# Model of optimum placement of servers and web-contents in content delivery systems

**Article** *in* Automatic Control and Computer Sciences · August 2006

**3 authors:**

Rasim Alguliyev
Institute of Information Technology,Azerbaijan National Academy of Sciences
**229** PUBLICATIONS **1,258** CITATIONS
SEE PROFILE

Ramiz Aliguliyev
Institute of Information Technology, Azebaijan National Academy of Sciences
**157** PUBLICATIONS **1,586** CITATIONS
SEE PROFILE

Mahammad Sharifov
Khazar University
**6** PUBLICATIONS **1** CITATION
SEE PROFILE

*ISSN 0146-4116, Automatic Control and Computer Sciences, 2006, Vol. 40, No. 4, pp. 28–34. © Allerton Press, Inc., 2006.*
*Original Russian Text © R.M. Alguliev, R.M. Alyguliev, M.G. Sharifov, 2006, published in Avtomatika i Vychislitel'naya Tekhnika, 2006, No. 4, pp. 42–50.*

# Model of Optimum Placement of Servers and Web-Contents in Content Delivery Systems

**R. M. Alguliev, R. M. Alyguliev, and M. G. Sharifov**

*Institute of Information Technologies National Academy of Sciences of Azerbaijan,*
*ul. F. Agaeva 9, AZ-1141 Baku, Azerbaijan*
*e-mail: rasin@science.az, a.ramiz@science.az*

**Abstract**—A new model of optimum placement of servers and Web contents in a Content Delivery Network that is intended to minimize the cost of delivery of content to the ultimate users is proposed. The model also takes into account the structure of the network and the weight of each Web content in the network nodes. A mathematical formulation of the proposed model reduces to a problem of linear integer programming. In the present study synthesis of a neural network for the solution of a problem of linear integer programming is also described.

**Key words:** Content Delivery Network, placement of servers and contents, linear integer programming, neural networks

## 1. INTRODUCTION

With the exponential growth in traffic on the World Wide Web (WWW), the problem of improving the quality of service of users is becoming ever more critical. New architectural technologies, such as caching [1, 2] and replication of servers [2, 3], have been developed and expanded to adapt to the rapid growth of the Internet, that is, increase the rate of responses requested by users.

Caching, which is used to increase productivity on the Web, plays an important role in the structure of the speed of Web clients. A cache is a local data storage device used by a client for temporary storage of copies of frequently requested Web contents. The disposition of the cache is a relatively simple task. Usually a cache is placed in the same computer as the client's or in a computer used jointly by several clients which is in the same local network Another approach to caching is to locate the cache servers at certain points of the global network and allow the client to find the closest one [4].

To improve users' QoS (Quality of Service) indicators a complex technology that is more modern than caching has appeared in recent years, termed *replicationi of servers*. With the appearance of this new technology caching has fallen by the wayside.

Two principal methods of replication of servers are used on the Web. First, deeply embedded Web sites utilize clusters of Web servers to reduce the response time. As a rule, this form of replication is transparent to clients. There is also a widely used nontransparent form of replication that involves creating an exact duplicate of the Web site on another server, called a *mirror site*. Here the client gains the ability to select the particular server by means of which he will gain access to the Web site.

Quite recently, the Internet has witnessed the appearance of another form of replication. This form follows a strategy of initiation of replicas of servers. In such an approach a collection of servers scattered throughout the Internet provides hosting of Web documents organized in the form of clients. A collection of such servers is called a Content Delivery Network (CDN) or Content Distribution Network [5, 6]. In other words, a Content Delivery Network is a group of so-called "authorized" distributed proxy servers that make it possible to increase (by comparison with centralized servers) access to content replicas stored on the servers.

Essentially, a Content Delivery Network is a superposed network through which frequently requested Web contents are relocated from central servers to the periphery of the global network closer to the ultimate user and situated on proxy servers; this approach makes it possible to more rapidly satisfy subsequent requests for the same contents. Hence, it may be concluded that a Content Delivery Network is not only an efficient and powerful tool for increasing the rate of response and reducing traffic on the Internet, but also solves yet another problem of no little importance, that of reducing the capital costs of content providers, that is, enable them to avoid having to create their own expensive network infrastructure.

Different technologies of replication and distribution of Web contents in a Content Delivery Network have now been developed [7–9]. For example, in RaDaR [7] the server tracks a number of requests from clients of an individual region. If such requests represent a significant fraction of all the requests to the server, a decision is made to place a copy of the document on a server of this region. Another popular approach is implemented in the Akamai system [8]. The idea underlying this system derives from the observation that since audio and video files embedded in a principal HTML page rarely change, it makes sense to cache or replicate them. The important element in the operation of Akamai and certain other systems, including CODIS [9], is to discover the nearest server of the Content Delivery Network.

At first glance, the principle underlying the creation of a Content Delivery Network seems quite simple. However, studies have shown that implementation of such a network involves a number of difficult problems, including determining an optimal arrangement of proxy servers [10, 11], the placement of Web contents on proxy servers [12–14], routing of requests [15, 16], and constant support of the currency of information on proxy servers and management of the delivery of dynamic data [17].

The present study is concerned with the question of optimal arrangement of servers and Web contents in a Content Delivery Network, with the goal of minimizing the cost of delivering contents to the ultimate users. The problem reduces to a problem of linear integer programming.

## 2. STATEMENT AND MATHEMATICAL FORMULATION OF PROBLEM

Suppose that a particular network consists of a set of nodes $N = \{1, \ldots, n\}$, where $n$ is the number of nodes. It will be assumed that there exists Web content in each node and that this content is frequently requested by other nodes. Our goal is to achieve an optimal arrangement of the set of Content Delivery Network servers, the number of which is $s$, among the potential $m$ nodes of the given network that minimizes the total cost of delivery of Web contents to the ultimate users upon redistribution of frequently requested Web contents in these nodes. The number of servers in the Content Delivery Network $s$ is always less than the number of potential nodes $m$ and the number of potential nodes is less than the number of nodes in the network, $s < m < n$.

By $v_{ij}$ we will denote the volume of a frequently requested Web content by the $j$-th node from the $i$-th node, $i, j = 1, \ldots, n$. In the general case, $v_{ij} \neq v_{ji}$, $v_{ij} \geq 0$ for any $i, j$. Let $c_{ij}$ be the cost of transmitting a unit of information to the $j$-th node from the $i$-th node, $c_{ii} = 0$.

Since one and the same Web content is requested by several nodes, in creating an optimal redistribution of Web contents among the nodes of a Content Delivery Network it is extremely important to determine the weight of the Web contents in each node of the network. To determine this weight, we introduce the following notation: $M$, total number of requests initiated in all nodes; $M_{ij}$, number of requests to $i$-th node initiated in $j$-th node; and $n_i$, number of nodes in which calls are made to the $i$-th node. Using the TF*IDF formula (Term Frequency, Inverse Document Frequency) [18], we determine the weight of calls of the $j$-th node to Web content situated in the $i$-th node:

$$w_{ij} = \begin{cases} \dfrac{M_{ij}}{M}\log_2\left(\dfrac{n}{n_i}\right), & \text{at} \quad n_i \neq 0 \\ 0, \quad \text{at} \quad n_i = 0 \end{cases} \qquad \forall i, j = 1, \ldots, n. \tag{1}$$

Thus, in light of the foregoing discussion the problem of minimizing the cost of delivering Web contents to the ultimate users is formulated thus:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} v_{ij} w_{ij} c_{ijk} x_{ijk} \longrightarrow \min, \tag{2}$$

where the coefficients $c_{ijk} = c_{ik} + c_{kj}$.

In the objective function (2) the Boolean variable $x_{ijk}$ is equal to one if the Web content requested by the $j$-th node from the $i$-th node is located on the Content Delivery Network server situated in the $k$-th node, and is equal to zero otherwise:

$$x_{ijk} \in \{0, 1\}, \quad \forall i, j, k. \tag{3}$$

In the distribution it is assumed that each Web content is situated on only one server of the Content Delivery Network, that is, the following condition must be satisfied:

$$\sum_{k=1}^{m} x_{ijk} = 1, \quad \forall i, j. \tag{4}$$

Since each server has a limited volume of memory, in assigning Web contents to the Content Delivery Network servers the following condition must be observed:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ijk} v_{ij} \le V_k S_k, \quad \forall k, \tag{5}$$

where $V_k$ is the volume of memory of a server situated in the $k$-th node of the network.

On the one hand, condition (5) imposes a constraint on the volume of the Web contents distributed among the servers of the Content Delivery Network, while on the other hand, it prevents assignment of Web contents to inactive nodes, that is, to those nodes at which no servers of the Content Delivery Network are situated.

The Boolean variable $S_k$ is equal to one if a server is located in the $k$-th node and is equal to zero otherwise:

$$S_k \in \{0, 1\}, \quad \forall k. \tag{6}$$

According to the definition of the two variables $x_{ijk}$ and $S_k$,

$$x_{ijk} \le S_k, \quad \forall i, j, k. \tag{7}$$

Since the number of servers is equal to $s$, the next condition is satisfied in a natural way:

$$\sum_{k=1}^{m} S_k = s. \tag{8}$$

We wish to emphasize that in order for the problem to be solvable, we must have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij} \le \sum_{k=1}^{s} V_k.$$

From the computational point of view, the process of solving problem (2)–(8) for large values of $n$ and $m$ is a simple matter. Several exact and approximate methods have now been developed for such problems. Of course, exact methods may be realized only for problems of comparatively low dimension. New tools have appeared for solving problems of optimization with high dimension to assist decision makers, for example, genetic algorithms or neural networks [19, 20]. Because of parallelism, these tools easily handle an “accursed” dimension. In the present study we will employ neural networks to solve the linear integer programming problem (2)–(8). To solve this problem a synthesis of a neural network with feedback is presented below.

## 3. SYNTHESIS OF NEURAL NETWORK FOR THE SOLUTION OF PROBLEM (2)–(8)

The principal advantage of neural networks lies in their ability to rapidly realize computations and, correspondingly, rapidly find desired solutions of a problem, both of which are achieved thanks to the high degree of paralleling of the computational process. The convergence time depends weakly on the dimension of the problem.

The technique used to synthesize a neural network with feedback incorporates the following stages:

—neural network interpretation of problem;

—construction of energy function;

—determination of parameters of problem, i.e., determination of the synaptic relations and external shift.

By a neural network interpretation of the problem is understood a determination of the architecture of the neural network, for which purpose a network of binary neurons that constitute a three-dimensional file $\mathbf{Z} = \|z_{ijk}\|$ of dimension $n \times n \times s$ is considered. In such a definition of the architecture of a neural network, with each Boolean variable $x_{ijk}$ there is associated an output signal $z_{ijk}$ of the $ijk$-th neuron, i.e., a one-to-one correspondence $(x_{ijk} = 1) \Leftrightarrow (z_{ijk} = 1)$, $\forall i, j, k$, is constructed. Following such a mapping, an excited state $z_{ijk} = 1$ of an output $ijk$-th neuron uniquely corresponds to a distribution such that a Web content requested from the $i$-th node is placed by the $j$-th node on the $k$-th server of the Content Delivery Network.

In synthesizing a neural network with feedback one difficult problem is to construct *the network energy function*. It should be noted that the definition of the energy function of a network as a function of the particular problem is not a trivial problem. The construction of such functions is more of an art than a science. The basic drawback of neural networks lies in the fact that they do not always guarantee that a global minimum will be found. The convergence of a network to such a minimum depends to a large extent on the complexity of the energy function corresponding to the problem which is to be solved, more precisely, on the complexity of the form of the surface specified by this function. In constructing the energy function, it is necessary to observe conditions such that, first, the function is convex and, second, at stationary points the function would assume its minimal values.

In view of the foregoing remarks, we construct the energy function of a synthesized neural network in the form of a sum:

$$E(\mathbf{Z}) = E_0(\mathbf{Z}) + E_1(\mathbf{Z}) + E_2(\mathbf{Z}) + E_3(\mathbf{Z}) + E_4(\mathbf{Z}) + E_5(\mathbf{Z}) \tag{9}$$

where each term corresponds to the objective function and the constraints of the problem (2)–(8).

Let us now pass on to the construction of the energy function of a neural network, i.e., to the construction of each term in formula (9).

Let us start with construction of the energy function $E_0(\mathbf{Z})$ corresponding to the objective function (2). This function is constructed in the following form:

$$E_0(\mathbf{Z}) = -\frac{C_0}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} a_{ijk} z_{ijk}, \tag{10}$$

where the coefficients $a_{ijk} = v_{ij} w_{ij} c_{ijk}$.

The next function

$$\begin{aligned} E_1(\mathbf{Z}) &= \frac{C_1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} z_{ijk}(1 - z_{ijk}) \\ &= -\frac{C_1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{m} \delta_{ip}\delta_{jq}\delta_{kr} z_{ijk} z_{pqr} + \frac{C_1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} z_{ijk}. \end{aligned} \tag{11}$$

corresponds to the parity of the variables $x_{ijk}$, i.e., condition (3). If this condition is satisfied, the function $E_1(\mathbf{Z})$ vanishes, which is the minimal value of this function. From condition (3) it follows that the file $\mathbf{Z}$ must contain precisely $n^2$ units. Consequently, it becomes necessary to construct an energy function that corresponds to this constraint. We construct this function in the form

$$\begin{aligned} E_2(\mathbf{Z}) &= \frac{C_2}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} z_{ijk} - n^2\right)^2 \\ &= \frac{C_2}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{m} z_{ijk} z_{pqr} - C_2 n^2 \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} z_{ijk} + \frac{C_2}{2} n^4. \end{aligned} \tag{12}$$

Note that if this condition is satisfied $\mathbf{Z}$ will contain precisely $n^2$ units and the function $E_2(\mathbf{Z})$ will assume its minimal value, or zero.

We next construct the energy function corresponding to constraint (4):

$$\begin{aligned} E_3(\mathbf{Z}) &= \frac{C_3}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{m} z_{ijk}-1\right)^2 \\ &= \frac{C_3}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{m}\delta_{ip}\delta_{jq}z_{ijk}z_{pqr}-C_3\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}z_{ijk}+\frac{C_3}{2}n^2. \end{aligned} \tag{13}$$

If constraint (4) is satisfied the function $E_3(\mathbf{Z})$ will assume its minimal value of zero. Here $\delta_{ip}$ is the Kronecker symbol.

The energy function corresponding to constraint (5) is written thus:

$$E_4(\mathbf{Z}) = \frac{C_4}{2}\sum_{k=1}^{m} h^2\left(\sum_{i=1}^{n}\sum_{j=1}^{n} z_{ijk}v_{ij}-S_kV_k\right), \tag{14}$$

where $h(u)$ is a penalty function that ensures that condition (5) will be satisfied. In determining the penalty function it is also necessary to take into account the property that if condition (5) is violated, the penalty function worsens the value of the function $E_4(\mathbf{Z})$.

On the basis of these considerations, we will define the penalty function $h(u)$ in the form $h(u) = \max(0, u)$ in which it may be shown without any special difficulty that the function satisfies the condition $h^2(u) = uh(u)$.

After this definition of $h(u)$, formula (14) assumes the following form:

$$E_4(\mathbf{Z}) = \frac{C_4}{2}\sum_{k=1}^{m}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} z_{ijk}v_{ij}-S_kV_k\right)h\left(\sum_{i=1}^{n}\sum_{j=1}^{n} z_{ijk}v_{ij}-S_kV_k\right). \tag{15}$$

We construct the function $E_5(\mathbf{Z})$ corresponding to constraint (7) by analogy to the construction of $E_4(\mathbf{Z})$:

$$E_5(\mathbf{Z}) = \frac{C_5}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} h^2(z_{ijk}-S_k) = \frac{C_5}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}(z_{ijk}-S_k)h(z_{ijk}-S_k). \tag{16}$$

The constants $C_0$, $C_1$, $C_2$, $C_3$, $C_4$, and $C_5$ present in formulas (10)–(16) are positive.

Substituting the terms (10)–(16) into (9), we obtain the final form of the energy function of the neural network:

$$\begin{aligned} E(\mathbf{Z}) = &-\frac{C_1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{m}\delta_{ip}\delta_{jq}\delta_{kr}z_{ijk}z_{pqr}+\frac{C_2}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{m}z_{ijk}z_{pqr} \\ &+\frac{C_3}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{m}\delta_{ip}\delta_{jq}z_{ijk}z_{pqr}-\frac{C_0}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}a_{ijk}z_{ijk}+\frac{C_1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}z_{ijk} \\ &-C_2n^2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}z_{ijk}-C_3\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}z_{ijk}+\frac{C_4}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}z_{ijk}v_{ij}h\left(\sum_{i=1}^{n}\sum_{j=1}^{n}z_{ijk}v_{ij}-S_kV_k\right) \\ &+\frac{C_5}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}z_{ijk}h(z_{ijk}-S_k)+\frac{C_2}{2}n^4+\frac{C_3}{2}n^2-\frac{C_4}{2}\sum_{k=1}^{m}S_kV_kh\left(\sum_{i=1}^{n}\sum_{j=1}^{n}z_{ijk}v_{ij}-S_kV_k\right) \\ &-\frac{C_5}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}S_kh(z_{ijk}-S_k). \end{aligned} \tag{17}$$

It is known that the canonical form of the energy function of a neural network with feedback is written thus:

$$E_{\text{canonic}}(\mathbf{Z}) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{m} W_{ijkpqr} z_{ijk} z_{pqr} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} B_{ijk} z_{ijk}, \tag{18}$$

where $W_{ijkpqr}$ is the synaptic weight between the input of the *ijk*-th neuron and the output of the *pqr*-th neuron while $B_{ijk}$ is the shift of the *ijk*-th neuron.

In formula (17) terms that are independent of the state of the neural network will not be taken into account in determining the parameters.

Comparing formulas (17) and (18) and equating their linear and quadratic components, we determine the parameters of the neural network thus:

$$W_{ijkpqr} = C_1\delta_{ip}\delta_{jq}\delta_{kr} - C_2 - C_3\delta_{ip}\delta_{jq}, \tag{19}$$

$$B_{ijk} = -\frac{C_0}{2}a_{ijk} + \frac{C_1}{2} - C_2 n^2 - C_3 + \frac{C_4}{2}v_{ij} + \frac{C_5}{2}, \tag{20}$$

where $i, j, p, q = 1, \ldots, n$, and $k, r = 1, \ldots, m$.

## 4. CONCLUSIONS

Content Delivery Network have recently emerged as a method of reducing the response time which Internet users experience through placement of multiple “proxy servers” closer to the clients. The key factor underlying the efficiency of intermediate networks is the use of such “authorizations.” Another key factor is the placement of the Web contents on the proxy servers of the Content Delivery Network. Existing methods of placement consider only simple network technologies and routing policies. In the present study a new method of optimal arrangement of proxy servers and the placement of contents on these servers is considered. The method takes into account both the structure of the network as well as the weight of each of the contents situated on the servers of the Content Delivery Networks. In most cases such problems reduce to problems of quadratic and linear programming. The proposed method reduces to a problem of linear integer programming, which is also its advantage. The study also p resents a synthesis of a neural network for the solution of a problem of linear integer programming

## REFERENCES

1. G. Barish and K. Obraczka, “World Wide Web Caching: Trends and Techniques,” *IEEE Comms.*, vol. 38, no. 5, pp. 178–184, 2000.
2. E. Tanenbaum and M. van Steen, *Distributed Systems. Principles and Parameters* [Russian translation], Piter, St. Petersburg, 2003.
3. S. Sivasubramanian, M. Szymaniak, G. Pierre, and M. van Steen, “Replication for Web Hosting Systems,” *ACM Computing Surveys* (*CSUR*), vol. 36, no. 3, pp. 291–334, 2004.
4. P. Backx, T. Hambrecht, B. Dhoedt, F. De Turck, and P. Demeester, “Optimizing Content Distribution Through Adaptive Distributed Caching,” *Comp. Comms.*, vol. 28, no. 6, pp. 640–653, 2005.
5. A. Vakali and G. Pallis, “Content Delivery Networks: Status and Needs,” *IEEE Internet Comput.*, vol. 7, no. 6, pp. 68–74, 2003.
6. B. Molina, C. E. Palau, and M. Steve, “Modeling Content Delivery Networks and Their Performance,” *Comput. Comms.*, vol. 27, no. 15, pp. 1401–1411, 2004.
7. M. Rabinovich and A. Aggrawal, “RaDaR: A Scalable Architecture for a Global Web Hosting Service,” in: *Proc. 8th Intern. World Wide Web Conference (WWW8)*, Toronto, May 11–14, 1999, pp. 1545–1561.
8. F. Leighton and D. Lewin, *Global Hosting System*, US Patent no. 6,108,703, August, 2000.
9. H. Hlavacs, M. Haddad, C. Lafouge, D. Kaplan, and J. Ribeirao, “The CODIS Content Delivery Network,” *Comput. Networks*, vol. 48, no. 1, pp. 75–89, 2005.
10. H. S. Bassali, K. M. Kamath, R. B. Hosamani, and L. Gao, “Hierarchy-aware Algorithms for CDN Proxy Placement in the Internet,” *Comput. Comms.*, vol. 26, no. 3, pp. 251–263, 2003.
11. D. Starobinski and T. Wu, “Performance of Server Selection Algorithms for Content Replication Networks,” in: *NETWORKING 2005: 4th Intern. IFIP-TC65 Networking Conference*, Waterloo, Canada, May 2–6, 2006. Lecture Notes in Computer Science (Springer-Verlag), vol. 3462, pp. 443–454, 2005.

12. N. Laoutaris, O. Telesis, V. Zissimopoulos, and I. Stavrakakis, “Local Utility Aware Content Replication,” *ibid.*, pp. 455–468.
13. S. Bakiras and T. Loukopoulos, “Combining Replica Placement and Caching Techniques in Content Distribution Networks,” *Comput. Comms.*, vol. 28, no. 9, pp. 1062–1073, 2005.
14. S. S. H. Tse, “Approximation Algorithms for Document Placement in Distributed Web Servers,” *IEEE Trans. Parallel and Distributed Systems*, vol. 16, no. 6, pp. 489–496, 2005.
15. S. Rangarajan, S. Mukherjee, and P. Rodriguez, “A Technique for User Specific Request Redirection in a Content Delivery Network,” *Proc. 8th Intern. Workshop on Web Content Caching and Distribution*, New York, September 29–October 1, 2003.
16. C. Yang and M. Luo, “A Content Placement and Management System for Distributed Web-server Systems,” *Proc. 20th IEEE Intern. Conf. Distributed Computing Systems* (*ICDCS’2000*), April 10–13, 2000, Taipei, Taiwan, IEEE Computer Society, 2000, pp. 691–698.
17. K. Walkowiak, “QoS Dynamic Routing in Content Delivery Networks,” in: *NETWORKING 2005: 4th Intern. IFIP-TC6 Networking Conference*, Waterloo, Canada, May 2–6, 2005, Lecture Notes in Computer Science (Springer-Verlag), vol. 3462, pp. 1120–1132, 2005.
18. G. Salton, Automatic Text Processing: The Transformation, Analysis and Retrieval of Information, Addison-Wesley, 1989.
19. Y. Takahashi, “A Neural Network Theory for Constrained Optimization,” *Neurocomputing* (*Netherlands*), vol. 24, no. 1–3, pp. 117–161, 1999.
20. A. I. Galushkin, ed., *Neural Mathematics. Book 6. A Textbook for Post-Secondary Educational Institutions* [in Russian], IPRZhR, Moscow, 2002.