USC VITERBI SCHOOL OF ENGINEERING

USC Autonomous Networks Research Group

ANRG

CIRCE – A runtime scheduler for DAG-based dispersed computing

Aleksandra Knezevic, Quynh Nguyen, Jason A. Tran, Praditpa Ghosh, Pranav Sakulkar, Bhaskar Krishnamachari and Murali Annavaram
(aleksank, quynhngu, jasontra, pradiptg, sakulkar, bkrishna, annavara@usc.edu)
Department of Electrical Engineering – Systems, Viterbi School of Engineering, University of Southern California, LA, CA

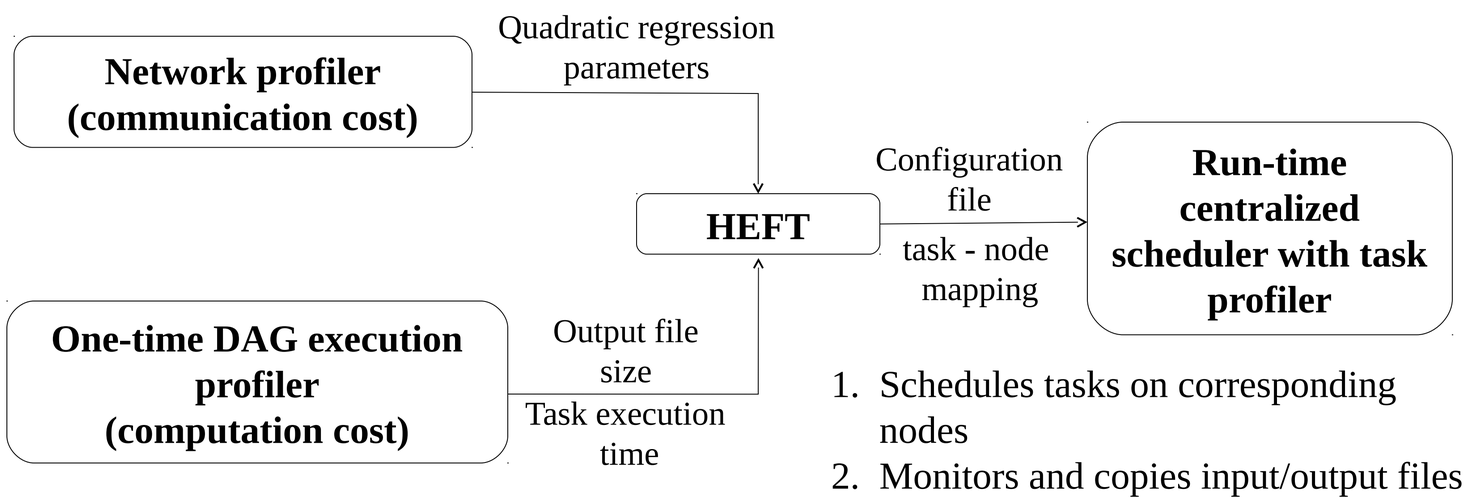
Introduction

- CIRCE (CentralIzed Runtime sChedulEr) is a runtime scheduling software tool for dispersed computing, written in Python. It can deploy pipelined computations described in the form of a Directed Acyclic Graph (DAG) on multiple geographically dispersed compute nodes at the edge and in the cloud.
- CIRCE has been released as an open source software tool, available for download at https://github.com/ANRGUSC/CIRCE.

Description

CIRCE System Components

CIRCE runs in several phases given in the Figure below.



1. Schedules tasks on corresponding nodes
2. Monitors and copies input/output files from node to node
3. Measures online task execution time & file size for feedback

1. One-time DAG execution profiler runs the whole DAG on each worker node and measures the execution time of each task and the size of the output data it passes to its child tasks.
2. Network profiler automatically schedules and logs communication information of all links between nodes in the network, which gives the quadratic regression parameters of each link representing the corresponding communication cost.

• A key innovation in this scheduler compared to prior work is the incorporation of a run-time network profiler which accounts for the network performance among nodes when scheduling.

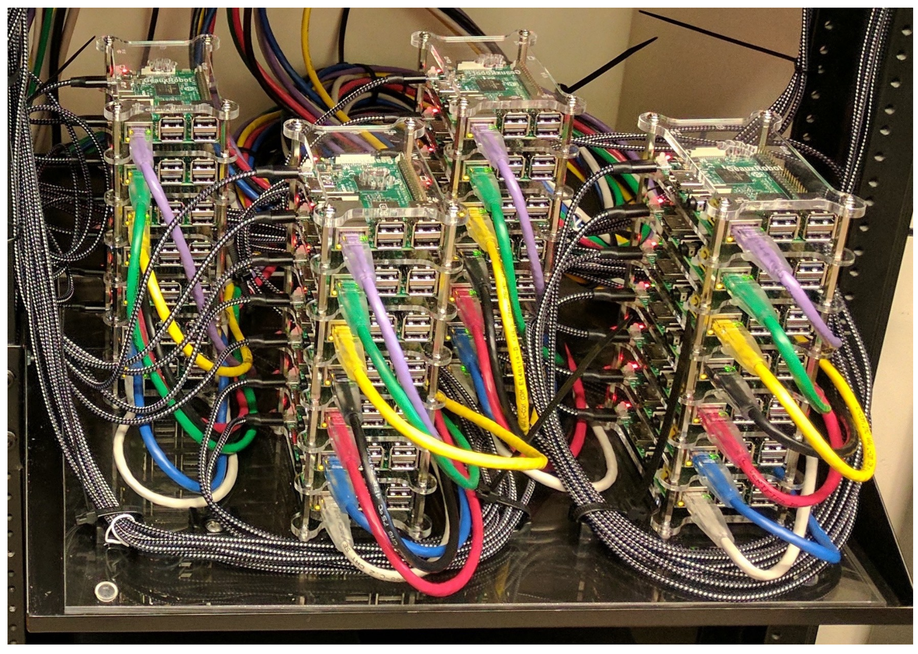
Static scheduler

• For scheduling DAG on distributed cluster we use classic **Heterogeneous Earliest Finish Time (HEFT)**.
• HEFT uses the data obtained from the Network profiler and the One-time DAG execution profiler.
• HEFT describes mapping of tasks to available computation nodes.
• We compare HEFT to random scheduler.
• Other researchers will be able to use the platform to evaluate their own scheduling algorithms as well.

CIRCE in Practice

Raspberry Pi Cluster

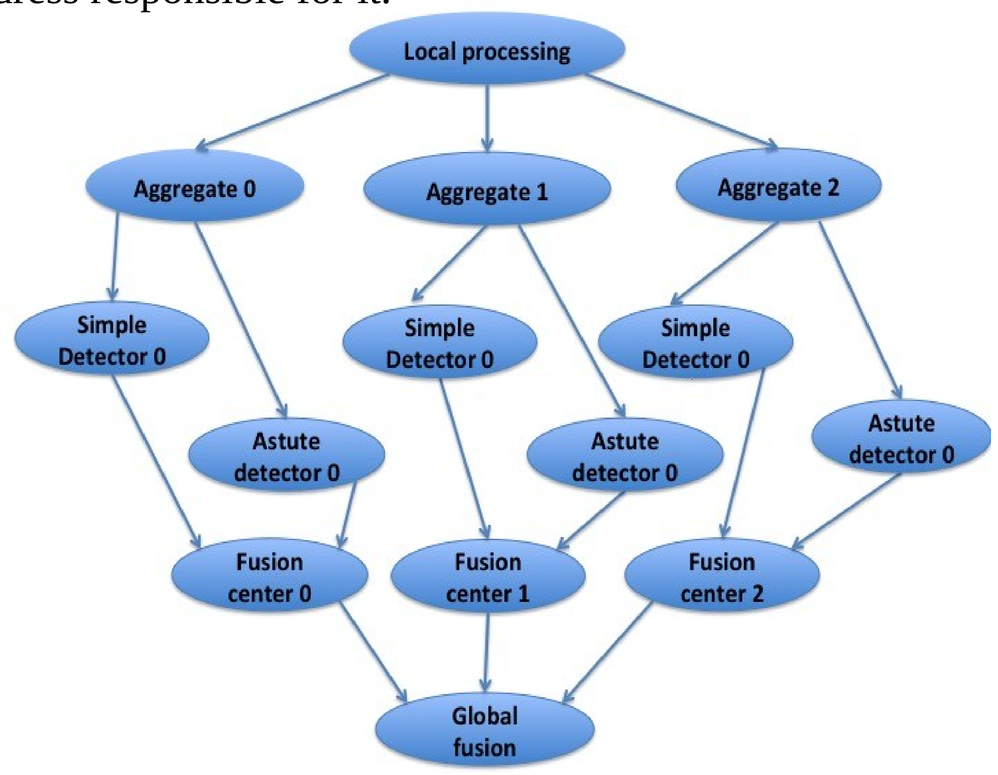
- A Raspberry Pi **edge cloud** used to deploy our software.
- One Raspberry Pi serves as **master** node.



- For realistic evaluation we are using a distributed network anomaly detection application.

Network Anomaly Detection – Task Graph

- The goal of this application is to detect anomalies in the network flow and identify the source or the destination IP address responsible for it.



This material is based upon work supported by Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001117C0053. The views, opinions, and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.