Model of Stream Processing Applications

Filip Nálepa

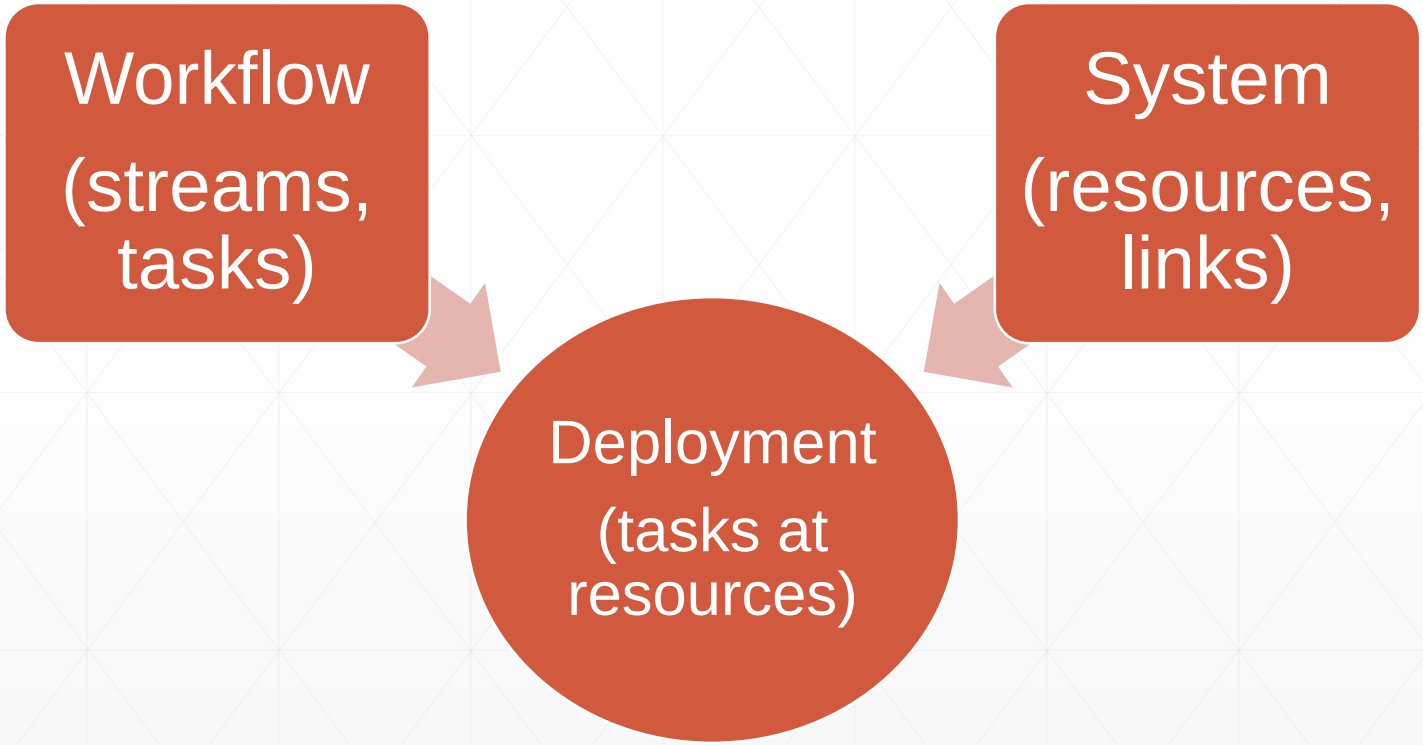
Outline

- Introduction
- Model
- Performance analysis

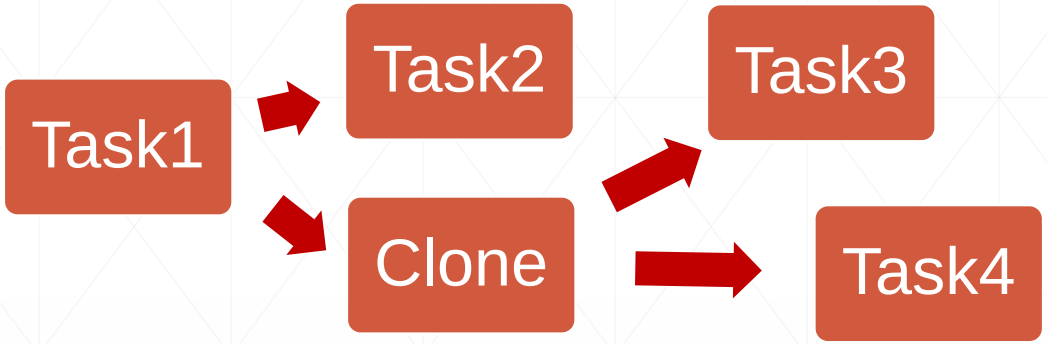
Introduction

- Stream processing application
  - Infinite streams of data
  - Tasks
  - Resources
- Focus: multimedia data, event detection
- Goal: performance analysis (delay)
- How: create a model

Model Structure



Workflow Model



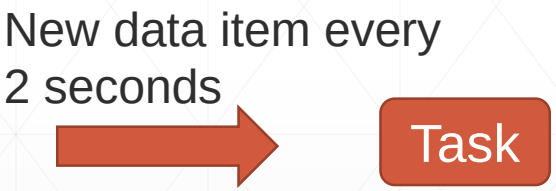
- Processing cost
- Waiting time
- Data size
- Output frequency

Processing Cost

- Multimedia data – variable cost
- Single value (average, maximum) – inaccurate

Processing Cost

- Multimedia data – variable cost
- Single value (average, maximum) – inaccurate
- Probability function – e.g., 1 s in 90 %; 4 s in 10 %

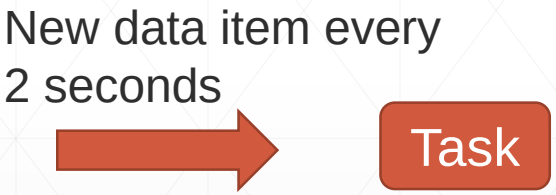


Possible cost sequences:

- 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1; maximum delay: 4 s
- 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1; maximum delay: 6 s

Processing Cost

- Multimedia data – variable cost
- Single value (average, maximum) – inaccurate
- Probability function – e.g., 1 s in 90 %; 4 s in 10 %



Possible cost sequences:

- 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1; maximum delay: 4 s
- 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1; maximum delay: 6 s

Solution: specify maximum number of 4s items in a row

Processing Cost Cumulative Function

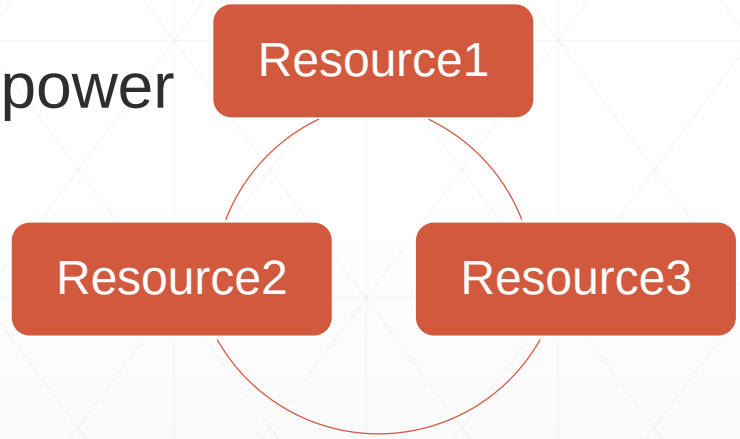
- cost(Δ) = x, where x is the maximum number of processing units (e.g., CPU cycles) needed to process any sequence of data items of length Δ
- Example:
  - Cost sequence: 4, 1, 1, 1, 1, 1, 1, 1, 1, 4, 4, 1, 1, 1, 1, 1
  - cost(1) = 4
  - cost(2) = 8
  - cost(3) = 9
- Analogically minimal cost

Workflow Model

- Processing cost – cumulative function
- Waiting time – external service, cumulative function
- Data size – cumulative function
- Output frequency – cumulative function
  - Time based – maximal/minimal number of items output per a given time interval
  - Processed items based – maximal/minimal number of items output per any sequence of processed data items of a given length

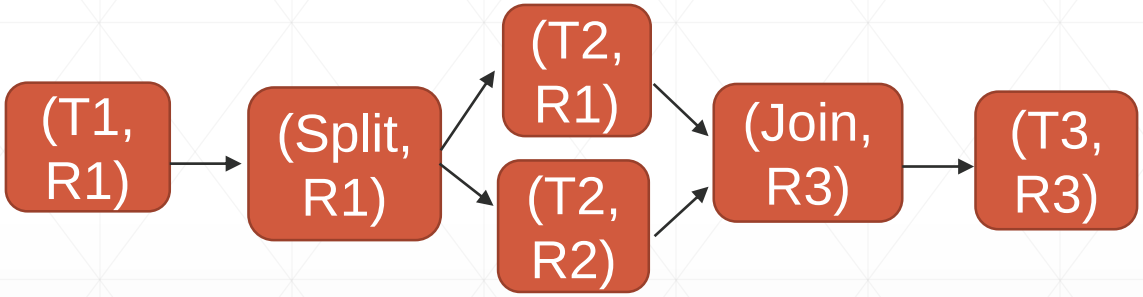
System Model

- Infrastructure
- Computational power
- Bandwidth



Deployment Model

- Tasks at resources



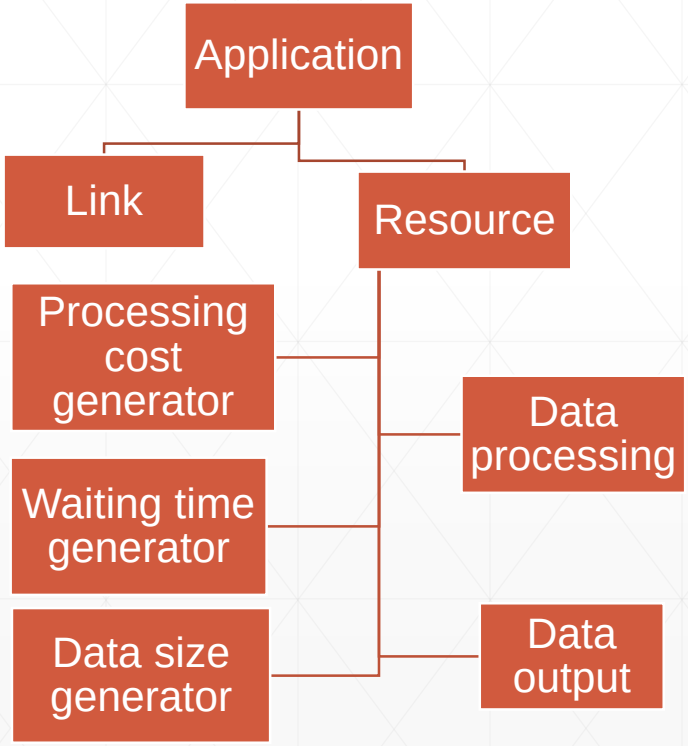
- Split – maximal/minimal number of data items for each outgoing edge per any sequence of a given length, i.e., cumulative functions
- Scheduling strategy

Performance Analysis

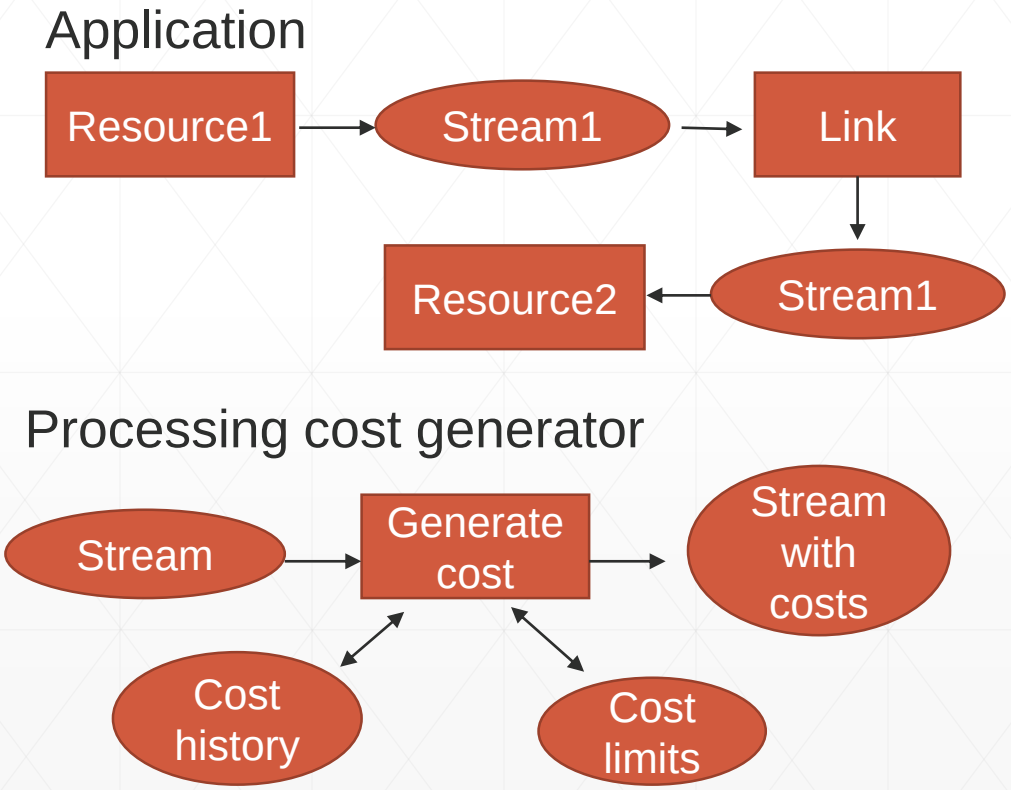
- Direct analysis of the model
- Model conversion to a well-known model of distributed systems
  - Colored Petri Nets (CPNs)
    - Formal state space exploration – computationaly intensive
    - Simulation – estimate

Model to CPN

- Hierarchical structure



Petri Net Examples



Summary

- Model of stream processing applications
  - Workflow model – tasks and streams
    - Processing cost, waiting time, data size, output frequency
  - System model – resources and links
    - Computational power, link bandwidth
  - Deployment model – tasks at resources
    - Scheduling strategy
- Performance analysis
  - Conversion to CPN
  - Simulation

Thank you for your attention