ESnet
ENERGY SCIENCES NETWORK

# Debugging Bad Performance in Huge Infrastructure: Using ML and AI


**Mariam Kiran (ESnet)**
**Anirban Mandal (RENCI)**
**George Papadimitriu (USC)**
**Cong Wang (RENCI)**
**Ewa Deelman (USC)**

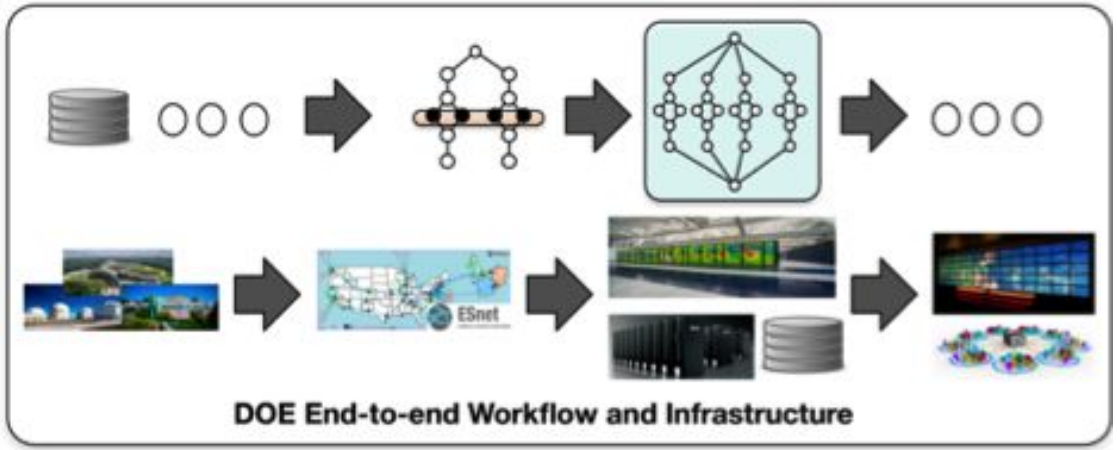


Huge Data Workshop
13-14 April 2020


U.S. DEPARTMENT OF ENERGY
Office of Science

BERKELEY LAB

# Complex Workflows using Distributed Facilities



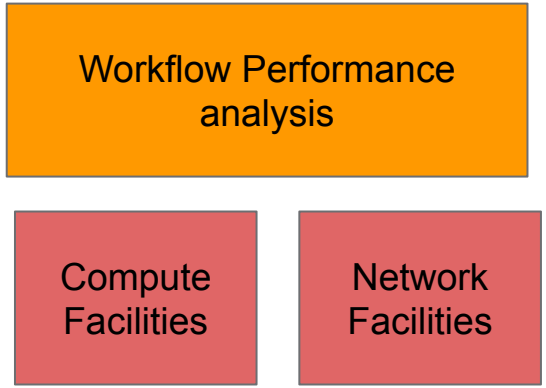

Workflow Performance analysis

Infrastructure Troubleshooting

- Distributed workflows massive computations and large data movements
- Workflow performance:
  - debug performance
  - performance optimization
- Infrastructure Troubleshooting:
  - Anomaly finding
  - root cause analysis

ESnet

# Challenge: Working with Distributed Facilities




- Current infrastructures are not designed to diagnose bad performance in distributed settings
- Distributed performance data repositories

- RQ: Can we use ML/AI to build “normal” profiles to debug problems

ESnet

# Focusing on Network Behaviors

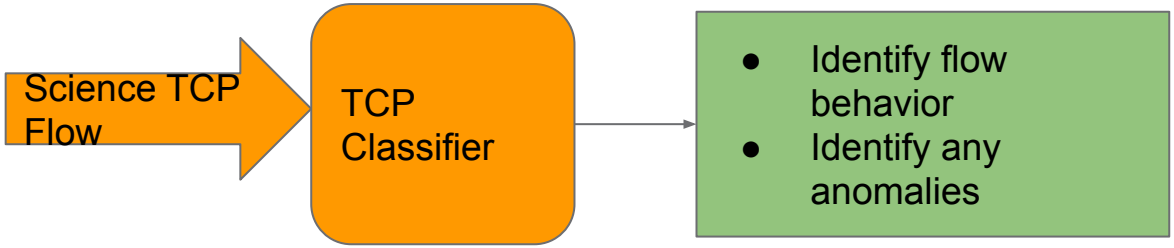
Science Workflows use TCP to guarantee data delivery, but what if something goes wrong

- Less Throughput
- Packet loss
- Packets reordered

Costing the credibility of experiments

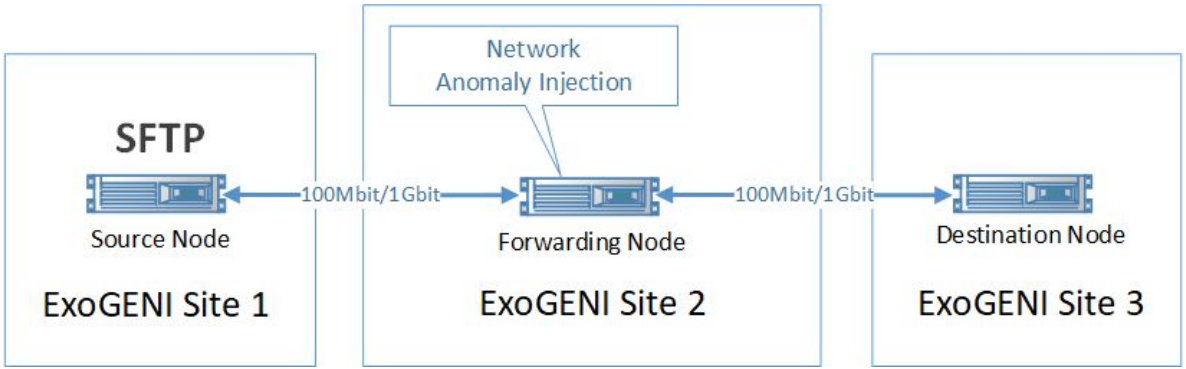
ESnet

# Building Classifiers for ML/AI

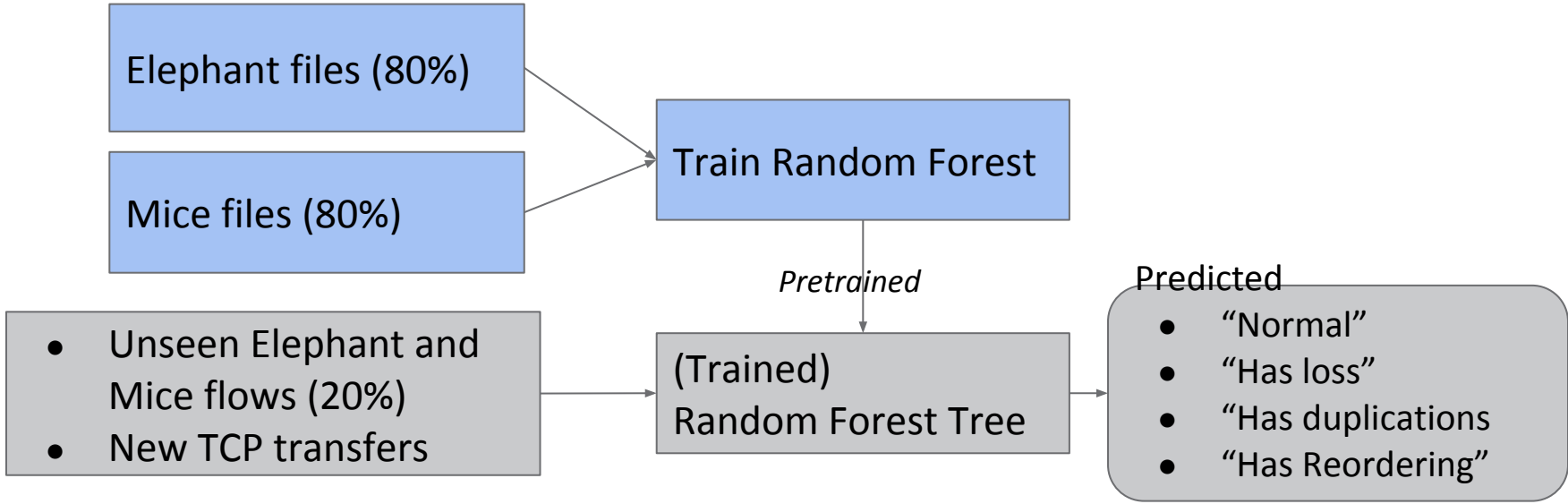



Build “normal” and “abnormal” TCP profiles

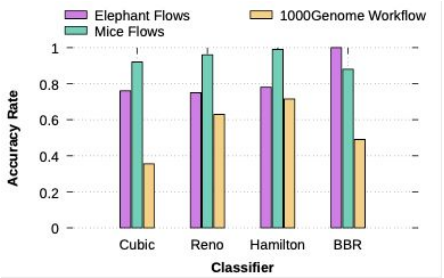
Train ML Classifiers to identify behavior patterns



# Deploying Classifiers at Network Edge



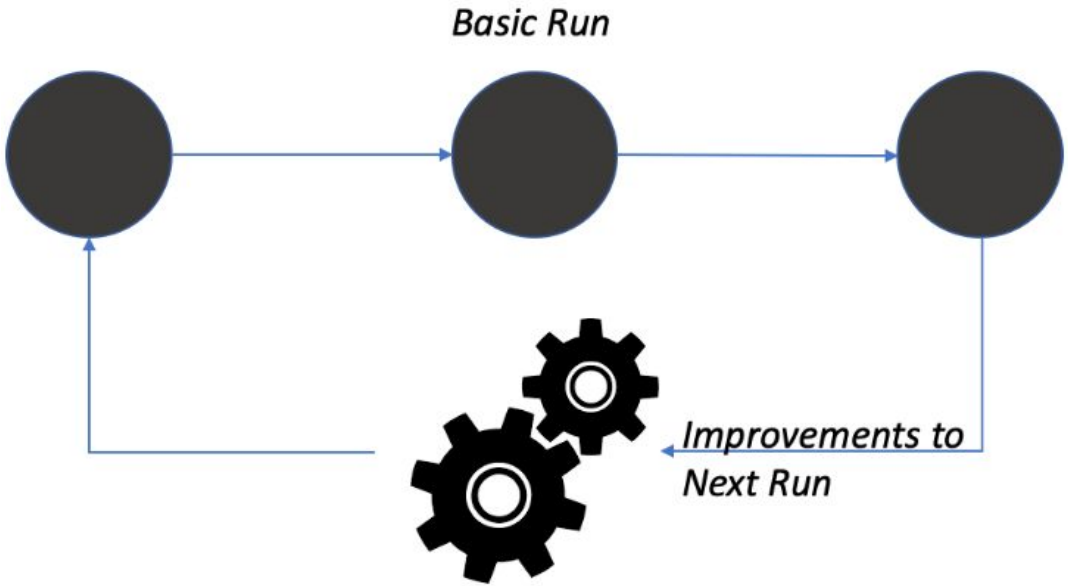

Challenge: Need labelled data sets for training classifiers




ESnet

# Next Phase: Self-Learning Classifiers

Adapting Deep Reinforcement Learning to Optimize Classifiers for accuracy




- Try new performance via trial-and-error
- Learn optimum profiles for workflows over time
- Can perform in unlabelled data settings
- Build recommender profiles!

ESnet

# Thankyou!

Email: <mkiran@es.net>

Multiple projects on ML/AI for Network Operations

Tying it back to Workflows can help scientists debug where problems are occuring

ESnet