PRIONN: Predicting Runtime and IO using Neural Networks

Michael R. Wyatt II[1,2,3] Stephen Herbein[1,2] Todd Gamblin[1] Adam Moody[1] Dong H. Ahn[1] Michela Taufer[2,3]
[1] Lawrence Livermore National Laboratory | [2] University of Delaware | [3] University of Tennessee Knoxville

Resource Prediction Challenges

- Next generation HPC schedulers rely on information about job resource usage to improve machine utilization [1]
- Analysis of job scripts via machine learning models can be used to improve knowledge of resource usage for HPC jobs [2]

Accurate Resource Usage Knowledge → Better Scheduling

- **Traditional job script analysis** (i.e., learning from parsed features of job scripts) **represents each job with a limited number of features**
  - Jobs look the same from the perspective of parsed features

```
#!/bin/bash
#PBS -l partition=cab
#PBS -l nodes=4
#PBS -l walltime=16:00:00
#PBS -q pbatch

cd $HOME/project_A
srun -n 64 ./GEOS -i Prop_bx_2a.xml
```

<cab, 4, 16, pbatch, GEOS> → F(x) = ? → Runtime: 12 hours

```
#!/bin/bash
#PBS -l partition=cab
#PBS -l nodes=4
#PBS -l walltime=16:00:00
#PBS -q pbatch

cd $HOME/dev_new
srun -n 64 ./GEOS -i Prop_bx_1b.xml
```

<cab, 4, 16, pbatch, GEOS> → F(x) = ? → Runtime: 6 hours

Traditional job script parsing cannot differentiate between these job scripts and provide accurate resource usage prediction

- Small differences between two job scripts often create large differences in resource usage
- Important additional information present in job scripts (highlighted in red) cannot be parsed with traditional methods

Image-Like Job Scripts

- We expand the representation of a job beyond parsed features:
  - **Entire job script text**
- The text in each job script is mapped to an image-like representation
  - We use the Word2Vec algorithm to transform each character in a job script to a vector (i.e., pixel)
- A 2D Convolution Neural Network (CNN) analyzes entire job scripts
  - CNNs automatically detect and learn features in unparsed text

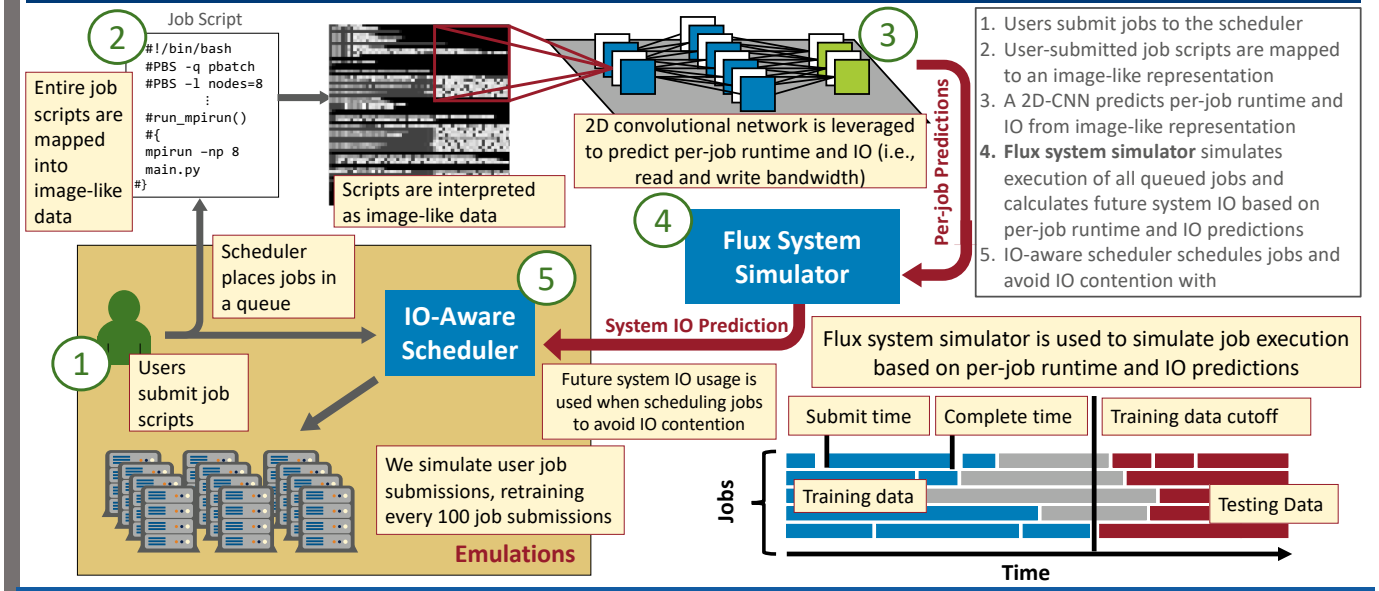
```
#!/bin/bash
#PBS -l partition=cab
#PBS -l nodes=4
#PBS -l walltime=16:00:00
#PBS -q pbatch

cd $HOME/dev_new
srun -n 64 ./GEOS -i bx_1b.xml
```

We use entire job scripts as input to deep learning models for predicting runtime and IO of HPC jobs. The text is automatically mapped to an image-like representation.

Runtime and IO Prediction Workflow for IO-Aware Scheduling

Can we leverage deep learning models to automatically predict job resource usage based on entire user-submitted job script inputs?

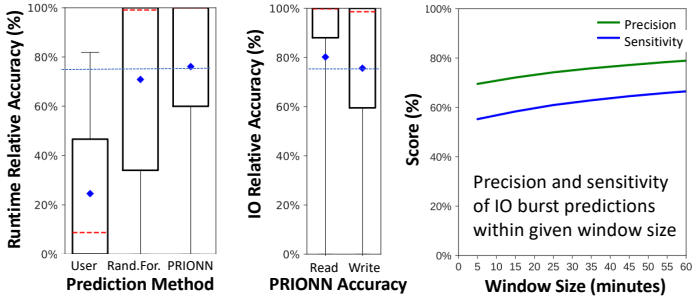


Forecasting System IO Contention

- We evaluate PRIONN's ability to predict per-job runtime and IO usage with 300,000 jobs run on Cab at LLNL during 2016
  - We measure the relative accuracy of predictions to the true resource usage for each job

$$Relative\ Accuracy\ (\%) = 100 * \left(1 - \frac{|pred - true|}{\max(pred, true)}\right)$$

- We use the **Flux** to simulate job execution with **PRIONN's** per-job runtime and IO predictions. From this, we obtain a forecast of future IO, including future **IO bursts** (i.e., high IO activity representing potential for IO contention)
- IO burst predictions enable the scheduler to avoid IO contention



PRIONN predicts per-job runtime (left) and IO usage (middle) with over 75% mean and 98% median accuracy. PRIONN + Flux system simulator predicts IO bursts with a precision of up to 80% with a window size of 60 minutes (right).

References:
[1] Herbein et al. "Scalable I/O-Aware Job Scheduling for Burst Buffer Enabled HPC Clusters." HPDC 2016; [2] McKenna et al. "Machine Learning Predictions of Runtime and IO Traffic on High-End Clusters." Cluster 2016.

LLNL-POST-749881
This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344
This material is based upon work supported by the National Science Foundation under grant no. CCF 1318445