CyberStar: Simple, Elastic and Cost-Effective Network Functions Management in Cloud Network at Scale

Tingting Xu, Bengbeng Xue, Yang Song, Xiaomin Wu, Xiaoxin Peng, Yilong Lyu, Xiaoliang Wang, Chen Tian, Baoliu Ye, Camtu Nguyen, Biao Lyu, Rong Wen, Zhigang Zong and Shunmin Zhu

南京大學 NANJING UNIVERSITY

Alibaba Cloud

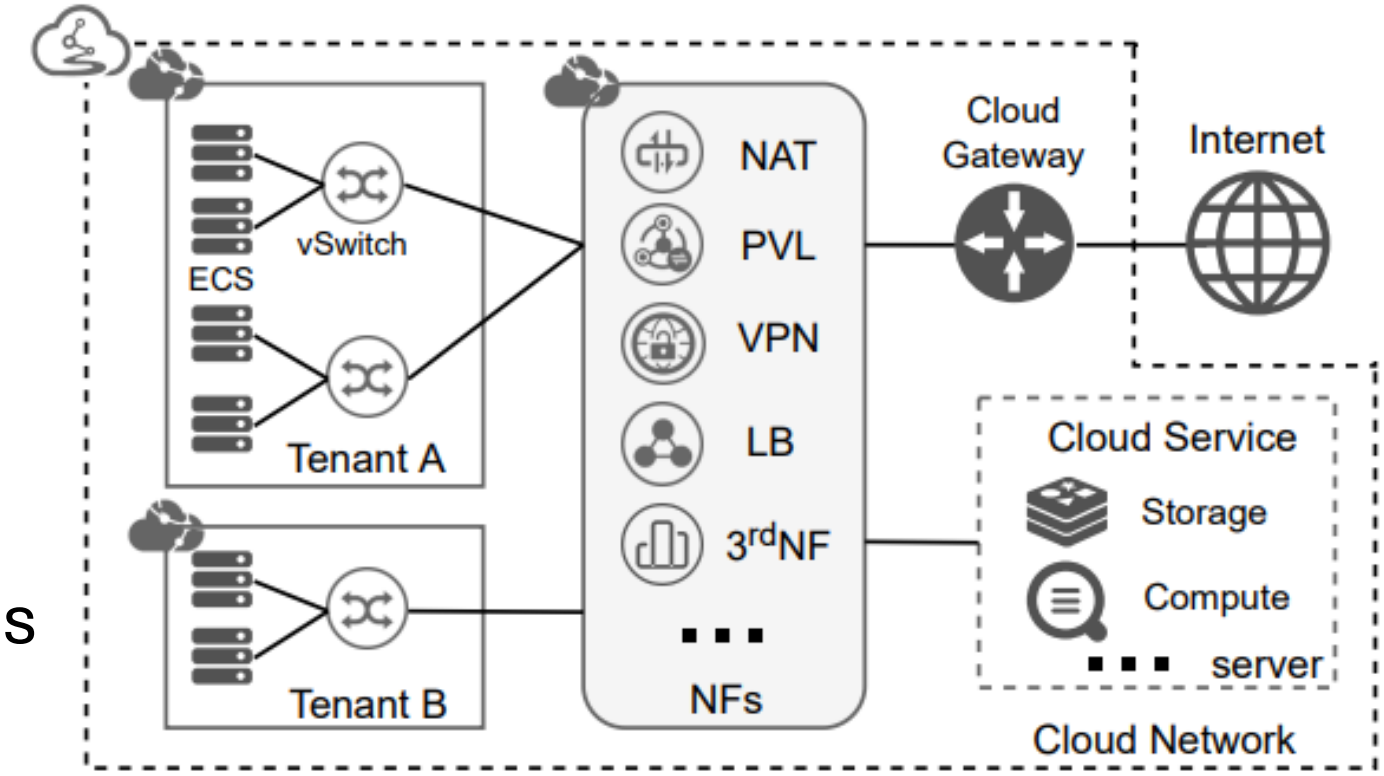
浙江大學 ZHEJIANG UNIVERSITY

清華大學 Tsinghua University

Network Functions in Clouds

❑ Network functions

- NAT (SNAT & DNAT)
- Private Link
- Virtual Private Network
- Load Balance



❑ Network functions in clouds

- Tenant VPCs
- Customer VPCs and on-premise datacenters
- Internet users and Cloud Service

Network Functions Deployment

- Hardware middleboxes
  - Long development cycles
  - Lack of programmability

- Bare-metal network functions
  - Monthly online cycle involves purchasing, constructing, configuring and verifying

- Reserve numerous devices for emergency events.
- Cost of maintaining such a large amount of infrastructure.

Network Functions Deployment

Demands

Cloud service providers are seeking **elastic** solutions that can dynamically respond to changing business demands.

- Elastic compute/container services (ECS)
  - *Potentially "infinite" computation resource*
  - *"Pay-as-you-go" price model*
  - *High availability*

CyberStar: elastic cloud-native NF management platform over ECSs.

Objects of CyberStar

Elastic Scalability

❑ Challenge: scale out & scale up

NFs	Capability
NAT	2 million connections; 100 thousand CPS;
LB	100 million connections; 1 million CPS; 100 thousand QPS
IPSec VPN	5~ 200 Mbps bandwidth

- Resource availability
- Complex internal execution
- State consistency requirement

Insights

→ “Infinite” resources

→ Decoupling NFs

Objects of CyberStar

High Resource Utilization

❑ Challenge

The average CPU utilization of Alibaba cluster is between 20% to 50%.

- Tenant distribution difference
  - Tenant A : traffic of 30 Mbps, 300K routes
  - Tenant B : traffic of 200 Gbps, just 7 routes
- Large fluctuations over time
  - Peak-to-average: 100:1

Insights

→ Resource Sharing

→ Auto-Scaling

Objects of CyberStar

Low Management Complexity

❑ Challenge

- Diverse resource configurations
  - Performance, cost, and availability.
  - Inherent delays and constraints.

Insights

A few types of **prevalent** and **low-configured** ECS instances

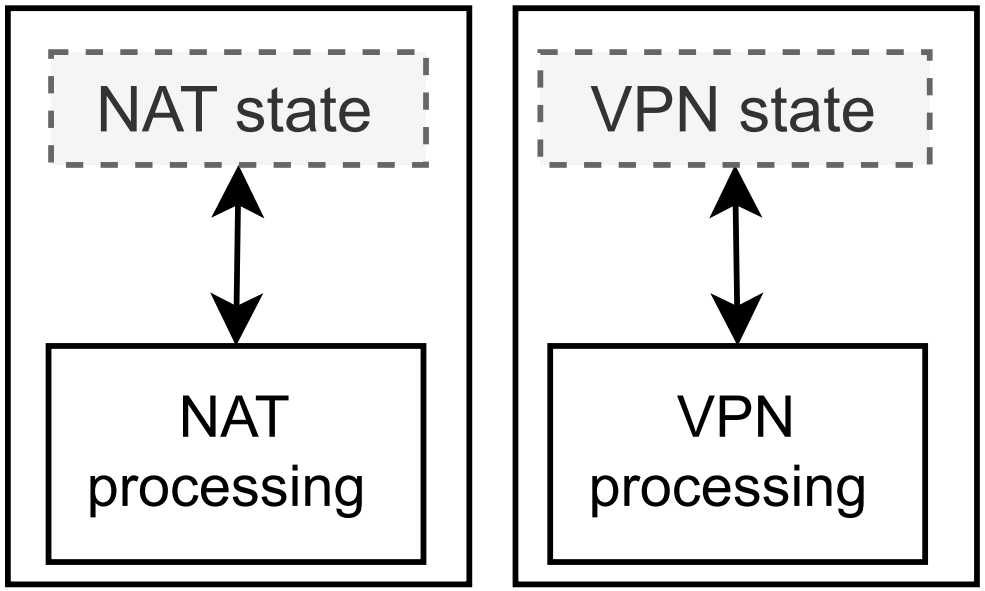
Entry Computing | General Computing | Enhanced Computing | Elastic Bare Metal | Accelerated Computing | Local Storage Enhanced

Instance Type	vCPU/Memory Ratio	Maximum Disk IOPS	Maximum PPS Capability	Pricing	
Universal Type(u1) ⓘ	1:1/1:2/1:4/1:8	60,000	2,000,000 pps	From $43.34/month	Buy Now
General Purpose(g7,g6,g5) ⓘ	1：4	600,000	g7：24000000 pps g6：6000000 pps g5：4000000 pps	From $73.69/month	Buy Now

Design Rationale

Elastic Scalability

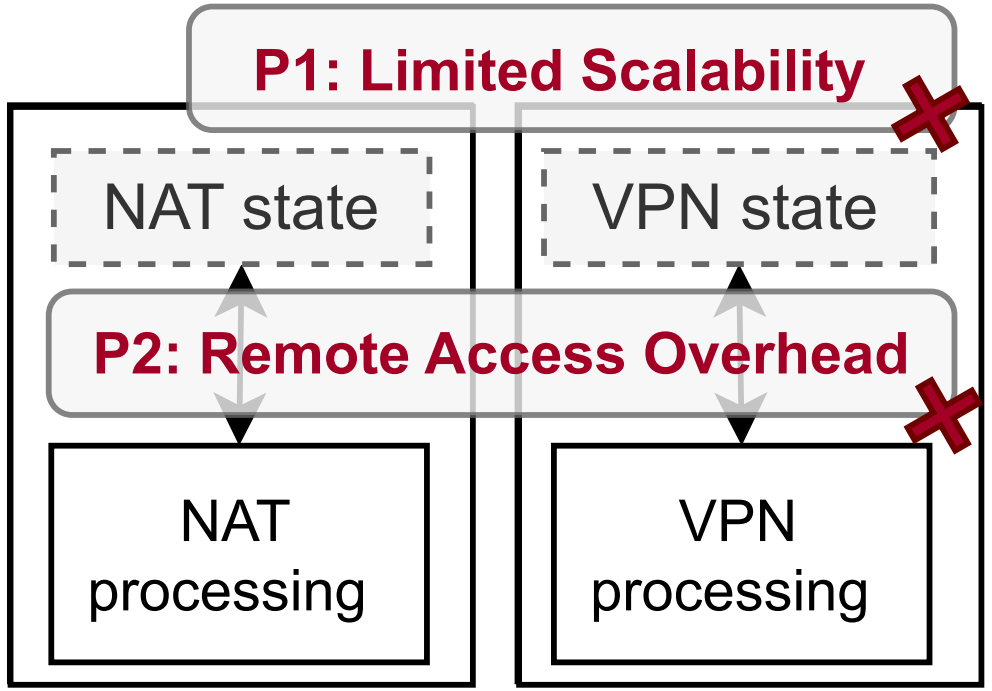
- Typical NF Architecture



Design Rationale

Elastic Scalability

- Typical NF Architecture



Design Rationale: Elastic Scalability

❑ Elastic Scalability using Disaggregated Architecture

- Partition NFs state and operations into lightweight components.
- Distribute them across massive ECS instances.

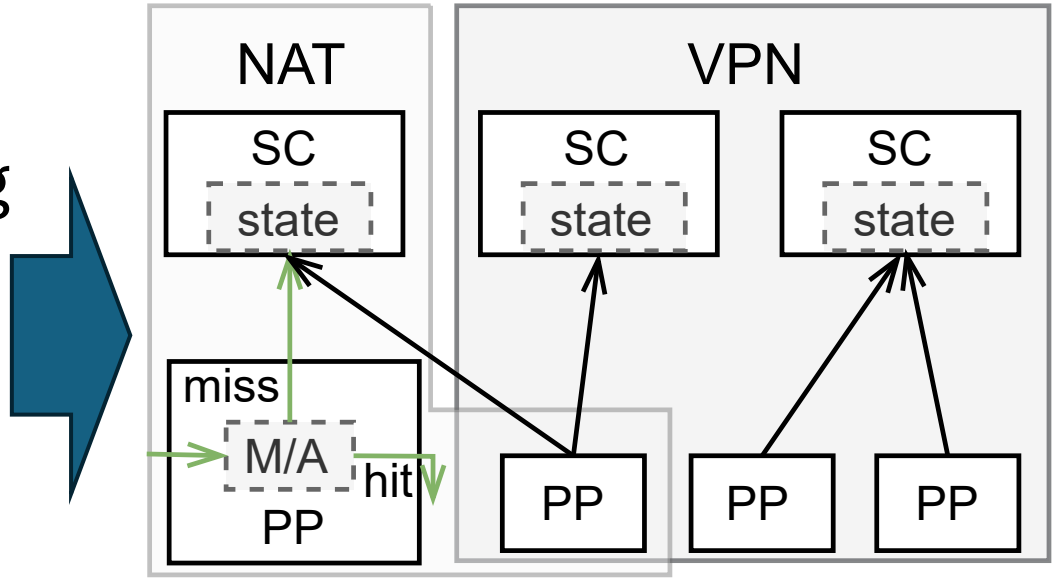
- NF-independent packet processing
- NF-specific computation
- NF-specific state management

Design Rationale: Elastic Scalability

❑ Elastic Scalability using Disaggregated Architecture

- Partition NFs state and operations into lightweight components.
- Distribute them across massive ECS instances.

- NF-independent packet processing
- NF-specific computation
- NF-specific state management



Design Rationale: Elastic Scalability

❑ Elastic Scalability using Disaggregated Architecture

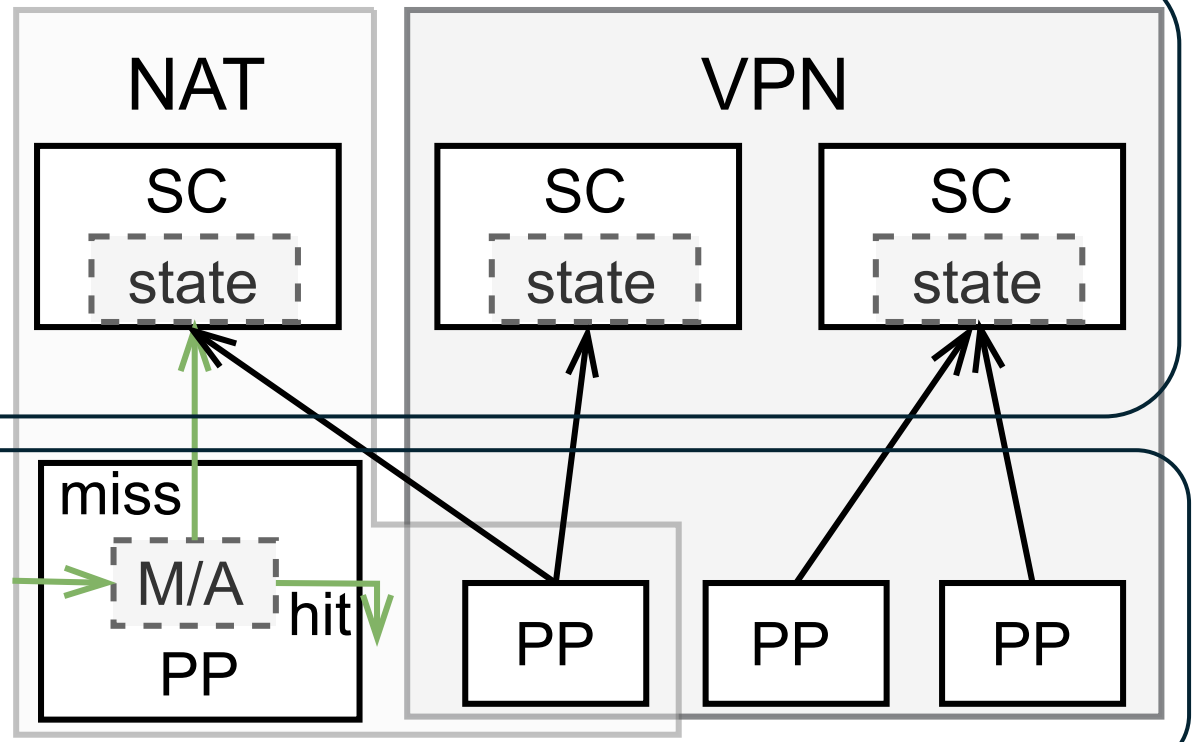
- Service computing (SC)

  NF-specific computation

  NF-specific state management

- Packet processing (PP)

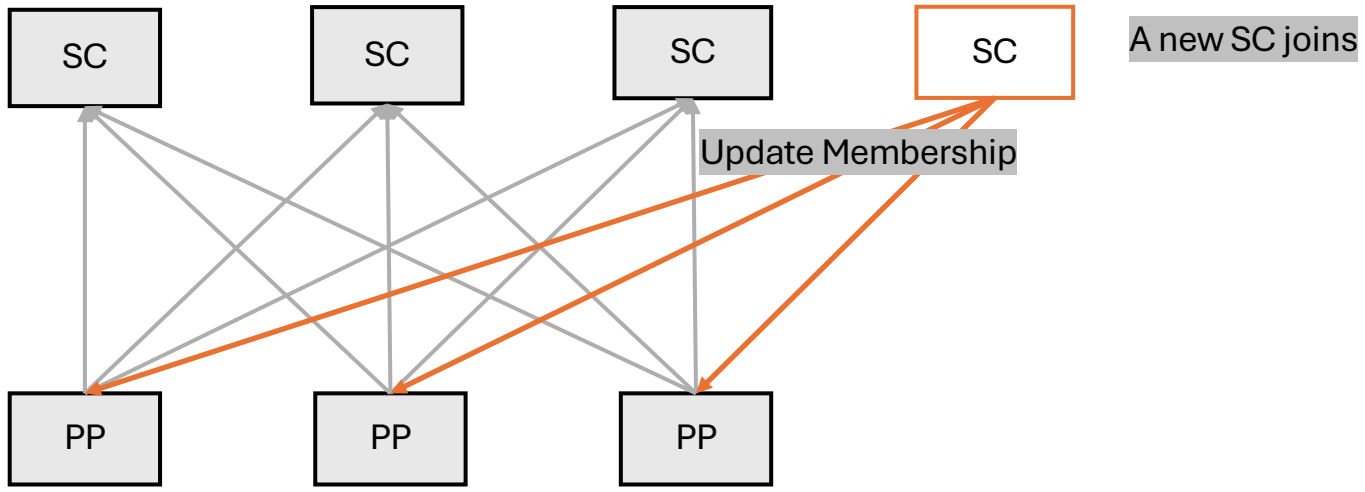
  A pipeline of Match-Action units



Design Rationale: Elastic Scalability

❑ Elastic Scalability using Disaggregated Architecture

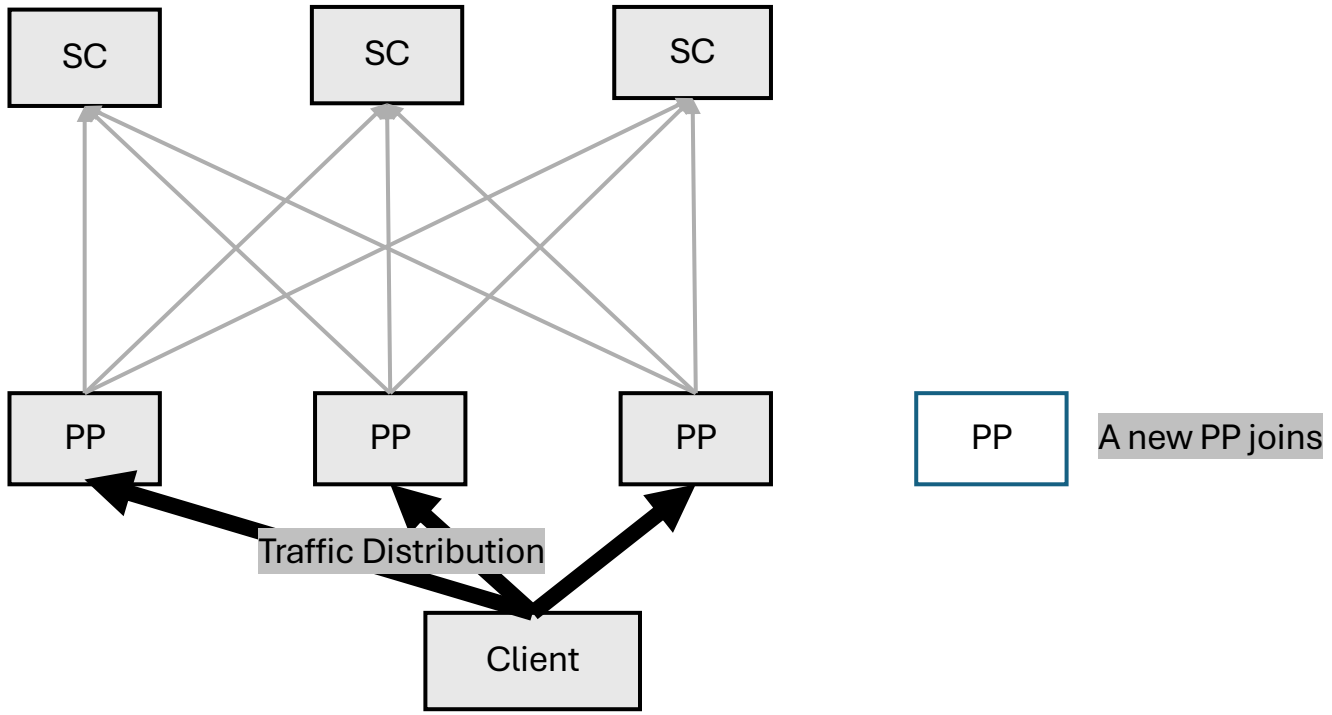
- Connections between SC and PP plane limits scalability.



Design Rationale: Elastic Scalability

❑ Elastic Scalability using Disaggregated Architecture

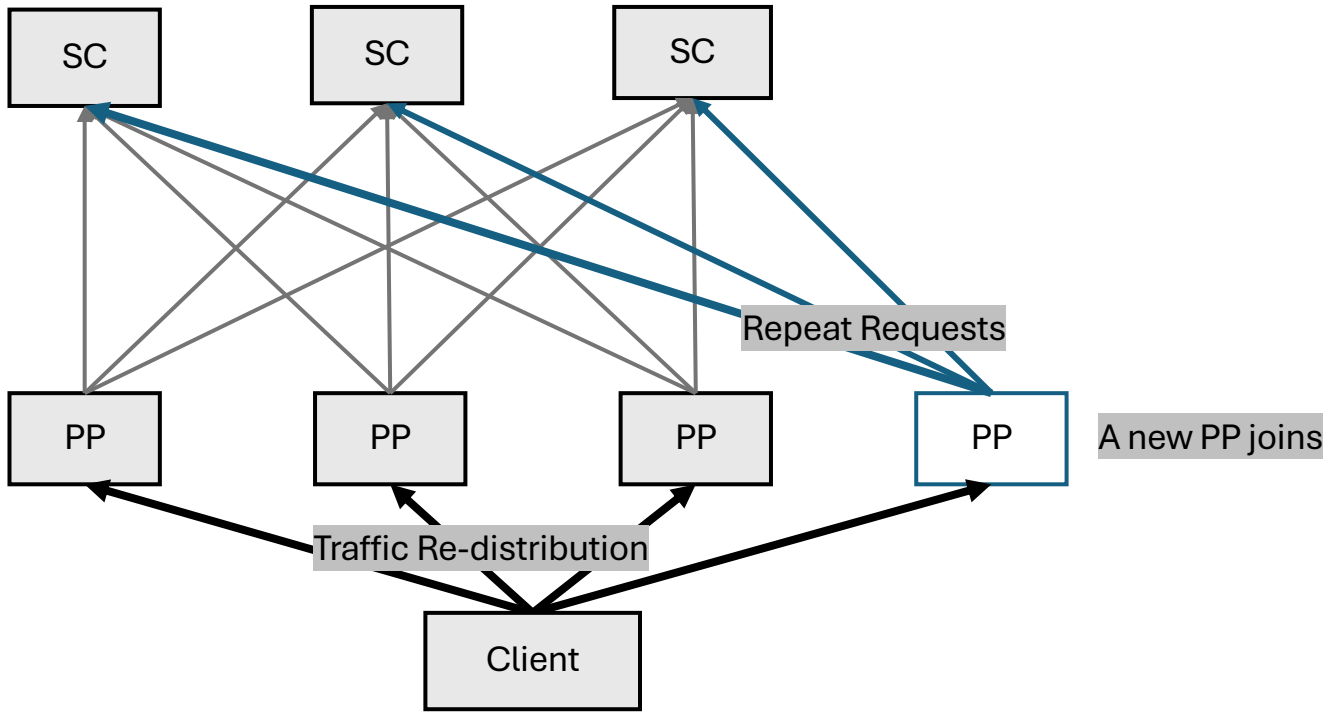
- Connections between SC and PP plane limits scalability.



Design Rationale: Elastic Scalability

❑ Elastic Scalability using Disaggregated Architecture

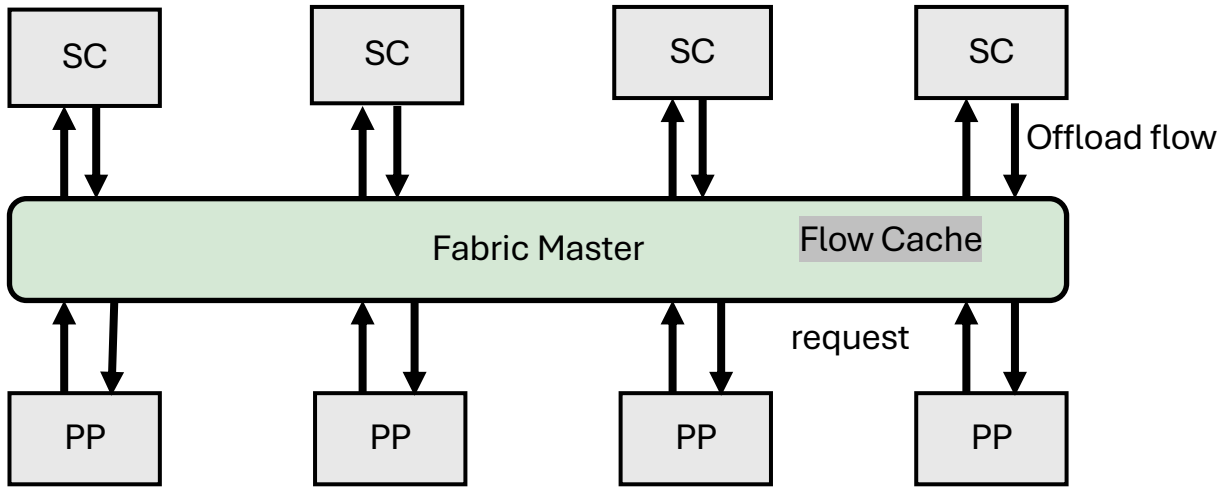
- Connections between SC and PP plane limits scalability.



Design Rationale: Elastic Scalability

❑ Elastic Scalability using Disaggregated Architecture

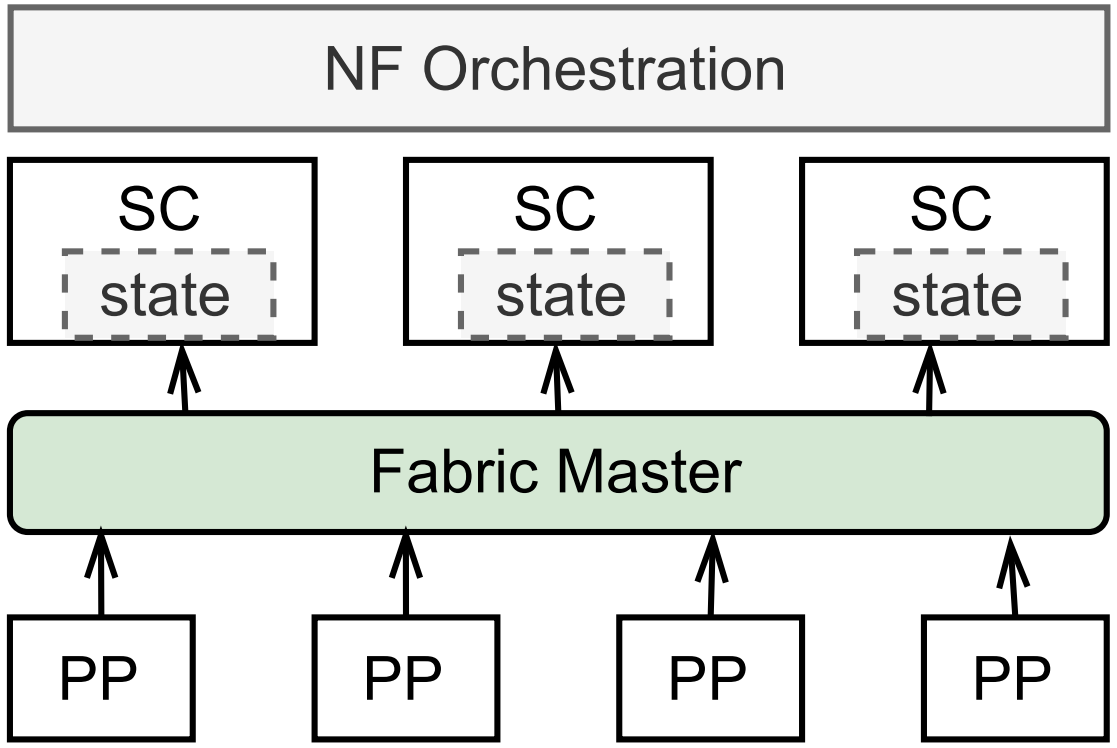
- Fabric Master (FM) decouples SC and PP plane.



CyberStar Architecture

Overview

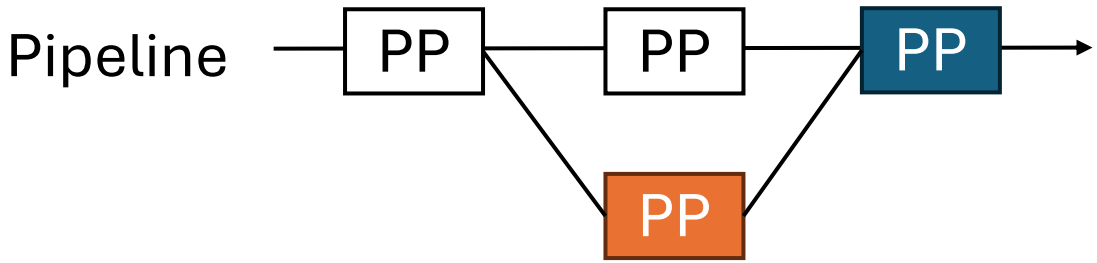
- ❑ NF Orchestration
  - Auto-Scaling mechanism
- ❑ Three-plane of NFs
  - Service Computing
  - Packet Processing
  - Fabric Master



CyberStar Architecture

Packet Processing

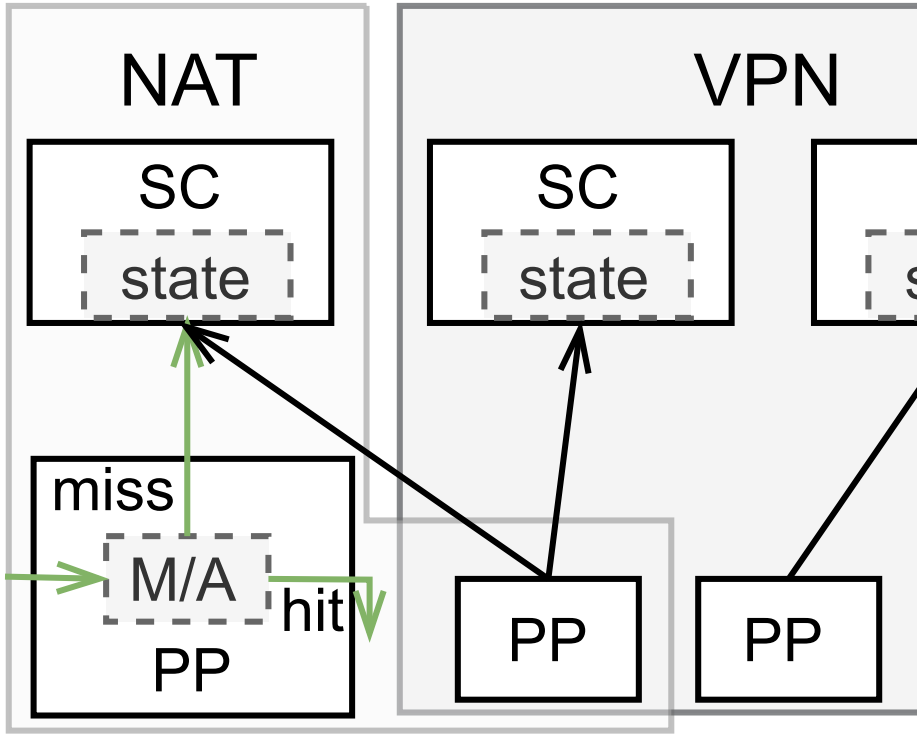
- PP: a series of match-action units.
  - Service chain, e.g., FW-NAT
  - Complex NFs, e.g., IPSec VPN
- Extend pipeline depth by adding new PPs along traffic path.
- Scale each stage of the pipeline separately.



CyberStar Architecture

Packet Processing

- PP: a series of match-action units.
  - Service chain, e.g., FW-NAT
  - Complex NFs, e.g., IPSec VPN

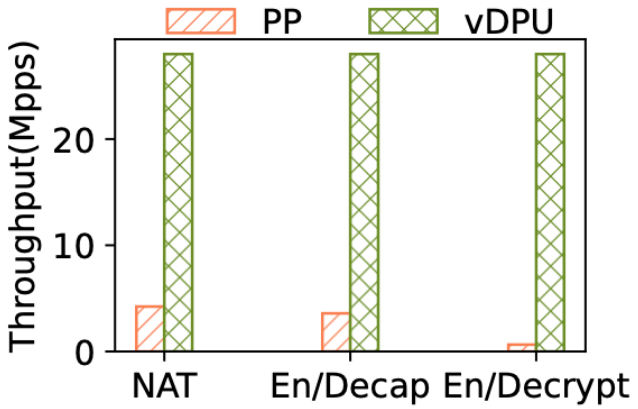


- Utilization
  - NF-independent units, can be shared by multiple NFs and tenants.
- Performance
  - Similar semantics with general-proposed hardware flow tables.

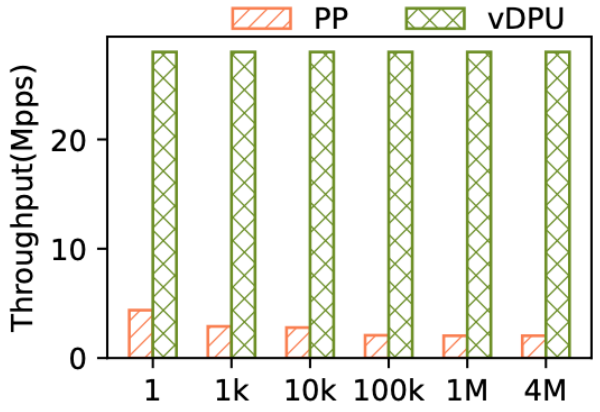
Evaluation of Performance

Packet Processing

- Performance improved by vDPU acceleration
  - Average delay as low as 20.587μs.
  - Throughput is 6.6× and 7.8× compared to ECS.
  - The side effect on software flow table is optimized by vDPU. It always maintains stable performance.



(a) Throughput of different actions



(b) Throughput with different number of flows

CyberStar Architecture

Service Computing

❑ Reliability

- Node crashing leads to state loss.
- State replica to prevent state loss.

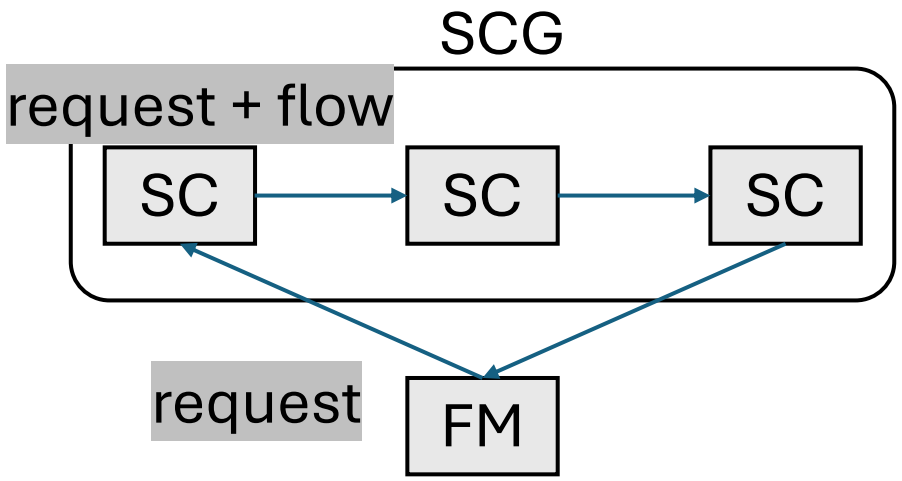
❑ Scalability

- State synchronization overhead increases with the SC plane scaling.
- Minimize the impact of state synchronization on service computation.

Service Computing

❑ Reliability

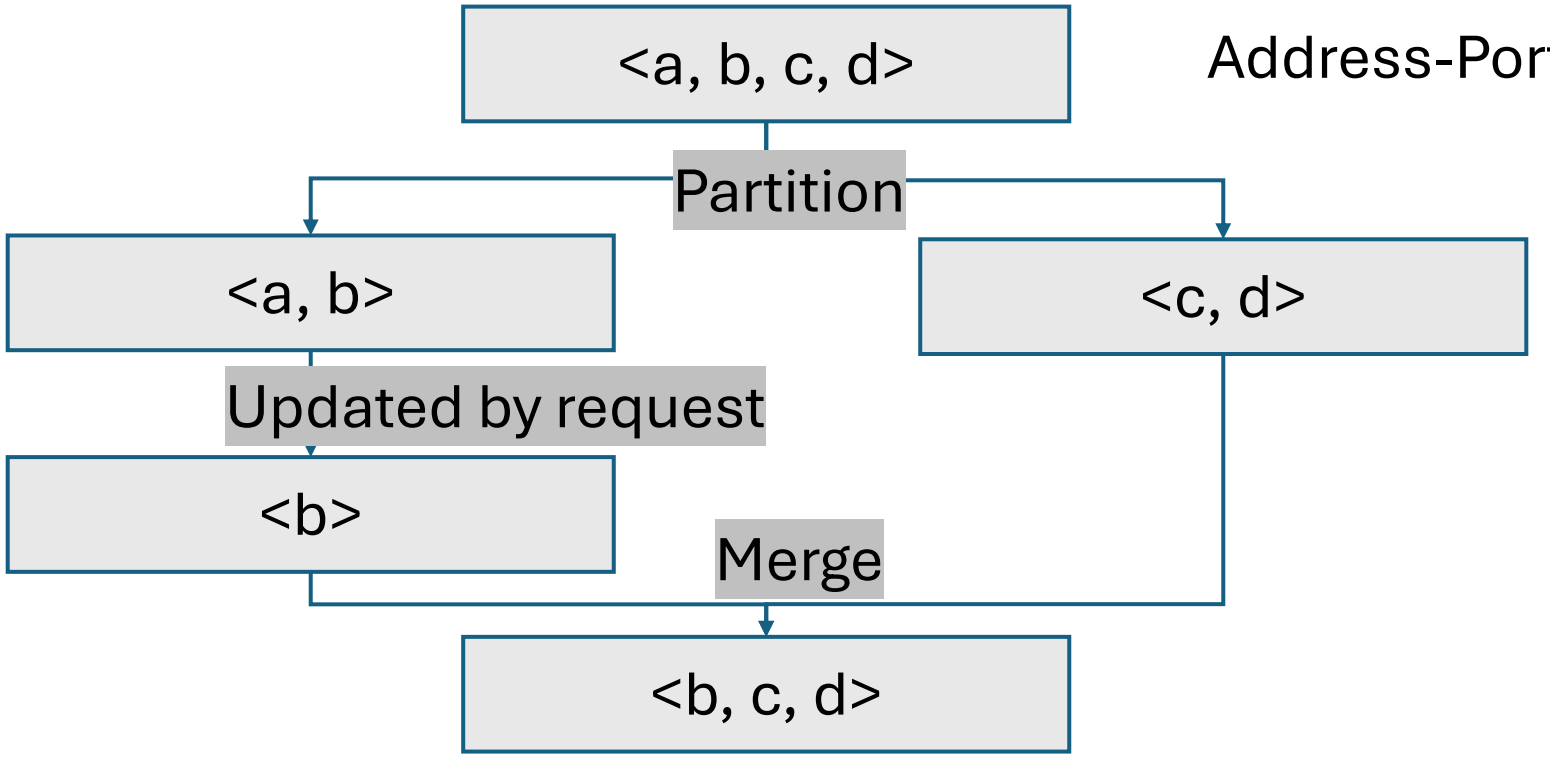
- Reliability Group, aka, SCG.
- Packet-pass-through within reliability group.



Service Computing

❑ Scalability

- State partition for shared states.



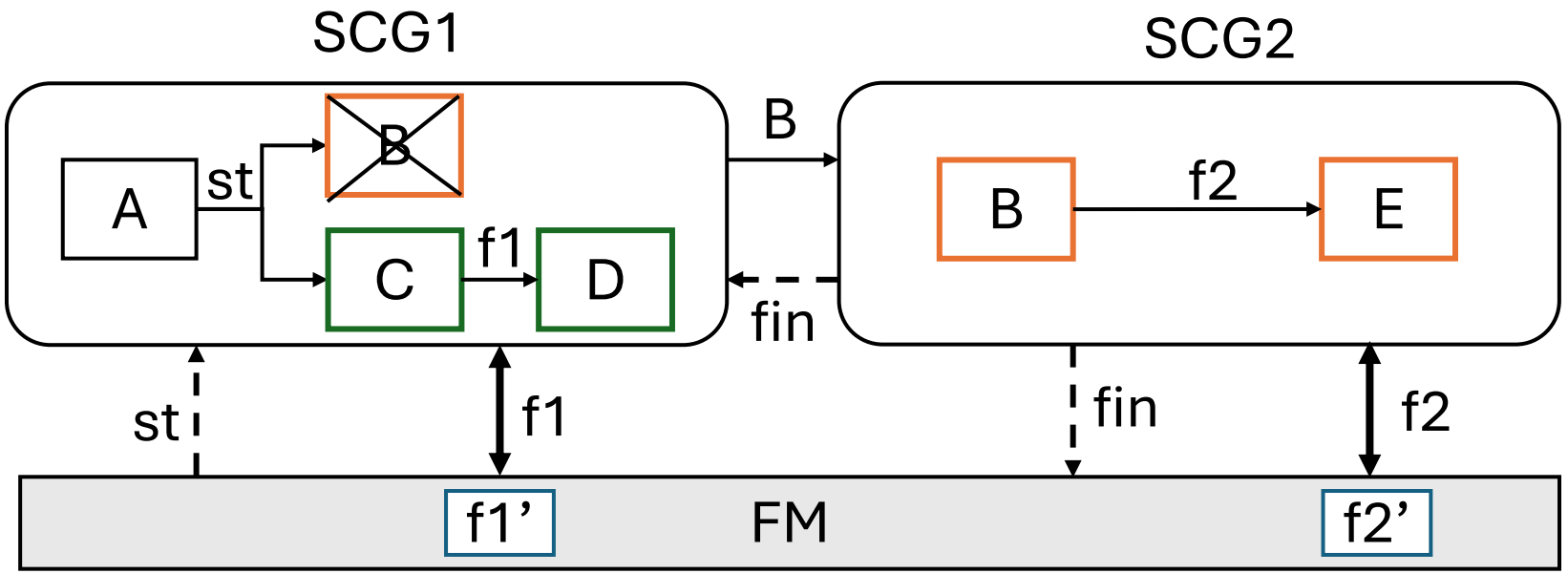
Address-Port Pool for SNAT

CyberStar Architecture

Service Computing

❑ Scalability

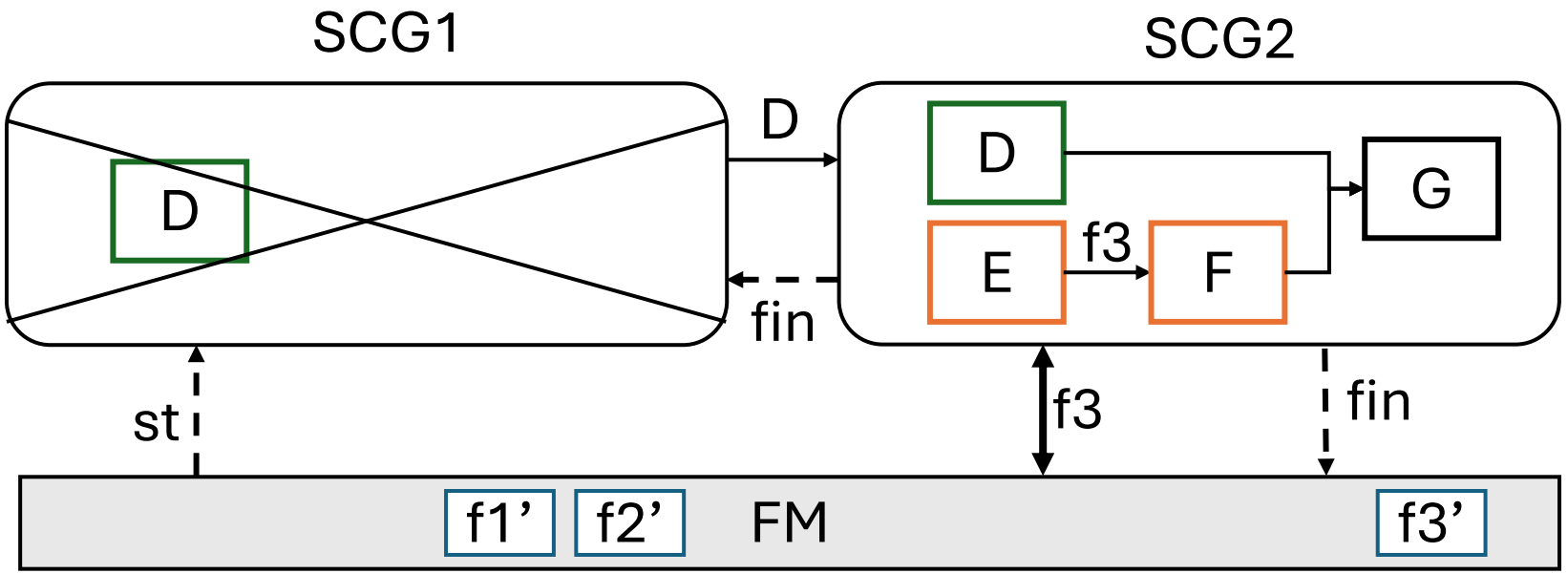
- State partition intra-SCGs
- State synchronization during **scaling out**



Service Computing

❑ Scalability

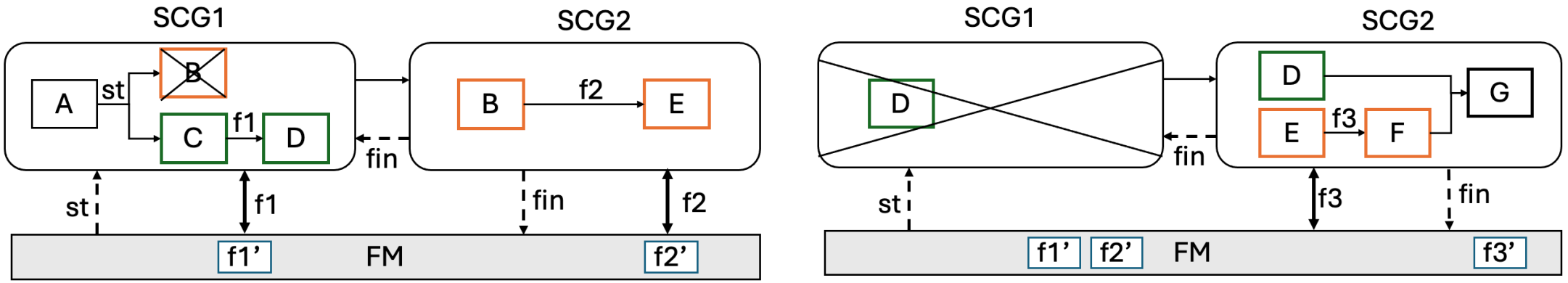
- State partition intra-SCGs
- State synchronization during **scaling in**



CyberStar Architecture

Fabric Master

- Decouple the SC and PP plane.



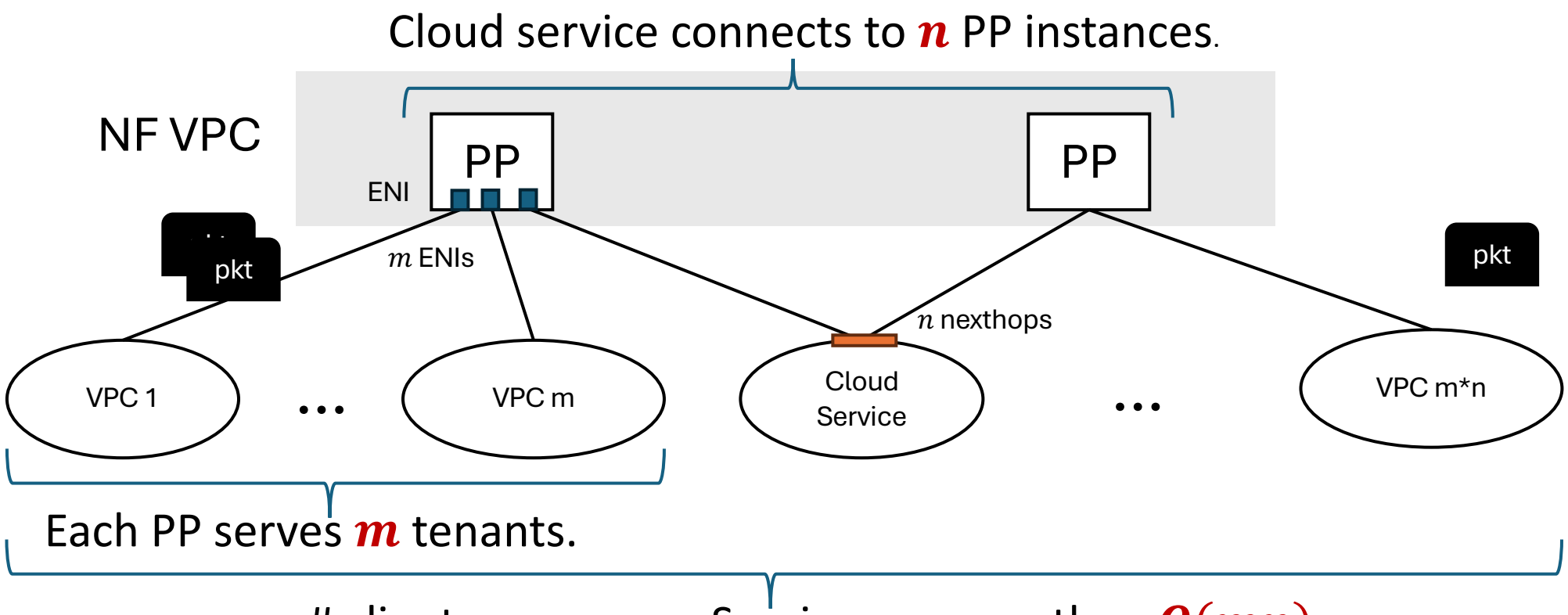
1. It delivers requests to SC plane, so the scaling of SC is transparent to PP plane.
2. It caches flows for repeat requests, so the scaling of PP is transparent to SC plane.

CyberStar Architecture

Elastic Network Interface (ENI)

Fabric Master

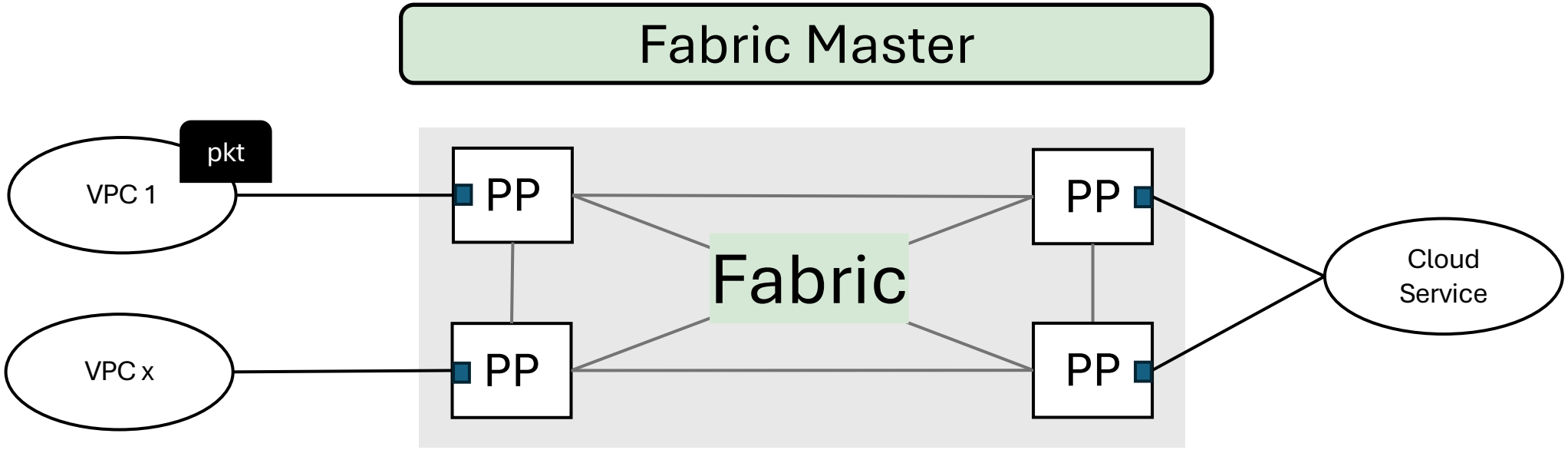
- Improve the tenant accessing cloud services through NFs.



CyberStar Architecture

Fabric Master

- Fabric improves the tenant accessing

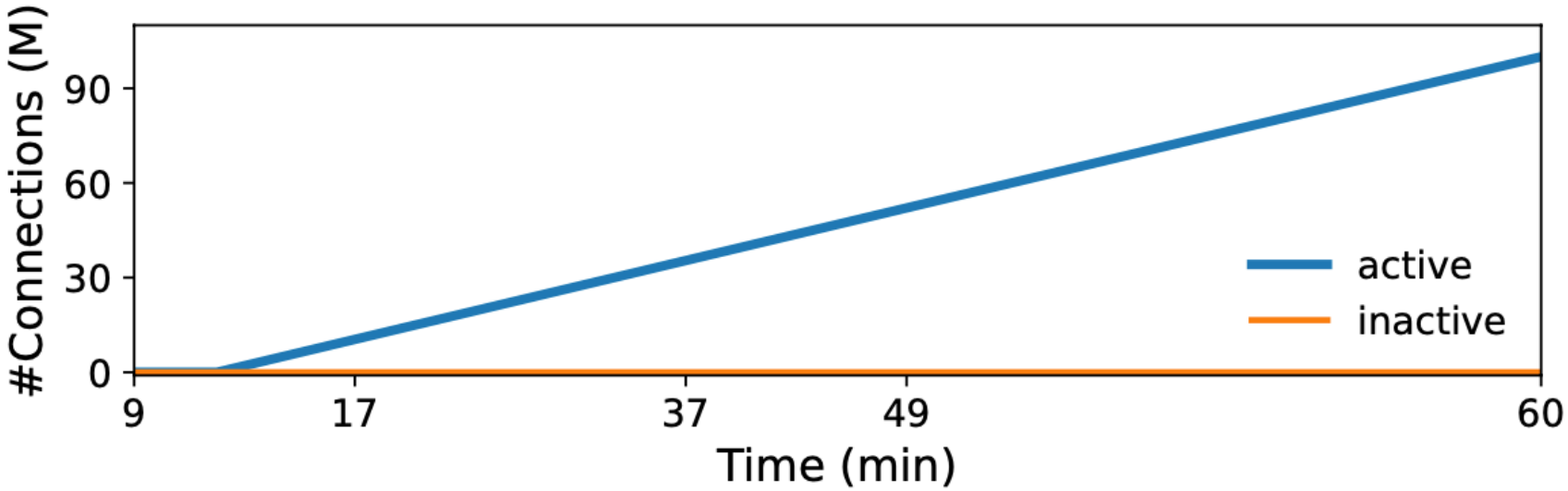


ECS instances belonging to the same VPC can connect without the requirement of extra ENI.

Evaluation of Scalability

Scalability

- 35 K connections per seconds, 100 million active connections

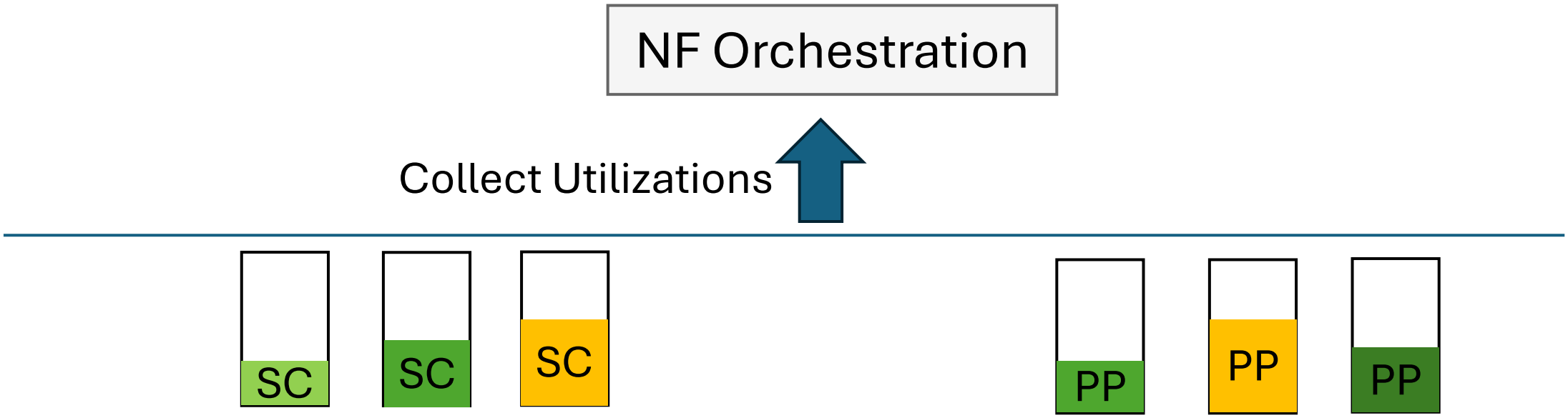


- 35 ECSs
- Each ECS is equipped with 32 vCPUs, 128GB memory, 15Gbps bandwidth

CyberStar Architecture

NF Orchestration

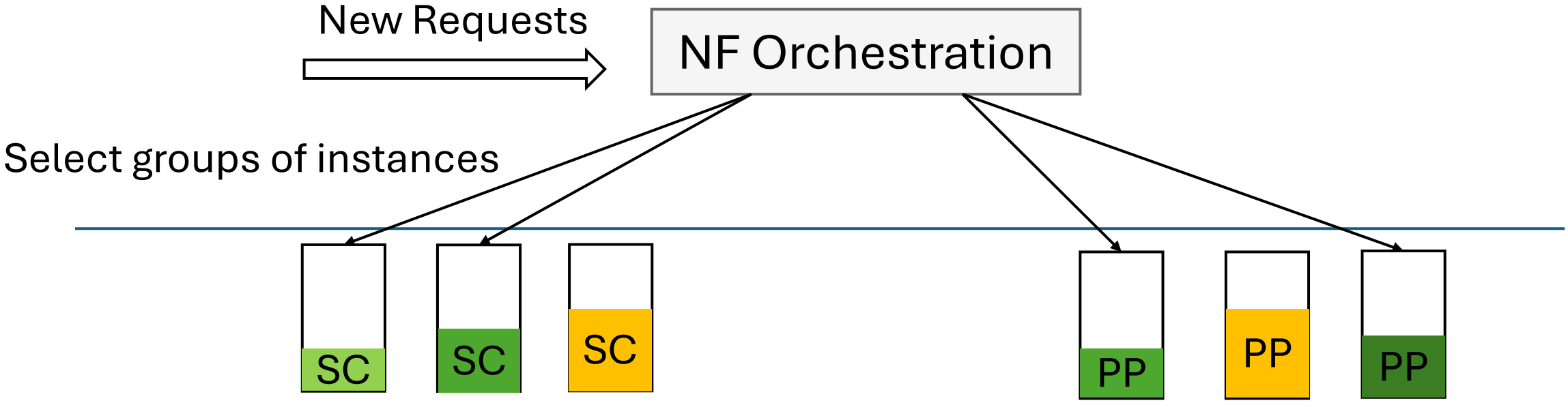
- Monitor long-term resource utilizations



CyberStar Architecture

NF Orchestration

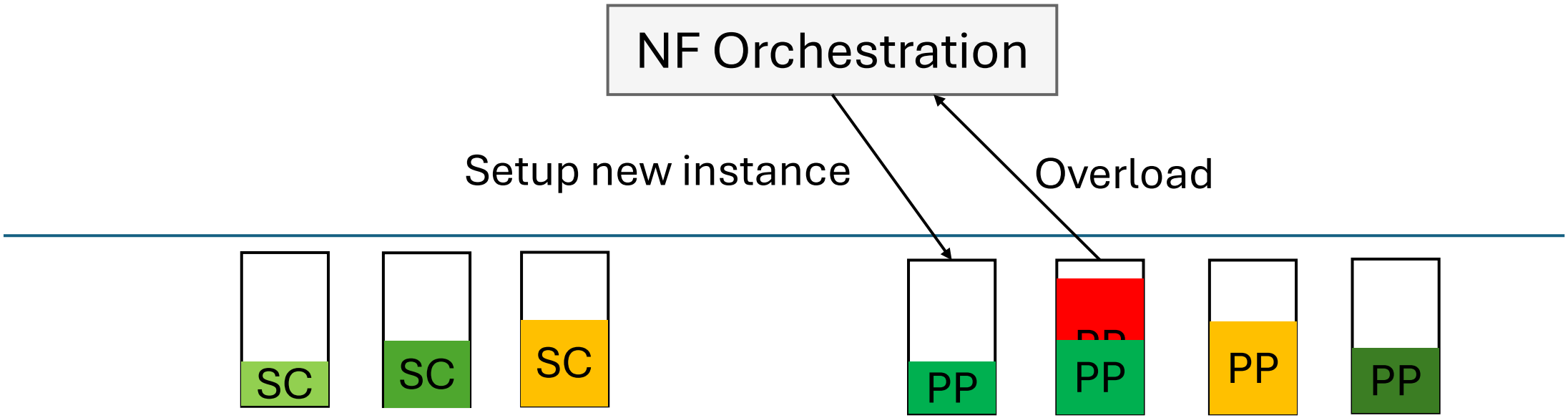
- ❑ Monitor long-term resource utilizations
- ❑ Determine how tenants' traffic is dispatched into ECSs



CyberStar Architecture

NF Orchestration

- ❑ Monitor long-term resource utilizations
- ❑ Determine how tenants' traffic is dispatched into ECSs
- ❑ Decide when scaling events are triggered

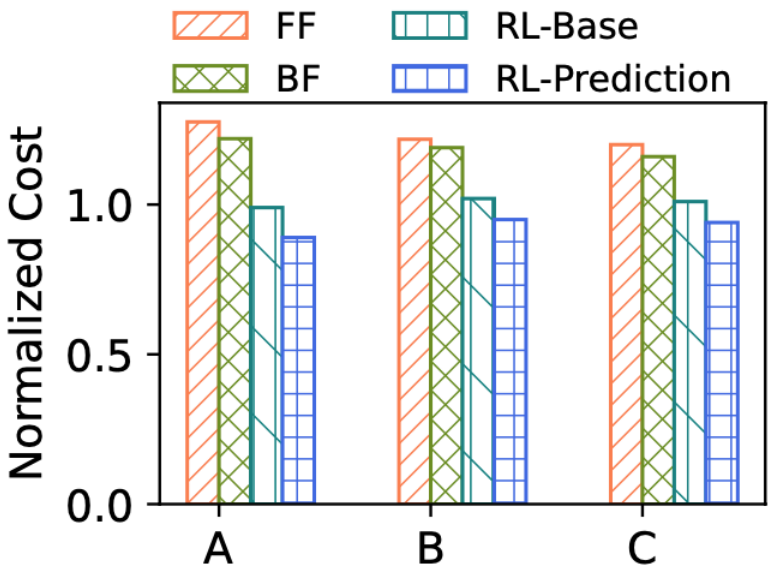


Evaluation of NF Auto-Scaling

Algorithm

❑ Alternatives

- First-Fit algorithm (FF), an online algorithm for the multi-dimensional vector bin packing.
- Weighted Best-Fit (BF) algorithm initially used in the production network.



The DRL-Base algorithm can achieve ~15%-25% lower cost compared to FF and BF.

Conclusion

CyberStar has been deployed for over four years and is publicly available in our cloud.

a. We introduce a new three-plane architecture to address scalability issues.
b. CyberStar facilitates the management of heterogeneous hardware resources in the cloud. We decompose the packet processing pipeline into NF-independent units, allowing the assembly of various NF types.
c. To improve utilization, CyberStar introduces an auto-scaling approach to minimize the cost of cloud providers.

南京大学
NANJING UNIVERSITY

Thanks for your attention!