



# Performance Prediction of On-NIC Network Functions with Multi-Resource Contention and Traffic Awareness

Shaofeng Wu<sup>1\*</sup> Qiang Su<sup>1\*</sup> Zhixiong Niu<sup>2</sup> Hong Xu<sup>1</sup>

<sup>1</sup>The Chinese University of Hong Kong <sup>2</sup>Microsoft Research

## Problem

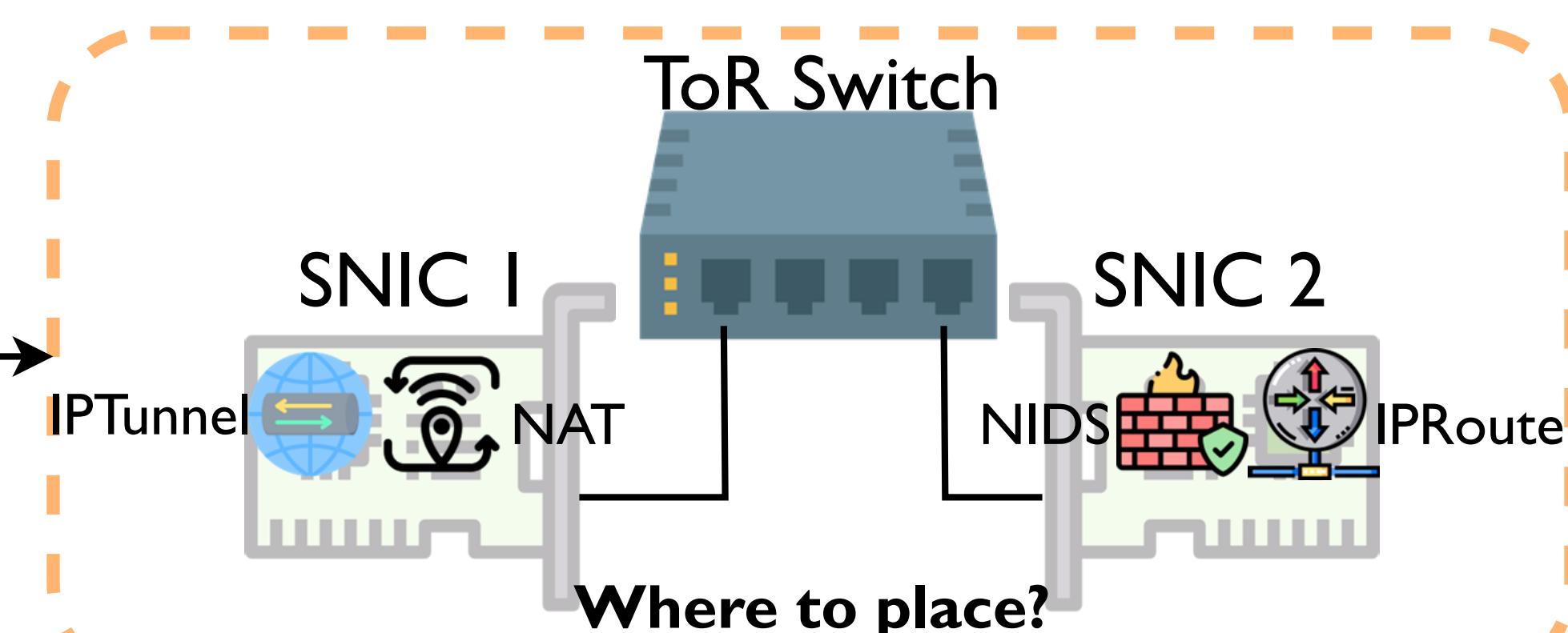
- How to predict NF performance under resource contention?

Goal:

- Minimize SLA violation
- Maximize co-locations

**NF Request**  
NF: FlowMonitor  
SLA: Tput  $\geq 1.0$  Mpps

Operator



✓ High resource utilization & Low TCO compared to monopolization

✗ Possible SLA violation

- Resource contention-induced throughput drop can be up to 60%
- Insufficient isolation on SmartNICs

FlowMonitor Tput: 1.2 Mpps (SLA satisfied)  
FlowMonitor Tput: 0.8 Mpps (SLA violated)

### Contention-Aware Scheduling:

Placing NFs based on contention-aware performance predictions

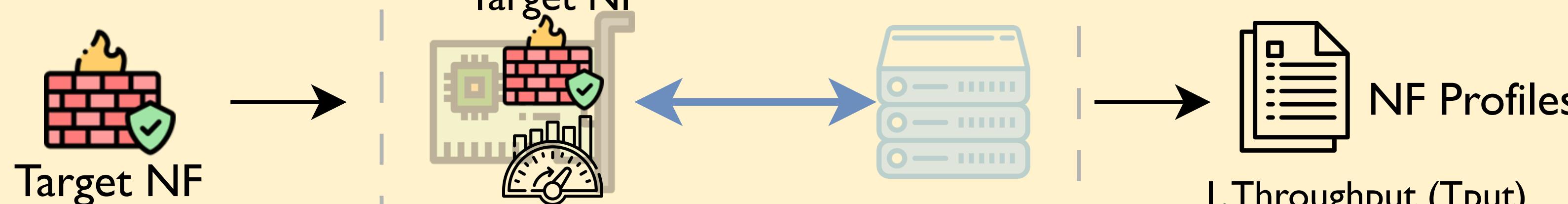
✓ Avoid SLA violation despite resource contention

$$f(\text{NIC}) = 1.3 \text{ Mpps} \quad f(\text{NIC}) = 0.7 \text{ Mpps}$$

- Our contribution: accurate performance prediction of on-NIC NFs under multi-resource contention and varying traffic attributes

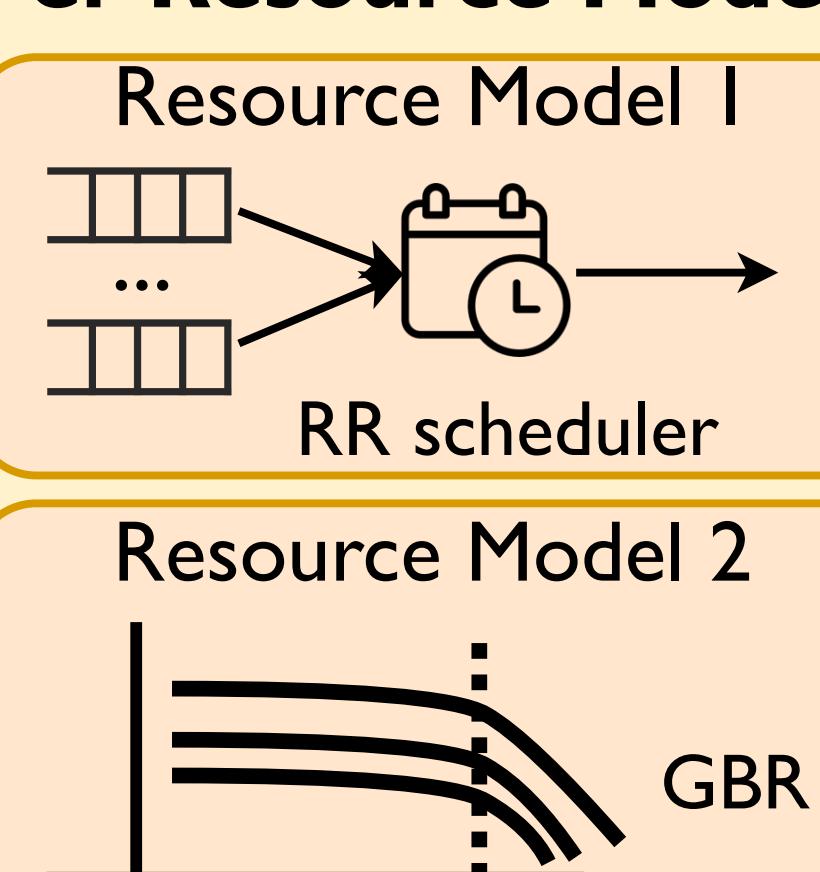
## Design Overview

### 1. Offline Profiling



### 2. Offline Training

#### Per-Resource Models



Traffic attributes (Flow count, match-to-byte-ratios, ...)

$$\text{Tput} = f(C, T)$$

### 3. Online Prediction

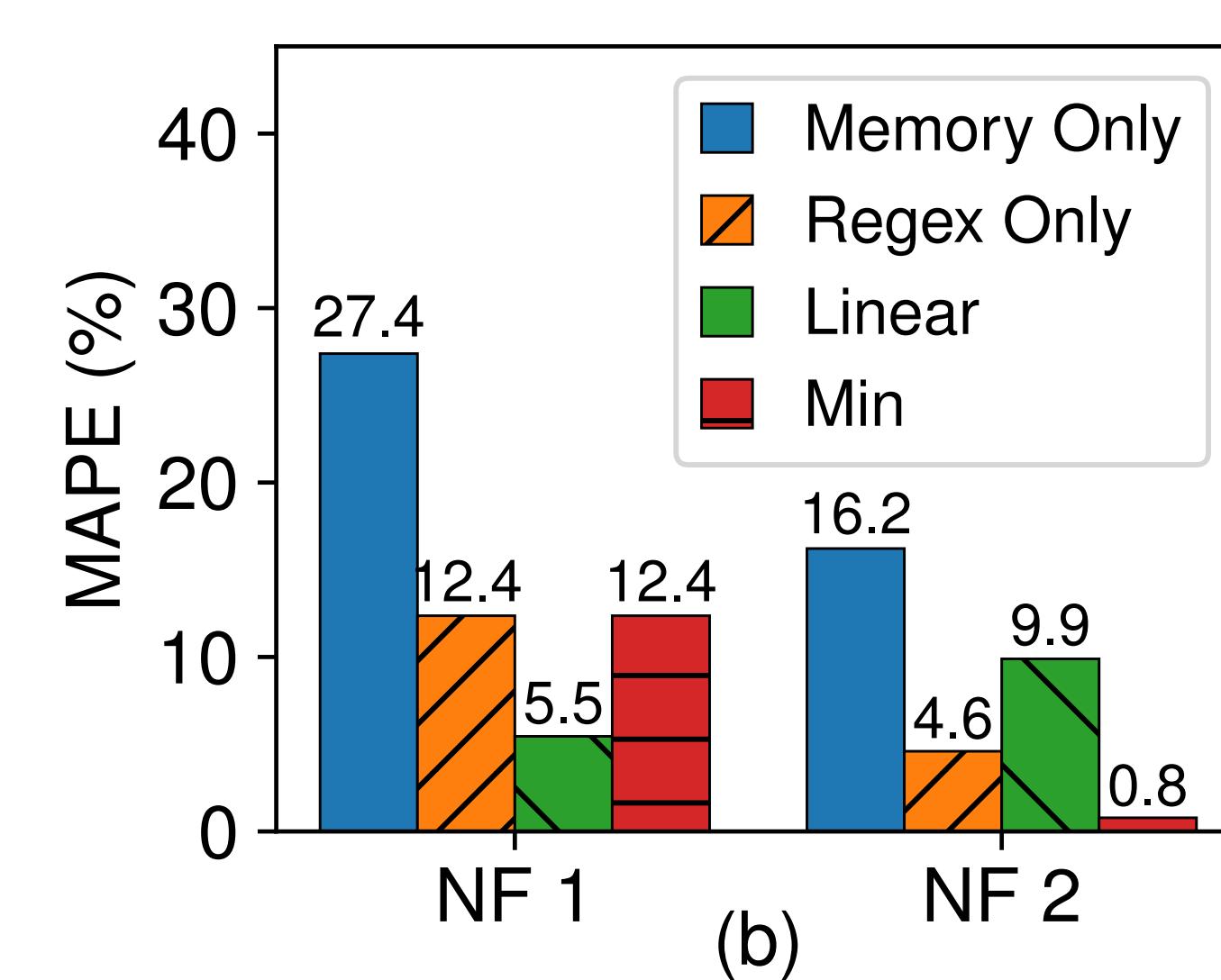
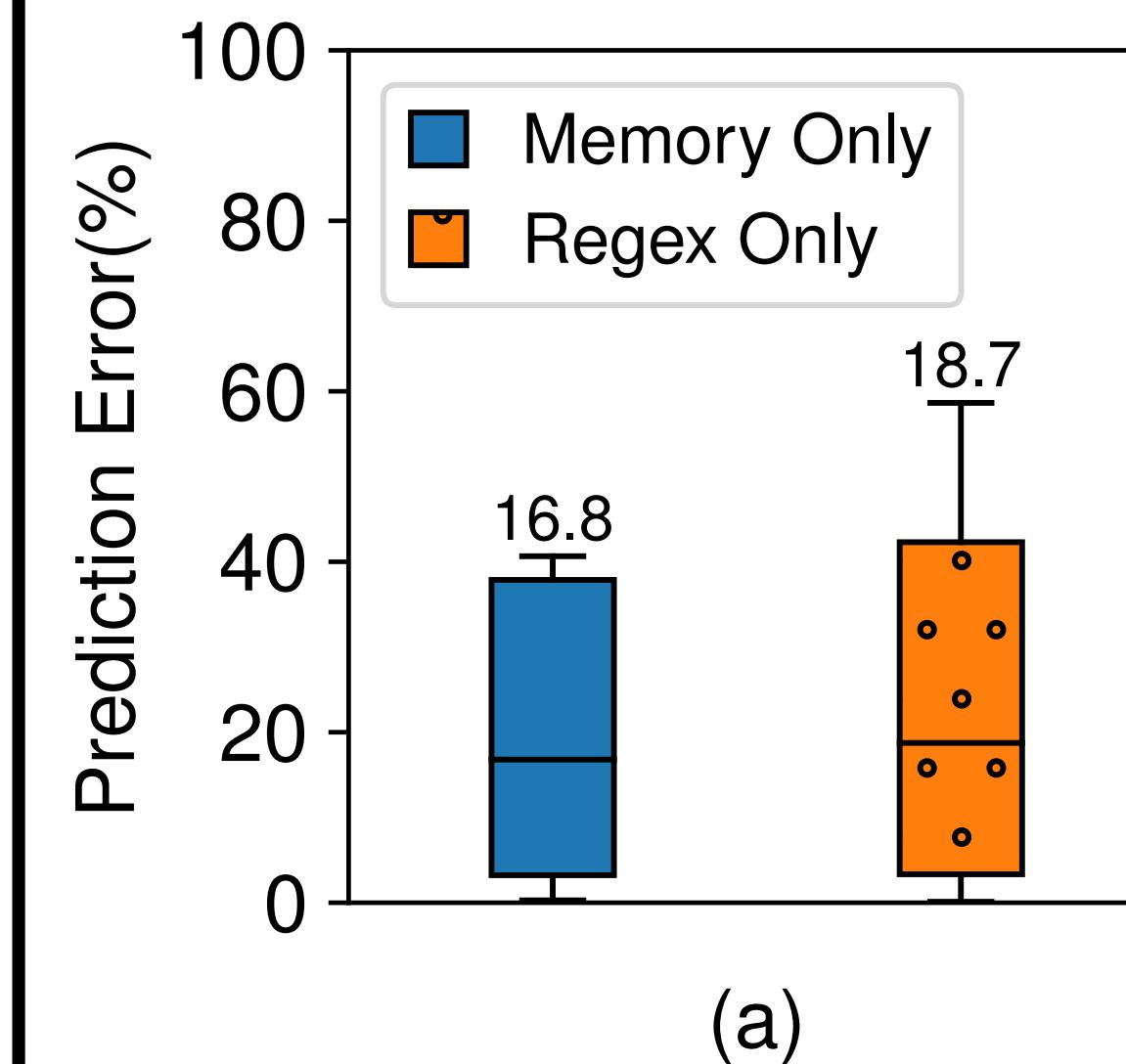
Target NF  
Competing NFs  
Tentative placement

$$\text{Tput} = f(C, T)$$

Prediction result  
Tput = X pps

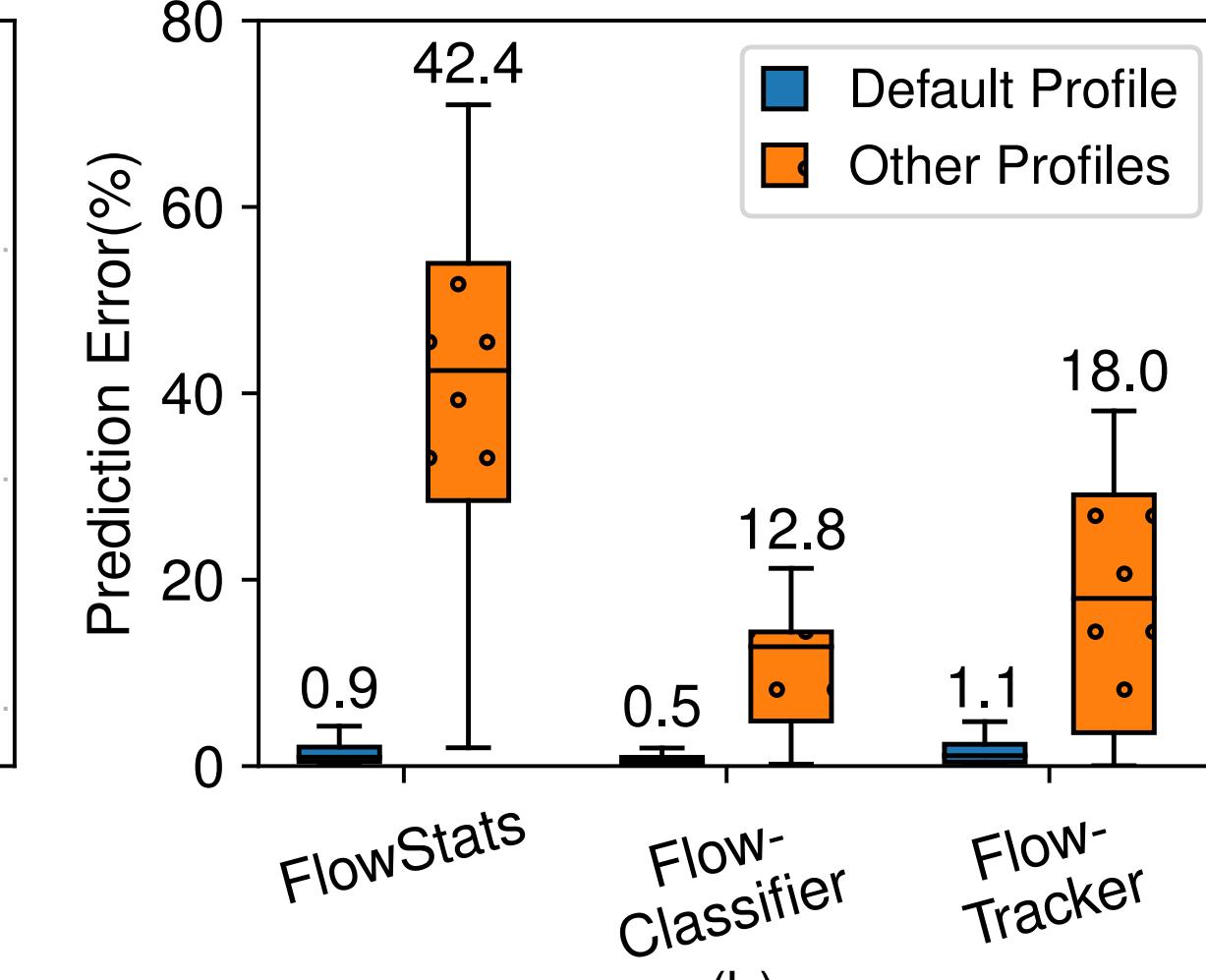
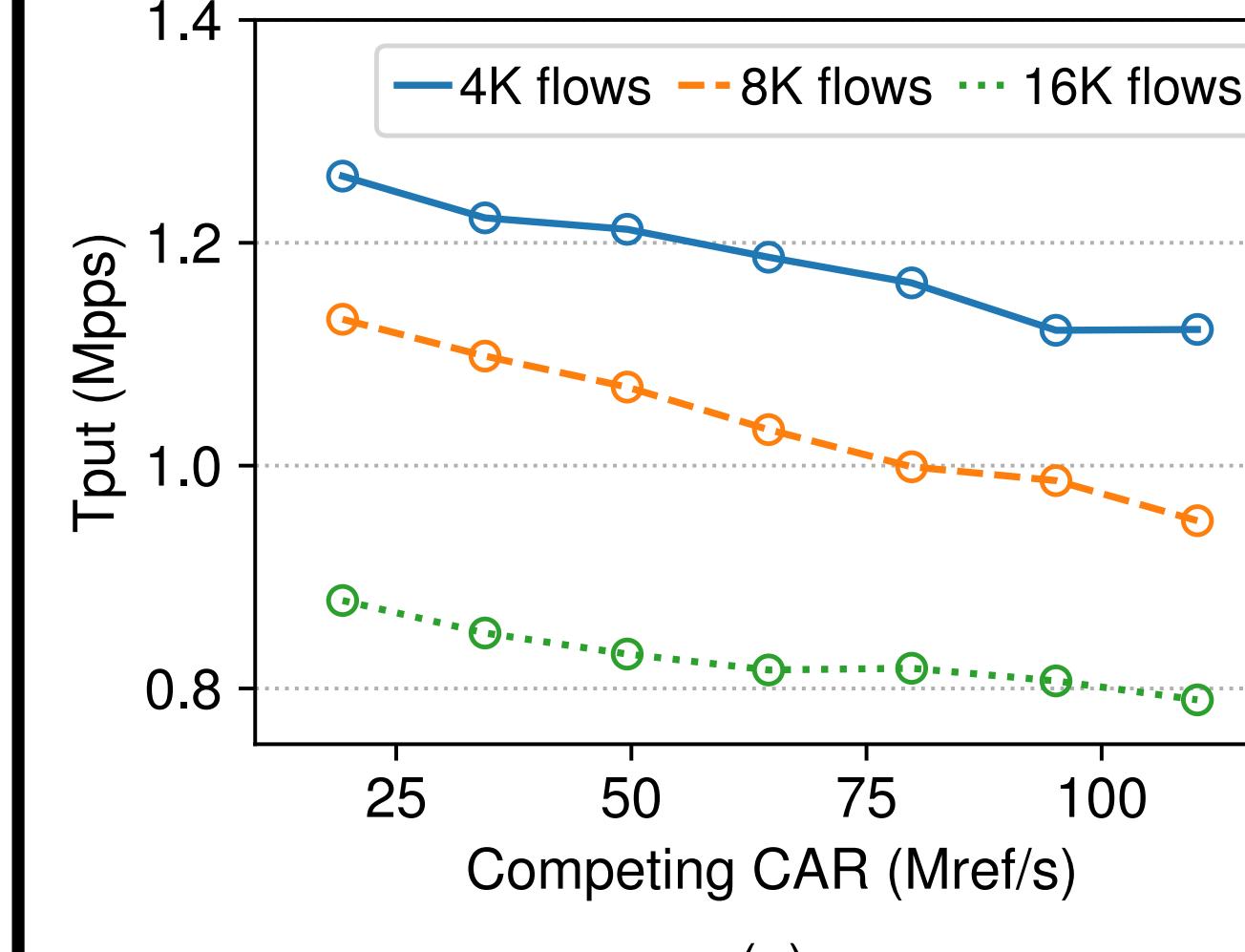
## C1. Multi-Resource Contention

- On-NIC NFs contend over multiple heterogeneous onboard resources



## C2. Varying Traffic Attributes

- NF performance depends on certain traffic attributes



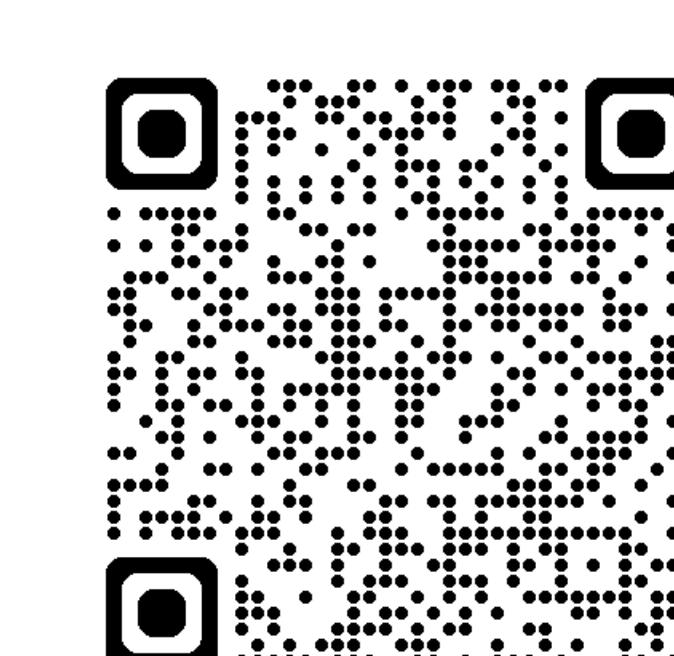
## Evaluation

- 3.7% prediction error & 78.8% accuracy improvement

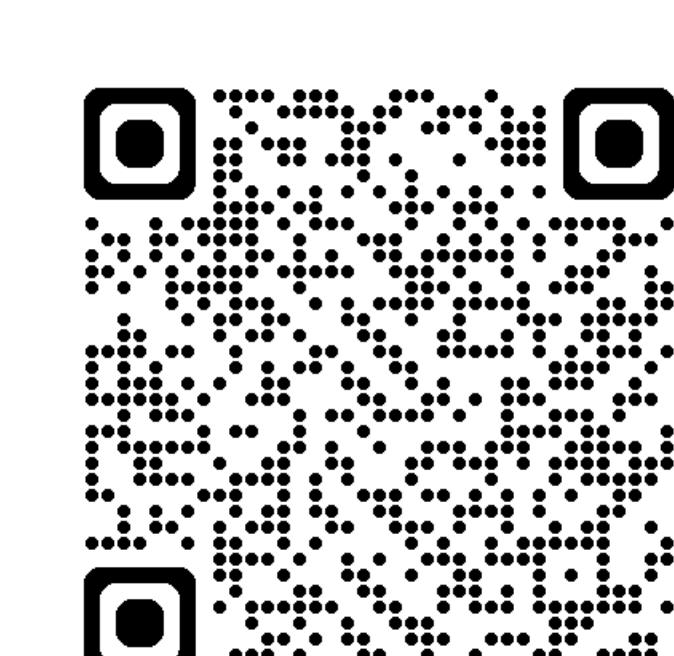
NF	SLOMO			Yala		
	MAPE (%)	±5% Acc. (%)	±10% Acc. (%)	MAPE (%)	±5% Acc. (%)	±10% Acc. (%)
ACL	1.3	100.0	100.0	<b>1.2</b>	100.0	100.0
NIDS	16.2	24.3	74.3	<b>1.5</b>	95.9	100.0
IP Tunnel	62.9	70.5	73.1	<b>3.8</b>	75.6	92.3
IP Router	4.2	<b>68.4</b>	98.2	<b>3.8</b>	66.7	100.0
FlowClassifier	7.5	28.1	73.7	<b>3.8</b>	63.2	100.0
FlowTracker	4.9	56.1	86.0	<b>3.9</b>	61.4	100.0
FlowStats	11.7	33.3	57.9	<b>4.3</b>	70.2	96.5
FlowMonitor	40.9	31.1	41.9	<b>4.5</b>	62.2	93.2
NAT	8.2	38.6	49.1	<b>6.4</b>	42.1	80.7

- Resource-efficient scheduling of on-NIC NFs, SLA violations reduced by 92.2%

Approach	Resource Wastage (%)	SLA Violations (%)
Monopolization	196.3	0
Greedy	19.0	16.5
SLOMO	-21.8	24.4
Yala	0.5	1.9



Paper



Github