

An Analysis of Selection of Mirror Sites for Parallel Processing

Pallavi Vishwakarma¹, Neha Khare²

^{1,2} Dept of Computer Science

^{1,2} Takshshila Institute of Engineering & Technology, Jabalpur MP, India

Abstract- With the recent development of internet, we are able to retrieve documents from multiple server sites, like the mirror sites, to increase the downloading speed, make better use of available network bandwidth and parallel processing speed of servers. Recent work by Rodriguez, Kirpal, and Biersack [1] studied how to use the existing HTTP protocol for retrieving documents from mirror sites in parallel to reduce the download time and to improve the reliability. The proposed approach utilizes to retrieve specific data in a mirror server site, which requires no changes on existing server and client settings. In this Work, we present a mathematical model that describes the problem of parallel download from multiple mirror sites. Based on the model, we present algorithms for selecting the best subset of mirror sites for parallel download. The proposed algorithm is implemented.

Keywords- Server selection, Parallel download, HTTP, Protocol, mirror site, HTTP, FTP, Meta data, NSGA-II, AMOSA.

I. INTRODUCTION

A mirror website is a site that is a reproduction of an officially existing webpage and is put under an alternate Universal Resource Locator (URL) and the substance of their separate servers ought to be synchronized with their lord servers. The normal favorable circumstances of reflecting are to permit quicker downloads for clients at a chaptericular land area, to advance flexibility of data, to adjust stack, to balance a sudden brief increment in rush hour gridlock and in this way enhance the accessibility of unique website [5]. Regardless of the favorable circumstances, reflecting has a few disadvantages like, extra cost to purchase servers, additional time is required to work them and irregularity when at least one mirrors are not synchronized with ace server (Main server). Mirror website determination is the issue of picking the best mirror webpage from an arrangement of accessible mirror locales for a web customer in such a design, to the point that specific parameters, for example, sit likelihood of the server, voyaging time, holding up time, bundle misfortune proportion and so on is improved. A package repository is usually an HTTP or FTP server from which clients can obtain packages and package Meta data [7]. Meta data is a set of data

that describes and gives information about other data. The package metadata is usually just a copy of the embedded package metadata for all packages on the repository. Package managers download the package metadata from a repository so that they know which packages are available from that repository.



Fig-1 Application of Meta data

This also provides the package manager with dependency information needed to perform dependency resolution. To facilitate convenient downloading of package metadata, most repositories store all of the package metadata in a small number of compressed files. In addition to the package metadata, almost all repositories have a root metadata file [9].

II. MIRROR

It is common for a distribution to have more than one server from which users can download packages and package metadata. There is usually a main repository for a distribution whose contents are copied by many separate mirrors. A mirror typically contains exactly the same content as the main repository and is updated via “rsync” or a similar tool. A mirror differs from a main repository in that a mirror is not intended to have packages directly added to it or removed from it by its administrators. Packages are added or removed only on the main repository and the mirrors later obtain the changes when copying the main repository [7]. A mirror can be public (available for anyone to use) or private (restricted to a specific organization). A mirror may also be endorsed by a

distribution for public use, typically when that the distribution is in contact with the mirror maintainers. This type of mirror is called an official mirror (the terminology used outside of this document varies by the distribution). Official mirrors are by definition public because the distribution is endorsing their use to the public. It should be noted that some distributions do not use official mirrors hosted by outside organizations. One example is tiny distributions that can support all of their clients by a small number of repositories that the distribution directly controls. Another example is a distribution that requires users to pay. These costs often are used to support a set of internally maintained mirrors for the distribution. Alternatively, a distribution may allow or require each organization using the distribution to set up their own private mirrors for the organization’s own use [12].

III. MIRROR SITE

A mirror site is a website or set of files on a computer server that has been copied to another computer server so that the site or files are available from more than one place. A mirror site has its own URL, but is otherwise identical to the principal site. Load-balancing devices allow high-volume sites to scale easily, dividing the work between multiple mirror sites. A mirror site is usually updated frequently to ensure it reflects the contents of the original site.



Fig-2 Mirror Site Explanation

Obtaining a Mirror

To evaluate the feasibility of controlling mirrors of popular distributions, a fictitious company (Lockdown Hosting) with its own domain, website and fictitious administrator (Jeremy Martin) were used as the organization maintaining the mirrors. A server with a monthly bandwidth quota of 1500GB was leased for \$200/month through the Planet (www.theplanet.com). Setting up a public mirror for each distribution involved acquiring the packages and metadata from an existing mirror and then notifying the distribution maintainers that the mirror was online and available for public usage [8]. The distributions varied in terms of the degree of automation in the public mirror application and approval process as well as whether newly listed mirrors have traffic immediately and automatically directed to them.

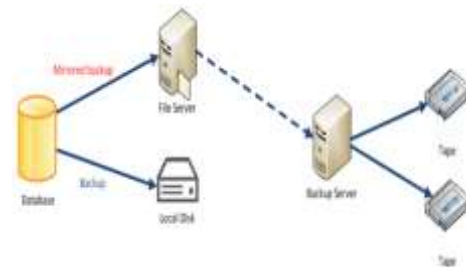


Fig-3 Obtaining a Mirror

IV. WORK TO BE DONE

Yuehao Yan and Mengru Ma et al [2018], Load Balancing assumes a vital job in lessening system misfortunes and improving the influence supply limit of the circulation organize. In this paper, the information structure of the rearranged model of the circulation arrange is improved, and the data portrayal of the hub is improved to the data depiction of the edge to lessen the information extra room. Going for the issue that the heap offsetting with the tie-switch as the center is anything but difficult to fall into the interminable circle, the limitation fulfillment issue show for the heap adjusted dissemination arrange recreation is right off the bat utilized in this paper. The limitations of system recreation with burden adjusting as the objective is built up and a heap adjusting file reasonable for the imperative fulfillment issue is proposed. The issue of dissemination organize remaking is communicated as the scientific issue with the base burden adjusting record as the goal work. The model demonstrates the possibility of the calculation utilized [1].

Table 1: Results for Balancing Rate given in Base Paper

Parameter	Disconnected Switch	Load Balancing Rate
Before Apply the Method	4, 11, 16, 19, 24, 32, 39, 47	3.3889
After Apply the Method	5, 12, 15, 24, 25, 32, 39, 46	1.4071

Munawwar Ali Abbas and Ibrahim Hussain et al [2018], Results are plotted to break down the connection of entropy age for some touchy parameters. It is finished up from these outcomes that there is a critical solid negative relationship exists among M and its entropy [3].

Chung-Ming Chen and Soo-Young Lee et al [1995], two sorts of allotments might be considered for recreated &a, in particular, total and disperse, for a 2D work. On the off chance that all PES in each gathering, spoke to in a similar

example in Fig. 2 structure a consistent locale, it is known as the total sort distribution. On the off chance that each gathering is additionally separated into subgroups of a similar size and the comparing subgroups of all gatherings structure a bunch. It is known as the disperse type designation [28].

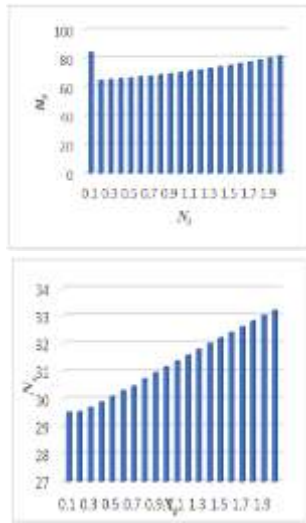


Fig-4 Entropy Calculation

V. PROPOSED WORK

F. Hrži, V. Jansky, D. Sušanj, G. Gulan, I. Kožar and D. Ž. Jerievi et al [2018], the information entropy H, is defined for N discrete states, each state i with its own probability P_i [7].

Information entropy

Let N= Total number of Mirror Web Site

Then entropy is maximum; if all the nodes have equal load.

$$H_{max} = -\sum_{k=1}^N \frac{1}{N} \log_2 \frac{1}{N}$$

$H_{max} = 1$.

Let M= Total number of Mirror Web Site

Then entropy is maximum; if all the nodes have equal load.

$$H_{max} = -\sum_{k=1}^M \frac{1}{M} \log_2 \frac{1}{M}$$

Selection of Mirror web sites / Cluster

For example if we are tossing a coin, then there may be 2 outcomes Head & Tail; it means $P=1/2$ and $Q=1/2$. Now entropy for Binary Numbers [H & T] is-

$$H_{max} = -[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}]$$

$$H_{max} = -\log_2 [2].$$

$$H_{max} = 1 \text{ Outcome / Toss}$$

VI. CONCLUSION

This shows that Average or Entropy is maximum if all the probabilities are equal. It means to increase the Speed of Downloading, Selection of Mirror web sites should be such that all the Servers have handled equal traffic load. This is the condition of balancing the Load. If balancing is used then all the Servers share equal amount of downloading Access. Therefore, the Accessing Probabilities of all the Servers are almost equal. Now the propagation will be Congestion free because it can select any of the Mirror web sites.

We have presented measurements of the performance of replicated web servers which have been selected for propagation. We have found that performance observed for a given web site can vary widely from mirror to mirror. However, the set of servers that a client must visit to achieve good performance is fairly small. Once a client has found a good server, the performance gets better.

Yuehao Yan and Mengru Ma et al [2018] says that Load Balancing assumes a vital job in lessening system misfortunes and improving the influence supply limit of the circulation organize. In this paper they found out the results for Balancing Rate [1].

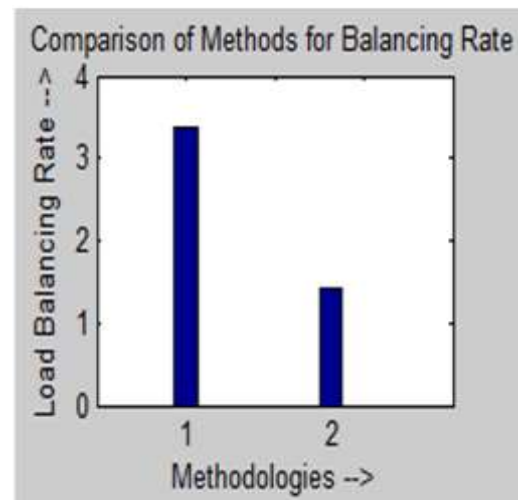


Fig-5 Comparison of Methods for Balancing Rate

REFERENCES

- [1] Yuehao Yan and Mengru Ma, “Load Balancing Distribution Network Reconfiguration Based on Constraint Satisfaction Problem Model”, 2018 China International Conference on Electricity Distribution, Tianjin, 17-19 Sep. 2018.
- [2] Simranjit Kaur and Tejinder Sharma, “Efficient Load Balancing using Improved Central Load Balancing Technique”, Proceedings of the Second International

- Conference on Inventive Systems and Control (ICISC) 2018.
- [3] Munawwar Ali Abbas and Ibrahim Hussain, “Correlation of Entropy Generation of Peristaltic Flow of Nanofluid Under the Influence of MHD”, International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018.
- [4] Tao Huang, Qiufeng Cao, Jia Liu, Dapeng Gong, Shifeng Li, Zhonghai Yang and Bin Li, “A multistage depressed collectors design tool for traveling wave tubes based on non-dominated sorting genetic algorithm II”, IEEE International Vacuum Electronics Conference (IVEC), 2018.
- [5] Zhonghu Xu, Kai Xing, Xuefeng Liu and Jiannong Cao, “Backward Thinking of Routing with High Uncertainties: Causal Entropy based Routing in Multi-agent Networks”, IEEE Symposium on Privacy-Aware Computing, 2018.
- [6] Sajib K. Biswas, Rishi Jagdev and Pranab K. Muhuri, “Energy Efficient Scheduling in Multiprocessor Systems using Archived Multi-objective Simulated Annealing”, International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018.
- [7] F. Hrži, V. Jansky, D. Sušan, G. Gulan, I. Kožar and D. Ž. Jeri_evi, “Information Entropy Measures and Clustering Improve Edge Detection in Medical X-Ray Images”, MIPRO Opatija Croatia, May 21-25, 2018.
- [8] Nupur Giri, Rahul Ravindran, Riya Suchdev, and Yash Tanna, “A Mirror Neuron based Neural Network for Website Classification”, International Conference on Computers, Communications, and Systems, 2015.
- [9] Naif Alasmari, “Optimising Cloud-Based Hadoop 2.x Applications”, IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion), 2018.
- [10] M. Rajiullah, A. C. Mohideen, F. Weinrank, R. Secchi, G. Fairhurst and A. Brunstrom, “Understanding Multistreaming for Web Traffic: An Experimental Study”, ISBN 9783-901882944, IFIP, 2017.
- [11] Pradipsinh K. Chavda and Prof. Jitendra S. Dhobi, “Web users Browsing Behavior Prediction by Implementing Support Vector Machines in MapReduce using Cloud Based Hadoop”, 5th Nirma University International Conference on Engineering (NUiCONE), 2015.
- [12] Dweepna Garg and Khushboo Trivedi, “Fuzzy K-mean Clustering in MapReduce on Cloud Based Hadoop”, IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2014.
- [13] Hyeokju Lee, Myoungjin Kim, Joon Her, and Hanku Lee, “Implementation of MapReduce-based Image Conversion Module in Cloud Computing Environment”, IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion), 2012.