OdinFS: Scaling PM performance with Opportunistic Delegation

Diyu Zhou Yuchen Qian Vishal Gupta Zhifei Yang Changwoo Min Sanidhya Kashyap

Existing PM file systems cannot fully utilize the performance of PM

Persistent Memory (PM) = disk + memory

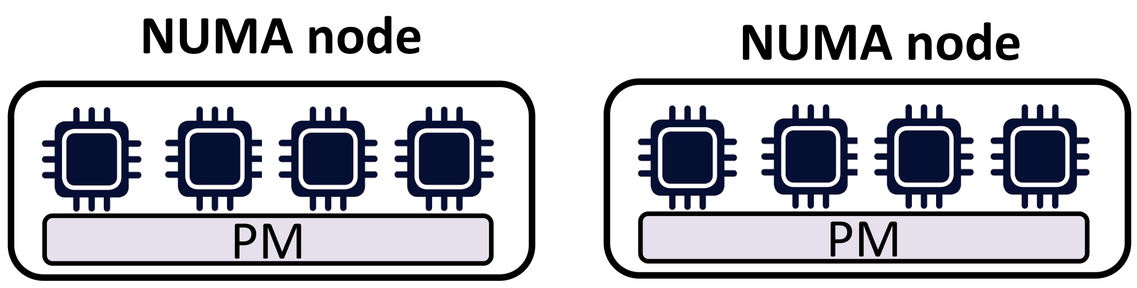
Disk

- Persist data across power cycles

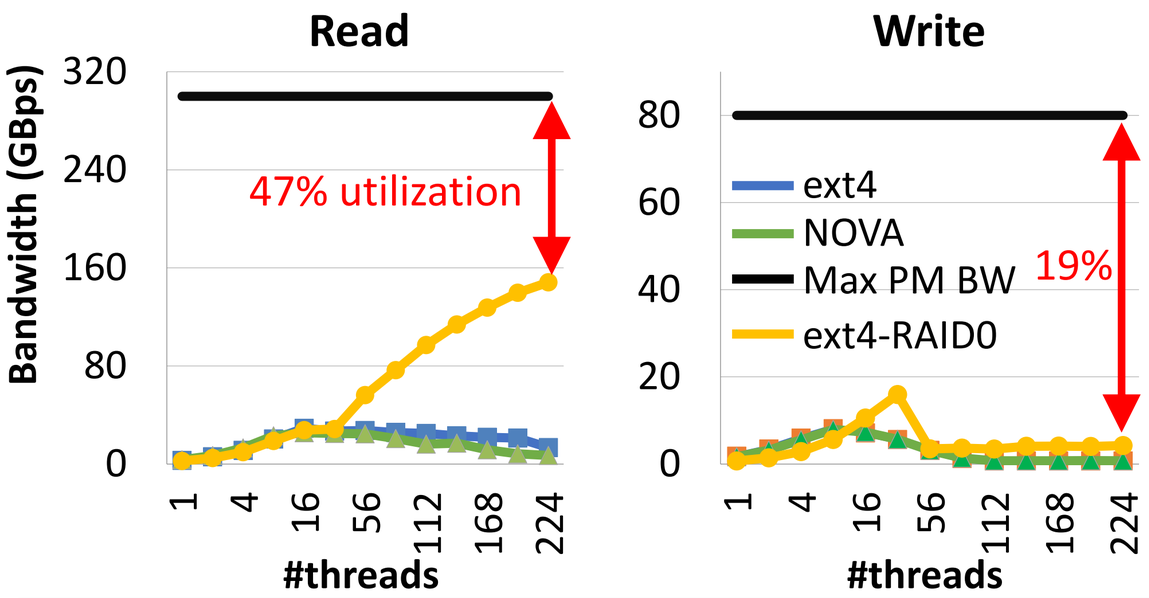
Memory

- Byte addressable
- Access latency in 100s of nanoseconds

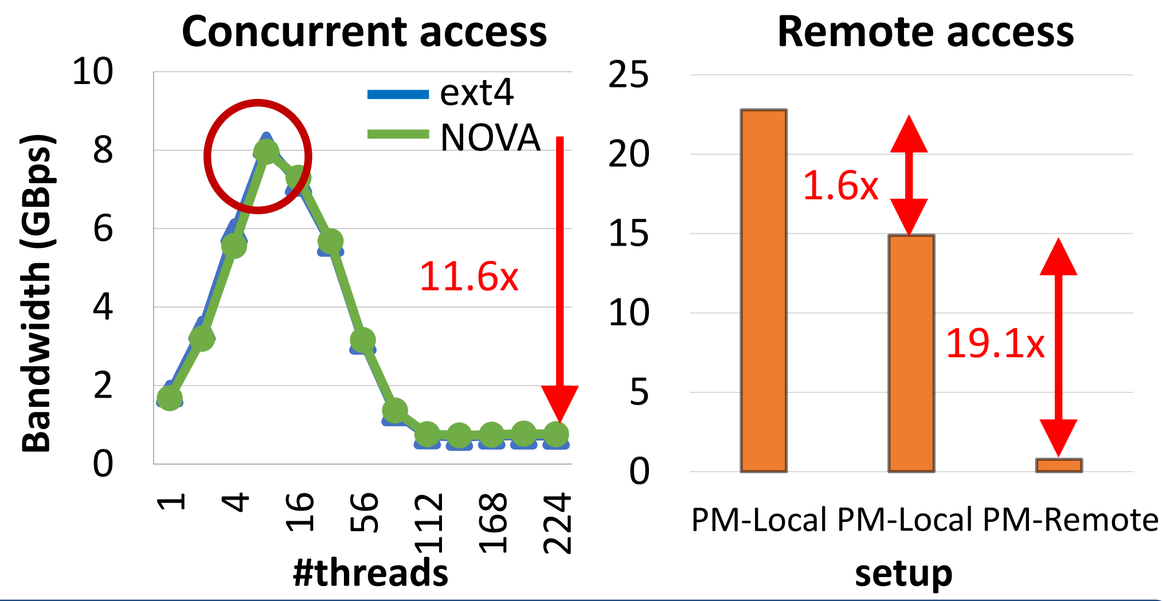
Servers have PMs on multiples NUMA nodes



Existing file system cannot utilize PM



PM performance anomalies



OdinFS: Controlled, localized, and parallel PM access with access delegation

Design overview

- **Limit concurrent access**

  Maximize PM performance on each NUMA node

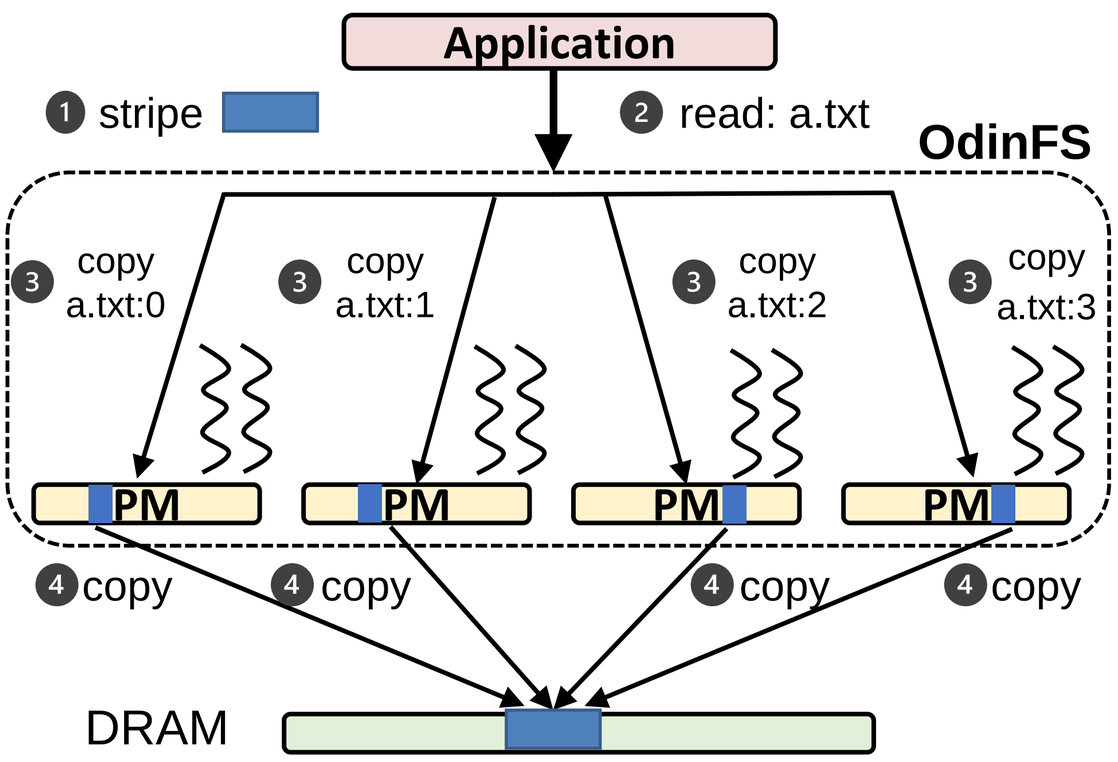
- **Localized PM access**

  Minimize PM NUMA impact

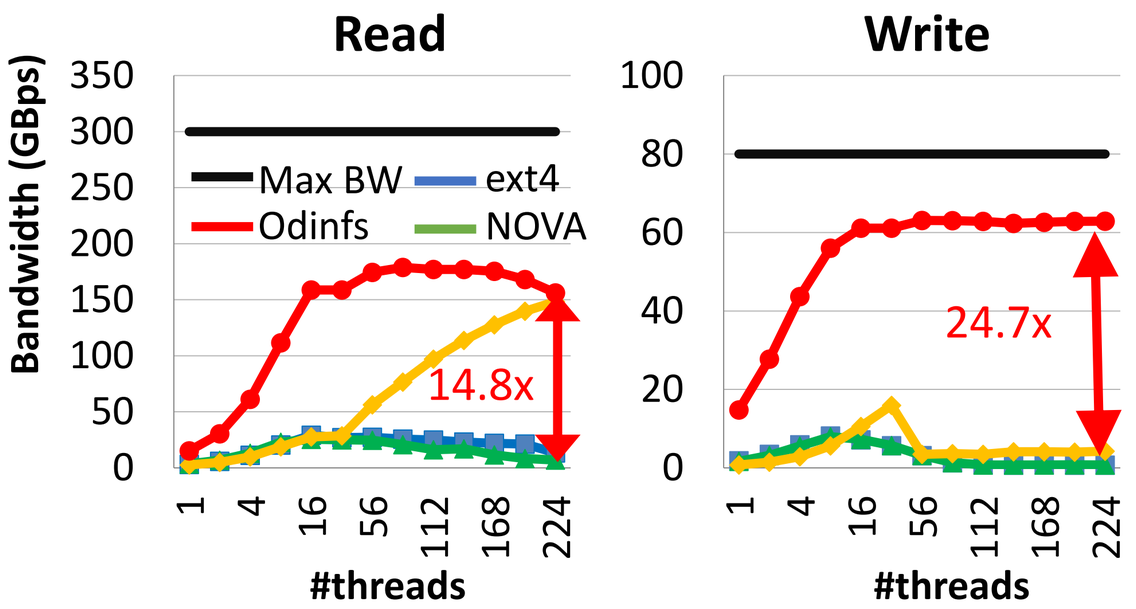
- **Efficient use of the aggregated PM bandwidth**

  Applications on single NUMA node can benefit

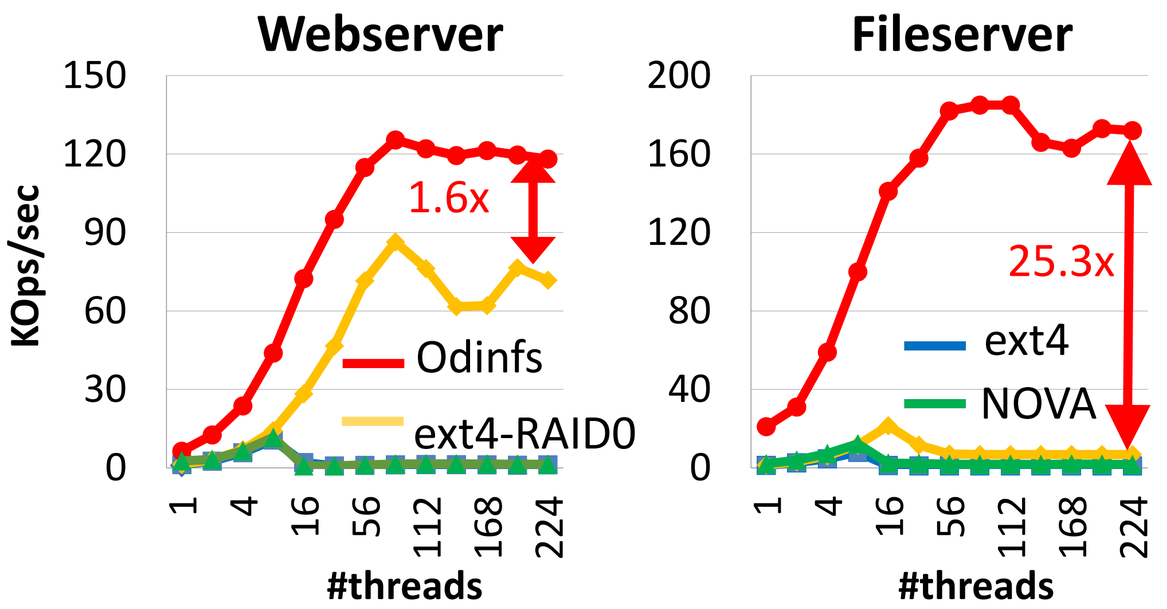
PM access with delegation



Microbenchmarks: FIO



Macrobenchmarks: Filebench



Conclusion: Decouple PM access from application threads to scale PM performance