A Visual Analytics System for Optimizing the Performance of Large-Scale Networks in Supercomputing System

Takanori Fujiwara\textsuperscript{1}, Jianping Kelvin Li\textsuperscript{1}, Misbah Mubarak\textsuperscript{2}, Caitlin Ross\textsuperscript{3}, Christopher D. Carothers\textsuperscript{3}, Robert B. Ross\textsuperscript{2}, and Kwan-Liu Ma\textsuperscript{1}

\textsuperscript{1}University of California, Davis
\textsuperscript{2}Argonne National Laboratory
\textsuperscript{3}Rensselaer Polytechnic Institute
About This Work

• Originally presented at PacificVAST 2018
  - available online

• A visual analytics system designed for analyzing the time-series network performance data
  - obtained from CODES or others
Complex Problems

Cosmological simulation [ANL]

Hurricane simulation [NASA, NCAR]

Splitting in Subdomains

Domain decomposition

Running in Parallel

Calculate with parallel processes on a certain network topology
New Supercomputing Systems

<table>
<thead>
<tr>
<th>System</th>
<th>Compute Nodes</th>
<th>Network Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theta (ANL)</td>
<td>&gt; 3,000</td>
<td>Dragonfly</td>
</tr>
<tr>
<td>Sierra (LLNL)</td>
<td>&gt; 3,400</td>
<td>Fat Tree</td>
</tr>
<tr>
<td>Cori (NERSC)</td>
<td>&gt; 10,000</td>
<td>Dragonfly</td>
</tr>
<tr>
<td>Trinity (LANL)</td>
<td>&gt; 19,000</td>
<td>Dragonfly</td>
</tr>
<tr>
<td>Aurora (ANL)</td>
<td>&gt; 50,000</td>
<td>Dragonfly</td>
</tr>
</tbody>
</table>

- A large number of compute nodes
- Low-diameter network topologies
- Efficient routing strategies (e.g., adaptive routing)
Hierarchical low-diameter network
- Each router is connected to multiple terminals (compute nodes)
- Routers in each group are also fully connected by local links
- All groups are fully connected by global links
Adaptive Routing

- Non-adaptive routing (Static)
  - Routers always send packets along the fixed paths (e.g., a shortest path)

- Adaptive routing (Dynamic)
  - Routing paths are updated based on some network information (e.g., choosing paths with lower traffics)

[Diagram showing network nodes and packet flows]
Adaptive Routing

- **Non-adaptive routing (Static)**
  - Routers always send packets along the fixed paths (e.g., a shortest path)

- **Adaptive routing (Dynamic)**
  - Routing paths are updated based on some network information (e.g., choosing paths with lower traffics)
Network Performance Analysis and Challenges

- Fast communication between compute nodes is crucial for high efficiency of parallel applications
  - need to identify communication bottlenecks for different applications, routing strategies, etc

- Difficult to identify the communication bottlenecks
  - large number of compute nodes
  - complex network topology
  - dynamically changing routing strategies
Related Work

• Visualizing network congestion on the Dragonfly network
• Both are focusing on the structural characteristics of network performance

[Bhatele et al., 2016]  [Li et al., 2017]
Related Work

- Analyzing temporal behavior of cloud computing
- The behavior lines show behavioral similarity of multivariate time-varying performance data
- Their method cannot analyze the structural characteristics

[Muelder et al., 2017]
Problem: Performance Analysis from Structural and Temporal Behavior

- How can we analyze the network performance both structural and temporal aspects?

- How can we find the potential bottlenecks from a large, complex, dynamic network behaviors?
Design Requirements

• To understand the cause of the bottlenecks from a large, complex, dynamic network behaviors, the system should:

1. Show **temporal** network behaviors and patterns

2. Depict information related with the **physical** network topology

3. Provide a **succinct summary** from large scale data to help the user identify the bottlenecks
Our Visual Analytics System
Our Visual Analytics System

Behavior Overview
Our Visual Analytics System

Behavior Overview

Behavior Detailed Views
Our Visual Analytics System

Behavior Overview

Behavior Detailed Views

Behavior Similarity Views

Topological Views
Behavior Overview

- Shows an overview of the network behaviors
  - one selected statistical metric of an network entity for each time point across time
    - e.g., the mean of network traffics on terminal links

- The user can select a time range with a slider to show more detailed information in the other views
Behavior Detailed Views

- Show the details of network behavior in the selected time range
  - e.g., traffics on terminal and global links
- Compare the behaviors of the different network entities and understand the relationship of cause and effect
Behavior Detailed Views: Time-Series Clustering

- Time-series clustering for helping the user find interesting patterns from complicated time series data

- A user can choose the clustering and time-series similarity measure based on the tasks
  - clustering methods:
    - k-means, k-medoids [Kaufman and Rousseeuw, 2009], complete linkage
  - similarity measures:
    - Euclidean distance, dynamic time-warping [Berndt and Clifford, 1994]
    - time warp edit distance [Marteau, 2009]
Behavior Similarity Views

- Show the time-series similarities of each behavior
- Supplement the behavior detailed views
Behavior Similarity Views: Dimensionality Reduction

- Place similar behaviors close together
  - using multidimensional-scaling (MDS) or t-SNE [Maaten and Hinton, 2008]

- Useful for finding the patterns in a small set of behaviors
  - e.g., outliers and anomaly behaviors
Topological Views

- Show summaries of network behaviors with physical network information within selected segmentations
  - inspired by "temporal summary images" [Bryan et al., 2017]
Topological Views: Segmentation of Behaviors

- Utilize the change point detection for automatic segmentation
  - E-Divisive method [Matteson and James, 2014]
    - can detect multiple change points
- The user also can manually change segmentation
Topological Views: Behavior Summaries

- Calculate the mean traffics on each network link for each segmentation
- Show the mean values with the visualization method developed by Li et al. (2017)
Topological Views: Behavior Summaries

- Calculate the mean traffics on each network link for each segmentation
- Show the mean values with the visualization method developed by Li et al. (2017)
Topological Views: Behavior Summaries

- Calculate the mean traffics on each network link for each segmentation
- Show the mean values with the visualization method developed by Li et al. (2017)
Case Study
Dataset

• **System**
  - Simulation of Theta at Argonne National Laboratory with the CODES network simulation toolkit [Cope et al., 2011]

• **Network Topology**
  - Dragonfly
    (9 groups, each group has 96 routers, each router has 4 terminals)

• **Routing strategy**
  - Adaptive routing

• **Application**
  - AMG solver application [Yang et al., 2002]  the multigrid application [Bell et al., 2012]

• **Experimental settings**
  - number of MPI ranks: 1,728 and 13,824
Case Study 1: Relationships between Terminal and Global Links' Traffic

The mean traffics on terminal links
Case Study 1: Relationships between Terminal and Global Links Traffic

Select a clear peak with the slider
Case Study 1: Relationships between Terminal and Global Links Traffic

Show traffics on terminal and global links

Traffic on terminal links

Traffic on global links
Case Study 1: Relationships between Terminal and Global Links Traffic

Apply k-medoid clustering for both plots.
Case Study 1: Relationships between Terminal and Global Links Traffic

Two clusters have heavy traffics on terminal links

Heavy traffics on global links
Case Study 1: Relationships between Terminal and Global Links Traffic

Show only the terminals on the first peak and related global links.

The first pink peak relates to the heavy traffics on global links.
Case Study 1: Relationships between Terminal and Global Links Traffic

Show only the terminals on the second peak and related global links.

The second peak mostly involves the orange traffics.
Case Study 1: Relationships between Terminal and Global Links Traffic

The adaptive routing helps reduce the communication bottlenecks on global links.
Dataset

- **System**
  - Simulation of Theta at Argonne National Laboratory with the CODES network simulation toolkit [Cope et al., 2011]

- **Network Topology**
  - Dragonfly
    (9 groups, each group has 96 routers, each router has 4 terminals)

- **Routing strategy**
  - Adaptive routing

- **Application**
  - AMG solver application [Yang et al., 2002], the multigrid application [Bell et al., 2012]

- **Experimental settings**
  - number of MPI ranks: 1,728 and 13,824
Case Study 2: Network Behavior When Using Many MPI ranks

Select time range where networks are full for a long time

Visualize traffics on global and terminal links with clustering
Case Study 2: Network Behavior When Using Many MPI ranks

Most of local links have high traffics
Case Study 2: Network Behavior When Using Many MPI ranks

The workload on global links is imbalanced.

Performance can be improved by utilizing unused global links (e.g. random job placement).
Conclusions and Future Work

• Presented a visual analytic system for understanding the dynamic behavior of a large-scale supercomputer’s networks
  - enables identifying bottlenecks visually and interactively
  - coupled with time-series analysis methods
Questions?

tfujiwara@ucdavis.edu

https://takanori-fujiwara.github.io

This research has been sponsored in part by the U.S. National Science Foundation and Department of Energy. This research was sponsored by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR) under contract DEAC02-06CH11357.