APNet 2024

vSwitchLB: Stratified Load Balancing for vSwitch Efficiency in Data Centers

Xin Yin, Enge Song, Ye Yang, Yi Wang, Bowen Yang,
Jianyuan Lu, Xing Li, Biao Lyu, Rong Wen
Shibo He, Yuanchao Shu, Shunmin Zhu

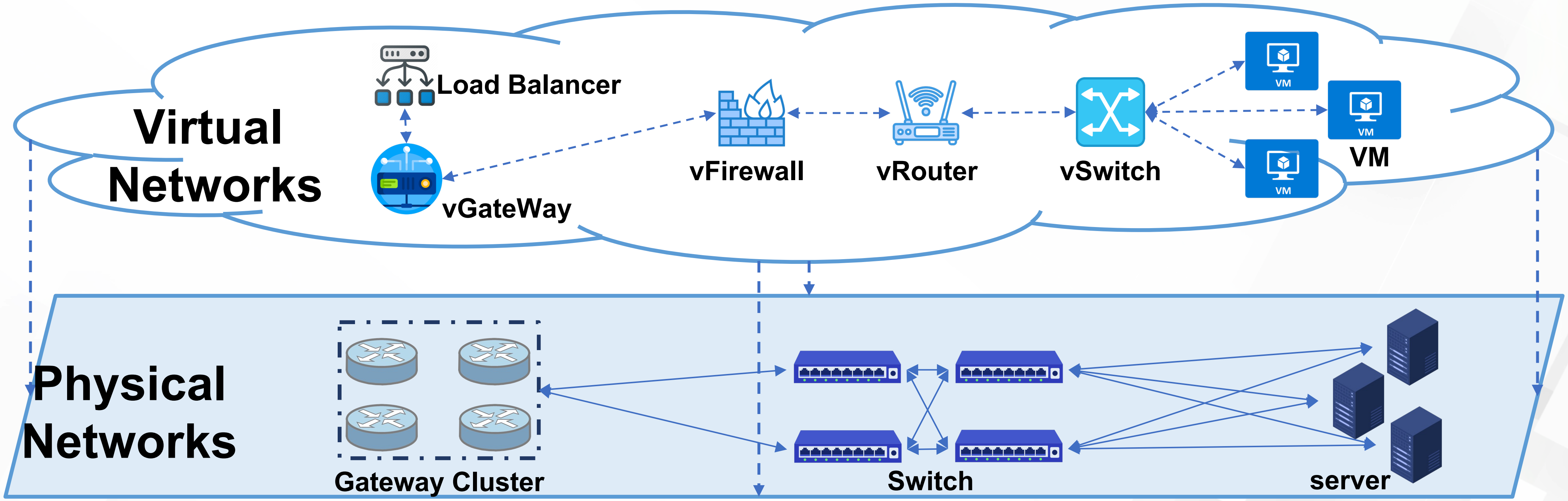
ZHEJIANG UNIVERSITY 1897

Alibaba Cloud

BACKGROUD

The virtual switch (vSwitch) is a critical component of Network Function Virtualization Infrastructure (NFVI), essential for facilitating communication between virtual machines (VMs) and between VMs and external networks

vSwitches load imbalance arising within multicore presents challenges.

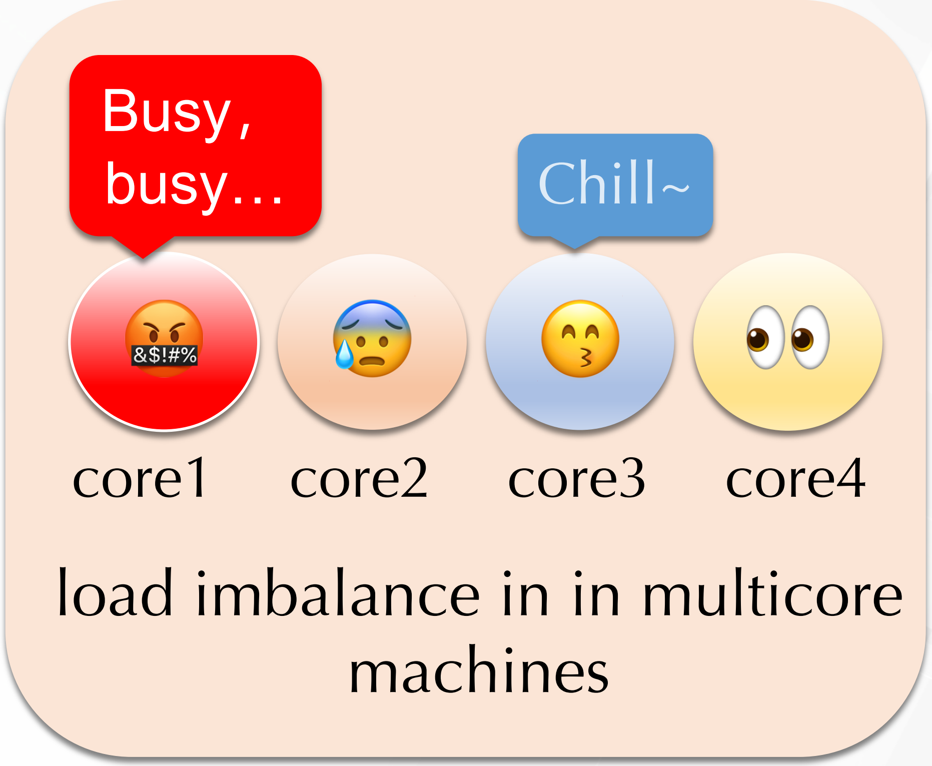


BACKGROUD

Q： What makes vSwitch load imbalance unique?

Q： Why can't we simply apply existing load balancing methods?

A： Traffic scheduling must be specifically designed and tuned in accordance with traffic granularity and the particularities of the vSwitches.



load imbalance in in multicore machines

What makes vSwitch load imbalance unique?

Multi-CPU Task Scheduling

Load balancing: Using complex scheduling algorithms and strategies to achieve load balancing and efficient task execution.

Scheduling units: Tasks, Processes, Threads.

Context information: Maintaining the context and related state.

VS

Multicore Traffic Scheduling

Load balancing: Optimizes data channels, transmission paths, and traffic control strategies for efficient data transmission.

Scheduling units: Traffic-class (different protocols), Queue, Flow, Packet, etc.

Context information: Maintaining the coherence and consistency of data flows.

Why can't we simply apply existing load balancing methods?

Static assignment

- RSS
- Intel’s Ethernet Flow Director
- Mellanox’s ASAP2

Centralized scheduling

- Shenango [NSDI 19]
- Shinjuku [NSDI 19]

Dynamic reassigning

- Metron [NSDI 18]
- RSS++ [CoNEXT 19]
- Dyssect [INFOCOM 22]

✗ Without regard to the state or occupancy of the individual cores.

✗ Dedicated cores for packet distribution become a bottleneck.

? Dynamic adjust with regard to the state or occupancy of the individual cores.

Dynamic reassigning

Academic work

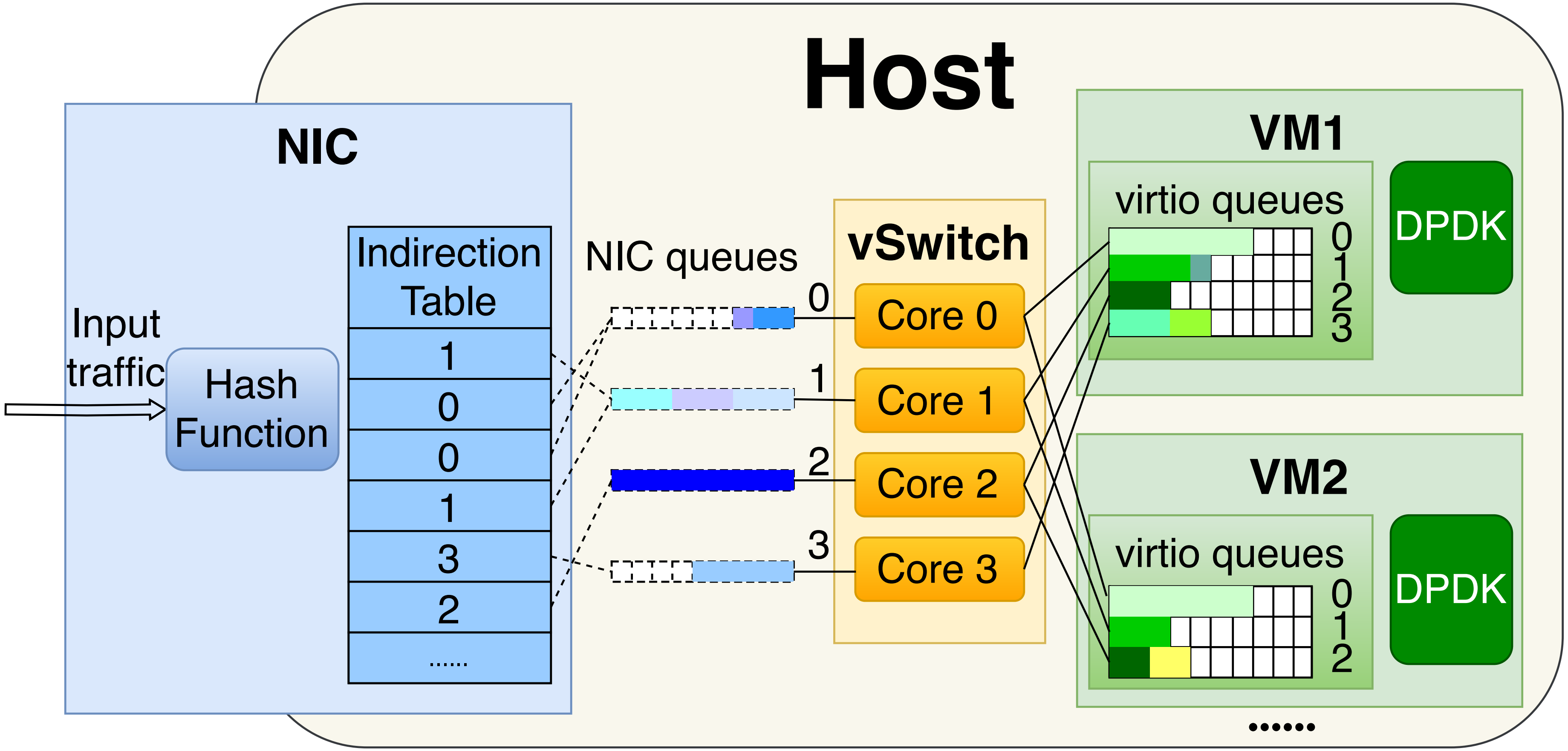
Metron	Traffic-class	Unable to load balance traffic which cannot be split
RSS++	Flow-based	Swamped by multiple large-volume or 'elephant' flows
Dyssect	Flow-based	Allocating offload cores is impractical in vSwitches

Technological solutions from industry community

OVS-DPDK	Queue-based	Fail to handle high loads in individual buckets
DLB	Packet-based	Binary operational mode, decrease in throughput

These methods excel in their specific contexts, but struggle finer-grained traffic imbalances.

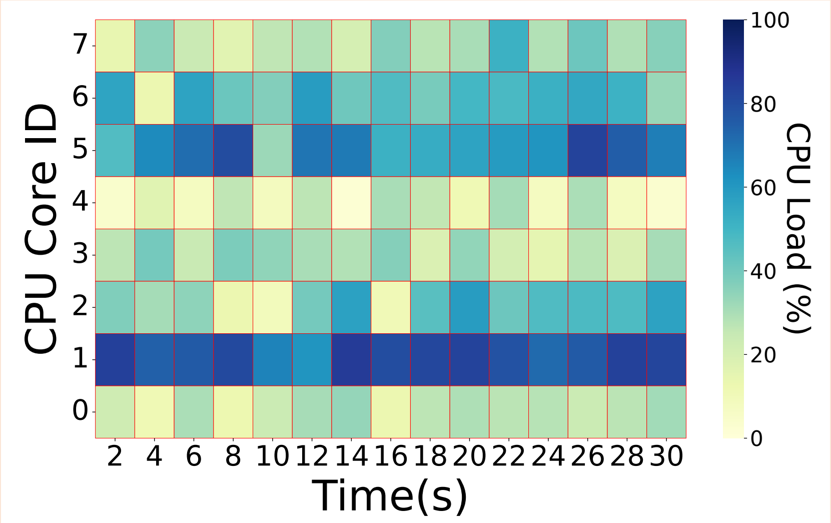
MOTIVATION



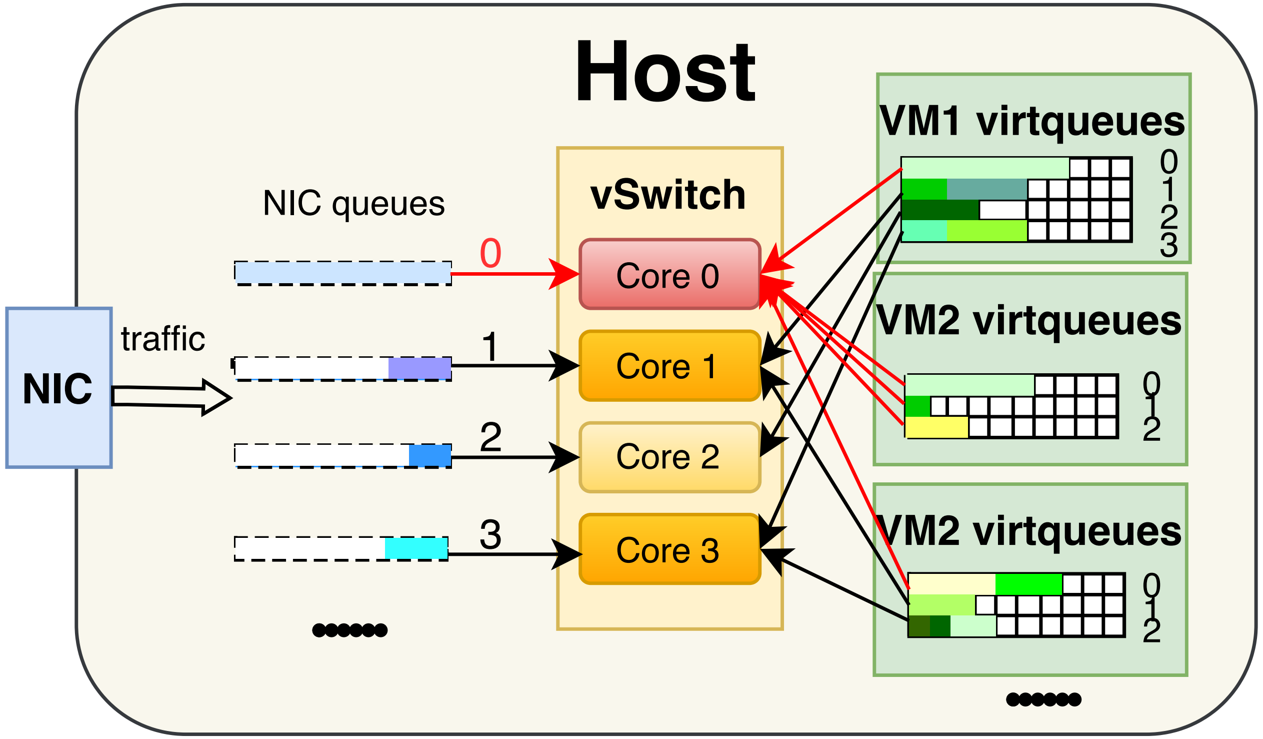
The architecture of the vSwitch datapath

MOTIVATION: Type I

🚨Type I Load Imbalance.



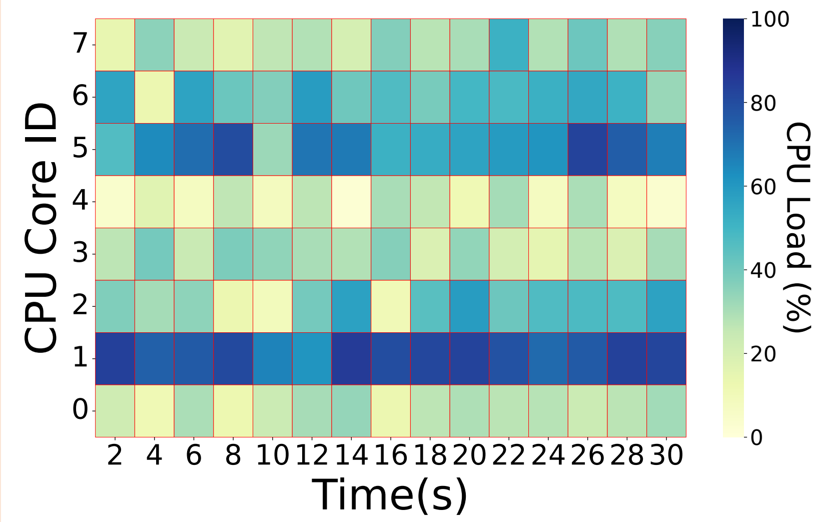
Workload Imbalance Across CPU Cores of vSwitch (a) Spatial.



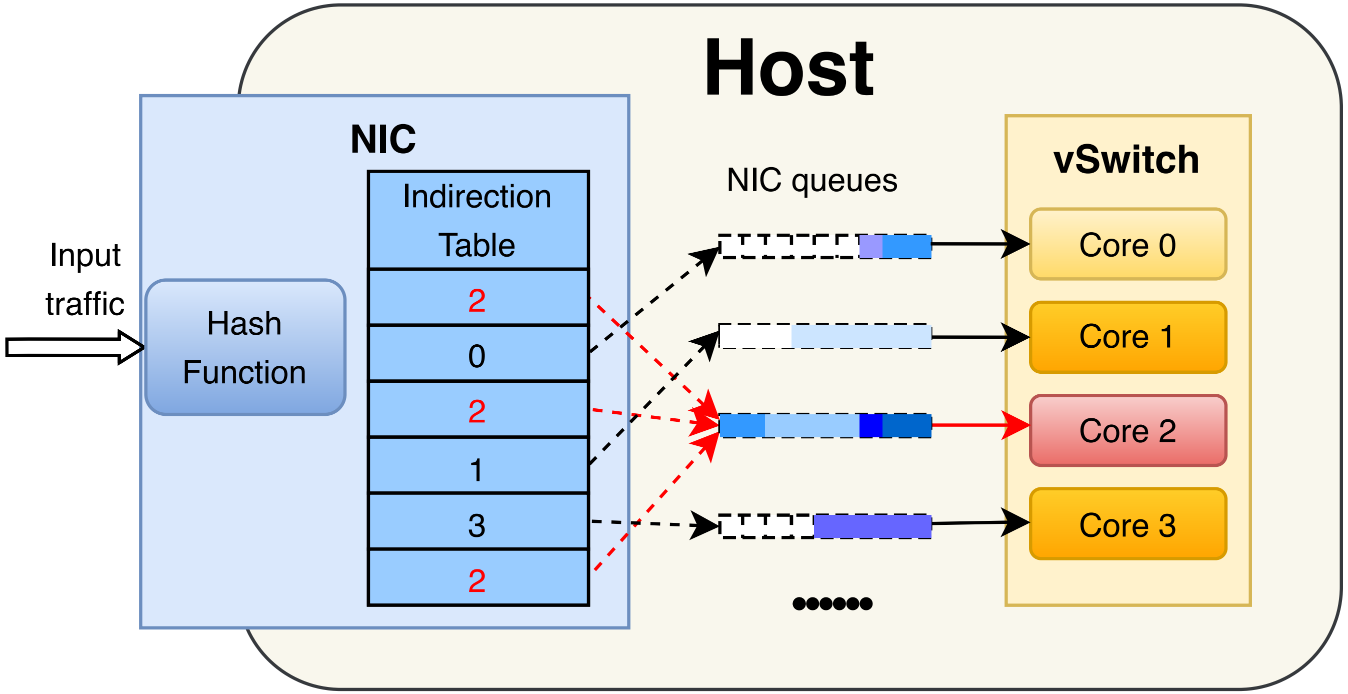
Traffic Imbalance in Virtual Queues Leading to load imbalance

MOTIVATION: Type II

🚨Type II Load Imbalance.



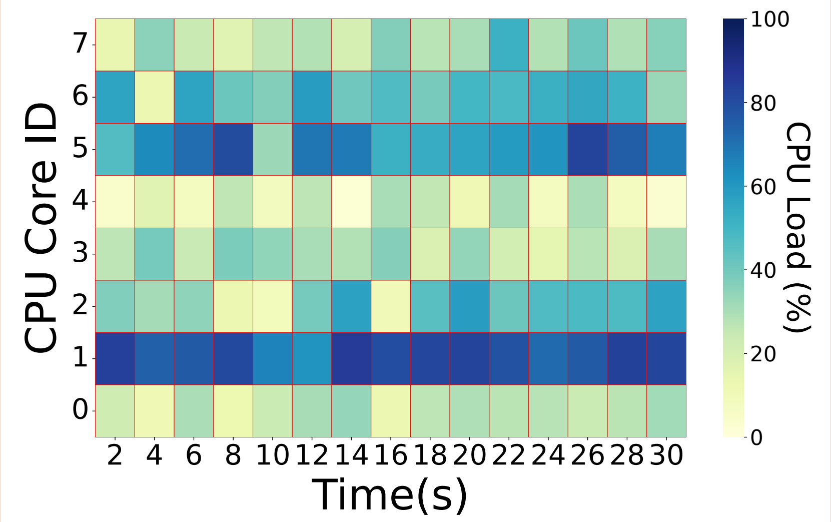
Workload Imbalance Across CPU Cores of vSwitch (a) Spatial.



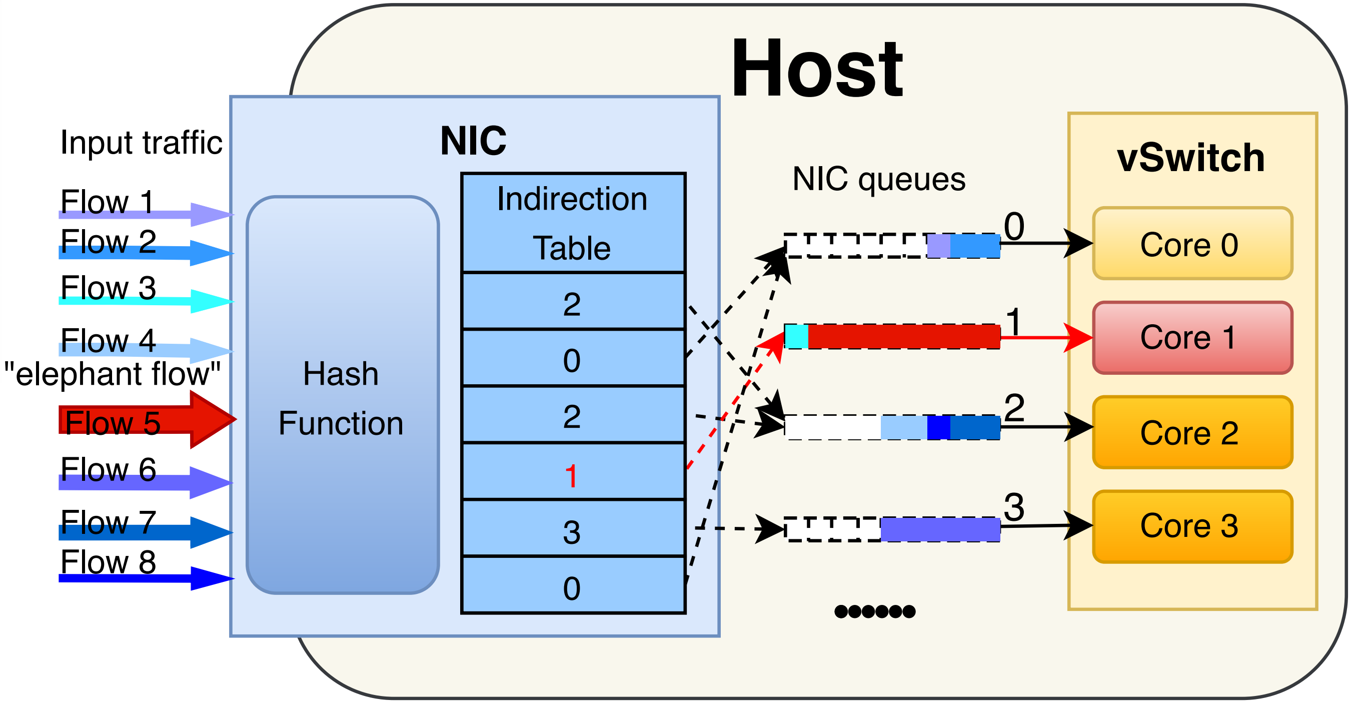
Traffic Imbalance in RSS buckets Leading to load imbalance

MOTIVATION: Type III

🚨Type III Load Imbalance.



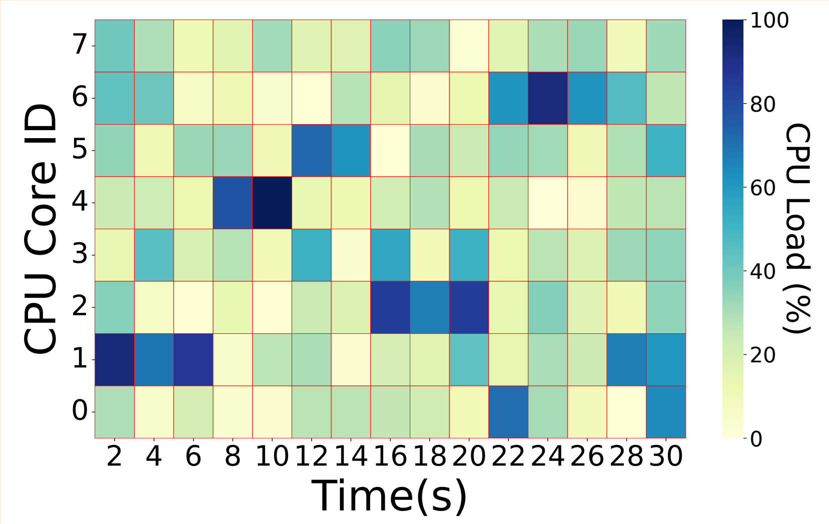
Workload Imbalance Across CPU Cores of vSwitch (a) Spatial.



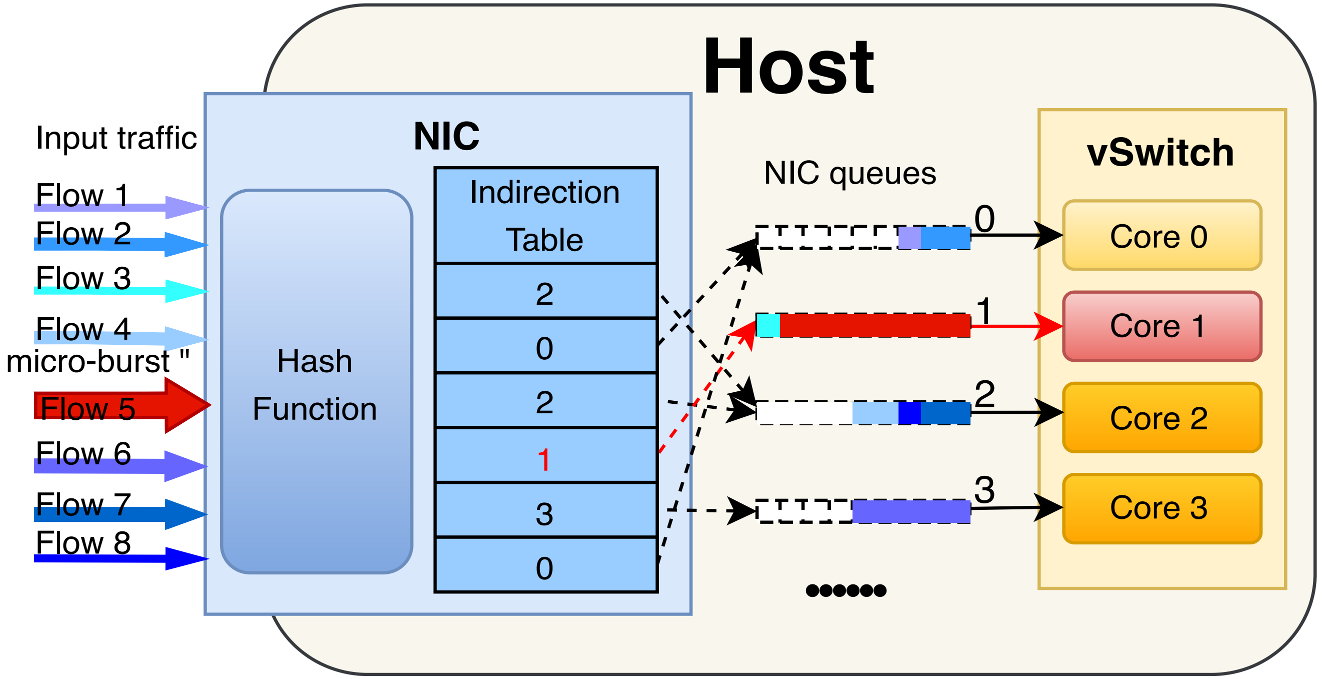
Heavy hitter Leading to load imbalance

MOTIVATION: Type IV

🚨Type IV Load Imbalance.

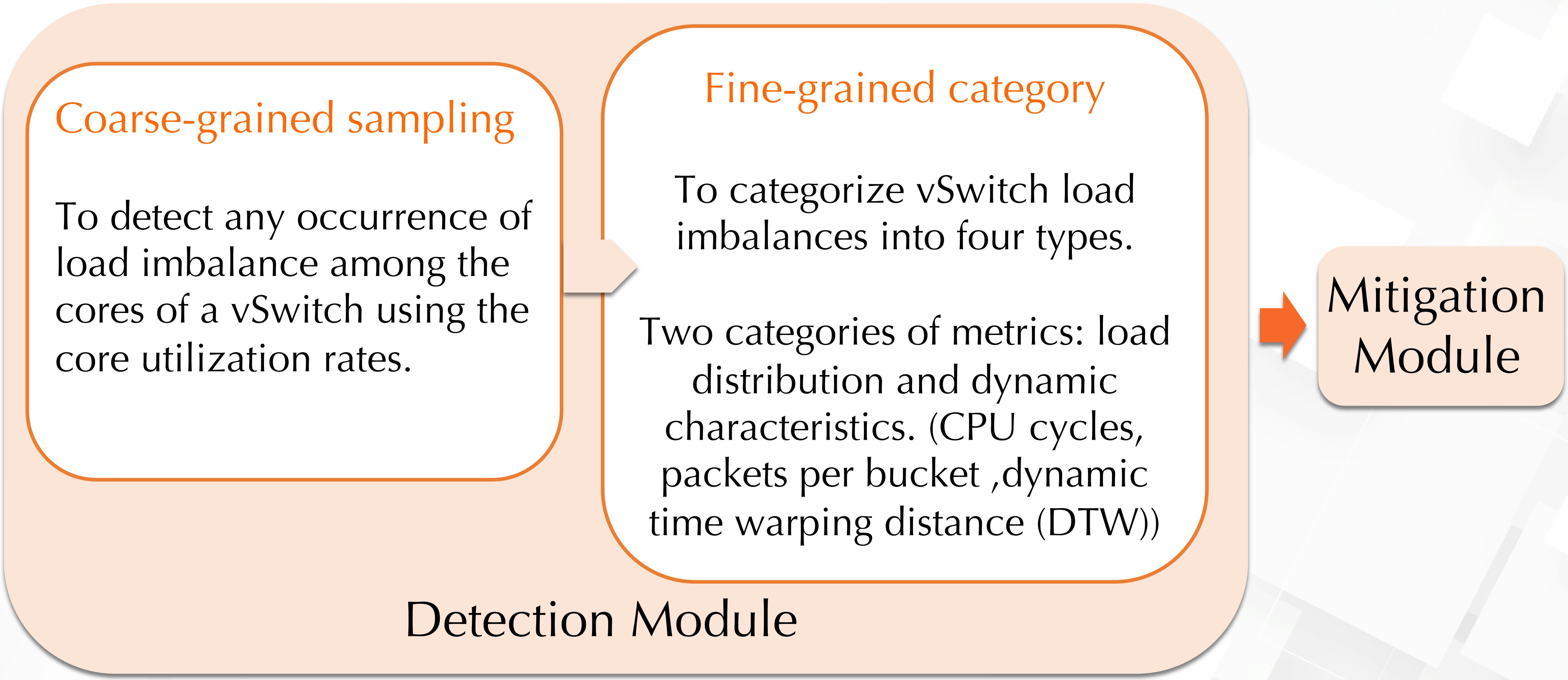


Workload Imbalance Across CPU Cores of vSwitch (b) Dynamic.



Micro-burst Leading to load imbalance

SYSTEM DESIGN: vSwitchLB – Detection Strategy



Coarse-grained sampling

To detect any occurrence of load imbalance among the cores of a vSwitch using the core utilization rates.

Fine-grained category

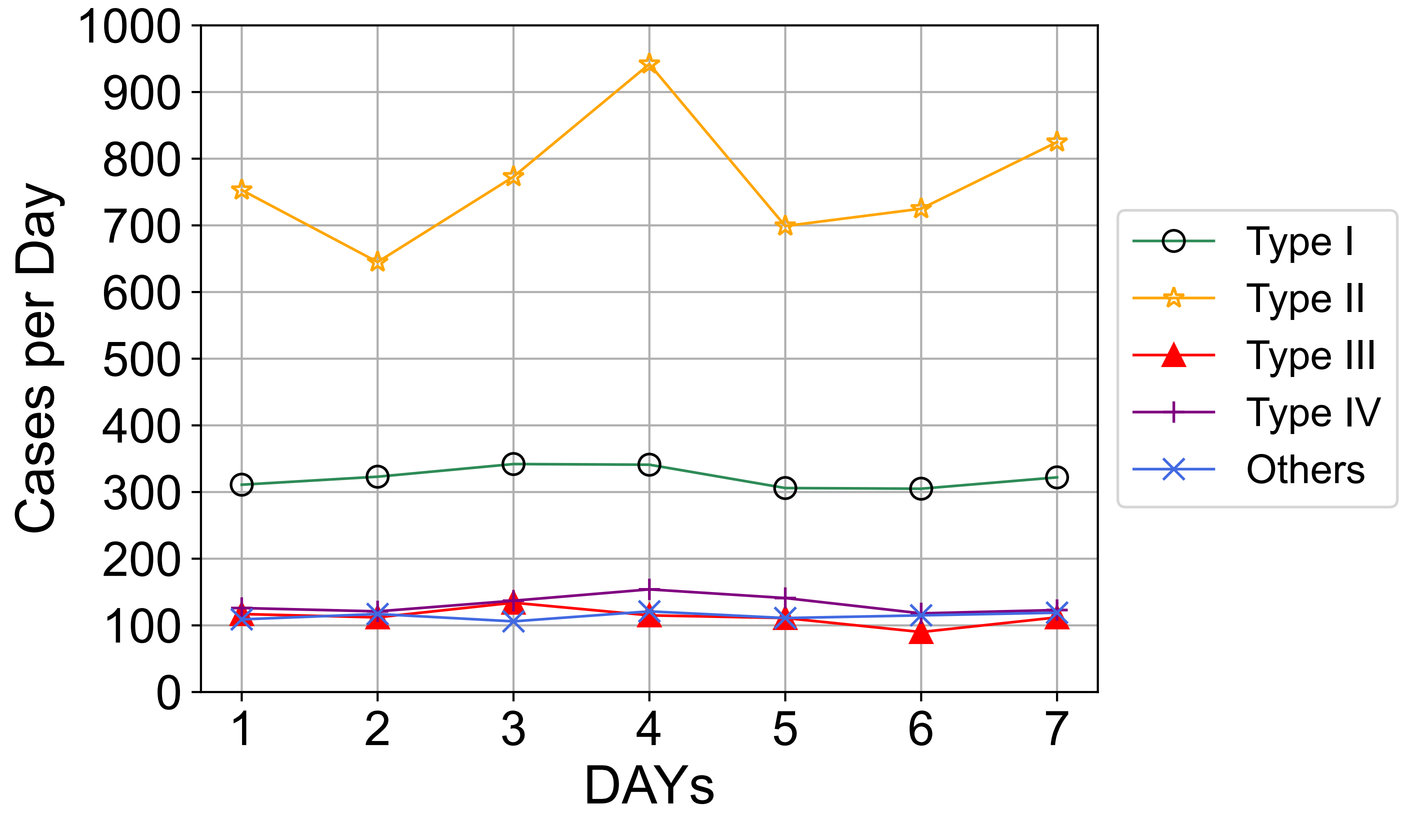
To categorize vSwitch load imbalances into four types.

Two categories of metrics: load distribution and dynamic characteristics. (CPU cycles, packets per bucket ,dynamic time warping distance (DTW))

Detection Module

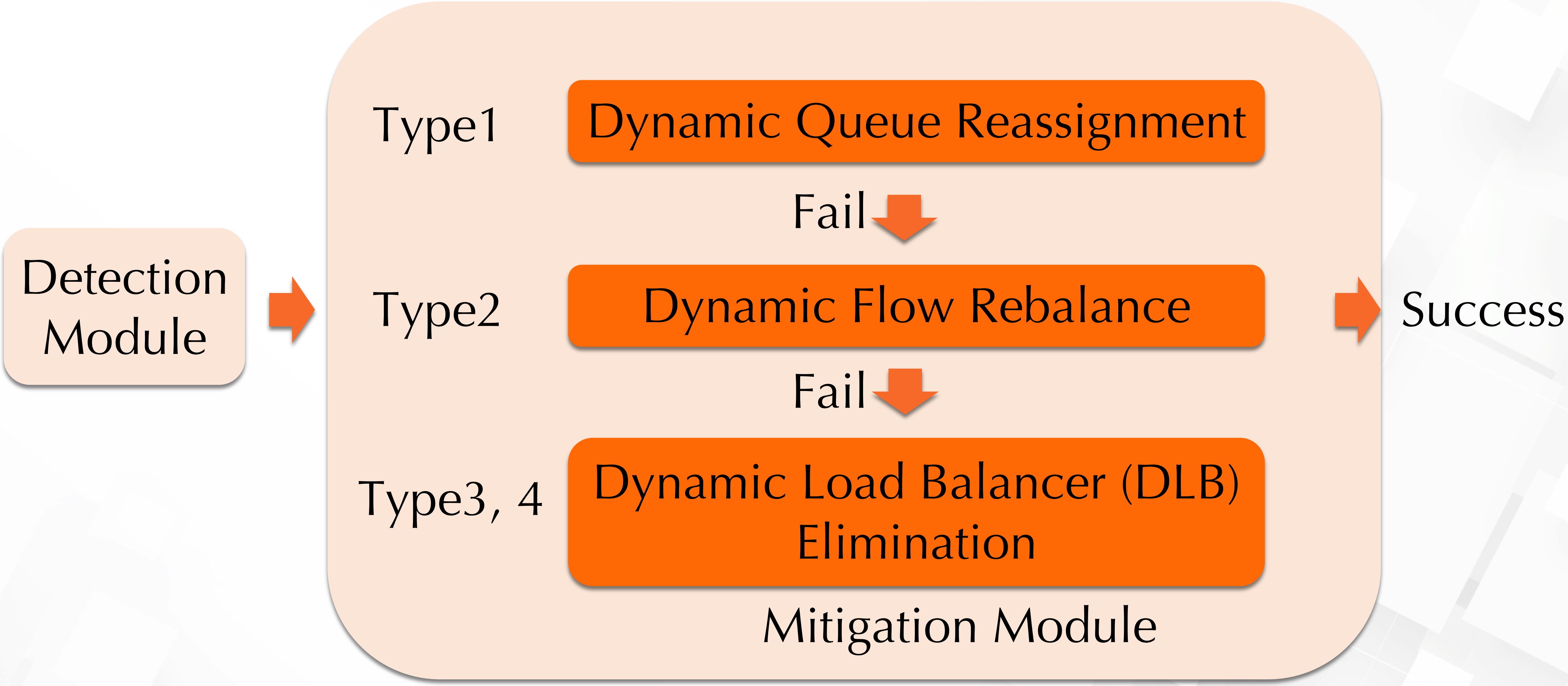
Mitigation Module

DATA COLLECTION



Occurrences of Different Types of Load Imbalance in 800,000 vSwitches Over 7 Days in Alibaba Cloud

SYSTEM DESIGN:
vSwitchLB – Stratified Load-balancing Mechanism



Detection Module

Type1

Dynamic Queue Reassignment

Fail

Type2

Dynamic Flow Rebalance

Fail

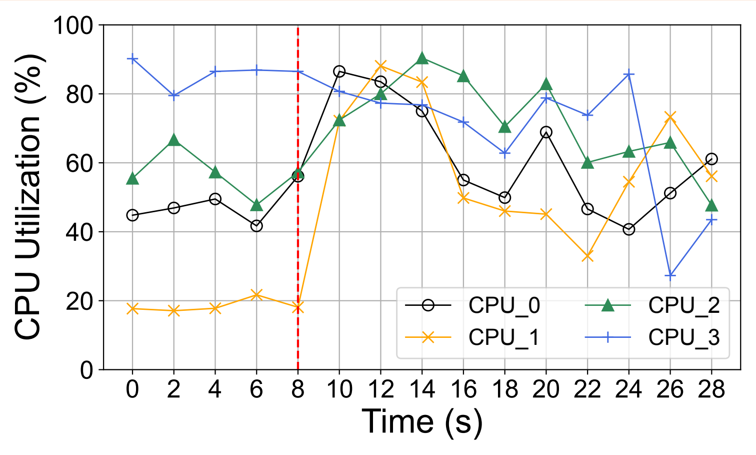
Type3, 4

Dynamic Load Balancer (DLB) Elimination

Mitigation Module

Success

PRELIMINARY EVALUATION

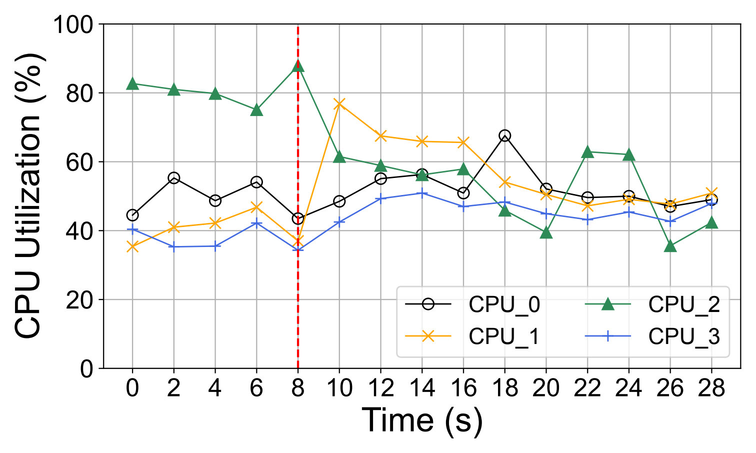


Mitigating Type I load imbalance with dynamic flow balance

Unbalanced → Balanced

Maximum $U_{max} < 80$

Range R = $(U_{max} - U_{min}) < 40$

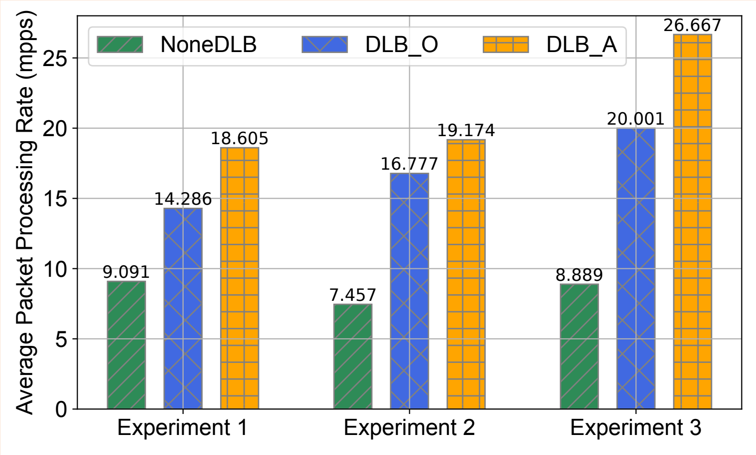


Mitigating Type II load imbalance with dynamic flow balance

Unbalanced → Balanced

Maximum $U_{max} < 80$

Range R < 40



Balancing heavy hitter to multicores with dynamic load balancer

Unbalanced → Balanced

Almost evenly balanced elephant flows across 4 cores

CONCLUSION

We have pinpointed **four cases of vSwitch load imbalance** in our cloud, stemming from unequal traffic distribution across **virtual queues** and **RSS buckets,** as well as from traffic patterns like **heavy hitters** and **micro-bursts.**

To solve this, we present vSwitchLB, a framework with a load imbalance **detection module** and dedicated techniques for addressing each specific type of imbalance.

Our preliminary evaluation shows that vSwitchLB can accurately **classify** and then **mitigate** different load imbalances.

vSwitchLB: Stratified Load Balancing for vSwitch Efficiency in Data Centers

Q & A

22232140@zju.edu.cn