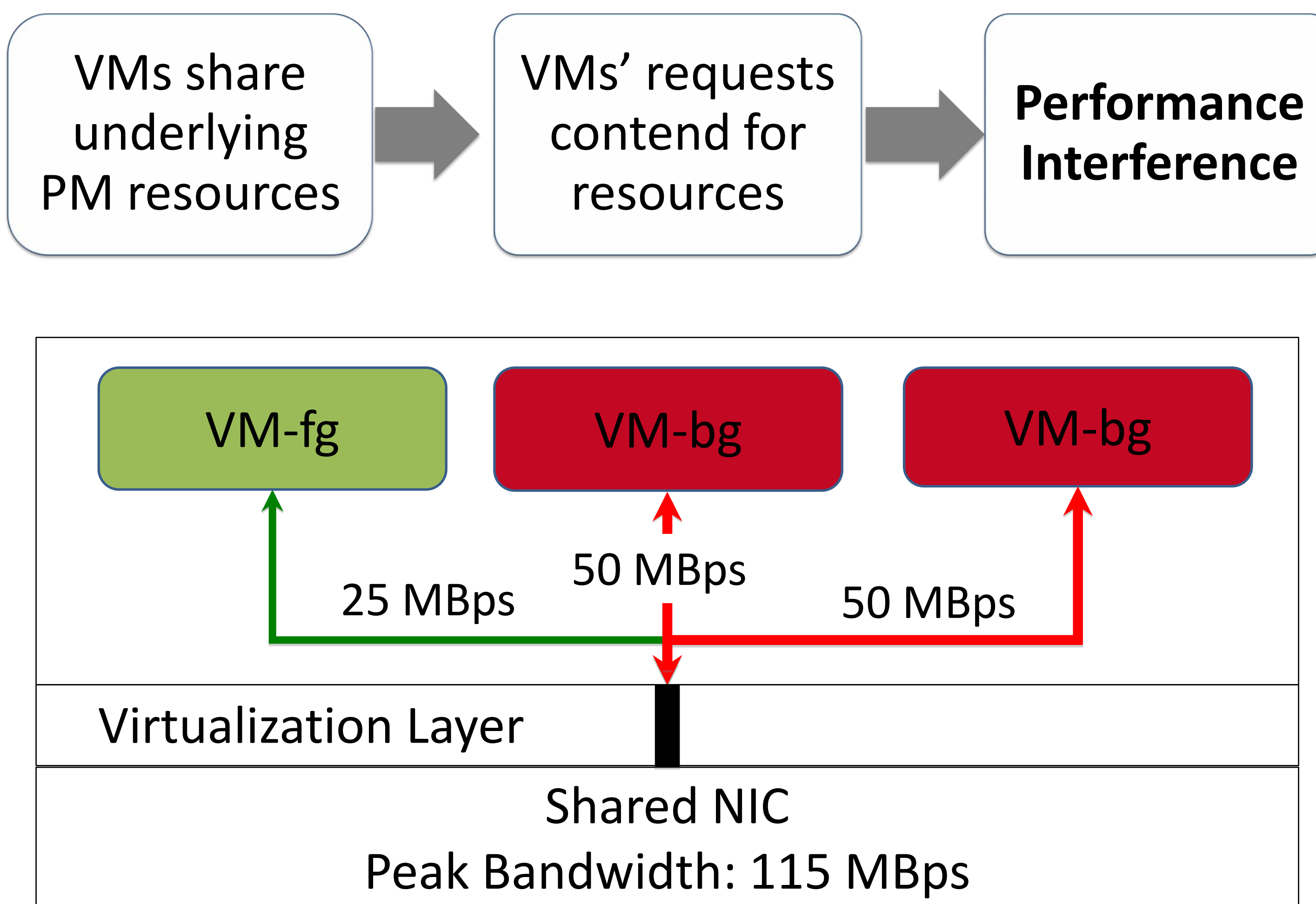


# DIAL: Dynamic Interference-Aware Load Balancing

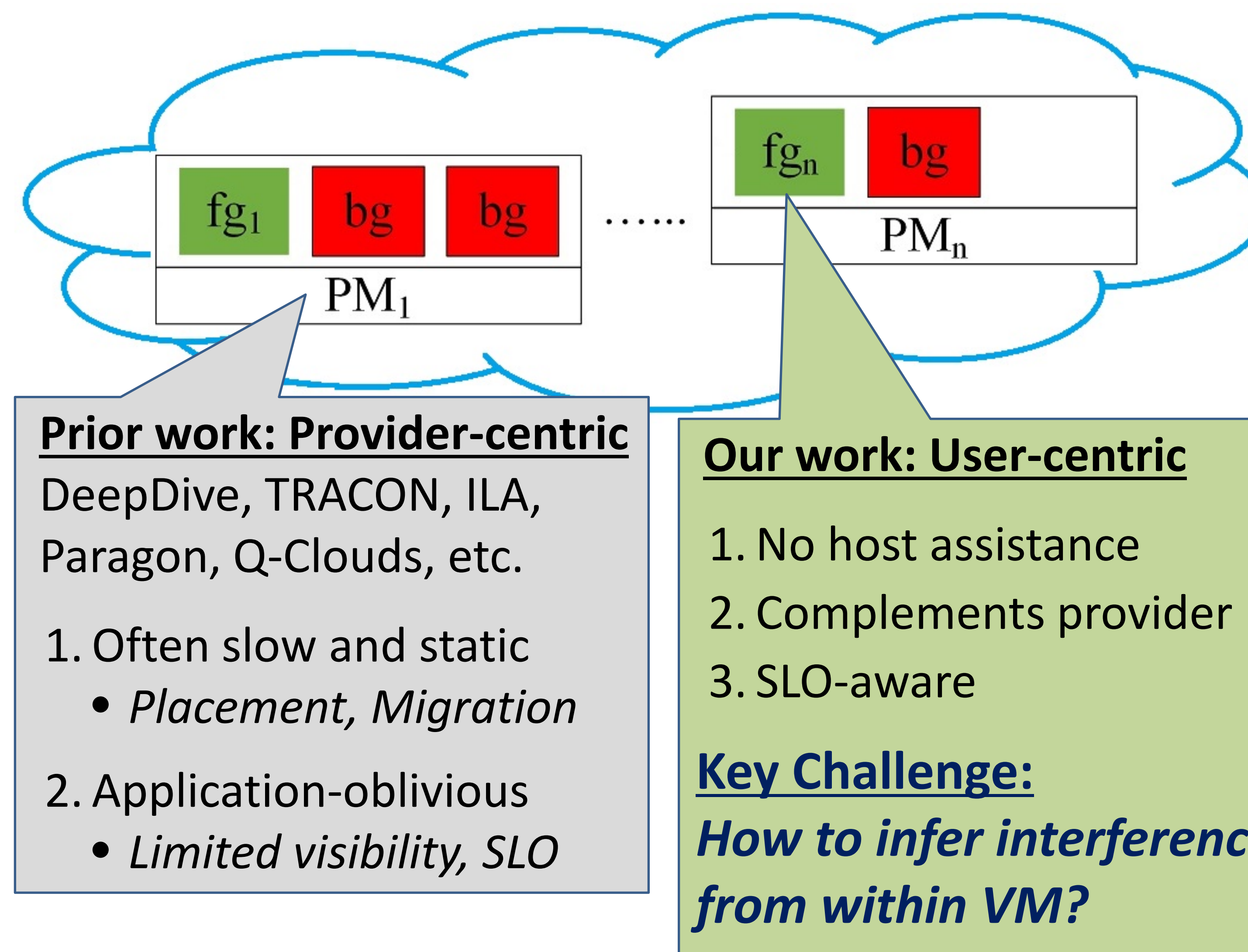
PI: Anshul Gandhi, PACE Lab, Stony Brook University

Supported by NSF CNS-1464151 (CRII) and NSF CNS-1617046 (SMALL)

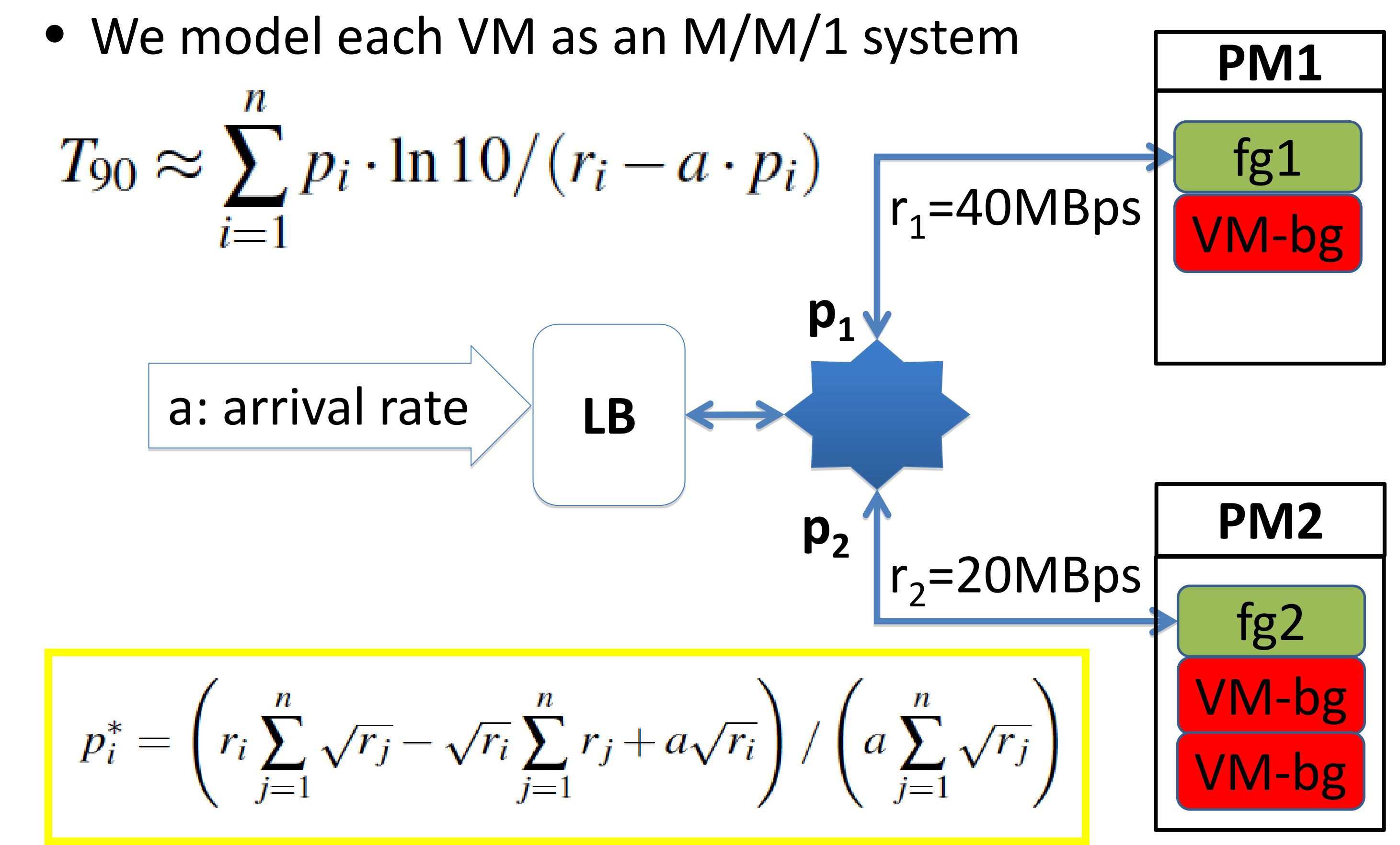
## OVERVIEW: Performance Interference



## PRIOR WORK vs OUR APPROACH

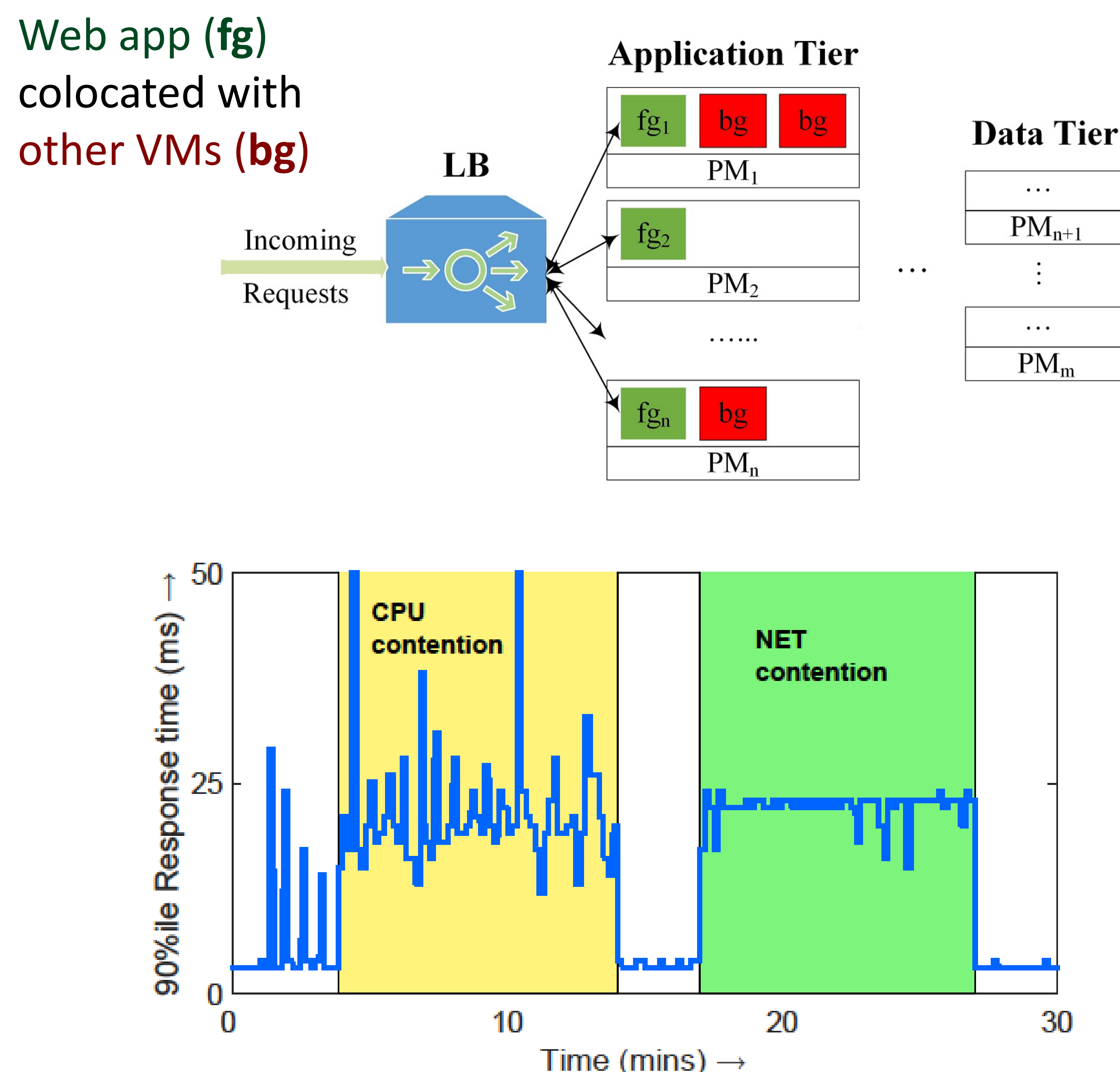


## CONTRIBUTION 2: Interference LB



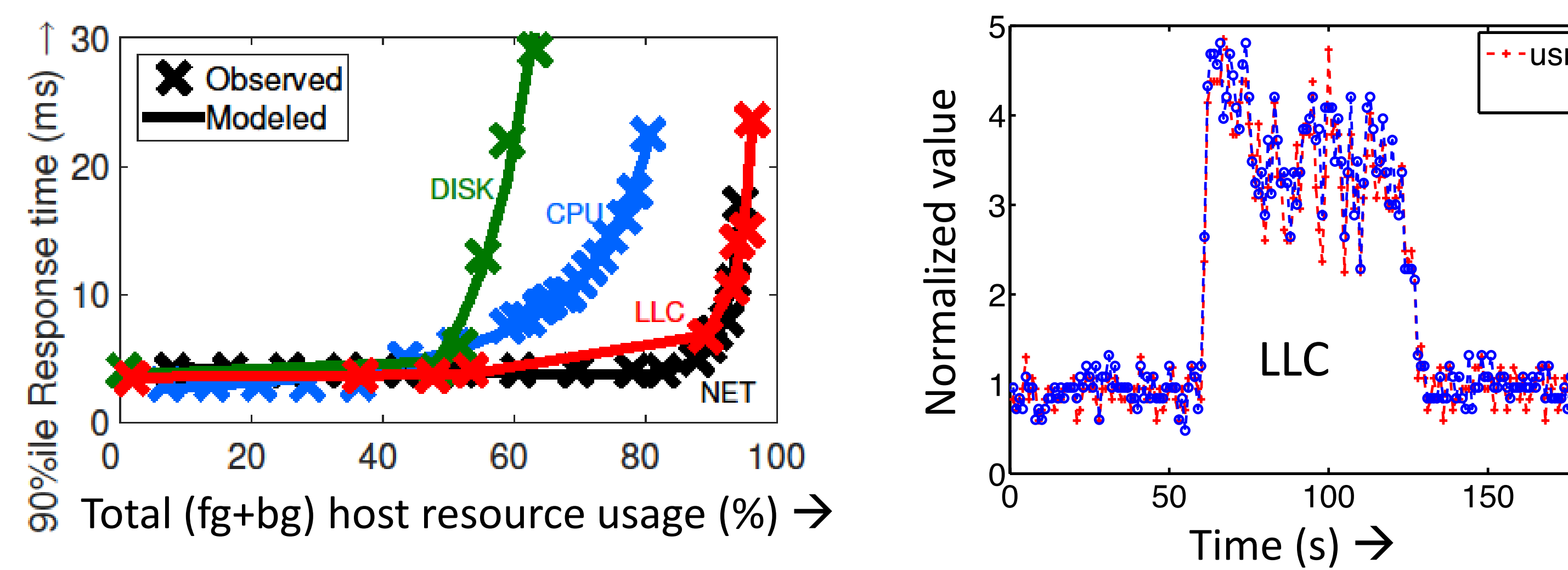
DIAL: Monitor → Detect → Classify → Estimate → LB

## PROBLEM: Tail Latencies



Tail latencies increase significantly under interference

## CONTRIBUTION 1: Interference Model



Observations:

- Non-linear increase in latencies (interference detection)
- Resource-specific curves (interference classification)
- Infer congestion from latency (interference estimation)

$$T_{90} = c_0 + \frac{c_1}{(1-load)} + \frac{c_2}{(1-load)^2} \quad load = \frac{fg\_load + bg\_load}{peak\_load}$$

Modeling Results:

- Detection:  $T_{90} > 5ms$  (5.7% false positive rate)
- Classification: Decision Trees (92.2% accurate)
- Inference: Queueing + Regression (93.9% accurate)

## DIAL EVALUATION RESULTS

