Oak Ridge National Laboratory

Towards the Development of an Exascale Network Digital Twin

John K. Holmen, Oak Ridge National Laboratory (ORNL)

Md Nahid Newaz, Oakland University

Srikanth Yoginath, ORNL

Matthias Maiterth, ORNL

Amir Shehata, ORNL

Nick Hagerty, ORNL

Christopher Zimmer, ORNL

Wes Brewer, ORNL

EXPERIENCE ORNL

MEET. EXPLORE. LEARN.

U.S. DEPARTMENT OF ENERGY

Motivation

- Resiliency a fundamental challenge for exascale systems
  - ~60M components in Frontier
- Component counts and complexity lead to more and new failures
- Important to ensure system functionality, performance, and usability
- Talk captures investigation of network digital twins

https://www.flickr.com/photos/olcf/52117623843/in/album-72177720299483343/

Frontier

- 9,408 HPE Cray EX235a nodes
- Theoretical peak of 2 Exaflop
  - Compute similar to 194,544 PS5s
- 74 cabinets weighing 8,000 pounds each
  - Total weight similar to a Boeing 747
- 90 miles of network cables
  - Perth to Wedge Island
- 700 PB of Lustre storage
  - 25 Mt. Everests of DVDs

https://www.flickr.com/photos/olcf/52117623763/in/album-72177720299483343/

Digital Twins

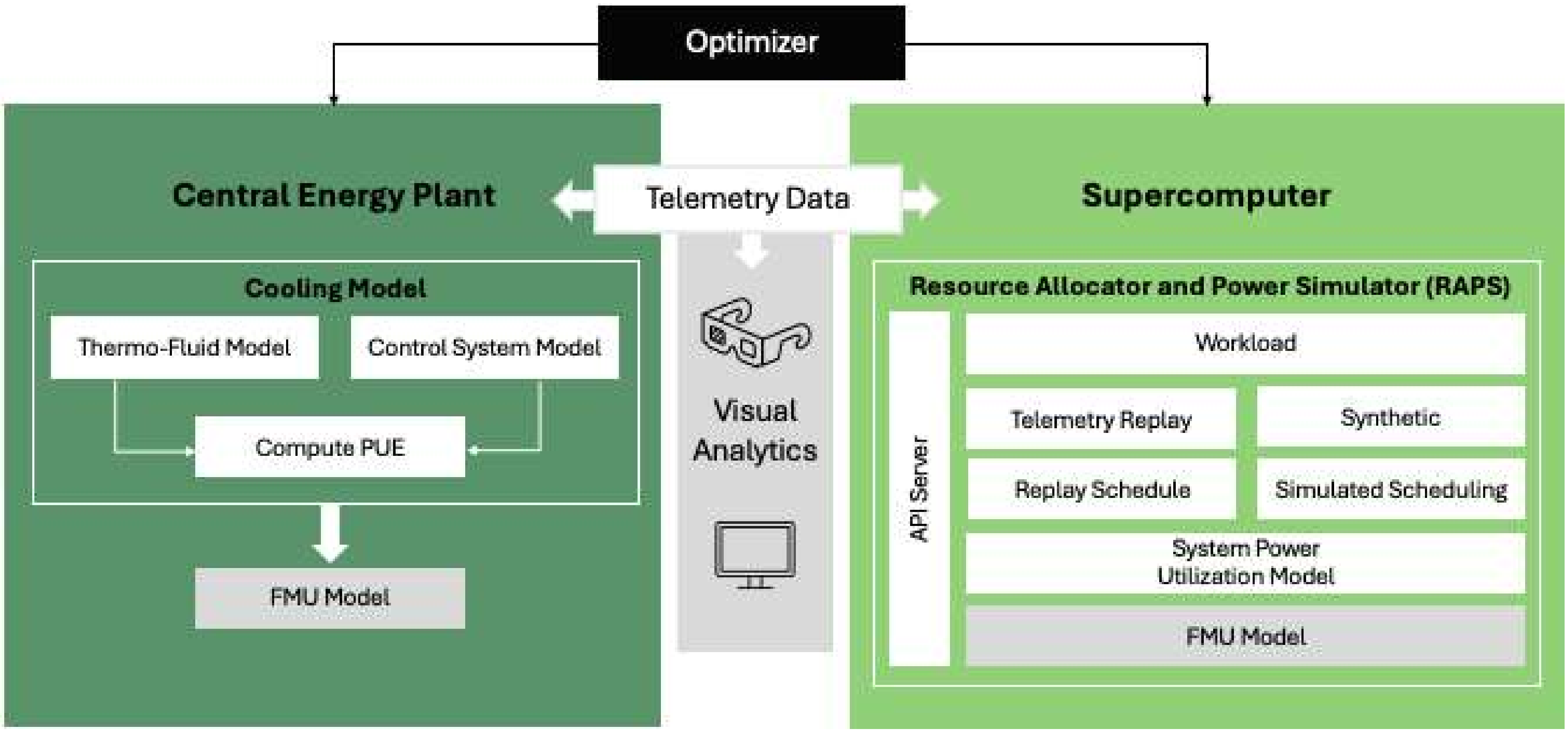
- National Academy of Sciences, Engineering, and Medicine (NASEM) define a digital twin as:
  - "A digital twin
    - is a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems),
    - is dynamically updated with data from its physical twin,
    - has a predictive capability, and
    - informs decisions that realize value.
  - The bidirectional interaction between the virtual and the physical is central to the digital twin."

ExaDigiT

- Multidisciplinary international collaboration
  - Academia, HPC centers, industry
- Community-driven effort to design open-source framework for digital twins of liquid-cooled supercomputers
- Effort guided by 8 working groups
- Differing progress across groups, e.g.,:
  - Established cooling and power models
  - Early investigation for network model

EXADIGIT

Architecture Overview



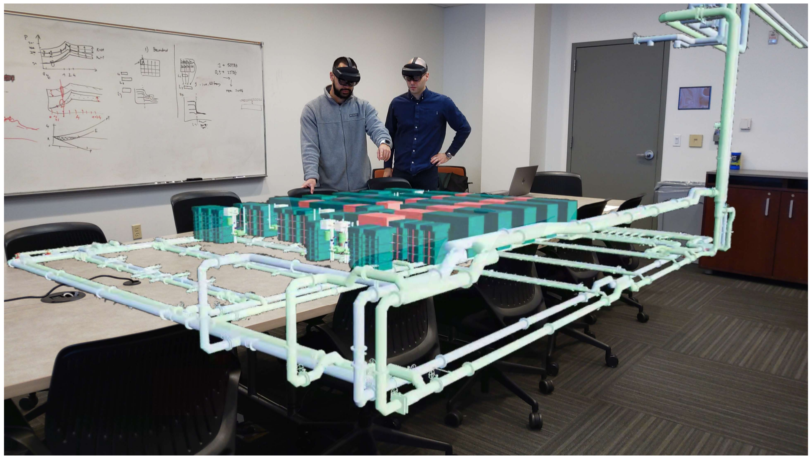
Resource Allocator and Power Simulator (RAPS)

- Enables replay and simulation of jobs
- Shows jobs running through queue
- Captures cooling, power, etc.
- Calculates efficiency, carbon emissions, etc.
- Replayed 183 days of Frontier data

OAK RIDGE National Laboratory

Frontier Augmented Reality (AR) Environment

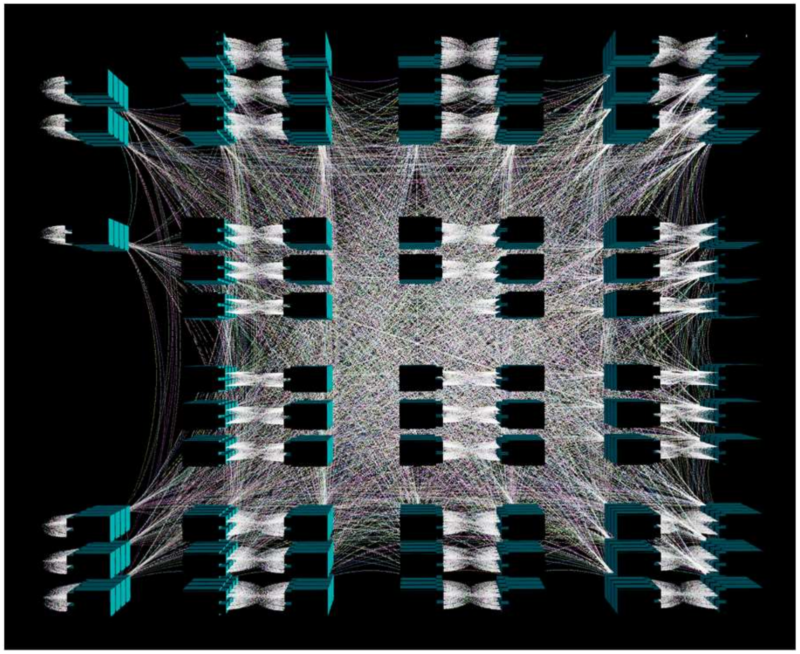
- Interactive scene visualizing Frontier
  - Implemented using Unreal Engine 5
  - Interaction using Microsoft HoloLens
- Color-coded visualization of various data gathered on Frontier
- Filters allow viewing of:
  - Cabinet interior (i.e., nodes),
  - Cooling infrastructure,
  - Network infrastructure,
  - Power infrastructure, etc.



OAK RIDGE National Laboratory

AR Network Infrastructure

- Recently added network visualization capabilities
- Connections can be color-coded
- Currently displaying all connections
  - Filter logic needed
- Network-related data not yet incorporated
  - Working to identify and gather such data as a part of this work



Network Digital Twins

- Mimics network infrastructure and transmission of data
  - Dynamically updated with data from its physical twin
- Example Use Cases:
  - Understanding and mitigating network congestion
    - e.g., congestion studies and routing optimization
  - Application fingerprinting
    - e.g., characterize and model workloads
  - Improve parallel discrete event simulator models (e.g., those in SST)
    - e.g., validate network models at first-of-kind scales

Target Use Case

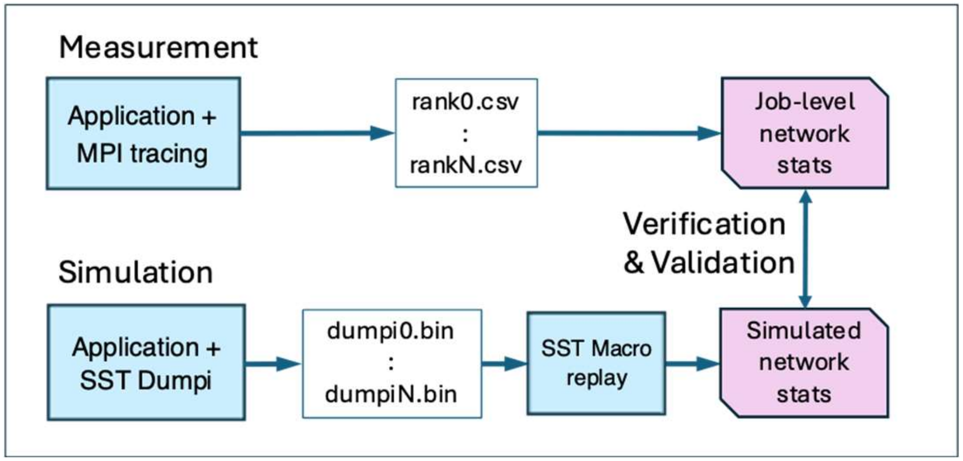
- Short-Term: Extend RAPS functionality
  - RAPS uses CPU and GPU utilization to model system power
  - Improve workload modeling by adding network utilization
- Long-Term: Application fingerprinting
  - Aim to collect:
    - Message sizes and counts for transfers across network
    - Network-related hardware power consumption
  - Combine the two for application fingerprinting
    - i.e., identify types of workloads stressing the network

Tools

- SST/macro (Structural Simulation Toolkit)
  - Platform to simulate full-scale machines and evaluate changes
  - Models to estimate processing and network component performance
- SST DUMPI Trace Library
  - Trace collection and replay tools for MPI applications
  - SST/macro uses trace data to simulate machine variations
- Custom MPI Tracers
  - fi_hook utility and PMPI used to trace calls
- Also explored CrayPat and Darshan+autoperf
  - Too coarse-grained (e.g., summarized statistics after run)

Goal

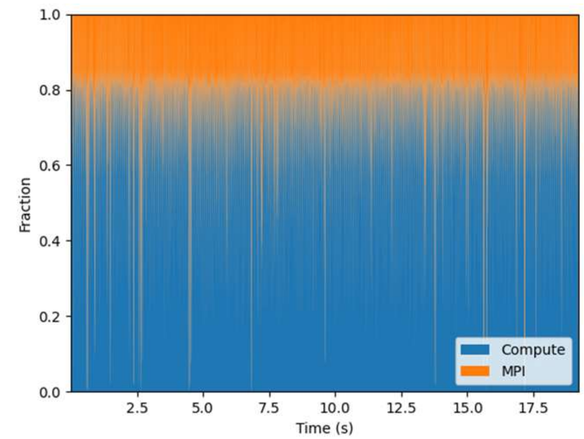
- Trace applications with SST DUMPI
- Replay traces with SST/macro
- Trace applications with MPI tracer(s)
- Validate simulations

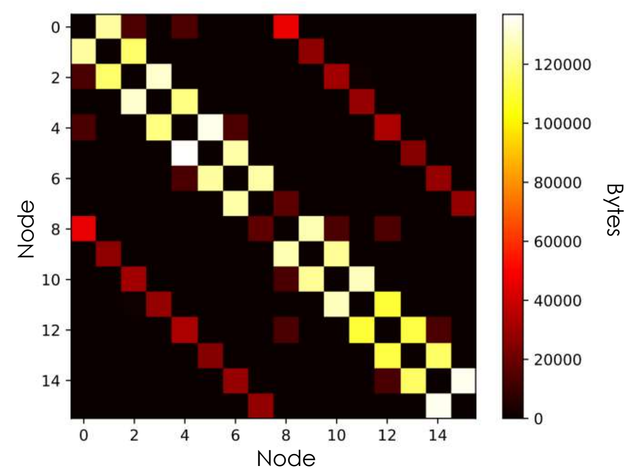


OAK RIDGE National Laboratory

Progress

- Generated fixed-time quanta charts with SST (top)
  - Time-dependent histogram showing split between compute and MPI
  - 1764 rank NAS Parallel Benchmark run
    - Block tri-diagonal solve
- Generated spyplots with SST (bottom)
  - Visualizes message counts or bytes between network endpoints
  - 16 node miniVite run
    - Graph analytics benchmark tool





OAK RIDGE National Laboratory

Progress (cont.)

- Explored CrayPat and Darshan
  - Moved to tracing for more fine-grained data
- Evaluating two approaches
  - Collecting traces using libfabric's hook fabric provider utility, fi_hook
  - Collecting traces using PMPI
    - https://github.com/hagertnl/mpi-trace
- Example output:

```
[Rank 63] MPI_Irecv started 1713984327.205686331, ended 1713984327.205687046 (elapsed 0.000000715), moved 1572864 bytes from source 59
[Rank 63] MPI_Irecv started 1713984327.205688477, ended 1713984327.205688715 (elapsed 0.000000238), moved 1572864 bytes from source 62
```

- Validation of SST a work in progress

Challenges

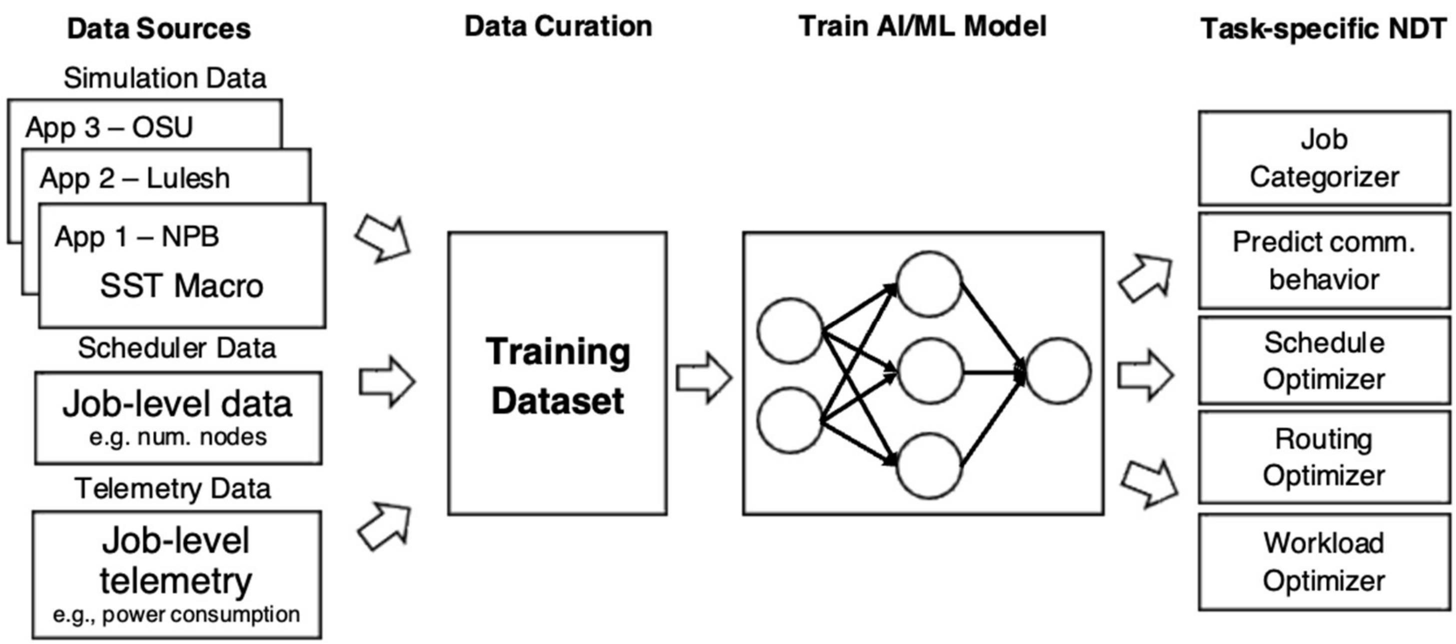
- Data Availability
  - Cooling and power data readily available
    - Network data hasn't been as readily available
  - Manual data collection is time intensive
    - Complicated by uncertainty in which tools are useful
- Data Visualization
  - Node-level details easier to visualize
    - e.g., GPU interconnectedness not needed when visualizing power
  - Network visualization complex due to interconnectedness
    - Difficult to gather meaningful insight from full view

Challenges (cont.)

- Differing Scales
  - Digital twin operates on 15-second intervals
  - Network time scale much faster
    - e.g., 100-350 ns for switching technology
  - Unclear how to incorporate network data
    - Aggregate data to align with 15-second interval?
    - Operate network data on separate interval?
- Simulation Time
  - Explored use of simulation to inform network digital twin
  - Tens of hours to simulate runs with SST/macro
    - Turn-around time not feasible

ExaDigit Integration Goals

- Explore AI/ML as simulation alternative
  - Train on validated simulation data



ExaDigit Integration Goals (cont.)

- Extend Resource Allocator and Power Simulator capabilities
  - Capture network-related data
  - Study relationship between network and power
- Extend augmented reality scene
  - Visualize messages sent/received
  - Filter network components shown to meaningful subsets
- Explore ways to passively gather network data
  - Eliminate need to manually gather data
  - Capture network-related data across all jobs

Conclusions

- Network digital twin important to overall digital twin
  - Frontier's scale helpful for validation
- Unclear how to best integrate a network digital twin
  - Separate operating intervals?
- Challenging to find tools aligned with goals
  - Suggestions?
- Tracers found most helpful
  - Anticipate manual data collection and processing

ExaDigiT Collaborations

- Monthly meetings and standalone working group meetings:
  - Visual Analytics
  - Application Fingerprinting
  - Power & Cooling
  - Use Cases & Architectures
  - Documentation
  - Networking
  - AI/ML/RL
  - Verification, Validation, and Uncertainty Quantification (VVUQ)
- If interested, contact Wes Brewer: brewerwh@ornl.gov

Questions?

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

holmenjk@ornl.gov