CODES: Using parallel discrete-event simulation to model performance and reliability of Extreme-scale and distributed data-intensive systems

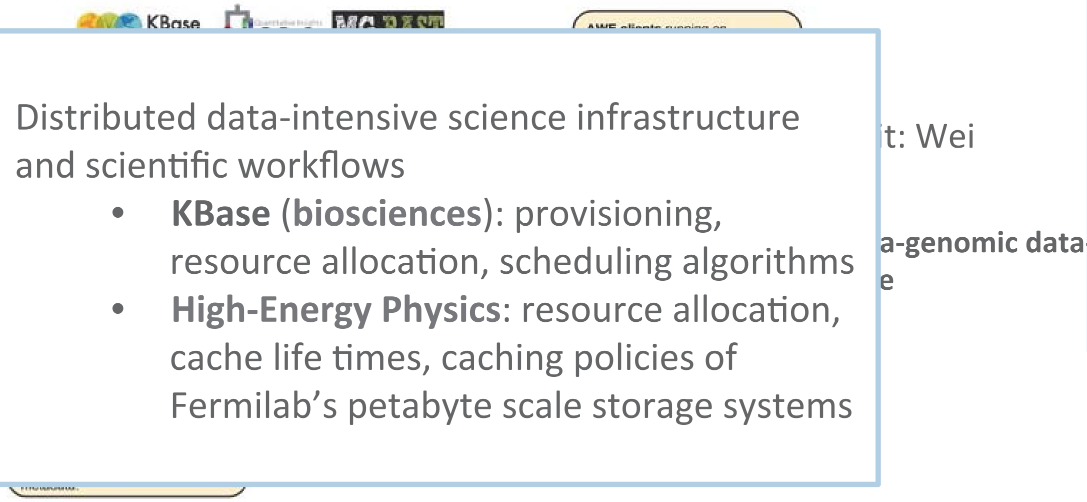
Misbah Mubarak, Argonne National Laboratory (ANL)

Workshop on Modeling and Simulation of Systems and Applications (MODSIM)
Aug 12-14, University of Washington, Seattle

ANL Team: Robert Ross (PI), Phil Carns, Kevin Harms, John Jenkins, Misbah Mubarak, Shane Snyder

RPI Team: Christopher Carothers (PI), Justin LaPre, Elsa Gonsiorowski, Caitlin Ross

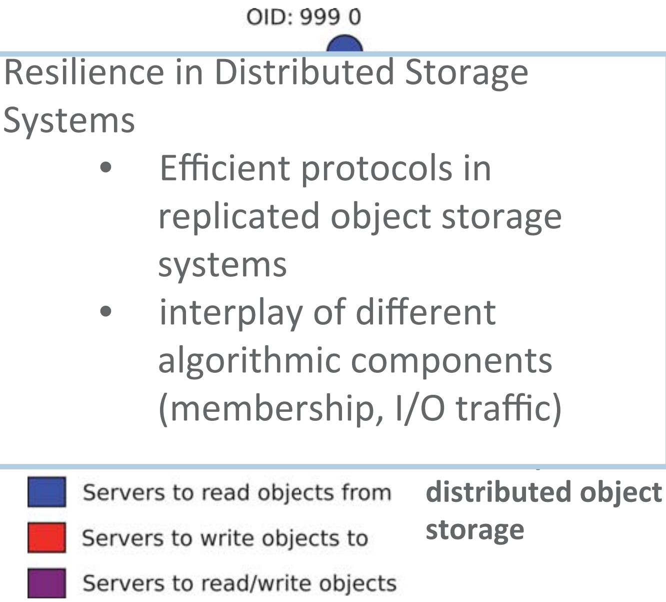
Understanding Performance and Reliability issues in HPC and distributed systems



Distributed data-intensive science infrastructure and scientific workflows

- **KBase** (**biosciences**): provisioning, resource allocation, scheduling algorithms
- **High-Energy Physics**: resource allocation, cache life times, caching policies of Fermilab's petabyte scale storage systems

Data movement and access on data-intensive science platforms

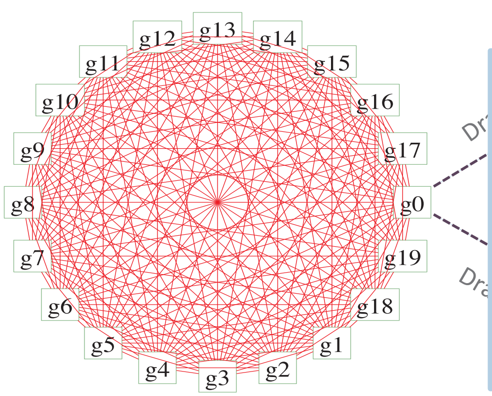


Resilience in Distributed Storage Systems

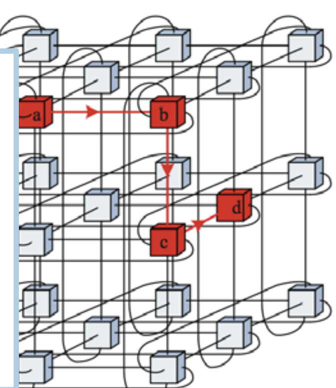
- Efficient protocols in replicated object storage systems
- interplay of different algorithmic components (membership, I/O traffic)

Resilience in distributed data-intensive storage systems

Dragonfly Network Topology



Torus Network Topology

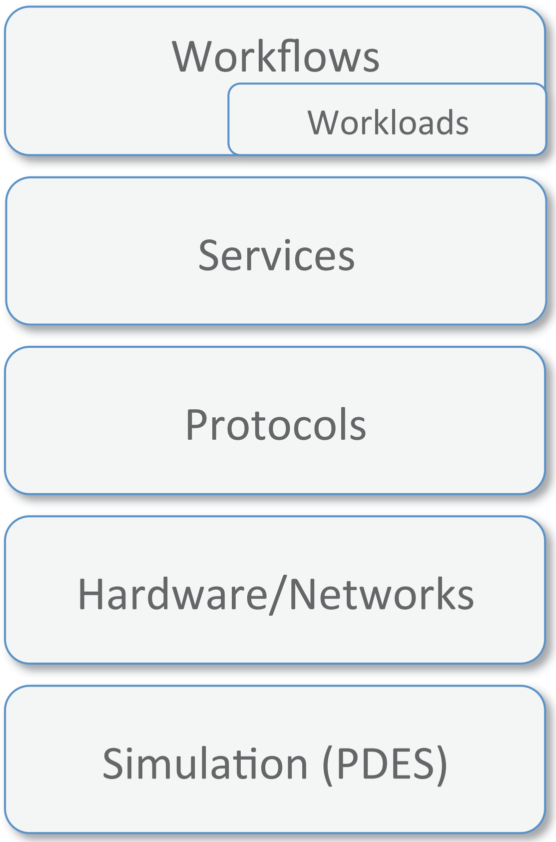


Massively parallel torus, dragonfly networks.

- Key questions: configuring link bandwidth, dimensionality, buffer space, routing etc. under different workloads?
- Effect of NVRAM Burst Buffer placement on Extreme-Scale HPC Systems

Design space of extreme-scale HPC network and Storage Systems

CODES Simulation framework for extreme-scale HPC & Distributed Systems



- **Detailed simulation of storage and network components, scientific workloads and the surrounding environment.**
- Modular simulation components
  - Pluggable I/O and network workload components
  - Pluggable high-performance network models
- Incrementally develop simulation capability, validating approach and components along the way
- **Goal: provide a simulation toolkit to the community to enable broader investigation of the design space.**

Enabling CODES: parallel discrete-event simulation

- **Discrete event simulation (DES)**: a computer model for a system where changes in the state of the system occur at *discrete points* in simulation time
- *Parallel* DES allows execution of simulation on a parallel platform
  - Enables much richer simulations (e.g., more events) than otherwise possible
- Rensselaer Optimistic Simulator System (ROSS) provides PDES capability for CODES
  - **Optimistically schedules events. Rollback realized via reverse computation**
  - Users provide functions for *reverse computation* – undo the effects of a particular evnt on the LP state
  - Logical processes (LPs) model state of the system
  - **Scalable: can efficiently execute simulations with millions of network nodes at a flit-level fidelity**

Simulation pre-reqs for co-design

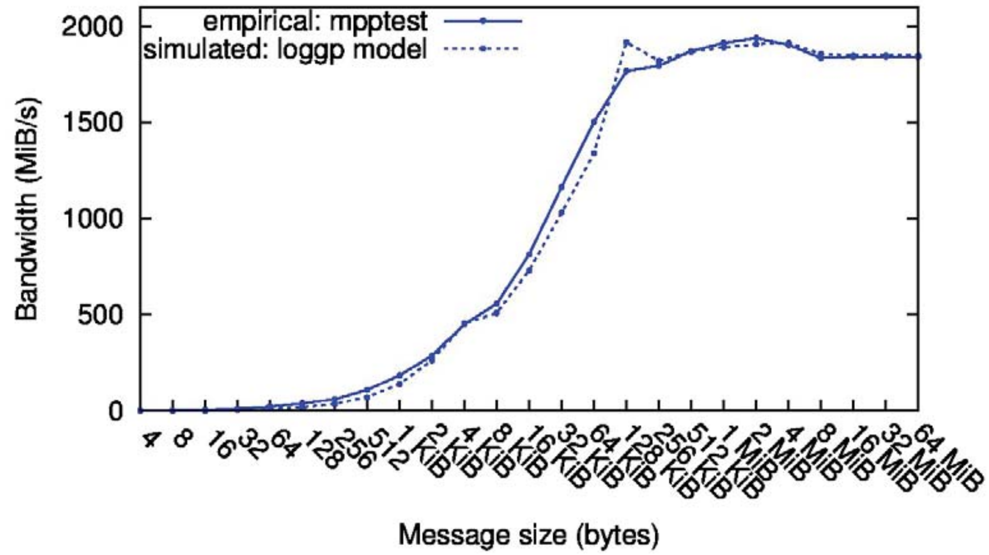
- Cost (Time + Memory)
- Fidelity + Accuracy
- Scalability

* Courtesy: Christopher D. Carothers, RPI

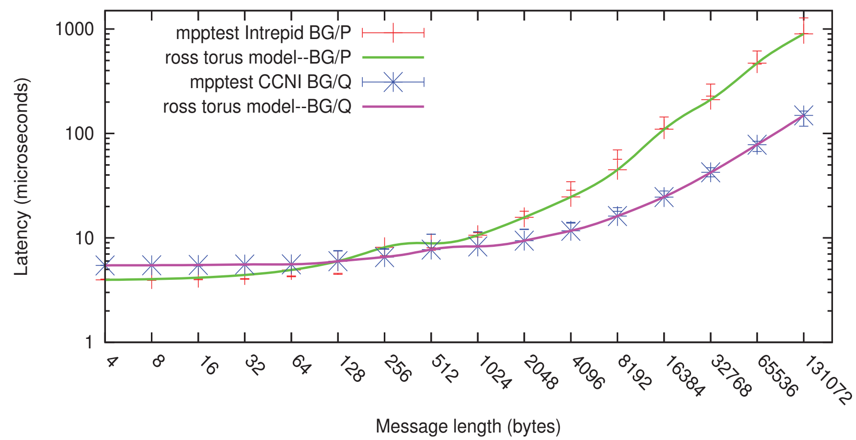
CODES Pluggable Network and I/O Models

- Pluggable Network models:
  - Models are decoupled from higher levels via a consistent API
  - **Analytic** – based on LogGP
  - Flit-level simulation of **torus**, **dragonfly** topologies at extreme scale
  - Flit-level **fat-tree** and **SlimFly** models are being developed
- Pluggable Network and I/O workloads
  - **I/O workload component:** Allows arbitrary workload consumer components to obtain IO workloads from a diverse range of input sources including characterization tools like Darshan.
  - **Network workload component:** Allows workloads from the Design Forward* program to be replayed on pluggable network models

*http://portal.nersc.gov/project/CAL/doe-miniapps.htm



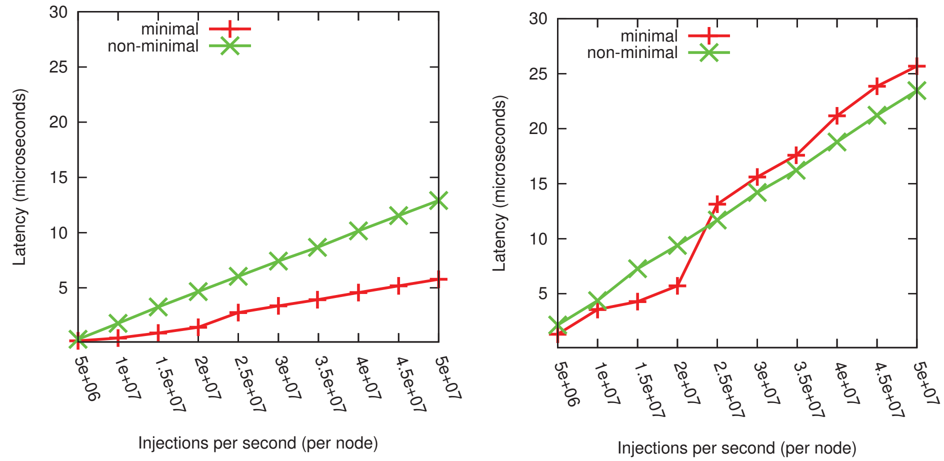
MPI message latency of CODES logGP model and mpptest



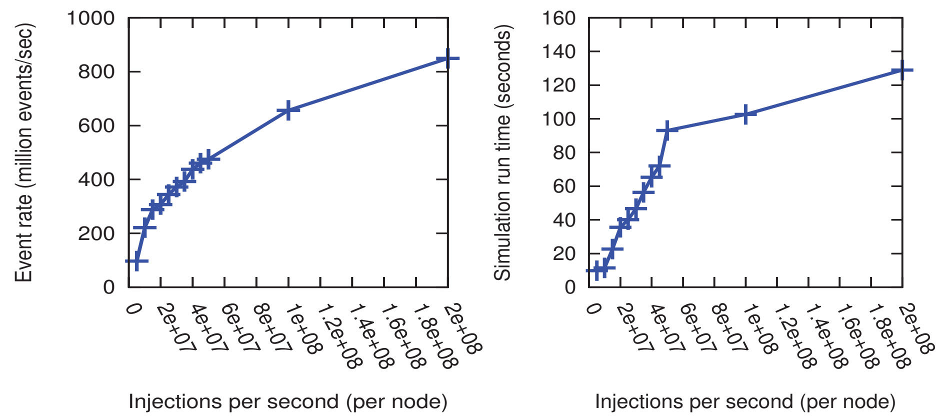
MPI message latency of CODES torus model vs. mpptest on Argonne BG/P (1 mid-plane) and RPI Amos BG/Q (1 rack) networks, 8 hops, 1 MPI rank per node

Example CODES Network Model Plug-in: Dragonfly

- ***Hierarchical topology***: several groups connected via all-to-all links
- Used in Edison (Cray XC30) and Cori architectures
- CODES dragonfly model supports minimal, non-minimal and adaptive routing
- Shown to scale on millions of network nodes with a simulation event-rate of up to 1.33 billion events/second on IBM Blue Gene/Q
- **Key Questions:** Effect of dimensionality, network configuration and link bandwidth on performance?



Average (left) and maximum (right) latency of a 1.3M node dragonfly network model with global traffic pattern (minimal & non-minimal routing)



Million-node dragonfly simulation performance on 1 rack of Mira Blue Gene/Q

Design space of extreme-scale HPC network and Storage Systems

CODES Service Models: High-Energy Physics (HEP) Science Data Facility

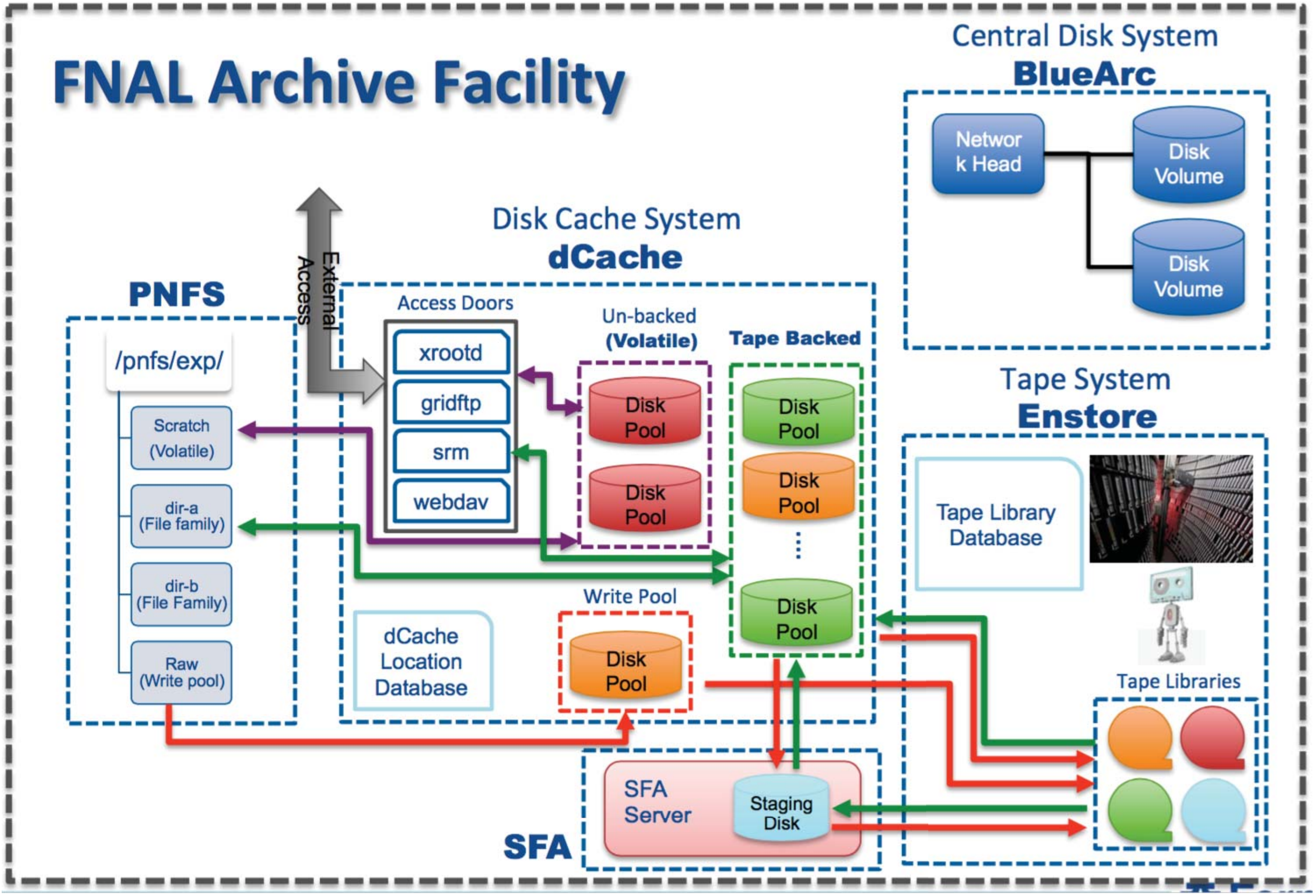


Image Credit: Andrew Norman, Adam Lyon @ Fermilab

CODES Service Models: High-Energy Physics (HEP) Data Facility

- Experimental/observational data processing pipeline @ Fermilab
- Work underway to develop an end-to-end simulation of Fermilab's HEP scientific data facility
  - Validate the simulation against actual data facility logging information
  - Use realistic HEP workflows scenarios
- Research Questions:
  - How best to configure existing peta-byte scale storage systems? E.g. how to increase cache life times, trying different cache policies
  - How to quantify the value for deploying new hardware? E.g. adding more archival devices for e.g. tapes.

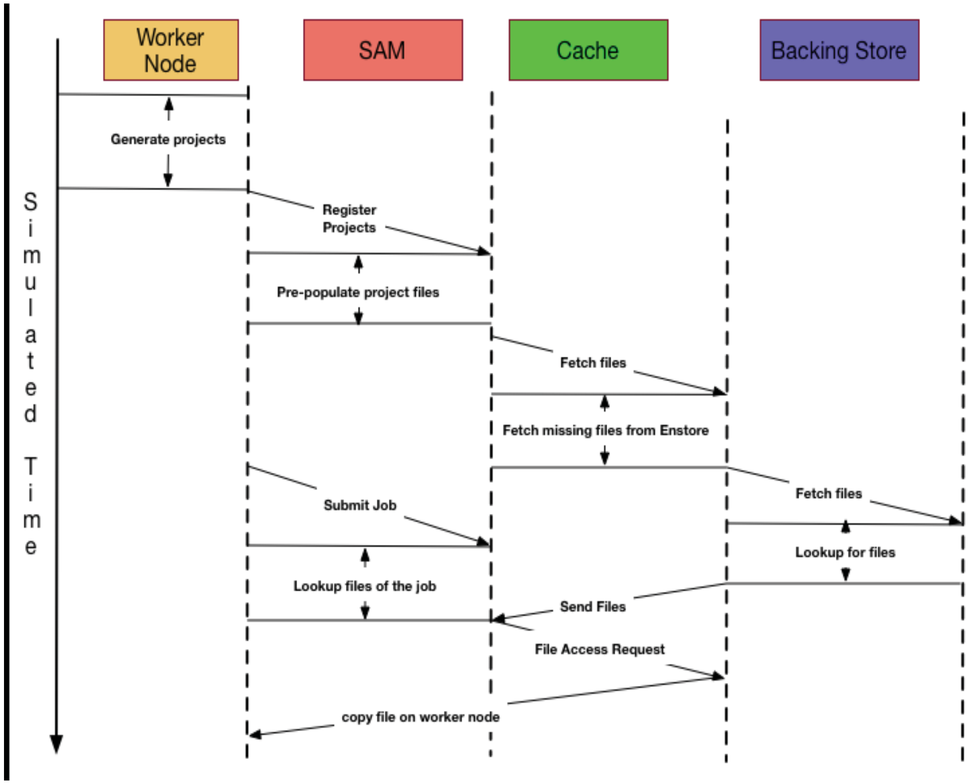
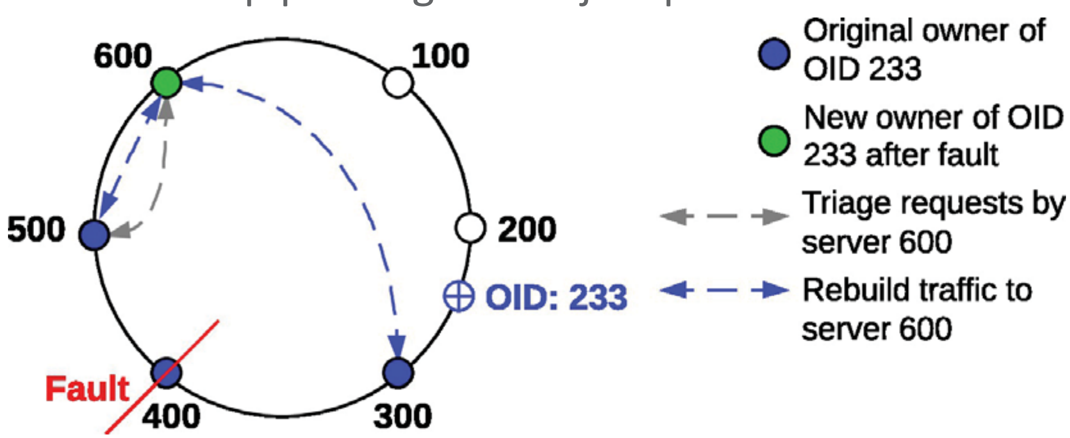


Fig: Initial simulation model of the Fermilab data-access facility

Example CODES Resilience Model: Distributed Object Storage Rebuild Simulation

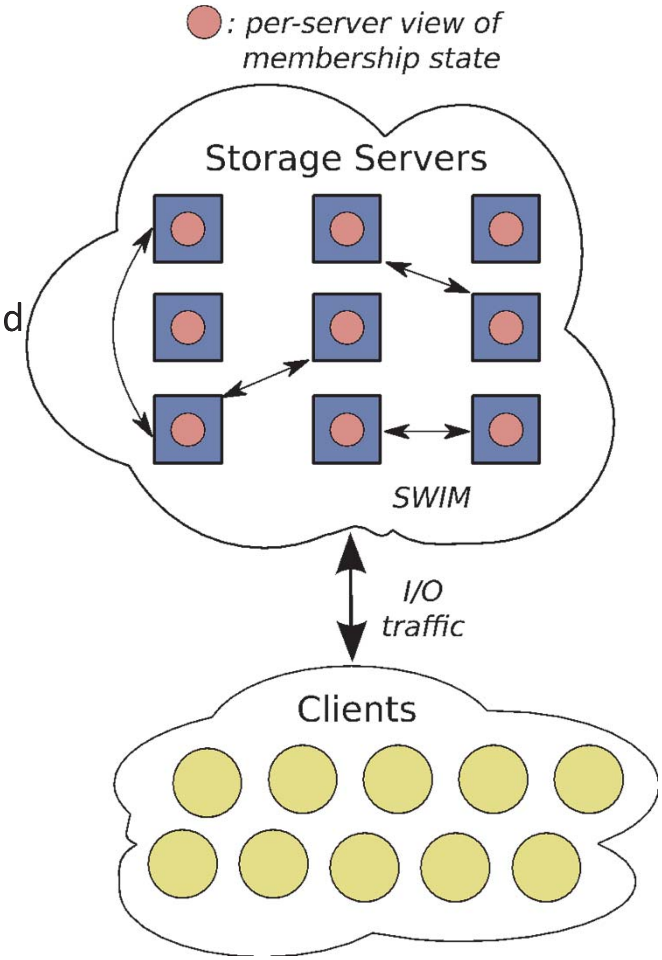
- Some protocols of interest in a replicated object storage system:
  - Rebuild – we've detected a server down, how do we re-replicate data?
  - Forwarding – how do we propagate write data between servers efficiently?
- Example: rebuild (lead: Phil Carns, ANL)
- Work explores advanced placement algorithms e.g. multi-ring hash algorithm
- Design questions: Determine the effect of algorithm/configuration on rebuild? (e.g., how many rings to use?)

Question: How do pipelining and object placement affect rebuild performance?



Example CODES Resilience Models: Group Membership with SWIM protocol

- **Storage system resilience → detection and dissemination of group membership updates**
- **S**calable **W**eakly-consistent **I**nfection-style Process Group **M**embership Protocol [1]
- Developed a high-resolution model of the SWIM protocol
  - Use simulation to evaluate behavior that can't be predicted using analytical models
  - Individual network message costs are calculated using the LogGP network model
- Scale:
  - O (thousands) of file servers
  - Tolerate transient errors < 15 seconds
  - Take action (confirm failure) within 30 seconds
- Questions:
  - How fast do membership updates propagate through the system?
  - How much network traffic are we willing/able to incur?



[1] Das, A., Gupta, I., Motivala, A.: Swim: Scalable weakly-consistent infection-style process group membership protocol. In: Proceedings of the 2002 International Conference on Dependable Systems and Networks. pp. 303–312. DSN '02, IEEE Computer Society Press, Washington, DC, USA (2002)

Summary: Work being done with CODES

- **Network models**
  - Torus
  - Dragonfly
  - Fat tree (IIT, UIUC)
  - Slim Fly (In Progress)
- **Storage models**
  - I/O protocols
  - Data placement
  - Fault Detection/ Response
  - Burst buffers
- **HPC workload models**
  - IOWA– modeling large-scale I/O workloads
  - Network --- Design Forward network trace replay
- **Workflow processing / Data management**
  - Bioinformatics: Kbase usecase, meta-genomic server model
  - High Energy Physics (HEP) : Fermilab's peta-scale storage system (In Progress)
- ...

Modsim Workshop Questions

- What is the major contribution of your research?
  - High-fidelity and large-scale parallel discrete-event models of high-end computing components like HPC networks, I/O and network workloads, scientific workflows and data-intensive architecture
  - Modular components can be combined for HPC and data-intensive architectural evaluation and co-design.
- What is the bigger picture for your research area? What are the gaps?
  - Ability to execute high fidelity network models over 12 hr to 24 hr time-scales
  - Lack of flexible network and storage behavior descriptions
  - Reverse computation is coded by hand for CODES models.
    - Compiler efforts: LLVM (RPI)
- What is the bigger picture for your research area? (i.e., identify synergistic projects, complementary projects in technical sense, etc)
  - Node-level simulators such as SST (SNL), Manifold (GaTech), Booksim (Stanford)

Acknowledgements

- **Sponsor**

  This material was based on work supported by the U.S. Department of Energy, Office of Science, ASCR Program under contract DE-AC02-06CH11357. The research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is a DOE Office of Science User Facility.

- **Acknowledgements**
  - Wolfgang Gerlach, Folker Meyer (ANL)
  - Adam Lyon, Andrew Norman, James Kowalkowski (Fermilab)
  - Wei Tang (Google)

Repo access and SoC

- **CODES developer access**
  - http://www.mcs.anl.gov/research/projects/codes/
- **Discussion forum**
  - codes-ross-users@lists.mcs.anl.gov
- **1st Summer of CODES workshop**
  - http://www.mcs.anl.gov/research/projects/codes/publications/

Questions?