On the Workload Deployment, Resource Utilization and Operational Cost of Fast Optical Switch based Rack-scale Disaggregated Data Center Network

Xiaotao Guo, Fulong Yan, George Exarchakos, Xuwei Xue, Bitao Pan, Nicola Calabretta
Institute of Photonic Integration, Eindhoven University of Technology, Eindhoven, the Netherlands
x.guo@tue.nl

Abstract: We investigate operational performance of a novel rack-scale disaggregated network. Results show that the disaggregated network achieves 30.6% higher workloads acceptance rate, 12.9% higher resource utilization, and 33% more power saving compared with the server-centric.

1. Introduction

The rapidly increasing of big data applications ranging from computing intensive (e.g. Hadoop and Spark) to network intensive (e.g. web search and media streaming) have diversified IT resource (CPU, memory, storage and network) requirements. To serve these emerging applications, flexible resource provision is demanded in data center (DC) networks. However, the fixed IT resource embedded in the server results in low resource utilization, high upgrade cost, and power waste in current DC [1]. As a promising solution, disaggregated networks like "dRedBox" [2] releases all the IT resource from the server in order to provision fine granularity IT resource. Nonetheless, there are several challenges to implement the disaggregated network. Powerful network interconnection with low latency, high bandwidth and high scalability should be developed to replace the high speed bus in the server.

To address the network interconnection, low latency and scalability in the disaggregated DC network, recently we proposed a novel rack-scale disaggregated architecture based on fast optical switch (FOS) [3]. The proposed rack-scale disaggregated architecture based on FOS outperforms the server-centric DC network, but at higher initial capital cost. To fully evaluate the overall costs, the operational performance of disaggregated network should be also assessed in terms of workloads acceptance rate, IT resource utilization and power consumption, which analyzes whether the initial capital costs of the disaggregated network is sustainable.

In this work, we investigate and numerically assess the operational costs of the disaggregated architecture based on FOS in terms of workload acceptance rate, resource utilization and power consumption. To compare the operational performance of disaggregated and server-centric architecture in a realistic scenario, the resource demands of various realistic applications are collected from measurement and employed in the performed simulations. Results show that with various request rates, the FOS based rack-scale disaggregated architecture can increase up to 30.6% acceptance rate compared with server-centric architecture. Moreover, rack-scale disaggregated network can achieve up to 12.9% higher resource utilization and 33% less power consumption.

2. Disaggregated network and cost-efficiency analysis

Fig. 1 shows the architecture of the FOS based rack-scale disaggregated network. All the hardware in the server are disaggregated into CPU, memory and storage kinds of resource pools. The FOS is employed for the high bandwidth and low latency connection between the CPU and the memory pools. The memory nodes and storage nodes are connected by an electrical packet switch (EPS). To compare the power consumption between the FOS based rack-

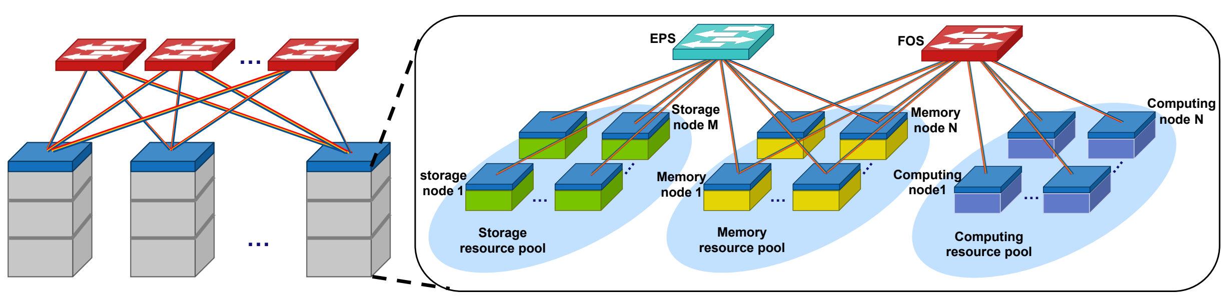

Fig. 1. Architecture of FOS based disaggregated network

scale disaggregated and the server-centric architecture, the power model is analyzed in this section. For the server-centric DC network, the power of single rack is the sum of the servers' power and the EPS (as TOR). As the main IT resource in DC, the power of server is calculated as [4]:

$$P_{server} = P_{idle} + P_{cpu} + P_{memory} + P_{disk} + P_{nic} \quad (1)$$

CPU makes a major impact on server power consumption based on its utilization:

$$P_{cpu} = P_{idle\_cpu} + (P_{\max\_cpu} - P_{idle\_cpu}) \times u_{cpu} \quad (2)$$

where P_{idle_cpu}, P_{max_cpu} and u_{cpu} denote the CPU idle power, the peak power and the utilization. Similar equation for the power consumption of the memory that can be described by the memory usage. Instead, the power consumption of the storage and the NIC remain almost the same when running the workloads [5]. Correspondingly, the power consumption of the FOS based rack-scale disaggregated architecture is calculated as:

$$P_{disaggregation} = P_{computing\_pool} + P_{memory\_pool} + P_{storage\_pool} + P_{FOS} + P_{EPS} \quad (3)$$

where $P_{computing\_pool}$, P_{memory_pool} and $P_{storage\_pool}$ denote the hardware power consumption of the resource pool effectively utilized by the running workloads. The details of FOS power model can be found in [6].

3. Simulation setup and results

In the operational performance comparison, we assume the storage and the network resource of disaggregated and server-centric architecture are both sufficient to run all the deployed workloads. Therefore, the resource demands for each workload request include CPU, memory resources and completion time. To compare the operational performance of the FOS based rack-scale disaggregated and server-centric architectures in the realistic scenario, the resource demands of various real applications has been experimentally measured. According to the application instances in the realistic DC network [7], we select in total 6 applications covering 3 categories: Hadoop, Spark, Web Search (WS), Instant Message (IM), Media Streaming (MS) and Cloud Storage. The server for measurement is configured with Intel Xeon Gold 48c processor and 128GB memory. The applications of Scientific Computing is deployed as one master and two worker, in which resource requirement of worker is measured. The rest four applications are set as one server-end and one client-end, while the resource requirements of server-end are recorded. The measured resources requirements are shown in Table 1. The statistic number of workload requests per minute is based on a Poisson distribution in the simulation, while the average request number per minute is defined as the request rate. Duration of 1000 minutes have been considered in the numerical assessment. Table 2 reports the power consumption of hardware components analyzed in section 2. The listed power are from parameter of commercial products [8, 9] and research results [6, 10-11].

Table 1. Resource requirements for realistic workloads

Resource requirements	Scientific Computing		Online Serving			Cloud Storage
	Hadoop	Spark	WS	IM	MS	
CPU(cores)	24	23	20	16	4	10
Memory(GB)	4.75	7.29	7.66	4.31	2.52	2.33
Completion time (minutes)	7	5	15	13	17	6

In the simulation, the server-centric architecture consists of 40 servers with of 48 cores CPU and 16 GB memory each. For a fair comparison, the simulated FOS based disaggregated rack architecture also consists of 40 48 cores computing nodes and 40 16 GB memory nodes. The workload requests are generated base on the requirements in Table 1. The ratio of different workload types is set to 17% Scientific Computing, 39% Web Search, 22% Instant Message, 10% Streaming Media and 12% Cloud Storage [7].

The operational performance results of the disaggregated and server centric architectures for a request rate of 20 are

Table 2. Power of hardware components

Components	Power (W)	Components	Power (W)	Components	Power(W)
Xeon 6230 20c CPU	119.6(idle) 446.4(max)	HDD Disk	6	EPS	350
16G RAM	4.6(idle) 6.14(max)	10G NIC	18	FOS	1000

reported in Fig. 2. Fig. 2(a) shows that the acceptance rate of FOS based disaggregated network is 78.1% when the server-centric architecture is 47.5%. The disaggregated network can accept all the workload requests with the probability of 0.268, while the probability of server-centric network is 0.014. When no workloads are deployed, the hardware resource could be set in the idle state and consumes little power. Fig. 2(b) shows the number of active resources. All the servers are in active state to deploy the workload requests. Instead, the disaggregated network requires in average only 34.9 active CPUs at even 30.6% more workload acceptance rate than server centric. Figure 2(c) and 2(d) summarize the total resource utilization and the power consumption, respectively. The FOS based disaggregated rack can achieve 10.9% higher resource utilization and 12.2% less power consumption compared to the server-centric rack architecture.

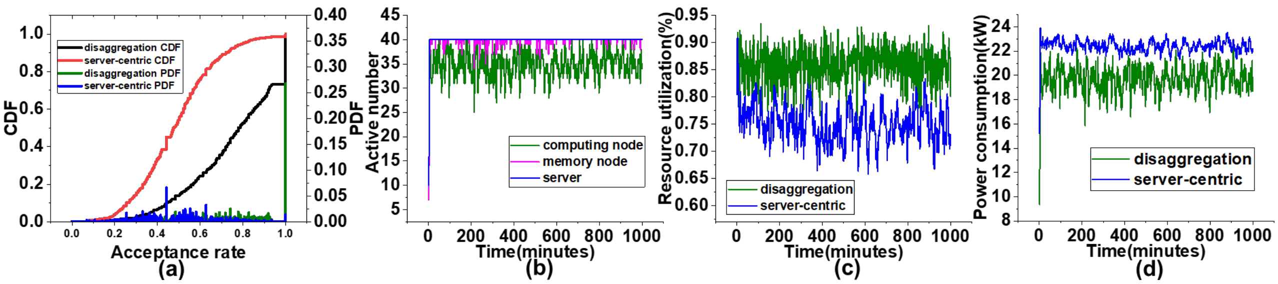

Fig. 2. (a) CDF and PDF of arriving workload acceptance rate (b) resource number in active state (c) resource utilization comparison (d) power consumption comparison

Table 3 shows the average operational performance of the FOS based disaggregated rack (FDR) and the server-centric (SC) for different average workloads request rates (10, 20 and 30). For a request rate of 10, the FDR can accept all the arriving workload, and uses 26.1% less hardware resources than the SC. The resource utilization of FDR is 12.9% more efficient than SC, and FDR can save 33% power consumption compared to SC. Increasing the average request rate to 30, similar to the SC, almost all the hardware resources in FDR are also in active state. Compared with the operational performance of FDR at request rate of 20, the workload acceptance rate decreases of 25.4%, while the resource utilization only improves of 0.5%.

Table 3. Operational performance with different request rate

Request rate	Acceptance rate (%)		Active resource			Utilization (%)		Power (kW)	
	SC	FDR	SC	CN	MN	SC	FDR	SC	FDR
10	82.1%	100%	39.5	22.7	29.2	70.4%	83.3%	21.74	14.56
20	47.5%	78.1%	39.9	34.9	39.7	75%	85.9%	22.42	19.69
30	36.2%	52.7%	40	37.2	39.9	75.1%	86.4%	22.45	20.69

4. Conclusion

We investigate and compare the operational performance of FOS based rack-scale disaggregated architecture with the server-centric architecture. To evaluate the disaggregated network in realistic scenario, resource requirements of realistic applications are measured and synthesized to generate the workload requests in the operational performance simulation. The simulation results demonstrate that FOS based rack-scale disaggregated network can accept up to 30.6% more workload requests than server-centric network with the same request rate. In addition, rack-scale disaggregated network achieves 12.9% higher resource utilization with 26.1% less active hardware resource, and reduces up to 33% power consumption compared with server-centric architecture.

5. References

[1] L. A. Barroso and U. Holzle, “The case for energy-proportional computing.” Computer, vol. 40, no. 12, pp. 33–37, Dec. 2007.
[2] Guo, Xiaotao, et al. “Performance assessment of a novel rack-scale disaggregated data center with fast optical switch.” OFC, 2019.
[3] G. Zervas, “Optically Disaggregated Data Centers with Minimal Remote Memory Latency” JOCN 10.2 (2018): A270-A285.
[4] Dayarathna M, Wen Y, Fan R. “Data center energy consumption modeling: a survey.” IEEE Commun Surv Tutorials 2016;18 (1):732–94.
[5] J. A. Aroca, et al. “A Measurement-based Analysis of the Energy Consumption of Data Center Servers.” Proceedings of the 5th international conference on Future energy systems, 63-74, 2014.
[6] Yan, Fulong, et al. "On the cost, latency, and bandwidth of LIGHTNESS data center network architecture." PS, IEEE, 2015.
[7] Thomas Barnett, Arielle Sumits, “Cisco Global Cloud Index 2015–2020,” Cisco Knowledge Network (CKN) Session, November 2016.
[8] Intel Xeon CPU specificationhttps://www.intel.com/content/www/us/en/products/processors/xeon.html
[9] Broadcom Ethernet Switches, https://www.broadcom.com/products/ethernet-connectivity/switching
[10] Makaratzis AT, “Energy modeling in cloud simulation frameworks.” Future Gener Comput Syst.2017; 79:715-725.
[11] W. Lin, “Experimental and quantitative analysis of server power model for cloud data centers,” Futur. Gener. Comput. Syst., 940–950, 2018.