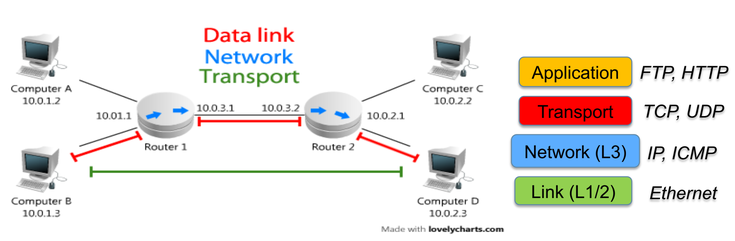
# Deep and Autonomic High-Performance Networks

Mariam Kiran[1,2], Nandini Krishnaswamy[2]

[1]Energy Sciences Network (ESNet) [2]Scientific Data Management Group (CRD)

## WAN-scale Network Resource and Traffic Optimization




Dynamic control of flow traffic for exascale science applications

- Achieve maximum application performance
- Minimize travel time
- Reduce traffic congestion
- Improve bandwidth utilization

Problem domain: Develop techniques to handle online and fast decision-making on very large-scale decentralized infrastructure

- > 50 sites
- >$10^2$ devices
- > $10^4$ flows per second per link
- > $10^5$ Bytes of moving data

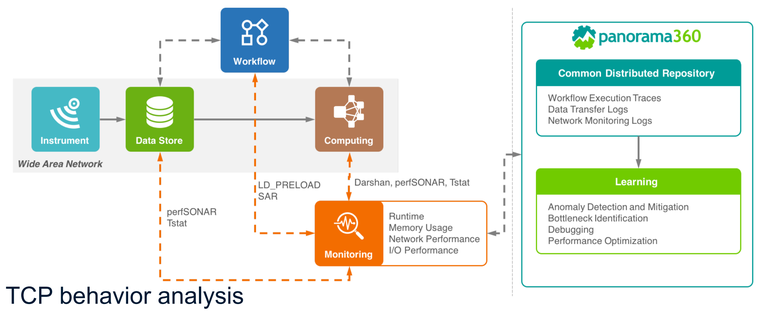
Impact

- Improve reliability of large and long running network transfers
- Enable intent-driven network resource allocation
- Enable complex science workflows requiring deadline-driven transfers
- Enable 100% utilization of network links
- Dynamically control and optimize network resource usage

## TCP Performance Improvement for End-to-End Workflows

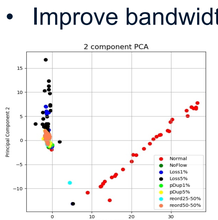
**Panorama 360: Performance Data Capture and Analysis for End-to-End Scientific workflows**

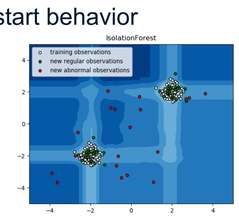
- End-to-End performance analysis of distributed and large scientific workflows
- Optimizing behavior of workflow manager, storage, compute and network during a workflow cycle
- Unsupervised feature extraction for studying TCP algorithm behavior




TCP behavior analysis

- Study TCP statistics for TCP cubic, TCP reno and TCP Hamilton for behavior differences using machine learning feature extraction
- Understand and predict packet loss, packet duplication and reordering to improve TCP flows
- Improve bandwidth utilization and TCP slow start behavior

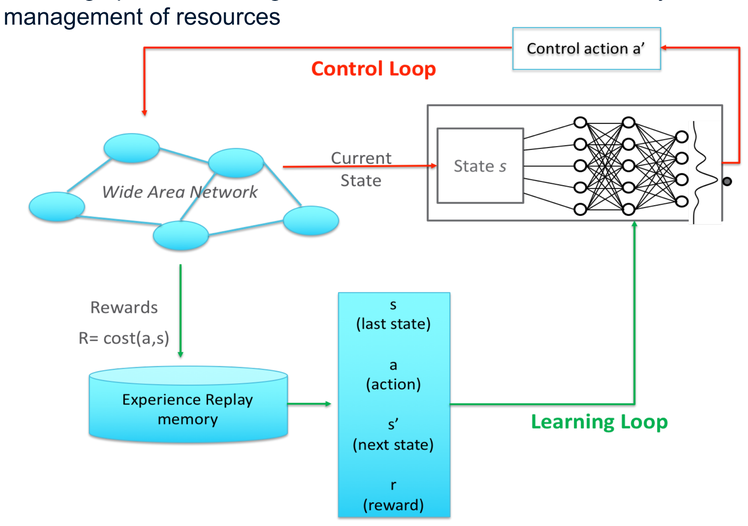


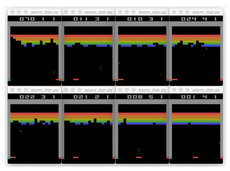

- PCA, Autoencoders and Isolation Forest for distinguishing good and bad transfers
- Developing new decision tree algorithms to extract abnormal transfer behavior features in large science transfers

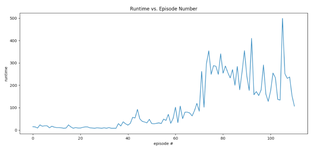
## Deep Learning and Control

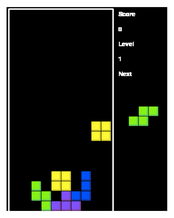
Learning optimal actions given current environment state for dynamic management of resources




- Learning through trial and error (state, action, Q(state,action))
- Deep neural networks to capture more detail through hidden states
- Improve the reward captured through many episodes played
- Multiple RL agents across the WAN layers
- Model-based and model-free learning
- Cost optimization function approximation



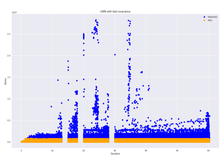



- First year explored unsupervised machine learning methods

*Anomalies in link performance: ARIMA*


*Predicting traffic topologies across DOE sites: Markov Models*

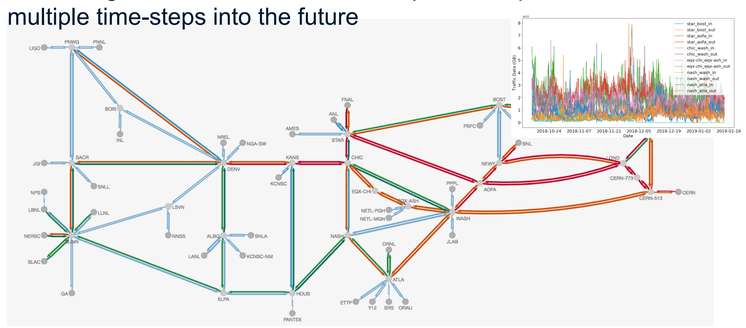


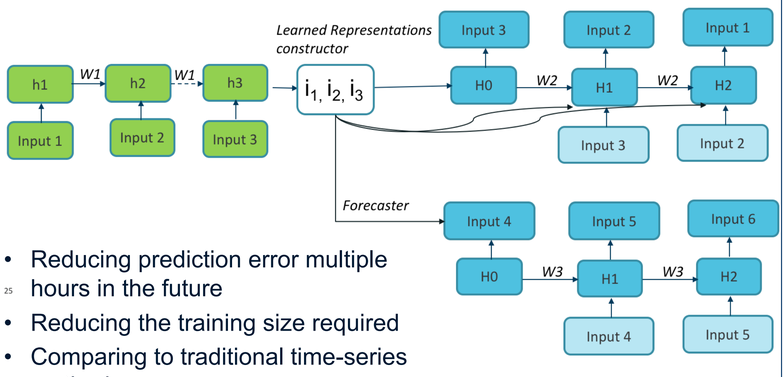
*Classifying flows across DOE sites: Gaussian Mixture Models*

- Future is exploring Reinforcement and Control algorithms

## Novel Time-Series Prediction Methods

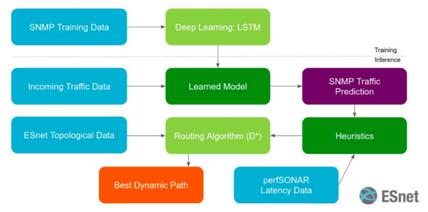
- Developing improved time-series prediction algorithms to capture and predict random user behavior on network use
- Advancing current LSTM models to improve the prediction error multiple time-steps into the future





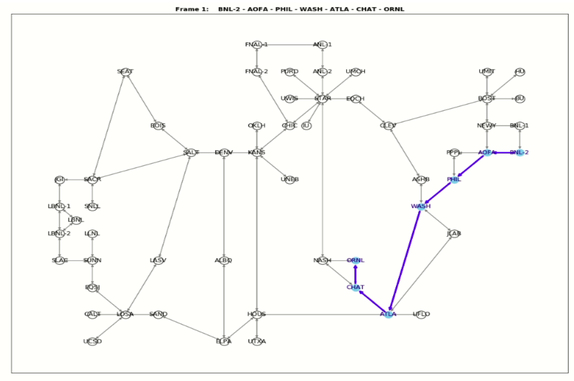

- Reducing prediction error multiple hours in the future
- Reducing the training size required
- Comparing to traditional time-series methods

## Scheduling and Optimizing Link Utilization




**DeepRoute***

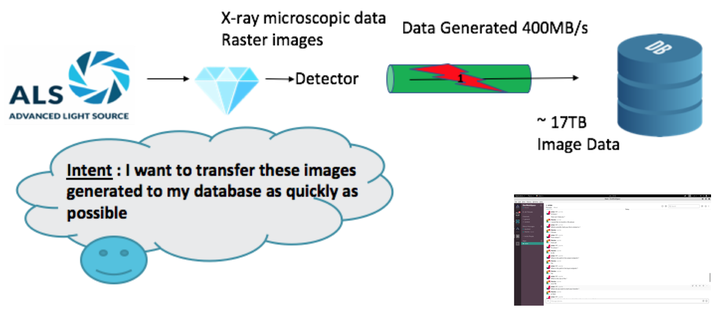
- Developing heuristic-informed path routing versus distance-vector (Bellman-Ford) and Link-state (Dijkstra) planning
- D* dynamic variant of A* route planning
- Current testing shows looping in the algorithm
- Advance based on utilized bandwidth for dynamic scheduling of transfers




- Finding optimal routes vs shortest possible routes
- Reducing flow completion times
- Utilize links to highest capacity (current cap 40% fault tolerance)

## Intent-based Network Research

**Tell me WHAT not HOW**




- Advancing intent-based projects INDIRA and EVIAN to enable superfacility use cases using Bots, ML and negotiation
- Planned network features:
  - Check availability of resources
  - Minimize transfer time
  - Optimizing configuration
  - Scheduling

Collaborations with NERSC, superfacility teams and ALS. This is ongoing work at LBNL

## Summary

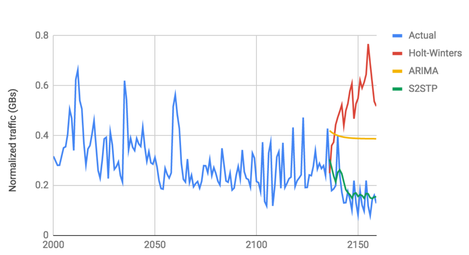
**Goal: Optimize and improve network resources by automating the network actions to make it an autonomic network**

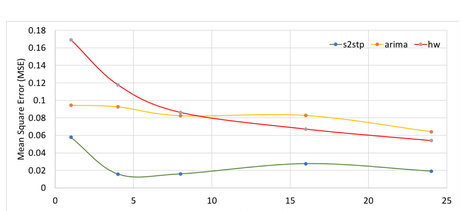
Real-time ML and control for network management

Next steps:

- Dynamic optimization of link utilization
- ML-based time-series prediction algorithms for predicting time-based behavior
- Intent-based resource allocation
- Autonomic network use cases with self-X properties (Self-configure, Self-optimize, Self-healing, Self-protect)

Collaborating Institutions: Renci, USC, ANL, ORNL, LLNL, FNAL, UCSD, UCDavis




*Funding: PANORAMA 0000232004 and Deep Learning : FP00006145*