



HAL
open science

Balancing Global and Local Fairness Allocation Model in Heterogeneous Data Center

Bingyu Zhou, Guangjun Qin, Limin Xiao, Zhisheng Huo, Bing Wei, Jiulong
Chang, Nan Zhou, Li Ruan, Haitao Wang, Zipeng Wei

► **To cite this version:**

Bingyu Zhou, Guangjun Qin, Limin Xiao, Zhisheng Huo, Bing Wei, et al.. Balancing Global and Local Fairness Allocation Model in Heterogeneous Data Center. 14th IFIP International Conference on Network and Parallel Computing (NPC), Oct 2017, Hefei, China. pp.136-139, 10.1007/978-3-319-68210-5_15 . hal-01705453

HAL Id: hal-01705453

<https://inria.hal.science/hal-01705453>

Submitted on 9 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Balancing Global and Local Fairness Allocation Model in Heterogeneous Data center

Bingyu Zhou^{1,2}, Guangjun Qin^{1,2}, Limin Xiao^{1,2}, Zhisheng Huo^{1,2}, Bing Wei^{1,2},
Jiulong Chang^{1,2}, Nan Zhou^{1,2}, Li Ruan^{1,2}, Haitao Wang³, Zipeng Wei³

¹ State Key Laboratory of Software Development Environment, Beihang University,
Beijing 100191, China

² School of Computer Science and Engineering, Beihang University, Beijing 100191, China

³ Space Star Technology Co., Ltd, Beijing, 100086
{zhouby, xiaolm}@buaa.edu.cn

Abstract. Aiming at the problem that the bandwidth allocation method in heterogeneous data center can't take into account the global and local fairness, we propose an allocation method called GLA that satisfies fair attributes and can improve the utilization of system resources by nearly 20% when the number of clients are more than six contrasting with the existing methods. The algorithm is also portable and scalable that can adapt to most of systems.

1 Introduction

The existing methods of bandwidth allocation[6,8] in the heterogeneous data center such as DRF, DRFH(e.g., [1-3]), BAA[4,5] and UAF[7] have a poor performance for that the utilization only can up to 70% in the best situation and the usage of resource is unbalanced. Through contrasting the distribution result among above methods, we found the situation that the allocation in the entire data center meet the fairness but on each server is unfair or the global allocation shows unfairness and the total throughput is at a low level for applying fair allocation algorithm on each server. These methods are restricted to fair policies and the efficiency of global and local, so we must balance the allocation of global and local to enhance the utilization rate. The allocation model called *Balancing Global and Local Fairness Allocation(GLA)* we proposed in the paper can solve this problem. It includes global allocation strategy and local allocation strategy., meanwhile setting the global parameters to coordinate the consistency of the two strategies.

In the second part, we introduce the GLA model and implementation architecture. The performance evaluation results will be described in Section 3. In the end, we will conclude our work.

2 GLA Allocation Model and Scheduling Architecture

The functions of GLA model are shown as follows:

$$\text{Maximize } \sum_{i \in D} \left(\sum_{l=1}^k (x_{ild} \times \frac{1}{m_{il}}) \right) + \sum_{j \in S} \left(\sum_{l=1}^k (x_{jls} \times \frac{1}{h_{jl}}) \right) \quad (3.1)$$

$$\begin{cases} \sum_{l=1}^k \varpi_{il} = 1 \\ \sum_{l=1}^k (\varpi_{il} \times h_{il}) = 1 \\ i \in D \text{ (or } i \in S) \end{cases} \quad (3.2)$$

$$\begin{cases} x_{ild} \times \left(\sum_{i \in D} + \omega \times \frac{C_{sl}}{C_{dl}} \times \sum_{j \in S} \frac{m_{jl}}{h_{jl}} \right) \leq C_{dl} \\ x_{ild} \times \left(\sum_{i \in D} \frac{h_{il}}{m_{il}} + \omega \times \frac{C_{sl}}{C_{dl}} \times \sum_{j \in S} \right) \leq C_{sl} \\ \forall l \in V, i \in D_l, j \in S_l \end{cases} \quad (3.3)$$

$$\begin{cases} x_{ild} \geq \frac{C_{dl}}{n}, x_{jls} \geq \frac{C_{sl}}{n} \\ x_{ild} \geq x_{jls} \times \frac{m_{jl}}{h_{jl}}, x_{jls} \geq x_{ild} \times \frac{h_{il}}{m_{il}} \\ \forall i \in D_l, j \in S_l \end{cases} \quad (3.4)$$

$$\begin{cases} \sum_{l=1}^k x_{ild} \geq \sum_{l=1}^k x_{jls} \times \frac{m_j}{h_j} \\ \sum_{l=1}^k x_{jls} \geq \sum_{l=1}^k x_{ild} \times \frac{h_i}{m_i} \\ \forall i \in D, j \in S \end{cases} \quad (3.5)$$

$$\begin{cases} \forall i, j \in D, \text{ if } \sum_{l=1}^k x_{ild} \times \frac{h_i}{m_i} \geq \sum_{l=1}^k x_{jld} \times \frac{h_j}{m_j}, \text{ then } \sum_{l=1}^k x_{ild} \leq \sum_{l=1}^k x_{jld} \\ \forall i, j \in S, \text{ if } \sum_{l=1}^k x_{ils} \times \frac{m_i}{h_i} \geq \sum_{l=1}^k x_{jls} \times \frac{m_j}{h_j}, \text{ then } \sum_{l=1}^k x_{ils} \leq \sum_{l=1}^k x_{jls} \end{cases} \quad (3.6)$$

The scheduling architecture of the heterogeneous data center has two layers(Fig.1), GLA is deployed in the second layer. The first layer scheduler assigns the requests of clients to the corresponding server in terms of the original weight, the second tier calculate the bandwidth allocated to each client based on the GLA algorithm. Different from other algorithm, the global parameters produced by GLA will be send to the first layer to assist the next allocation. The algorithm obeys the principle of conservation. The throughput of allocation subjects to the limitations that are presented into progressive from local to the whole.

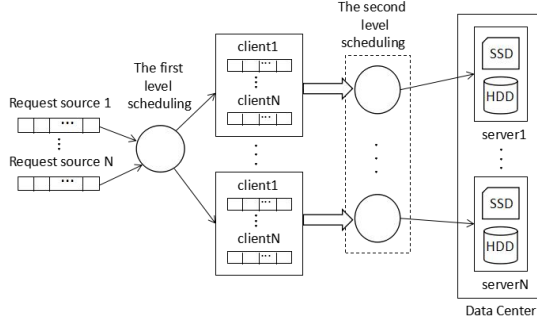


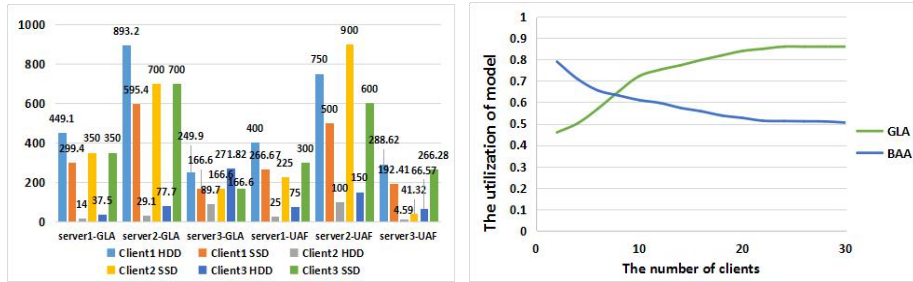
Fig. 1. The scheduling architecture of heterogeneous data center

3 Performance Evaluation

We use a parallel simulator called PFSSim and a Linux cluster test I/O performance of GLA. The cluster are configured as AMD Quad-core processor, 8GB RAM, 1TB Seagate 7200RPM hard drive, on which we configure three servers: $server_1$ (HDD: 500IOPS, SSD: 1000IOPS), $server_2$ (HDD: 1000IOPS, SSD: 2000IOPS), $server_3$ (HDD: 3000IOPS, SSD: 500IOPS). The workloads we used are from Umass Trace and Microsoft Exchange server. The test will be compared between BAA and UAF.

Fig.2(a) show the allocation of UAF and GLA on each server for PFSSim. It's obviously that the local allocation of UAF does not obey fair properties. For GLA, the hit ratios on each server for clients are $h_{11} = h_{12} = h_{13} = 0.4$, $h_{21} = h_{22} = 0.96$, $h_{23} = 0.65$, $h_{31} = h_{32} = 0.9$, $h_{33} = 0.38$, the global hit ratios are $\sum_{l=1}^3 (h_{1l} \times \varpi_{1l}) = 0.4 = h_1$, $\sum_{l=1}^3 (h_{2l} \times \varpi_{2l}) = 0.9 = h_2$, $\sum_{l=1}^3 (h_{3l} \times \varpi_{3l}) = 0.8 = h_3$. We also get $A_{1,hdd} \geq \{A_{2,hdd}, A_{3,hdd}\}$, $A_{2,ssd} \geq A_{1,ssd}$, hence the GLA is a fair strategy. Fig2(b) shows the throughput will increase with the number of clients arise, GLA has a best performance in the large-clients cases.

Fig.3 express the throughput of GLA are 20% high than that of UAF on Linux cluster, GLA model has practical significance.



(a) The allocation on each server

(b) The throughput variation of BAA and GLA

Fig. 2. The comparison of utilization between GLA, BAA and UAF

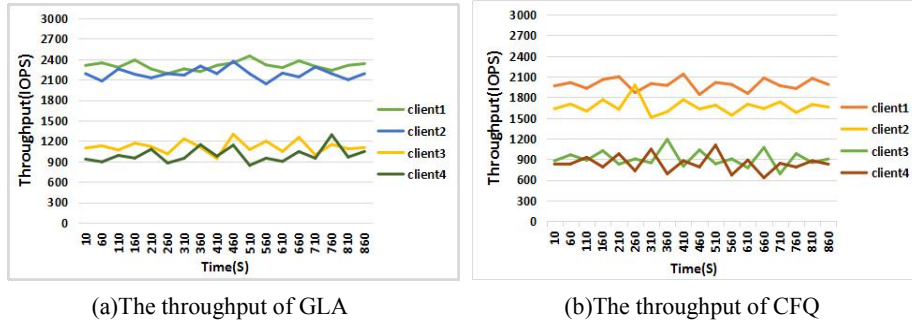


Fig. 3. The throughput GLA and CFQ for four clients

4 Conclusion

In this paper, we focus on the problem that the existing allocation methods can't balance global and local fairness, and propose a bandwidth allocation algorithm called GLA. Through experiments, we exhibit that GLA is fairness and improve the throughput of heterogeneous data center for 19.8% when the number of clients are more than six. GLA is more suitable for large-scale clients.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grant No. 61370059 and No. 61232009, Beijing Natural Science Foundation under Grant No. 4152030, the fund of the State Key laboratory of Software Development Environment under Grant No. SKLSDE-2016ZX-13, the State Administration of Science Technology and Industry for National Defense, the major projects of high resolution earth observation system under Grant No. Y20A-E03.

References

1. A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica.: Dominant resource fairness: Fair allocation of multiple resource types. In Proc.8th USENIX NSDI, Vol. 11. Berkeley, CA, USA(2011)
2. D. Dolev, D. G. Feitelson, J. Y. Halpern, R. Kupferman, and N. Linial.: No justified complaints: On fair sharing of multiple resources. In ITCS '13. New York, NY, USA(2012)
3. D. C. Parkes, A. D. Procaccia, and N. Shah.: Beyond dominant resource fairness: Extensions, limitations, and indivisibilities. In EC '12. New York, NY, USA(2012)
4. H. Wang and P. Varman.: Balancing fairness and efficiency in tiered storage systems with bottleneck-aware allocation. In Proc. 12th Usenix FAST. Santa Clara, CA(2014)
5. H. Wang and P. Varman.: Brief Announcement: Fairness-Efficiency Tradeoffs in Tiered Storage Allocation. In SPAA '14. Prague, Czech Republic(2014)
6. A. Gutman and N. Nisan.: Fair allocation without trade. In Proc.11th AAMAS. Valencia, Spain(2012)
7. Zhisheng Huo, Limin Xiao. (ed.): Hybrid Storage Throughput Allocation Among Multiple Clients in Heterogeneous Data Center. In HPCC '15. New York, USA(2015)
8. Yusuke, Ito., Hiroyuki, Koga.: A Bandwidth Reallocation Scheme to Improve Fairness and Link Utilization in Data Center Networks. In Perform'16. Sydney, New South Wales(2016)