A Novel Distributed Data Access Scheme Considering with Link Resources and Metric in Lambda Grid Networks

Ryota Usui, Hiroyuki Miyagi, Yutaka Arakawa, Satoru Okamoto, and Naoaki Yamanaka
Dept. of Information and Computer Science,
Keio University, 3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa, Japan 223-8522
Email: usui@yamanaka.ics.keio.ac.jp

***Abstract*—A link load balancing technique that uses a normalized link metric based on a new wavelength utilization and link metric is proposed. The new normalized link metric is a combination of a previous link metric and wavelength utilization. The shortest path algorithm selects the lowest normalized link metric, so that heavily loaded links are avoided if possible. Due to the development of WDM optical network technologies, a lambda grid system, which shares computing resources across a wide area optical network, has been proposed as a grid system. To reduce the influence of file damage and/or prevent particular storage servers from being over loaded, duplicate files are stored in several replica servers. Parallel downloading from these servers decreases downloading time and achieves load distribution. However, parallel downloading can create bottleneck links in the WDM network, since many connections may become used in common. In this paper, we propose a novel server selection method for parallel downloading in WDM networks. The server selection method creates link load balancing and suppresses bottleneck links. Computer simulations show that the proposed scheme can reduce bottleneck link number and the blocking probability by up to 90%.**

I. Introduction

With the advance of network technologies and high performance computing, research on grid computing has become very popular[1]. Grid computing is a technique to create a high performance virtual machine by combining high performance computers (HPCs), data storage devices, and various I/O devices. The high performance virtual machine makes it possible to execute large scale jobs such as scientific and engineering simulations and the high speed processing of large amounts of data. In grid computing, data size ranges from Tera bytes to Peta bytes because large volumes of calculation results are stored and shared among the components. In many cases, data are stored on many storage servers using a specific file format. This network category is known as a data grid[2]. In the future, file size will increase and file transfer times may reach several tens of minutes to several hours. To share these huge files which are geographically dispersed, lightpath-based communication is essential. To transmit huge data between end hosts (HPCs), gigabit or terabit class bandwidth light paths are needed. A lambda grid, which employs Wavelength Division Multiplexing (WDM) and lightpath technologies, is a very attractive approach to realizing the data transmission needed[3], [4]. WDM offers large transmission capacity to the network and also end hosts, so high speed data transport is possible.

In grid computing, to distribute storage server load and to reduce the influence of file damage, replica files are generally stored on many storage servers. Because the same file is stored on multiple storage servers, it can be downloaded in parallel from several of the storage servers which increases the data transfer rate; parallel downloading is supported by GridFTP[5], the standard file transfer protocol for grid computing. A method of parallel downloading in which the client connects to all servers has been proposed[6]. Unlike single-sourced downloads, parallel downloading generates many connections in the WDM network which means that there a greater chance that several connections will share the same link. Therefore, optimal storage server selection is required to improve the wavelength resource efficiency and to reduce the request blocking probability.

In this paper, a novel server selection method for parallel downloading in the lambda grid network is proposed. The proposed server selection method provides more efficient wavelength resource utilization and so new parallel downloading requests experience lower blocking probability. Computer simulation results show the effectiveness of the proposed method.

II. File transfer technique for lambda grid networks

A. Parallel downloading for lambda grid

Figure 1 shows an example of the lambda grid system. Many storage servers and client hosts are located at the edge of the WDM networks which use optical cross-connects (OXCs) and WDM links. Replicated files are held in many storage servers, and clients setup lightpaths to download files from the storage servers.

In parallel downloading, a file is divided into multiple blocks of appropriate size that are then downloaded to the client in parallel using different lightpaths from multiple storage servers. Two types of file division methods are used: static parallel downloading and dynamic parallel downloading. In the former, the file is divided into fixed-size blocks in advance and predetermined quantities of blocks are downloaded from preset servers. The download time can, however, become excessive if the link between one of the servers and the client becomes congested. Dynamic parallel downloading[7], [8], allows the client to download more blocks more servers whose links are relatively uncongested. As a result, the number of blocks

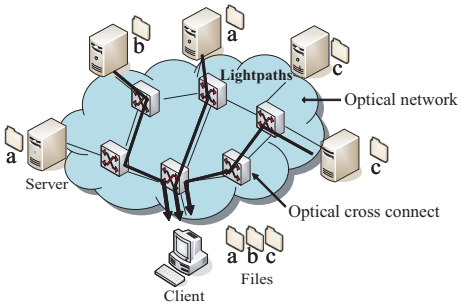


Fig. 1. Overview of Lambda grid system

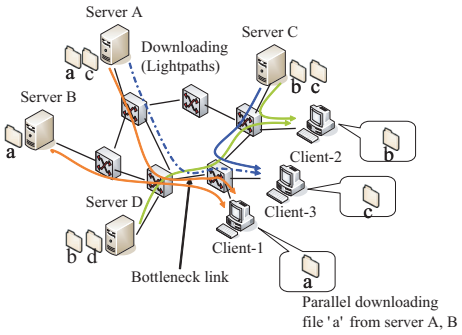


Fig. 2. An example of bottleneck link in the conventional Lambda grid network

downloaded from each server is dynamically controlled. This avoids the problem of the static parallel downloading method. This paper considers only the dynamic parallel downloading method hereafter.

B. The server selection method for lambda grid

The major goals of parallel downloading for the lambda grid are server load balancing and download speed enhancement. However, published conventional studies address only download time. This paper considers both server load balancing and download time. Round trip time (RTT) is a popular method of evaluating the distance between server and client[9]. It is one of the most scalable evaluation methods, because the packets injected into the network are so small that they are unlikely to degrade network throughput. However, if [9] is applied to the lambda grid, the shortest routes or smallest metric values are always selected as the routes between clients and servers[10]. When multiple clients download from neighboring servers, their traffic tends to focus on one link near the servers. As a result, a new lightpath may be blocked because all wavelengths of the link are occupied. This link is defined as a bottleneck link, see Figure 2. In Fig.2, each link offers three wavelengths. If enough lightpaths pass through the same link, the link will become a bottleneck link. As shown in Fig.2, client-1 downloads file 'a' in parallel from servers A and B, and client-2 downloads file 'b' from servers C and D at the same time. In this case, if client-3 requests the download of file 'c' from servers A and C, the setup of a new lightpath set between server A and client-3 is blocked.

Figure 3 shows an example of the conventional server selection method for parallel downloading[9], [10]. The

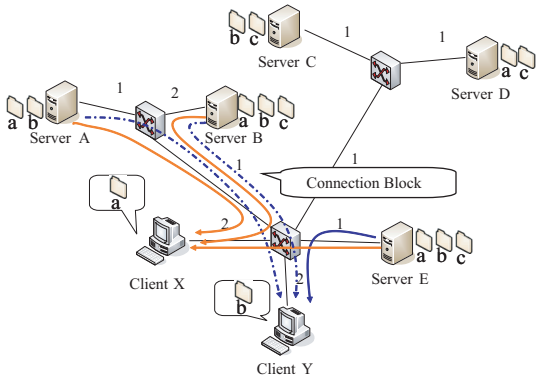


Fig. 3. Bottleneck link and server selection method for the parallel downloading in conventional network

number written on each link is the link cost metric. When client X requests file 'a' from three parallel downloading servers, the three shortest paths between client X and the servers are located. As a result, server A, server B and server E are selected and three lightpaths are established for client X. When client Y requests the download of file 'b' from three parallel downloading servers while client X is downloading file 'a' from server A, server B, and server E, the three shortest paths between client Y and the three selected servers are identified. Server A, server B and server E are selected again. However, client X is already using two wavelengths links between A, B and client Y. Therefore, client Y's request is blocked because the wavelength resources are fully occupied, so parallel downloading cannot be established via the specified number of wavelengths (in this example: 3).

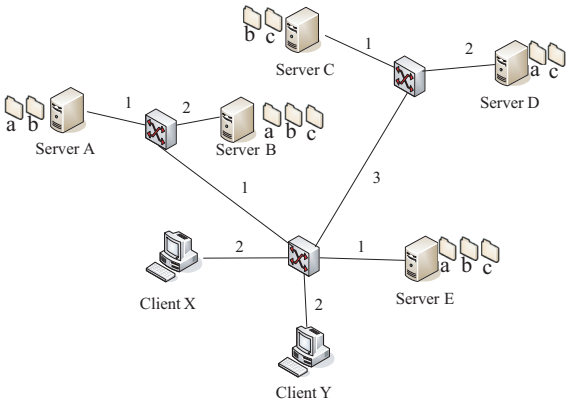
III. Proposed Scheme

In the conventional server selection method described in Section II-B, a client selects $'k'$ nearest servers by the shortest path search method[11] which does not consider network resource availability in server selection. Therefore, some links may be selected so often that they become bottleneck links. To reduce the number of bottleneck links and to enhance overall efficiency of the network resource utilization; i.e. wavelength utilization, a novel server selection method that distributes parallel downloading routes over the whole WDM network is required. The method proposed herein considers both link wavelength availability and link cost metric in server selection. A normalized link cost is calculated from wavelength availability and a direct cost metric. The normalized link cost is used to identify the least cost servers. The normalized cost is defined as follows.

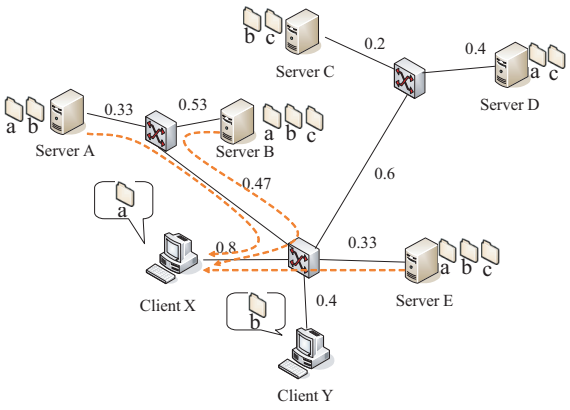
$$\alpha \times \frac{Link\,Metrics}{Max\,Link\,Metrics} + (1-\alpha) \times \frac{Number\,of\,Used\,Wavelength}{Number\,of\,Maximum\,Wavelength} \quad (1)$$

where $Link\,Metrics$ is the original metric cost of each link. $Max\,Link\,Metrics$ is the largest metric value in the network. $Number\,of\,Used\,Wavelength$ is the current number of wavelengths used in each link. $Number\,of\,Maximum\,Wavelength$ is the maximum

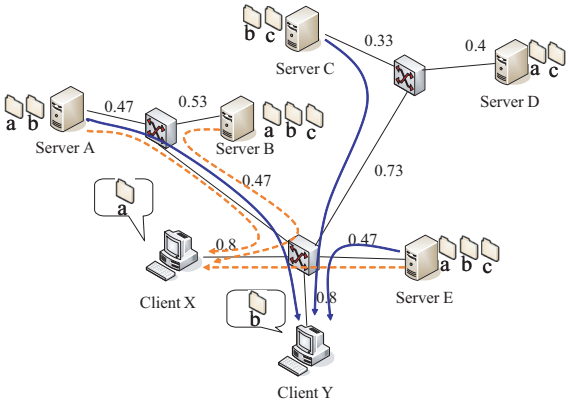
number of wavelengths carried (active and inactive) by each link. α is a weighting parameter that defines the balance between direct cost and wavelength availability. The proposed method may yield longer downloading times than the conventional server selection method because sometimes the nearest servers are not selected. However, the elimination of bottleneck links allows for successful lightpath setting between the client and servers. Lightpaths have guaranteed bandwidth, so parallel download times can also be guaranteed.



(a) initial state



(b) client X is downloading



(c) client X and client Y are downloading

Fig. 4. Server selection method for proposed scheme which avoid bottleneck link

Figure 4 shows an example of the proposed server selection method. In this case, $alpha = 0.6$, the number of wavelengths in each link is set to three. The numbers written on each link are direct costs in (a) and normalized costs in (b) and (c). Fig.4(a) shows the initial state of the network considered. Here, the initial normalized costs equal the direct costs. This is because, the $Number\ of\ Used\ Wavelength$ is zero for all links. In Fig.4(a), client X requests the download of file 'a' in parallel from three servers. The three servers that have minimum total normalized costs (relative to client X) are selected. Servers E, A, and B are selected since their total normalized costs are 3, 4, and 5, respectively. While client X is downloading file 'a', the normalized cost of each link is changed because of the change in the Number of Used Wavelengths. As shown in Fig.4(b), when client Y request the download of file 'b' in parallel from three servers, the servers that have minimum total normalized costs (relative to client Y) are selected. Servers E, A, and C are selected since their total normalized costs are 0.77, 1.2, and 1.2, respectively. While files 'a' and 'b' are being downloaded, the normalized cost of each link is changed as shown in Fig.4(c). The completion of file downloading changes the normalized costs on the links.

As described above, the proposed scheme uses the normalized cost, which is calculated from both wavelength availability and a direct link metric, to select optimal servers so as to reduce the blocking probability. The proposed scheme can reduce the probability of bottleneck links and distribute the downloading load across the whole network.

TABLE I
PERFORMANCE EVALUATION

Server	20
Client	20
Optical cross Connect	50
Number of wavelength (Access Link)	32
Number of wavelength (Others)	16
Bandwidth of wavelength	1Gbps
File Size	10Gbyte
Topology	Random
Number of link	200
Transport Protocol	TCP/IP

IV. PERFORMANCE EVALUATION

Computer simulations were conducted to determine the blocking probability of download requests, the mean download time, and the success rate of parallel downloading. The following lambda grid system was simulated. Each link consisted of bidirectional optical fiber, the total number of nodes including OXCs, servers, and clients was set to 90. The number of links was 200, and the random topology was used. Each access link (client to OXC) had 32 wavelengths and each OXC to OXC link had 16 wavelengths. Each wavelength offered the bandwidth of 1 Gbps, all files requested were 10 Gbytes, and TCP/IP was used as the transport protocol. Simulation parameters are summarized in Table I. α values range from 0 to 1.0. Note that setting α to 1.0 is the same as using the conventional scheme since wavelength availability is not considered in the server selection step. As α decreases, the importance of wavelength availability is increased.

A. Optimal α decision

Figure 5 shows the download request blocking probability evaluation for different α values. In this evaluation,

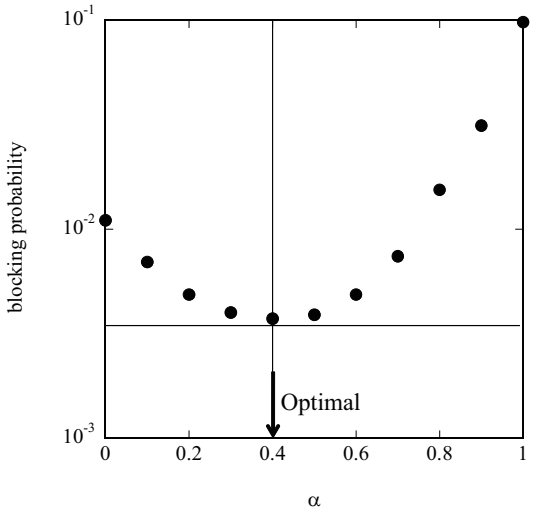


Fig. 5. The blocking probability and optimal value for weighting α (Request rate = 0.04/sec)

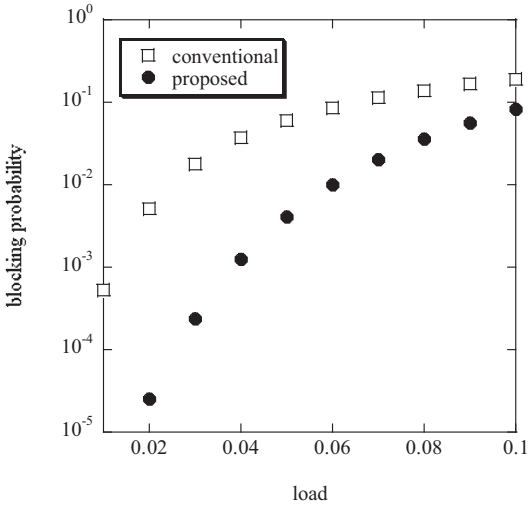


Fig. 6. The blocking probability reduction from conventional scheme as a function of request ratio

a request block (rejection) is defined as follows. First, a client requests parallel downloading from the specified number $'k'$ of servers. Next, if the requested $'k'$ lightpaths can not be set to $'k'$ servers, the client retries with $'k-1'$ servers. Finally, the download request is blocked if no lightpath can be set for downloading.

This evaluation used the request rate of 0.04 [requests/sec]. In Fig.5, the blocking probability is highest at $\alpha = 1.0$(conventional). When α is large, the client tends select the nearest servers. In this case, clients select specific servers according to the replica file location. Therefore, the probability that multiple lightpaths are established on the same optical fibers is high. This triggers bottleneck links and request blocking is frequent. As the value of α decreases, the blocking probability is decreased because server selection is executed with consideration of wavelength availability and direct link costs. In these results, the blocking probability is lowest at $\alpha = 0.4$. However, as in α smaller than 0.4, the blocking probability becomes high. Excessively small α values devalue the direct link cost, which results in very long lightpaths that consume the wavelength resources. Other load values yielded the same optimum α value of 0.4 and so this value was used in subsequent evaluations.

B. Blocking probability characteristics

Figure 6 shows the blocking probability characteristics. Here, the request rate was changed from 0.02 to 0.1 [requests/sec]. The proposed scheme can reduce the blocking probability compared to the conventional scheme. This is because the proposed scheme considers the available network resources as well as the available wavelengths. For the retransmission control scheme considered, the blocking probability must be under 10^{-1}. The proposed method can realize significantly lower blocking probabilities for the evaluated conditions.

C. Average downloading time characteristics

It is assumed that lightpaths are setup and used for data transfer between a server and the client. Therefore, once the lightpaths are set, the bandwidth between the client and the servers is guaranteed during downloading. The block size or the number of pieces for parallel downloading can be decided from the number of servers, guaranteed bandwidth of each lightpath, and the distance between the client and the server. That is to say, large blocks are set when the bandwidth is large and/or the distance is short. If there are many servers, small blocks would be used. Under a fixed value of α, increasing the number of servers decreases the average download time. However, the interface capabilities of the client may be a limiting factor.

Figure 7 shows the average downloading time characteristics with the number of servers as a parameter. In this evaluation, request blocking and retry were not considered. Therefore, for calculating the average download time, the number of servers actually assigned for downloading was used. When the number of servers is 1, the difference in downloading time between conventional and proposed methods is large. This is due to the difference in mean distance between the client and the server. However, as the number of servers is increased, downloading times becomes almost the same. In this evaluation, TCP/IP was used as the data transport protocol, so the transmission rate is limited by RTT. In the case of long transmission distances, the client cannot fully utilize the bandwidth so the downloading time of the proposed scheme becomes longer than that of the conventional scheme. As the number of servers is increased, the influence of distance is reduced because smaller blocks are used which yields shorter downloading times.

D. Success rate of the specified number of parallel downloading

Figure 8 shows the success rate versus the specified number of parallel downloads. Here success is defined as the the setting of $'k'$ lightpaths when the user requested parallel downloading from $'k'$ servers. The success rate is different from the blocking probability. The blocking probability is calculated as follows. $p(k) + p(k-1) \cdots + p(1)$,

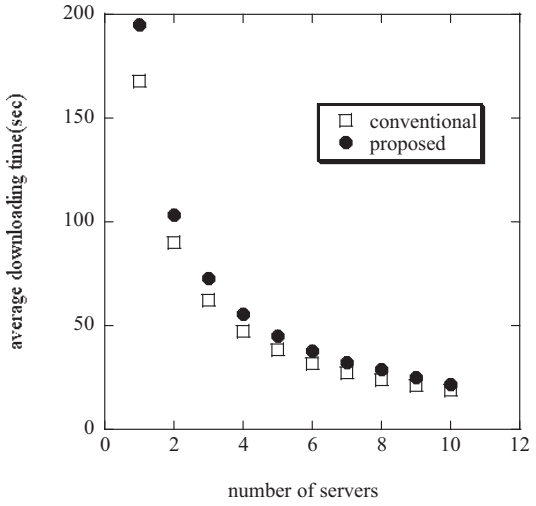


Fig. 7. The average downloading time characteristics of conventional and proposed scheme(Request rate = 0.04/sec)

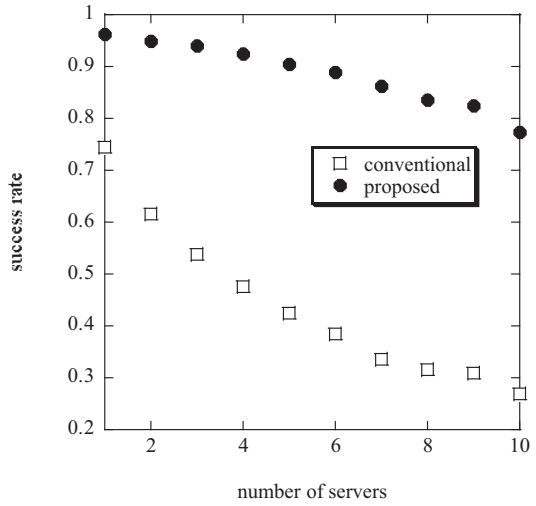


Fig. 8. Data parallel downloading success rate for conventional and proposed scheme(Request rate = 0.04/sec)

where $p(k)$ is the probability of failing given the number $'k'$ of servers. When the number of servers increases, many lightpaths exist in the network and therefore bottleneck links are increased. If many clients execute parallel downloading, the number of lightpaths increases in the network and blocking probability becomes high. As shown in Fig.8, the success rate falls as the number of servers increases. However, the success rate of the proposed scheme is over 0.7 in the case of 10 servers. This means that the proposed scheme can suppress the bottleneck links, and that efficient parallel downloading is possible.

V. Conclusion

In this paper, we proposed a novel server selection method for parallel downloading in the lambda grid network which considers both the available wavelength resources of each link and direct link costs; a weighting function is used to assigned the relative importance of one to the other. This method can suppress bottleneck links and enhance network resource utilization efficiency. Computer simulation results that showed the proposed scheme can achieved lower blocking probability and more servers can be used for downloading. They also showed that the proposed scheme can reduce the blocking probability by a factor of at least 10 and increase the success rate in offering the user-specified number of parallel downloading servers by about 30 percent.

Acknowledgment

This work was supported by Global COE Program "High-Level Global Cooperation for Leading-Edge Platform on Access Spaces (C12)" and by the Japan Society for the Promotion of Science's (JSPS) Grant-in-aid for Scientific Research(C) (19500063).

References

[1] I. Foster and C. Kesselman, "The grid: blueprint for a new computing infrastructure", *Morgan Kufmann*, Nov. 1998.

[2] K. Ranganathan and I. Foster, "Decoupling Computation and Data Scheduling in Distributed Data-Intensive Applications," *IEEE HPDC* 2002, pp. 352-358, Edinburgh, Scotland, July 2002.

[3] Simeonidou D., Nejabati R., Zervas G., Klonidis D., Tzanakaki A., and O'Mahony M. J., "Dynamic optical-network architectures and technologies for existing and emerging grid services,' *J. Lightw. Technol.*, vol.23, no.10, pp.3347-3357, Oct. 2005.

[4] S. Figueira, S. Naiksatam, H. Cohen, D. Cutrell, P. Daspit, D. Gutierrez, D. B. Hoang, T. Lavian, J. Mambretti, S. Merrill, and F. Travostino, "DWDM-RAM: Enabling Grid Services with Dynamic Optical Networks," *IEEE CCGrid* 2004, pp.707-714, Illinois, Chicago, Apr. 2004

[5] W. Allcock, J. Bester, J. Bresnahan, A. Chervenak, L. Liming, and S. Tuecke, "GridFTP: Protocol extensions to FTP for the Grid," *Internet Draft*, Mar. 2001.

[6] P. Rodriguez, A. Kirpal, E. W. Biersack, "Parallel-access for Mirror Sites in the Internet", in *Proc. of IEEE INFOCOM* 2000, pp.864-873, Mar 2000.

[7] P.Rodriguez and W.Biersack, "Dynamic parallel access to replicated content in the Internet", IEEE/ACM Trans. Netw., vol.10, no.4, pp. 455-465, Aug. 2002.

[8] J.Funasaka, N.Nakawaki, K.Ishida, and K.Amano, "A parallel downloading method to utilize variable bandwidth", IEICE Trans. Commun., vol.E86-B, no.10, pp. 2874-2881, Oct. 2003.

[9] A. Zeitoun, H. Jamjoom, M. El-Gendy, "Scalable Parallel-Access for Mirrored Servers", in *Proc. of The 20th IASTED Intl. Conf. on Applied Informatics*, 2002.

[10] E. Karasan and E. Ayanoglu, "Effects of wavelength routing and selection algorithms on wavelength conversion gain in WDM optical networks", *IEEE / ACM Transaction on Networking*, vol.6, pp.186-196, Apr. 1998.

[11] E. W. Dijkstra, "A note on two problems in connection with graphs," Numerische Mathmatik, vol. 1, pp.269-271, 1959.