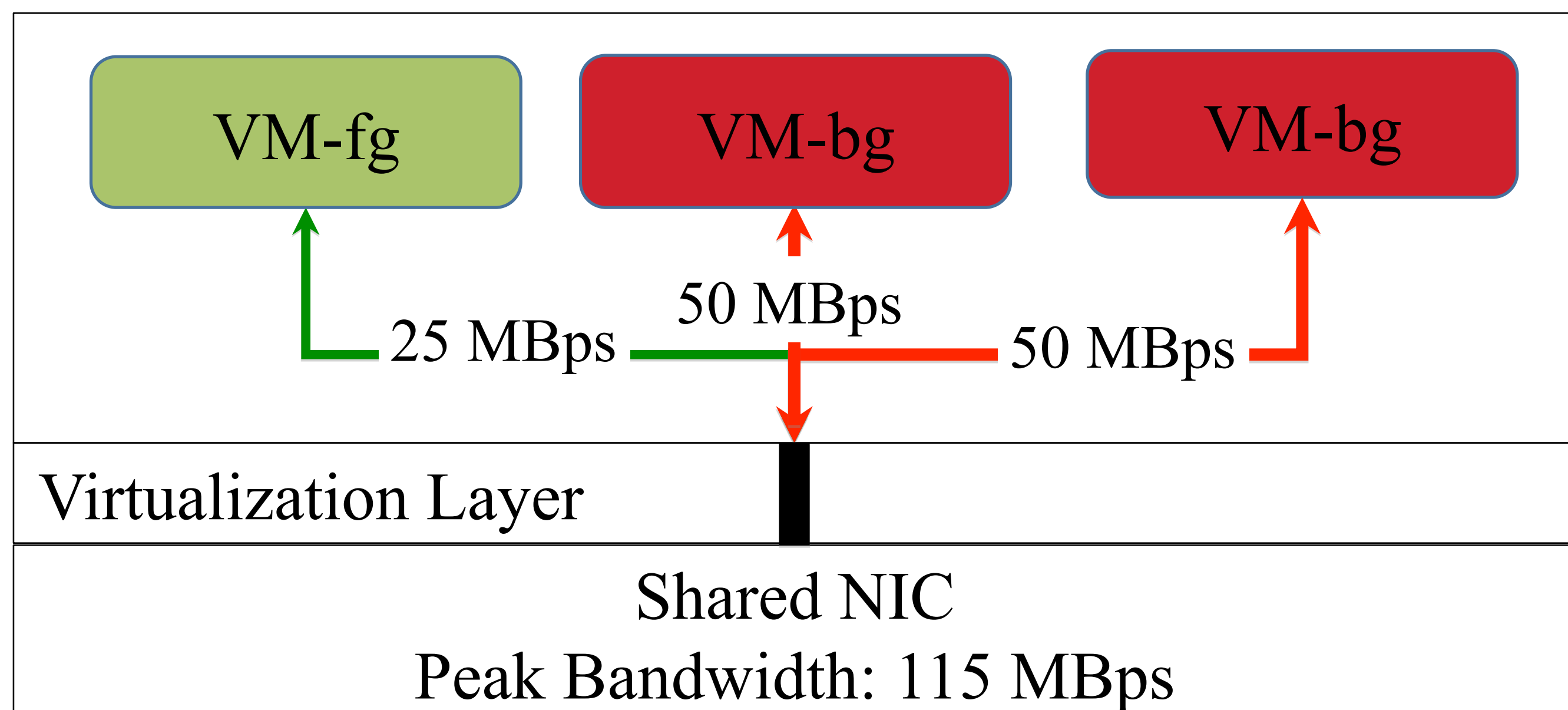
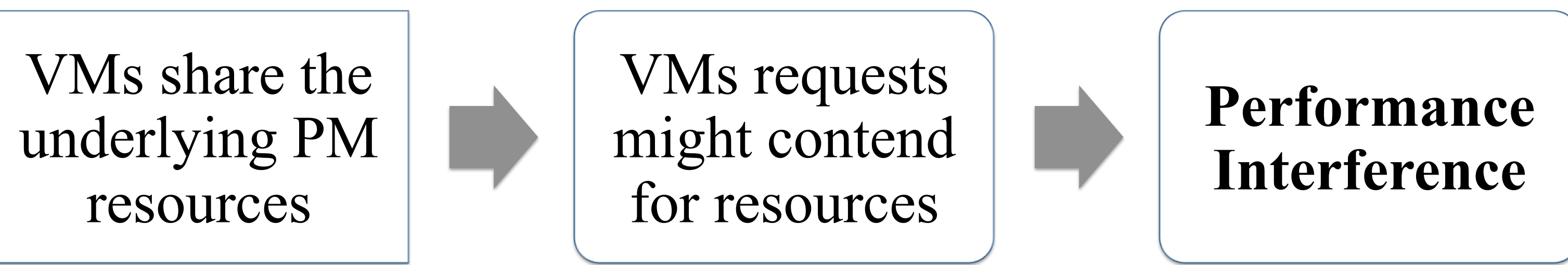


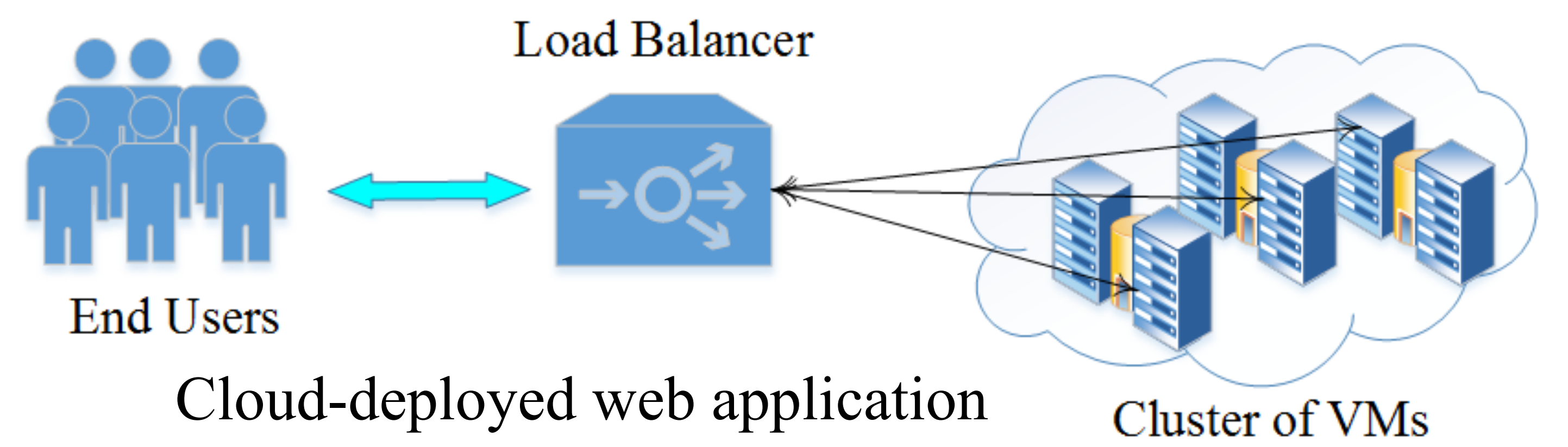
# Dynamic Interference-Aware Load Balancing

Seyyed Ahmad Javadi, Himanshu Rajput, Anshul Gandhi (Stony Brook University)

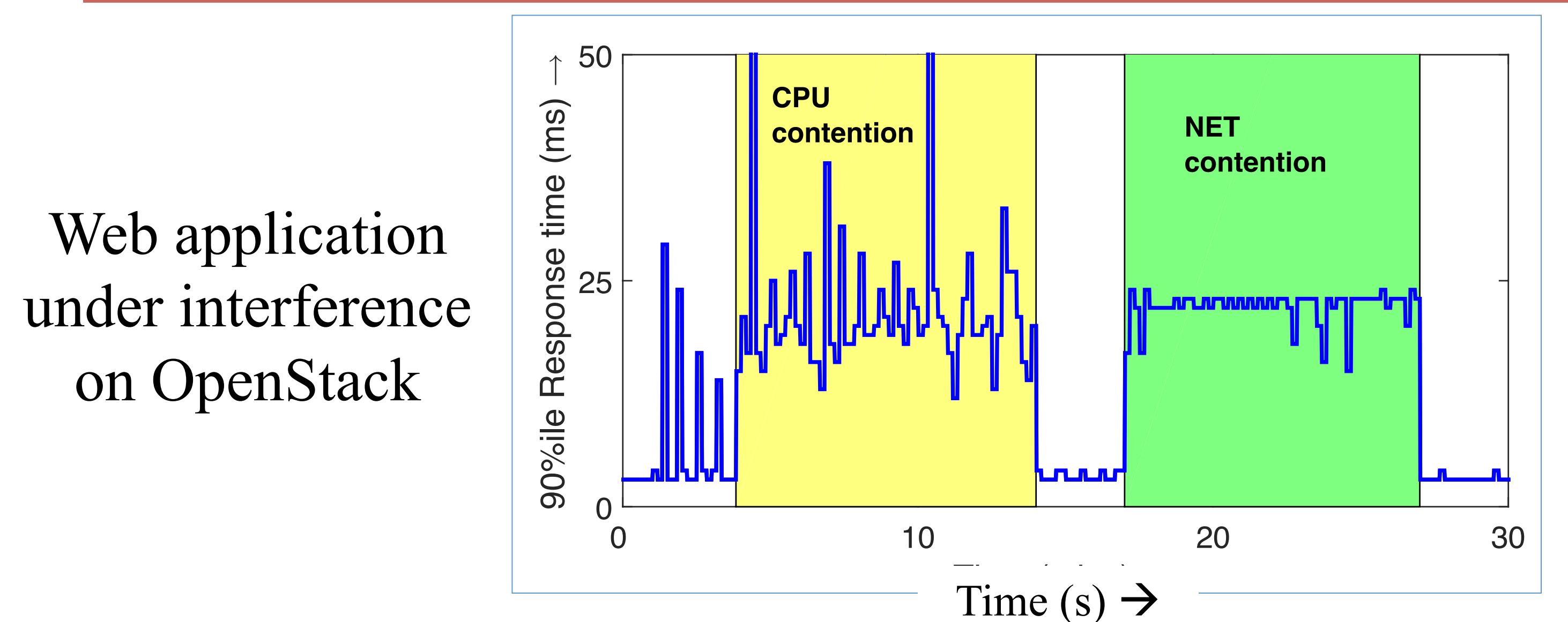
## Performance Interference



## Problem Statement

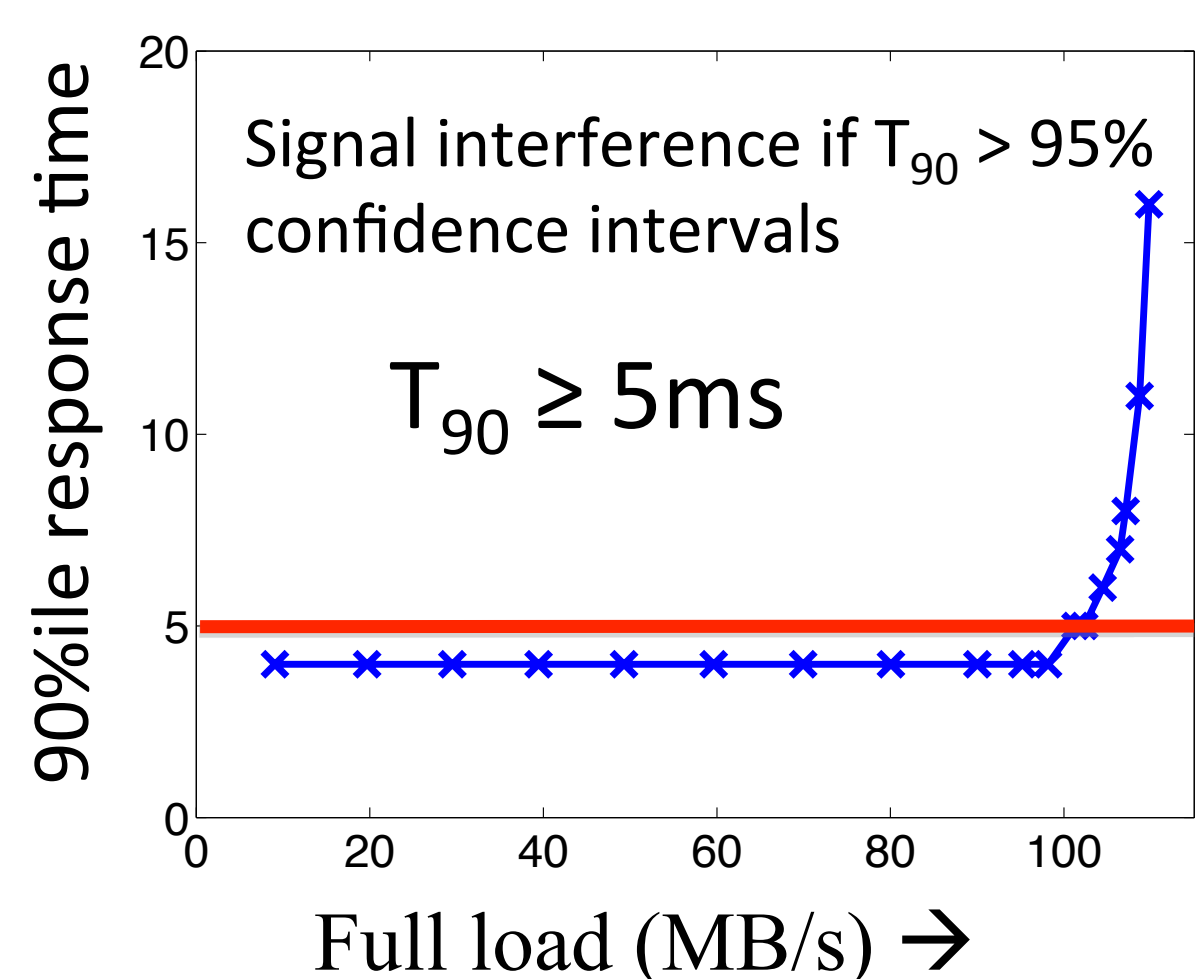


Tail response times are critical for web applications

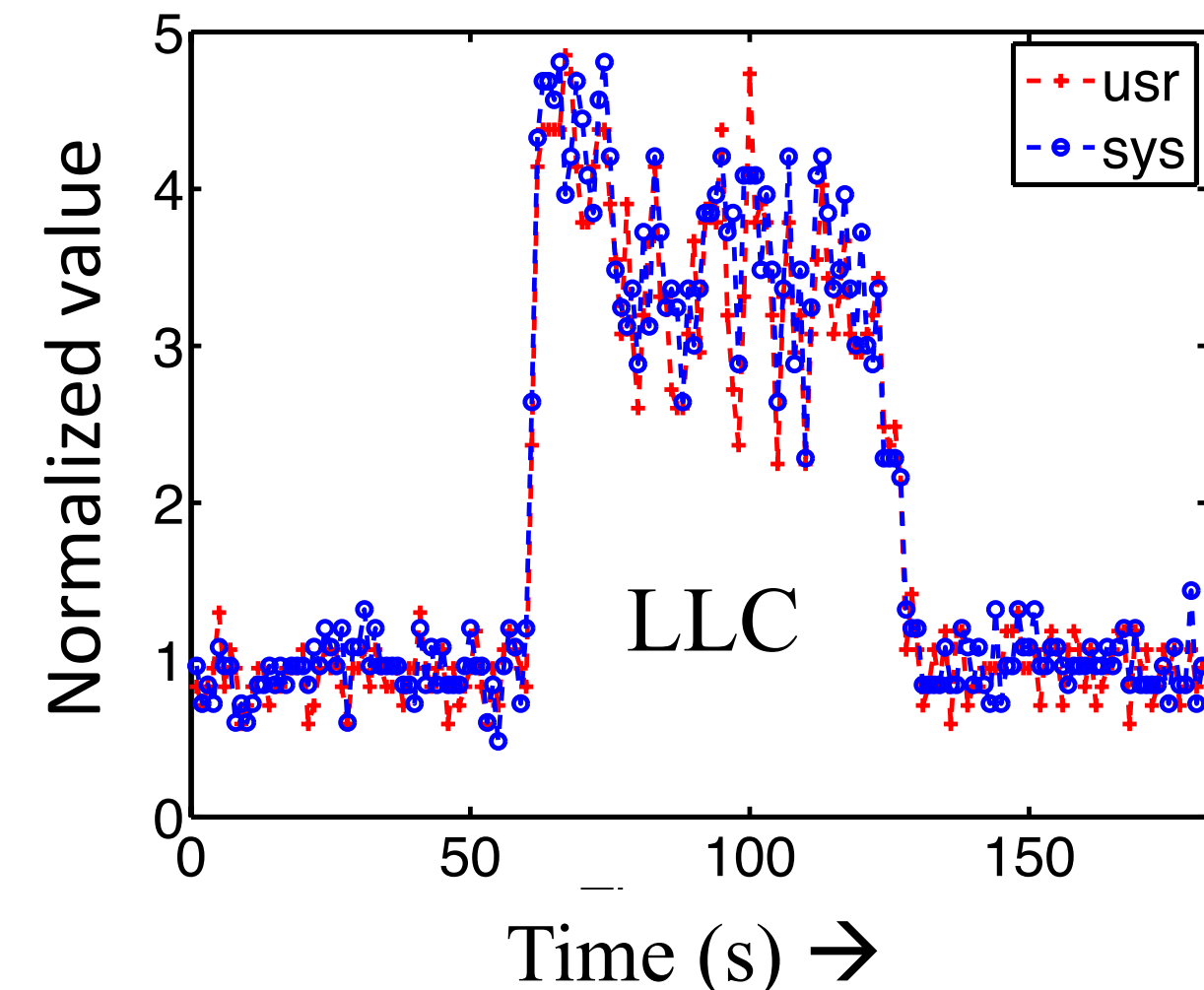


## Performance Modeling under Interference

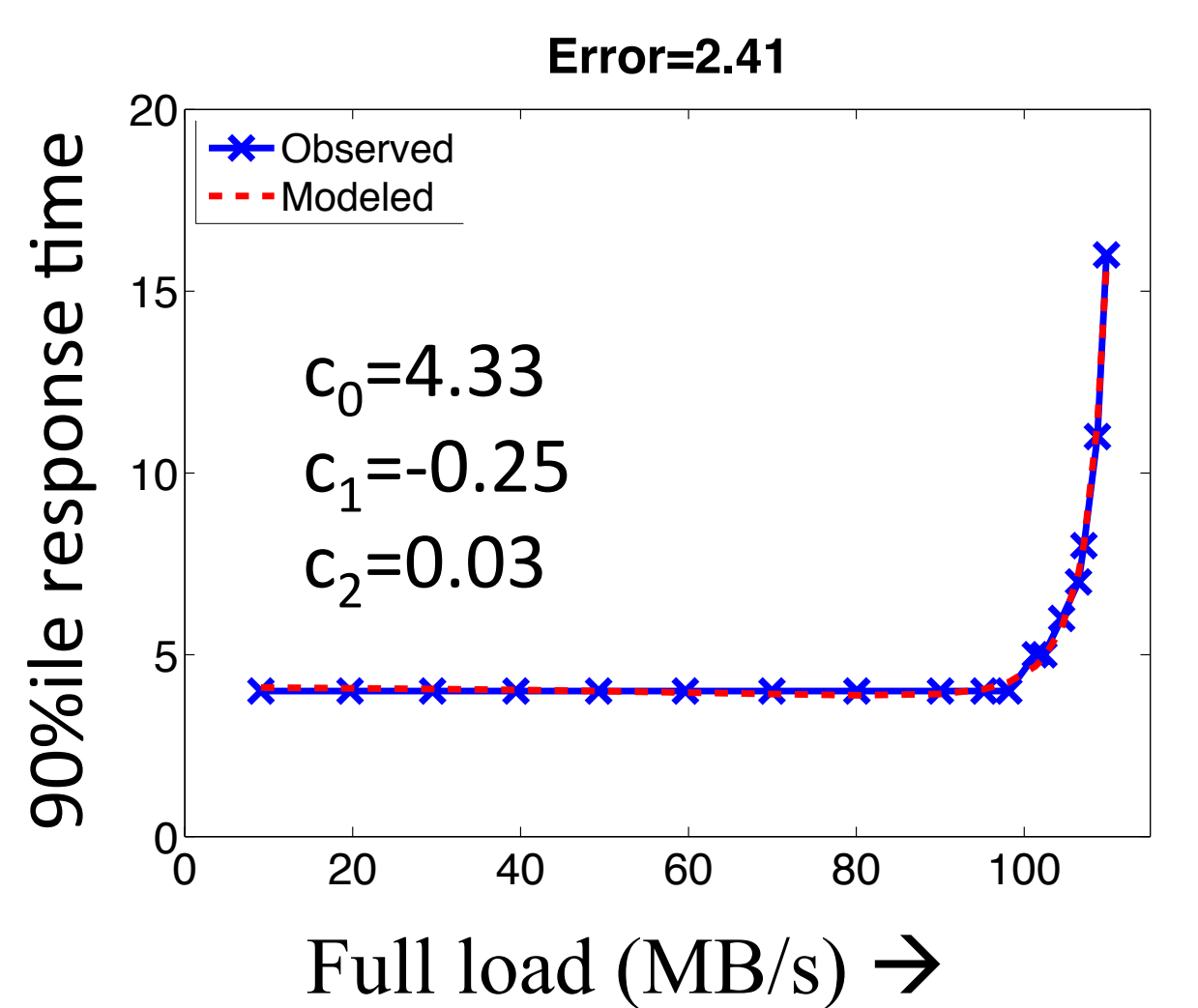
### Interference Detection



### Interference Classification



### Interference Estimation



$$T_{90} = c_0 + \frac{c_1}{(1-load)} + \frac{c_2}{(1-load)^2}$$

$$load = \frac{fg\_load + bg\_load}{peak\_load}$$

Online usage

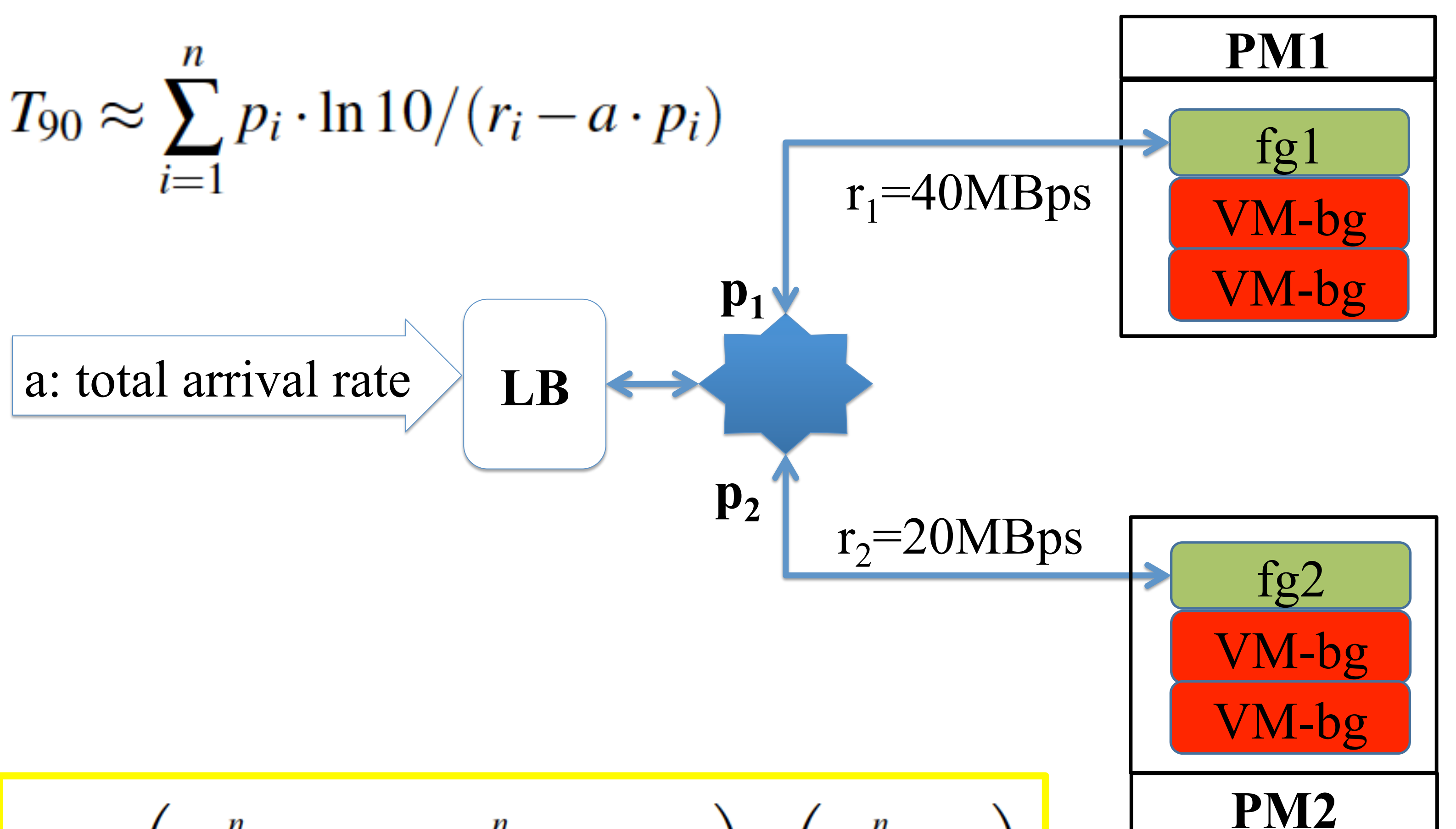
Monitor fg\_load and  $T_{90}$

Calculate bg\_load using  $T_{90}$  equation

## Interference-aware Load Balancing

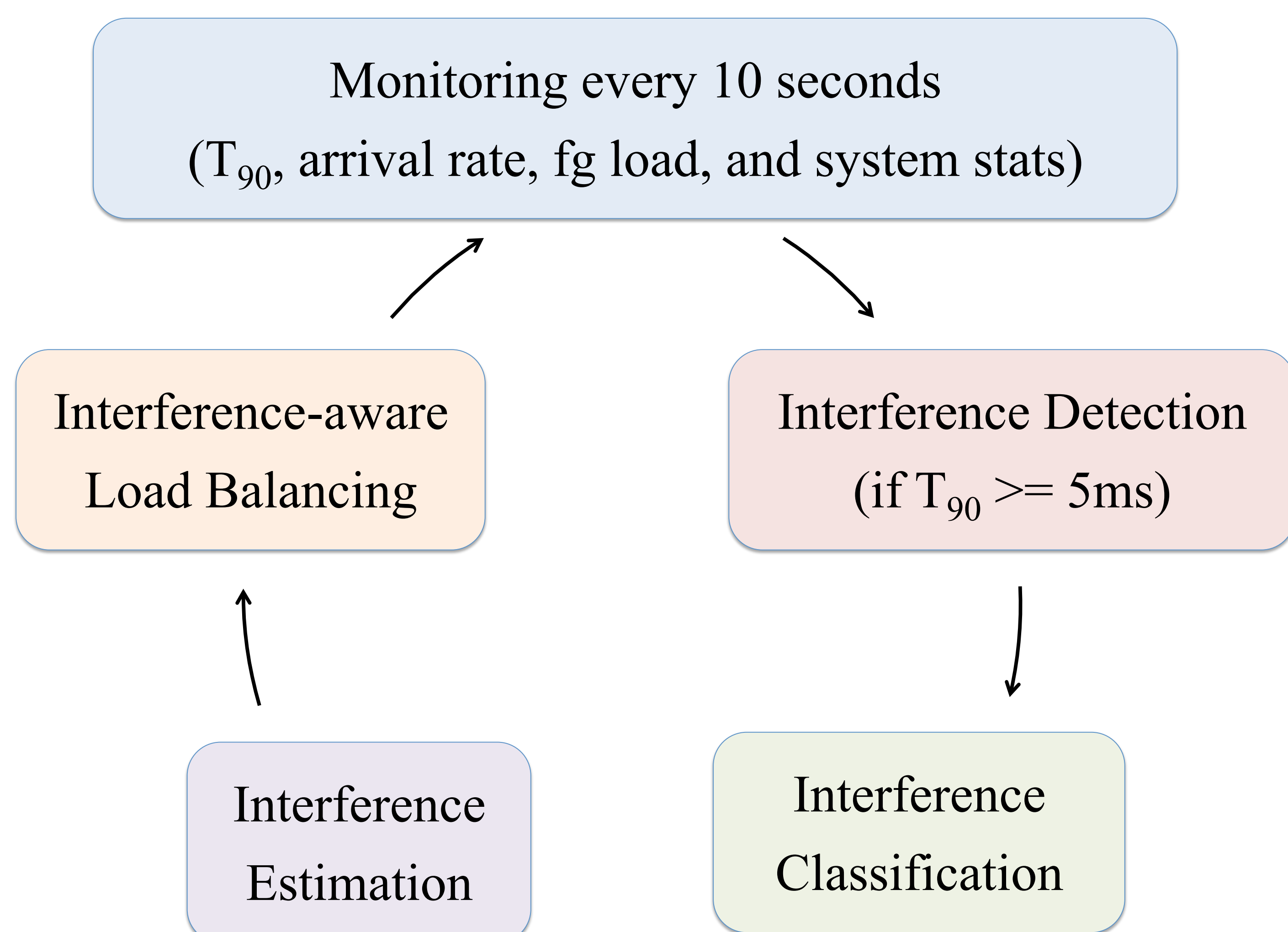
We model each VM as an M/M/1 system

$$T_{90} \approx \sum_{i=1}^n p_i \cdot \ln 10 / (r_i - a \cdot p_i)$$



$$p_i^* = \left( r_i \sum_{j=1}^n \sqrt{r_j} - \sqrt{r_i} \sum_{j=1}^n r_j + a \sqrt{r_i} \right) / \left( a \sum_{j=1}^n \sqrt{r_j} \right)$$

## DIAL



## Evaluation on OpenStack using CloudSuite

