Machine-Learning-Based Coflow scheduling in OTSS-enabled Datacenters

Speaker: Lin Wang

Research Advisor: Biswanath Mukherjee

UCDAVIS

Introduction

- **Coflow**
  - Represents a collection of independent flows that shai a common performance goa
  - Coflow's performance depends on its slowest flow
  - Coflow aware scheduling benefits distributed data processing applications.
- **Coflow analysis**

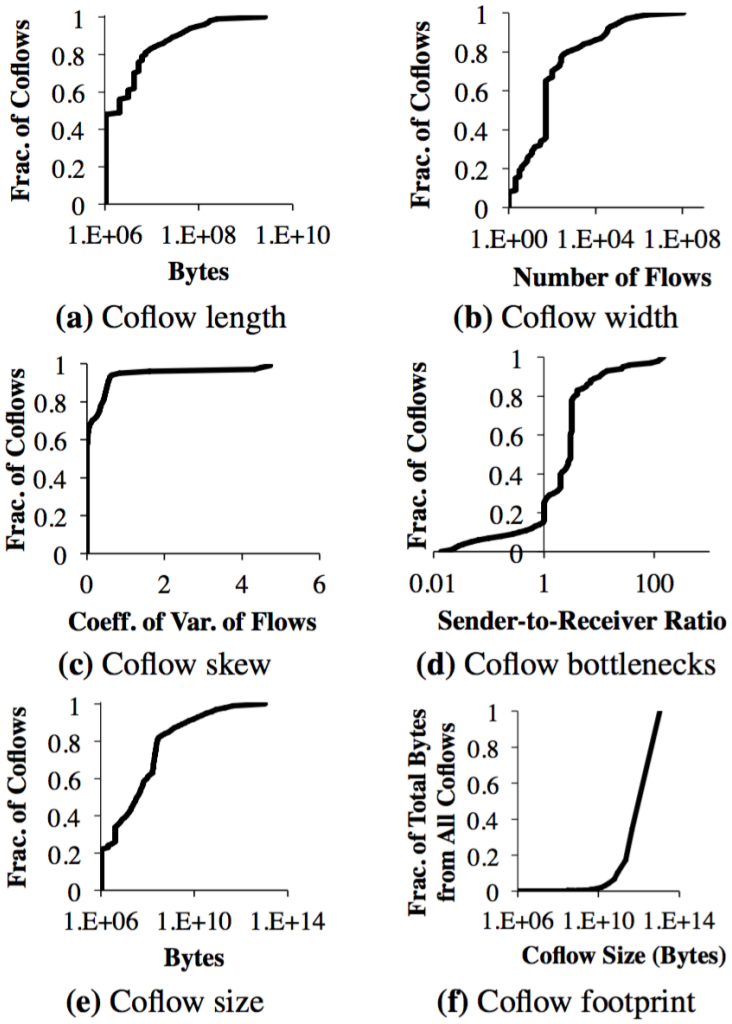


Figure 4: Coflows in production vary widely in (a) length, (b) width, (c) skew of flow sizes, (c) bottleneck locations, and (e) total size. Moreover, (f) numerous small coflows have tiny network footprint.

M. Chowdhury, et al., "Efficient coflow scheduling with varys," *ACM SIGCOMM Computer Communication*, vol. 44, no. 4, pp. 443-454, 2014.

Coflow

- **Coflow Category**
  - We consider a coflow to be short if its longest flow is less than 5 MB and narrow if it involves at most 50 flows.

Coflow Bin	1 (SN)	2 (LN)	3 (SW)	4 (LW)
Length	Short	Long	Short	Long
Width	Narrow	Narrow	Wide	Wide
% of Coflows	52%	16%	15%	17%
% of Bytes	0.01%	0.67%	0.22%	99.10%

Table 4: Coflows binned by width and length.

Coflow scheduling

- **Varys : Coflow scheduling algorithm**

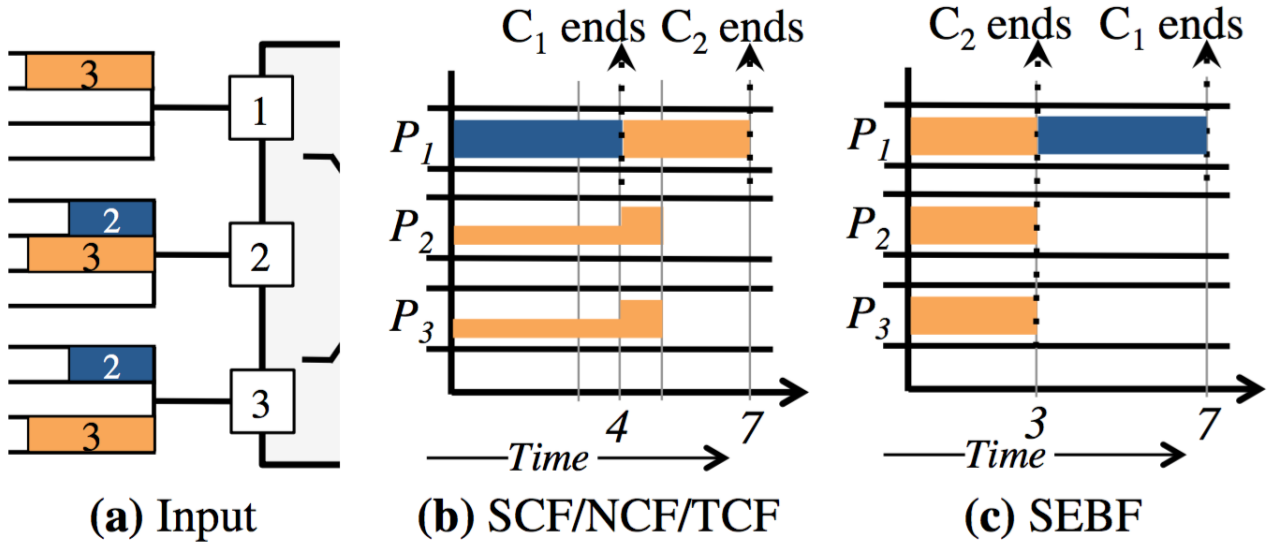
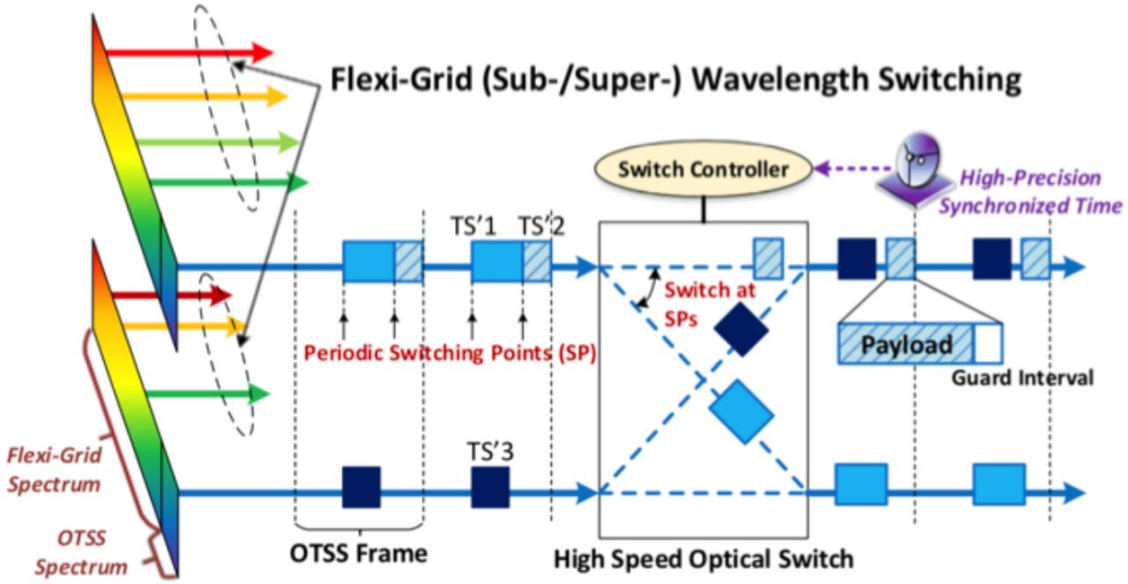


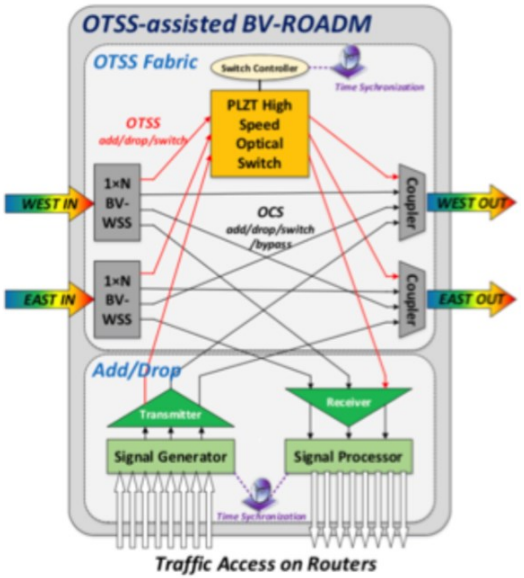
Figure 5: Allocations of *egress* port capacities (vertical axis) for the coflows in (a) on a 3 × 3 fabric for different coflow scheduling heuristics (§5.3.2).

OTSS Principle

- Designing a WDM-like TDM switching paradigm
  - WDM: all nodes have same frequency coordinate.
  - OTSS: all node should have same time coordinate!



OTSS switching principle



OTSS node architecture

Coflow scheduling

- **Coflow scheduling analysis**
  - Short Narrow (Width) coflows should be transmitted via OTSS switching paradigm.
  - Long Narrow (Width) coflows should be transmitted via WDM switching paradigm.
- **Coflow classification**
  - Using data traces from Facebook that were collected on a 3000-machine cluster with 150 racks.
  - Apply C4.5 Decision Tree (C4.5), Naïve-Bayes Discretisation (NBD) to detect whether a flow belongs to Short Coflows or Long Coflows.

M. Chowdhury, et al., "Leveraging endpoint flexibility in data-intensive clusters," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 231-242, 2013.

Coflow scheduling

- **Coflow features requirement**
  - The main limitation in choosing features is that calculation should be realistically possible within a resource-constrained datacenter network device. So, potential features need to fit the following criteria:
    - Packet payload independence.
    - No features spanning multiple flows.
    - Simple to compute.

- **Features**
  - Packet length (minimum, mean, maximum, and standard deviation).
  - Inter-arrival time between packets (minimum, mean, maximum, and standard deviation).
  - Number of packets in a flow.
  - Bytes of total packets in a flow.

M. Chowdhury, et al., "Leveraging endpoint flexibility in data-intensive clusters," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 231-242, 2013.

Machine-learning-based Coflow scheduling on OTSS

- **First step**
  - Detect whether a flow belongs to Short Coflow.
- **Second step**
  - Transmit flows belong to short coflows using OTSS switching paradigm.
  - Transmit flows belong to long coflows using WDM switching paradigm.
- **Third step**
  - Using switch controller to schedule flow transmitting.

amlwang@ucdavis.edu