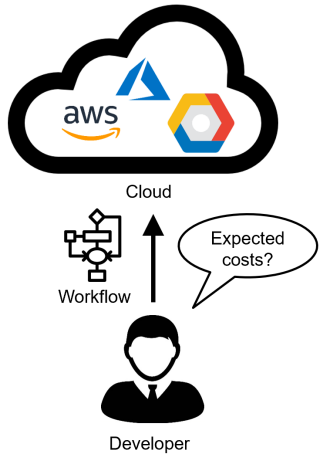
Predicting the Costs of Serverless Workflows

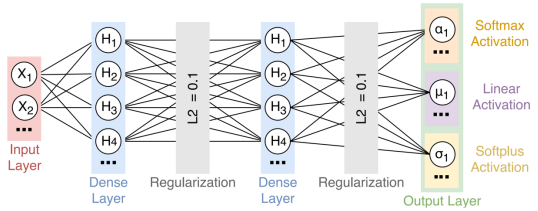
Simon Eismann, Johannes Grohmann, Erwin van Eyk, Nikolas Herbst, Samuel Kounev

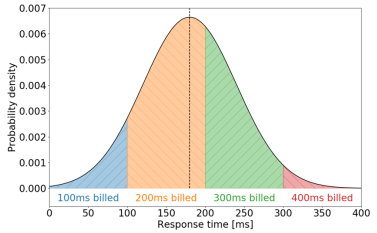
Pay-per-use makes estimating costs challenging



- Cost of serverless functions depends on [1, 2]:
  - Response time rounded to nearest 100ms
  - Function size (allocated memory/CPU)
  - Static overhead per execution
- Moreover, function response time depends on input [3]
  - Function execution in a different context changes cost
  - Makes estimation of costs for workflows challenging
- Existing approaches for cost estimation [4, 5, 6]:
  - Describe the response time as a static mean
  - Require user to estimate response time

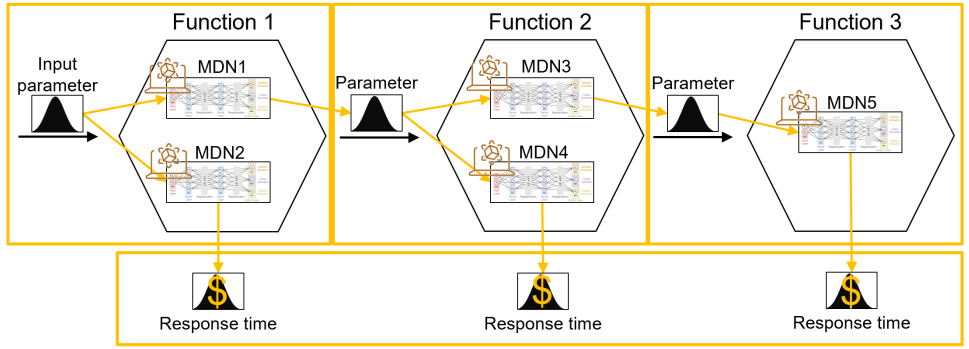
Function Response Time Distribution Prediction



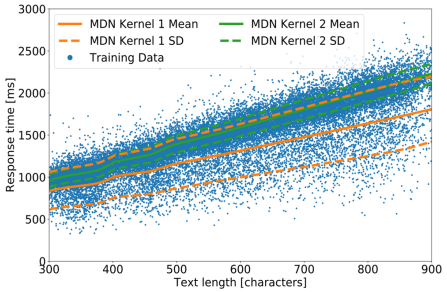


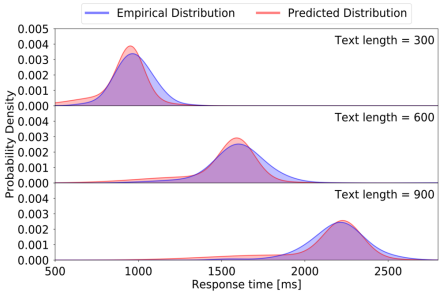
Approach

1. Model Workflow Structure
2. Integrate MDNs
3. Identify next node
4. Monte-Carlo simulation
5. Repeat steps 3+4
6. Calculate costs



Evaluation





Can we accurately predict the response time distribution of serverless function? ✓

Can we accurately predict the costs of a previously unobserved workflow? ✓

Is the overhead feasible for a production environment? ✓