R(t) of the random variable ζ

❐

7. Conclusions and Future Work

This paper has described an approach for effectiveness evaluation of the distributed system based on a stochastic simulation of the process of passing messages through the network. In this method it is assumed that stochastic models of processes of passing messages along the links and paths of the network are random variables with different Weibull distributions. Preliminary simulation results indicate the possibility of applying the proposed method to analyze the effectiveness of structurally and functionally complex systems.

Directions of further works should tackle the construction of adequate models of processes passing messages along the network paths. This is a difficult issue for two main reasons. The first is that it is usually not possible to directly observe the bit rates of passing messages along the entire paths. The second is related to the fact that the transmission of the data over a network is a multi-dimensional non-stationary process, subject to rapid and unpredictable fluctuations [10, 36, 39, 40, 43, 46]. These reasons make the problem of building the considered models difficult. For its solution one should use achievements of stochastic processes theory [4] and chaos theory [1, 2].

Attachment

The following algorithm is an implementation of the simulation method to generate samples, based on which the effectiveness analysis of distributed systems is kept.

Algorithm 1.

Input
- the net N, Eq. (2), represented in the form of the sets H and R and the vector $\mathbf{c}$;
- the routing matrix $\mathbf{A}$, Eq. (4), of the net N;
- the set S, Eq. (5);
- the sets L_s, Eq. (6), $s \in S$;
- the samples $\mathbf{y}_{sl}$, Eq. (11), for: $l \in L_s$, $s \in S$;
- the lengths m_{θ_s} [Mb] of the messages $\theta_s \in \Theta$, $s \in S$;
- the number $m_G \in \mathbb{N}$ of observation of the simulation experiment G.

Output
- the sample $\mathbf{z} = [z(0), z(1),..., z(m_G - 1)]^T$ that is the result of the simulation experiment G.

I. To conduct preprocessing

i. To conduct examining the randomization of samples $\mathbf{y}_{sl}$, $l \in L_s$, $s \in S$, using the Wald-Wolfowitz test [18, 27].
   If there are no grounds for rejecting null hypotheses stating about the randomization of these samples, to continue calculations from the point ii; otherwise to finish the algorithm.

ii. To conduct examining the goodness-of-fit of the samples $\mathbf{y}_{sl}$ with the Weibull distributions $W(\alpha_{sl}, \beta_{sl})$, $l \in L_s$, $s \in S$, using the Pearson χ^2 test [18, 27]. Let $\hat{\mathrm{F}}_{ls}(t, \mathbf{y}_{ls})$ be the distribution function from the sample $\mathbf{y}_{sl}$ that is an evaluation of the distribution function of the random variable $\psi_{sl} \in W(\alpha_{sl}, \beta_{sl})$, for: $l \in L_s$, $s \in S$.

If there are no reasons for rejecting null hypotheses stating about the goodness-of-fit of the samples $\mathbf{y}_{sl}$ with the distributions $W(\alpha_{sl}, \beta_{sl})$, $l \in L_s$, $s \in S$, to continue calculations from the point iii; otherwise terminate the algorithm.

iii. To construct pseudorandom number generators $\hat{\psi}_{sl}$ with distributions determined by the distribution functions $\hat{\mathrm{F}}_{ls}(t, \mathbf{y}_{ls})$, $l \in L_s$, $s \in S$ [17, 49].

iv. To construct the sets:

- $I = \{0, 1, ..., m_G - 1\}$ — the set of numbers of observations of the simulation experiment G;
- $L = \{1, ..., m_E\}$ — the set of numbers of links of the network N, Eq. (2);
- $P = \{p_1, p_2, \ldots, p_{m_P}\}$ — the set of paths developed on the basis of the routing matrix $\mathbf{A}$;
- $S_l = \{s \in S : a_{ls} = 1\}$ — the set of numbers of paths run across the link l, for $l \in L$.

II. Conduct processing

1. **For** $i \in I$ **do**
   begin
   - using pseudorandom number generators $\hat{\psi}_{sl}$ to generate the realizations y_{sl} of the random variables ψ_{sl}, $l \in L_s$, $s \in S$, satisfying the inequalities:

$$y_{sl} > 0, \quad l \in L_s, s \in S,$$

$$\sum_{s \in S_l} y_{sl} \leq c_l, \quad l \in L_s, s \in S;$$

   - to calculate the bit rates of passing messages along the paths with numbers from the set S from the formula:

$$y_s = \min_{l \in L_s} \{y_{ls}\}, \quad s \in S;$$

   - to calculate the length of the time interval of executing the task X from the formula:

$$z(i) = \max_{s \in S} \left\{ \frac{m_{\theta_s}}{y_s} \right\};$$

2. **end.**

□

Received November 7, 2012. Revised April 2, 2014.

REFERENCES

[1] Abarbanel H.D.I., *Analysis of Observed Chaotic Data*, Springer, 1996.

[2] Alligood K.T., Sauer T.D., Yorke J.A., *Chaos: An Introduction to Dynamical Systems*, Springer--Verlag, 1996.

[3] Blondel V.D. (ed.), Megretski A. (ed.), *Unsolved Problems in Mathematical Systems and Control Theory*, Princeton University Press, 2004.

[4] Box G.E.P., Jenkins G.M., Reinsel G.C., *Time Series Analysis. Forecasting and Control*, John Willey and Sons, Inc., 2008.

[5] Boxma O., Koole G., Liu Z., *Queuing-theoretic Solution Methods for Models of Parallel and Distributed Systems*, in *Performance Evaluation of Parallel and Distributed Systems Solution Methods*, CWI Tract 105 and 106, Amsterdam, 1994.

[6] Brandt S., *Analiza danych*, Państwowe Wydawnictwo Naukowe, 2002.

[7] Camarasu-Pop S., Glatard T., Da Silva R.F., Gueth P., Sarrut D., Benoit-Cattin H., *Monte Carlo Simulation on Heterogeneous Distributed Systems: A Computing Framework with Parallel Merging and Checkpointing Strategies*, Future Generation Computer Systems, 29, 3, March 2013, 728-738.

[8] Calafiore G.C., *Distributed Randomized Algorithms for Probabilistic Performance Analysis*, Systems & Control Letters, vol. 58, Issue 3, March 2009, 202-212.

[9] Casale G., *Exact Analysis of Performance Models by the Method of Moments*, Performance Evaluation, 68, 6, June 2011, 487-506.

[10] Deane J.H.B., Jefferies D.J., Smythe C., *Chaotic Traffic Flow in Local Area Networks*, Proceedings of the 11th European Conference on Circuit Theory and Design, Davos, Switzerland, 1993, 843-848.

[11] Diestel R., *Graph Theory*, Springer, 2010.

[12] Dirkx E., Verboven F., Tiberghien J., *Distributed Simulation of Computer Networks*, Microprocessing and Microprogramming, 30, 1-5, August 1990, 215-220.

[13] Dodson B., The *Weibull Analysis Handbook*, ASQ Quality Press, 2006.

[14] Fishman G.S., *Symulacja komputerowa. Pojęcia i metody*, Państwowe Wydawnictwo Ekonomiczne, 1981.

[15] Fishman G.S., *Discrete-Event Simulation: Modelling, Programming, and Analysis*, Springer, 2001.

[16] Fujimoto R.M., *Parallel and Distributed Simulation Systems*, Wiley, 2000.

[17] Gentle J.E., *Random Number Generation and Monte Carlo Methods*, Springer, Heidelberg, 2003.

[18] Gentle J.E. (ed.), Härdle W. (ed.), Mori Y. (ed.), *Handbook of Computational Statistics*, Springer, 2004.

[19] Harchol-Balter M., *Performance Modelling and Design of Computer Systems: Queuing Theory in Action*, Cambridge University Press, 2013.

[20] Hussain H., Malik S.U.R., Hameed A., Khan S.U., et al., A *Survey on Resource Allocation in High Performance Distributed Computing Systems*, Parallel Computing, 39, 11, November 2013, 709-736.

[21] Jean-Marie A., Lefebvre-Barbaroux S., Liu Z., An *Analytical Approach to the Performance Evaluation of Master–Slave Computational Models*, Parallel Computing, 24, 5-6, June 1998, 841-862.

[22] Kleijnen J.P.C., *Design and Analysis of Simulation Experiments*, Springer, 2010.

[23] Kushwaha R., *Methodology for Predicting Performance of Distributed and Parallel Systems*, Performance Evaluation, 18, 3, November 1993, 189-204.

[24] Ljung L., *System Identification. Theory for the User*, PTR Prentice Hall, 1999.

[25] Mazzeo A., Mazzocca N., Russo S., Vittorini V., A *Method for Predictive Performance of Distributed Programs*, Simulation Practice and Theory, 5, 1, 15 January 1997, 65-82.

[26] Magiera R., *Modele i metody statystyki matematycznej.* Część I. *Rozkłady i symulacja stochastyczna*, Oficyna Wydawnicza GiS, 2005.

[27] Magiera R., *Modele i metody statystyki matematycznej.* Część II. *Wnioskowanie statystyczne*, Oficyna Wydawnicza GiS, 2007.

[28] Medhi D., *Network Routing: Algorithms, Protocols, and Architectures*, Morgan Kaufmann, 2007.

[29] Mentis A., Katsaros P., Angelis L., A *Simulation Process for Asynchronous Event Processing Systems: Evaluating Performance and Availability in Transaction Models*, Simulation Modelling Practice and Theory, 29, December 2012, 66-77.

[30] Mitzenmacher M., Upfal E., *Metody probabilistyczne i obliczenia. Algorytmy randomizowane i analiza probabilistyczna*, Wydawnictwa Naukowo-Techniczne, 2009.

[31] Motwani R., Raghavan P., *Randomized Algorithms*, Cambridge University Press, 1995.

[32] National Research Council, *Network Science*, Washington, DC: The National Academies Press, 2005.

[33] Nikolaidou M., Anagnostopoulos D., A *Distributed System Simulation Modelling Approach*, Simulation Modelling Practice and Theory, 11, 3-4, 15 July 2003, 251-267.

[34] Pacut A., *Prawdopodobieństwo. Teoria. Modelowanie probabilistyczne w technice*, Wydawnictwa Naukowo-Techniczne, 1985.

[35] Panik M.J., *Advanced Statistics from an Elementary Point of View*, Elsevier Academic Press, 2005.

[36] Paxson V., Floyd S., *Difficulties in Simulating the Internet*, IEEE/ACM Transactions on Networking, 9, 4, 2001, 392-403.

[37] Peterson J.L., *Petri Net Theory and the Modeling of Systems*, Prentice Hall, 1981.

[38] Shestak V., Chong E.K.P., Maciejewski A.A., Siegel H.J., *Probabilistic Resource Allocation in Heterogeneous Distributed Systems with Random Failures*, Journal of Parallel and Distributed Computing, 72, 10, October 2012, 1186-1194.

[39] Snyder H., Kurtze D., *Chaotic Behaviour in Computer Mediated Network*, Information Processing & Management, 32, 5, 1996, 555-562.

[40] Srikant R., The *Mathematics of Internet Congestion Control*, Birkhäuser, Berlin, 2004.

[41] Stroock D.W., An *Introduction to Markov Processes*, Springer, 2005.

[42] Tikhonenko O., *Modele obsługi masowej w systemach informacyjnych*, Akademicka Oficyna Wydawnicza, 2003.

[43] Vattay G., Diriczi K., Fekete A., Kocarev L., Marodi M., Steger J., *Statistical Properties of Chaos in Communication Networks*, in L. Kocarev (ed.), G. Vattay (ed.), *Complex Dynamics in Communications Networks*, Springer, 2010.

[44] Veres A., Boda M., *Nonlinear Dynamics of TCP and its Implications to Network Performance*, in L. Kocarev (ed.), G. Vattay (ed.), *Complex dynamics in Communications Networks*, Springer, 2010.

[45] Wainer G.A. (ed.), Mosterman P.J. (ed.), *Discrete-Event Modeling and Simulation: Theory and Applications*, CRC Press, 2010.

[46] Willinger W., Paxson V., *Where Mathematics Meets the Internet*, Notices of the AMS, 45, 8, September 1998, 961-970.

[47] Wilson R.J., *Wprowadzenie do teorii grafów*, Państwowe Wydawnictwo Naukowe, 2012.

[48] Yuan J., Mills K., A *Cross-Correlation Based Method for Spatial-Temporal Traffic Analysis*, Performance Evaluation, 61, 2-3, July 2005, 163-180.

[49] Zieliński R., Wieczorkowski R., *Komputerowe generatory liczb losowych*, Wydawnictwa Naukowo-Techniczne, 1997.

Z. WESOŁOWSKI

Symulacyjna metoda szacowania efektywności systemów rozproszonych

Streszczenie. Artykuł jest poświęcony omówieniu symulacyjnej metody szacowania efektywności systemów rozproszonych opartych na przesyłaniu komunikatów. Do modelowania matematycznego tych systemów wykorzystano dwa paradygmaty, tj. paradygmat modelowania eksperymentalnego i paradygmat modelowania probabilistycznego. W pracy zaproponowano miary efektywności rozpatrywanej rodziny systemów. Przedstawiono algorytm symulacji stochastycznej szacowania efektywności systemów rozproszonych.

Słowa kluczowe: systemy rozproszone, symulacja stochastyczna, analiza efektywności systemów