Capturing I/O Dynamics in HPC Applications

Sagar Thapaliya[†], Kathryn Mohror[‡], Adam Moody[‡], and Purushotham Bangalore[†]
[†]University of Alabama at Birmingham, [‡]Lawrence Livermore National Laboratory

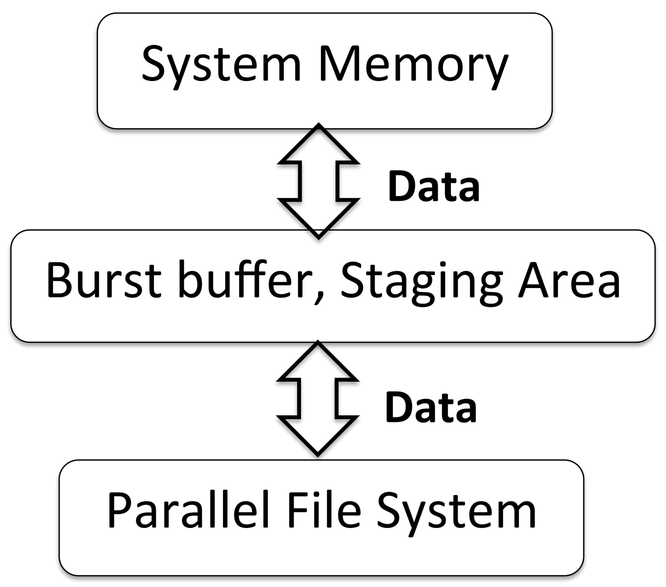
COLLEGE OF ARTS AND SCIENCES
Knowledge that will change your world

Motivation

- I/O challenge in HPC Systems
  - Limited I/O bandwidth (e.g. Parallel File System)
  - Scarce I/O resources (e.g. expensive Solid State Drives)
- Applications I/O needs
  - For checkpointing, data analysis and visualization
  - Dependent on application specific I/O behavior

Emerging Challenges

- Emerging I/O solutions
  - Multi-tiered storage
  - Co-processing
- Emerging needs
  - I/O resource management
  - Resource utilization
  - Quality of Service in I/O
  - Interference management



Data Flow in HPC Storage Hierarchy

I/O Dynamics

- Repetitive I/O Pattern
- Characterization
  - Output type
  - Interval in arrival
  - Data transfer size
  - Concurrency
- Variation
  - Applications
  - Job Scale
  - Output types
  - Processes
  - Simulation progress

Example I/O Pattern

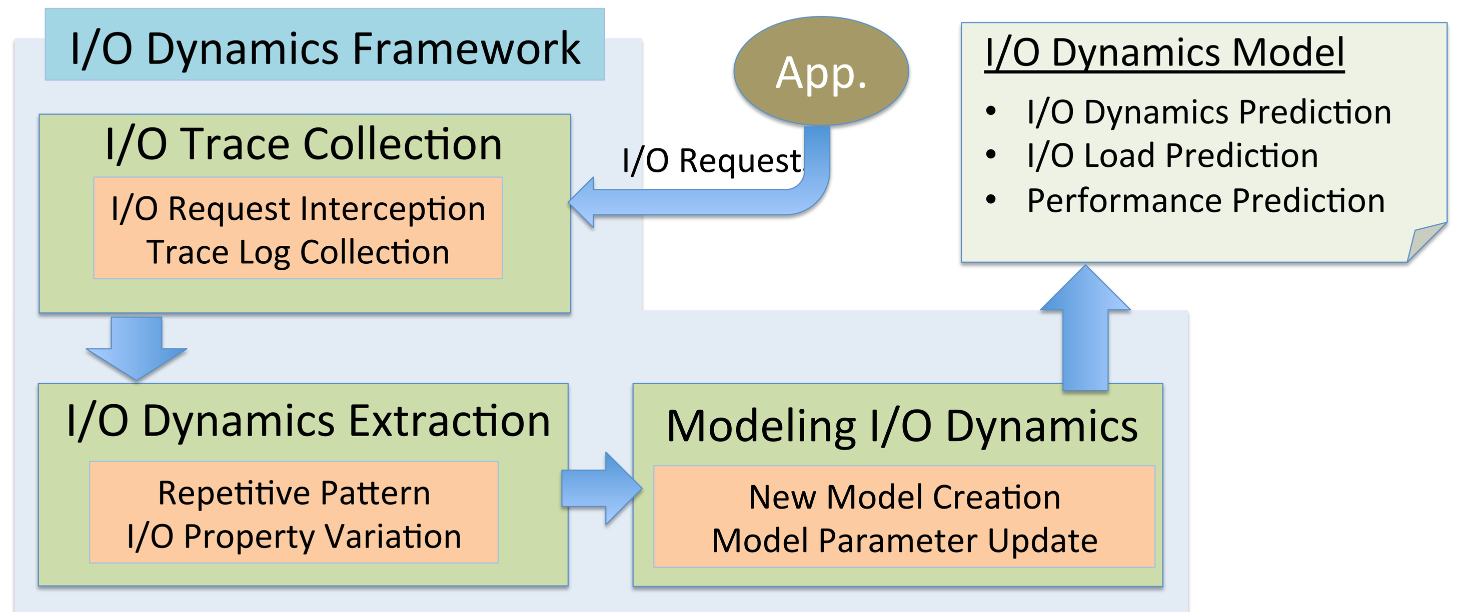
```
Start Program {
   Read Input data
   Perform Computation
   Simulation loop {
     Perform computation
     write checkpoint data // every i iterations
     write visualization data // every j iterations
     write performance log // every k iterations
   }
}
End Program
```

Why it matters?

- Application I/O performance prediction
- Storage system design and tuning
- Application specific I/O benchmarks
- Scheduling data movement
- Application aware I/O management (QoS)

I/O Dynamics Framework

- Generate I/O dynamics model for applications
- I/O trace based solution
- Offline or online I/O dynamics detection
- Online detection of I/O dynamics variation



Experiment Setup

Applications

- **ParaDiS**
  - Dislocation dynamics simulation
  - Experiment
    - Dislocation in Copper
    - Domain: 4x2x2
    - 16 procs.
  - Output
    - Checkpoint, Visualization
- **ENZO**
  - Simulation of cosmological structure formation
  - Adaptive Mesh Refinement (AMR)
  - Experiment
    - Unigrid dark matter-only cosmology simulation
    - Grid: 128 x 128 x 128
    - ?? procs.
  - Output
    - Checkpoint, Visualization

Testbed

Sierra cluster at LLNL

- 261.3 TFLOP/s machine
- Computer cluster
  - Node: Intel(R) Xeon(R) EP X5660 CPU, 12 cores
  - 24 GB RAM
  - 64 Nodes used
  - TOSS 2.0 operating system
- Storage System
  - Lustre Parallel File System
  - 480 OSTs
  - 1.8 PB
  - Peak bandwidth: 40 GB/s
- InfiniBand QDR (Qlogic) interconnect

Measurements

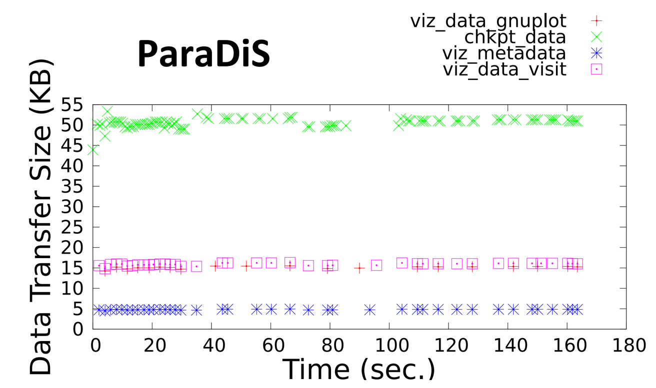
- **I/O requests**
  - Data transfer size
  - Arrival time
  - Output type
- **I/O load**
  - Total size of data requested for transfer during each time slice
  - Time slice = 1 second

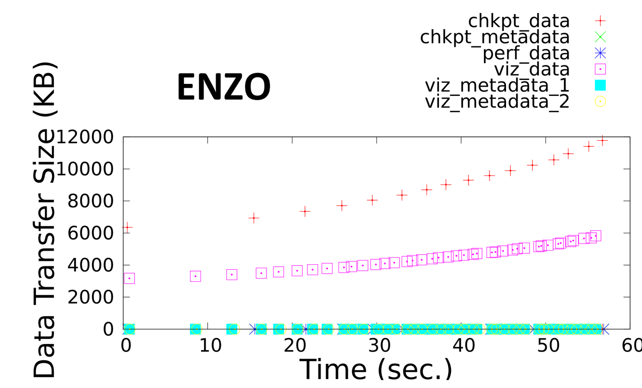
References

- P. H. Carns, R. Latham, R. B. Ross, K. Iskra, S. Lang, and K. Riley, "24/7 characterization of petascale I/O workloads," in *IEEE International Conference on Cluster Computing (Cluster '09)*, New Orleans, LA, Sep. 2009
- Ning Liu; Cope, J.; Carns, P.; Carothers, C.; Ross, R.; Grider, G.; Crume, A.; Maltzahn, C., "On the role of burst buffers in leadership-class storage systems," *Mass Storage Systems and Technologies (MSST), 2012 IEEE 28th Symposium on* , vol., no., pp.1,11, 16-20 April 2012
- Tran, N.; Reed, D.A., "Automatic ARIMA time series modeling for adaptive I/O prefetching," *Parallel and Distributed Systems, IEEE Transactions on* , vol.15, no.4, pp.362,377, April 2004

Results

Types and incidence of different output types





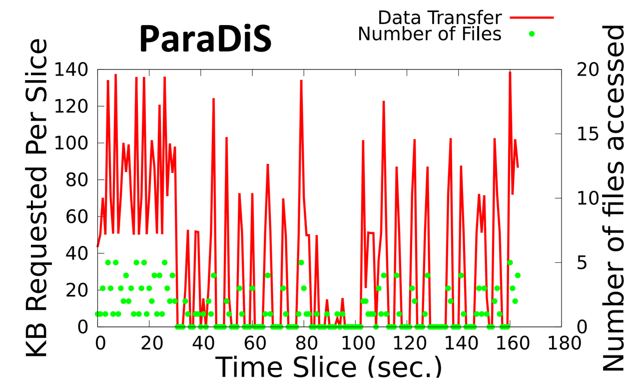
Observation

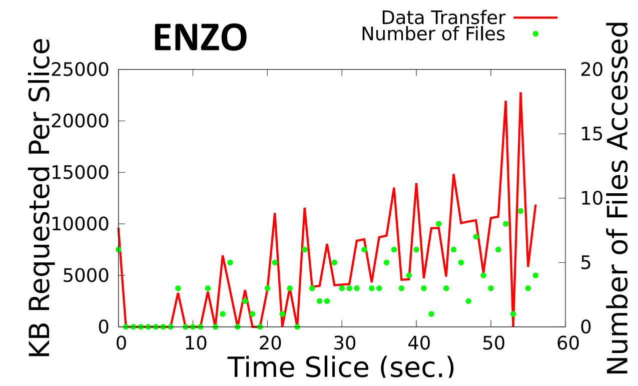
- Output interval and data transfer size vary across different output types and applications
- Outputs occur in repetitive pattern

Implication

- QoS needs for I/O varies across output types and applications
- Need: data transfer rate > repetitive data generation rate

I/O load generation in single process of each application





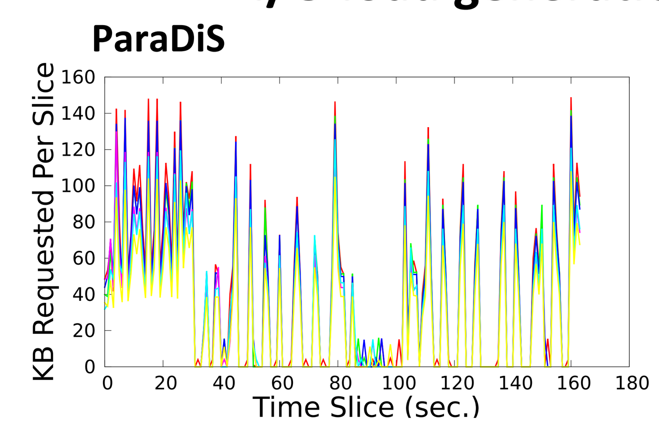
Observation

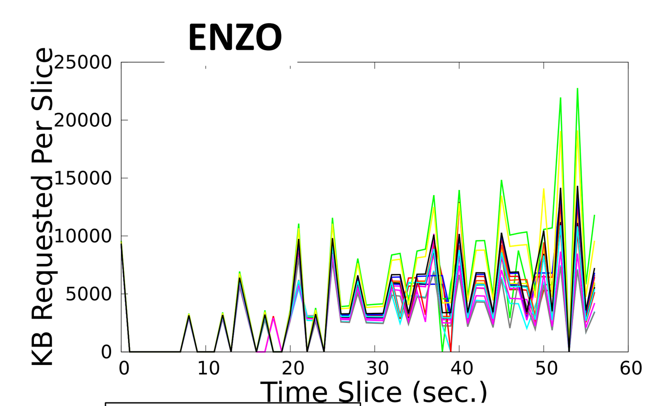
- Data transfer size varies with simulation progress
- ParaDiS: Max = 29.2 x Min, CoV = 0.46
- ENZO: Max = 6.9 x Min, CoV = 0.56

Implication

- I/O load changes during simulation
- I/O performance requirement changes too

I/O load generation across all processes





Observation

- I/O Dynamics varies across processes

Implication

- I/O performance need varies across processes

Conclusion and Future Work

- Conclusion
  - Application I/O dynamics affects system I/O load
  - QoS needs for I/O can be derived using I/O dynamics
- Future work
  - Modeling I/O dynamics for performance prediction
  - Application aware I/O management

This work is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.
This work is also supported in part by the National Science Foundation under Grant No. 12292820.

LLNL-POST-646170