Micro-Storage Services for Open Ethernet Drive

Hariharan Devarajan, Anthony Kougkas, and Xian-He Sun
Illinois Institute of Technology, Department of Computer Science, Chicago, IL
{hdevarajan, akougkas}@hawk.iit.edu, sun@iit.edu

***Index Terms*—Open Ethernet Drive, OED, Micro services, Storage, API, Micro kernels**

I. Introduction

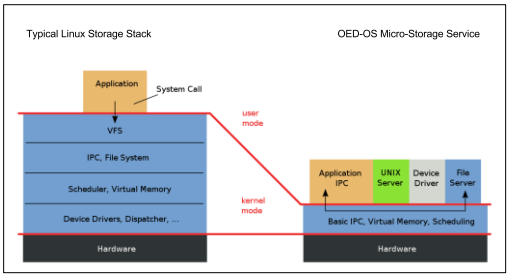
As we move towards the exascale era, data management is one of the greatest challenges which fuels the evolution of both hardware and software. ActiveStorage [1] and ActiveDisk [2] proposed taking advantage of the embedded processors on the storage servers. These enhancements in technology focus on improving performance but with a high energy cost [3]. Up to 40% of the total energy consumed in these clusters is by the storage nodes [4]. This ratio would increase due to two factors (a) the power consumption of computing resources has been getting a lot of attention resulting in more efficient utilization; and (b) data deluge (expected to increase tenfold) will also increase the relative contribution to power consumption.

One new technology that has the potential, to reduce the energy consumption in storage nodes, is the Open Ethernet Drive (OED). Its architectural characteristics enable "data-centric" storage services. The technology brings computation capabilities close to the data. Each OED can be treated as a storage node with the processor and RAM embedded close to the disk. This makes the OED an "intelligent" drive with the capability of not only storing data but also executing certain data-centric computations. In most implementations, each OED comes loaded with a full-fledged Linux OS. This enables programmers to run scientific applications and software natively. The OED technology offers a cost-effective mechanism to distribute storage services across several low-powered processors like ARM-based instead of a few more powerful server-graded CPUs (e.g., Intel Xeons or AMD Opterons). Having the capability to manage data close to its origin enables several performance optimizations through the device. Apart from the specification differences, OED technology is not just using Ethernet as a new connection interface, but it also moves the communications protocol from simple commands to read-and-write data blocks to a higher level of abstraction.

In our previous work, we had explored the capabilities of this device [6], [7] along with its performance characteristics. Our study showed that a collection of OED devices (i.e., a JBOD enclosure) is an ideal candidate for the next-generation, low-powered storage nodes offering competitive performance through higher degree of parallelism. In fact, an OED device consumes only 10% of the power required by a typical storage node to perform the same task (e.g., run a storage server). During the study we identified two major challenges: a) lack of simple and efficient software to manage OEDs, b) low-powered devices such as OED have a lot of noise from full stack of software from server nodes. These challenges demand finely tailored software to efficiently use the capabilities of the OED architecture while mitigating its weaknesses.

II. Our Approach

This study aims to solve the above-mentioned challenges by proposing a new middleware software which can be used to access the OEDs as a networked storage service. This software aims to provide a simple yet powerful API that will enable users to fully utilize the OED's computation and storage capabilities. Our solution will provide means to automatically parallelize I/O requests on top of an OED JBOD (i.e., collection of OEDs). Additionally, our library will enable offloading small data-intensive computations on these devices to minimize expensive data movements and achieve an efficient pipeline of operations. We also aim to develop a specialized micro kernel tailored for storage-only services and thus, deliver a new OED-OS software which will be lightweight and finely tuned. The final goal will be to design a lightweight virtualized storage service that runs on the OED technology and which will support a wide range of modern workloads running in existing HPC and cloud environments.



In a traditional storage system, the kernel mediates access to lower level device hardware by applications, to enforce process isolation as well as network and disk security. We envision a new system where applications entirely skip the kernel and thus, allowing most I/O operations to directly access the virtualized I/O nodes (i.e., OEDs) as the above figure shows. A system like this, will demonstrate improvements in access latency and throughput and will be an ideal platform to run modern parallel file system or object store server processes.

REFERENCES

[1] E. Riedel, G. a Gibson, and C. Faloutsos, “Active Storage for Large-Scale Data Mining and Multimedia,” *Proceedings*, no. February, pp. 62–73, 1998.

[2] A. Acharya, M. Uysal, and J. Saltz, “Active disks: programming model, algorithms and evaluation,” *SIGPLAN Not.*, vol. 33, no. 11, pp. 81–91, 1998.

[3] S. Mingay, “Green IT: The New Industry Shock Wave,” 2007.

[4] T. Dreibholz, M. Becke, and H. Adhari, “Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431,” *Tdr.Wiwi.Uni-Due.De*, vol. 109, p. 431, 2007.

[5] A. Kougkas, A. Fleck, and X. H. Sun, “Towards energy efficient data management in HPC: The open Ethernet drive approach,” in *Proceedings of PDSW-DISCS 2016: 1st Joint International Workshop on Parallel Data Storage and Data Intensive Scalable Computing Systems - Held in conjunction with SC16: The International Conference for High Performance Computing, Networking, Storage and Analysis*, 2017, pp. 43–48.

[6] H. Devarajan, A. Kougkas, and X. H. Sun, “Open Ethernet Drive Evolution of Energy-Efficient Storage Technology.” DataCloud’17, Chicago.