ONCache:
A Cache-Based Low-Overhead Container Overlay Network

Shengkai Lin[1], Shizhen Zhao[1], Peirui Cao[1], Xinchi Han[1],
Quan Tian[2], Wenfeng Liu[2], Qi Wu[2], Donghai Han[2], Xinbing Wang[1]

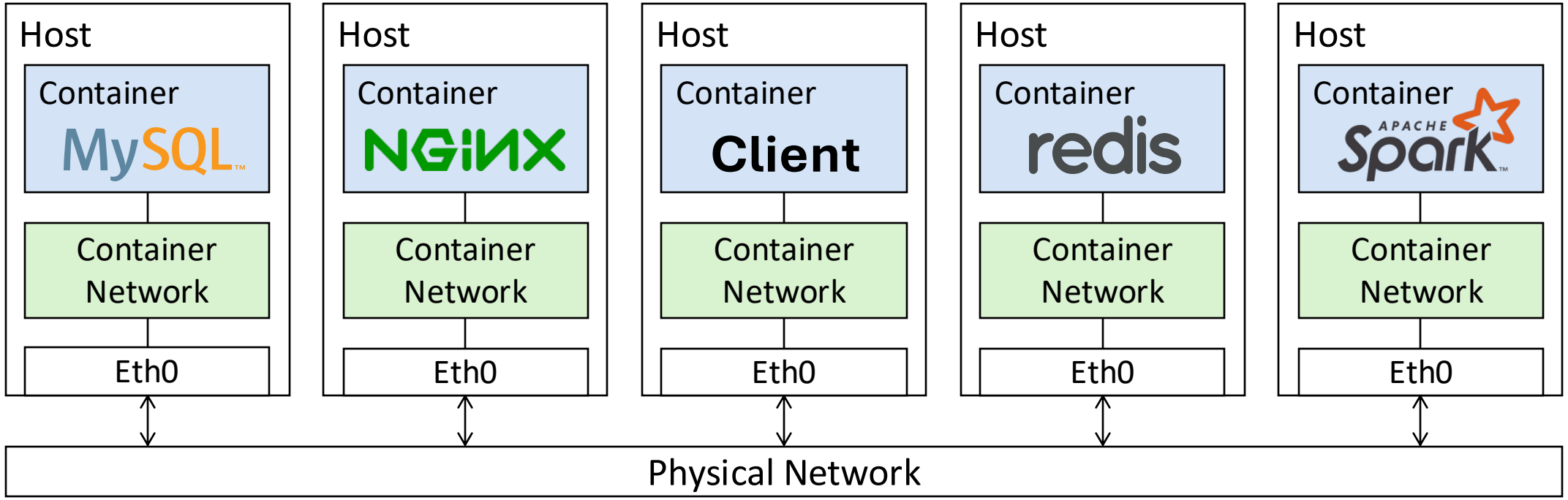
1

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

2

BROADCOM®

Containers and container networks

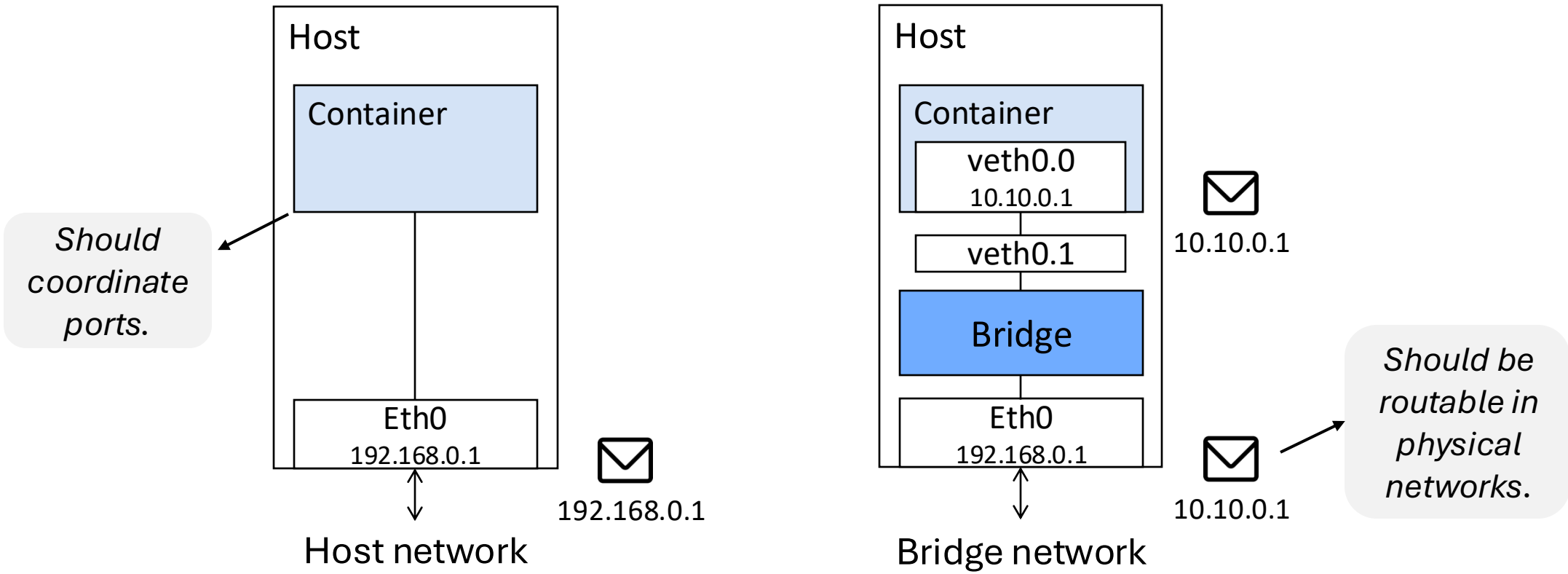


Container networks are critical to connect containers to physical networks.

In this talk…

- Revisiting container networks: None achieves low overhead, high flexibility, and high compatibility at the same time.
- ONCache: A Cache-Based Low-Overhead Container Overlay Network.
- ONCache saves CPU cycles and improves performance.

From the simplest: host and bridge networks



👍 **Low overhead, Good compatibility**

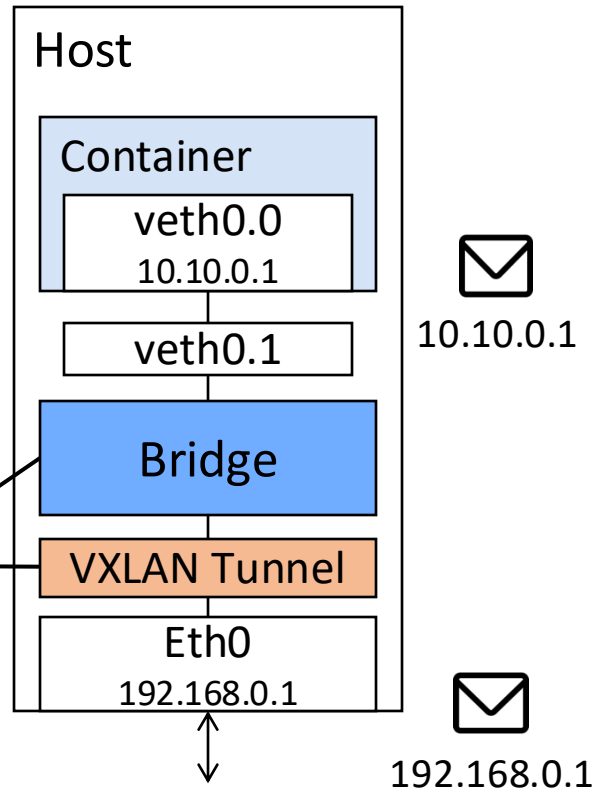
👎 **Poor flexibility**: Containers and physical networks are coupled.

Higher overhead: overlay networks

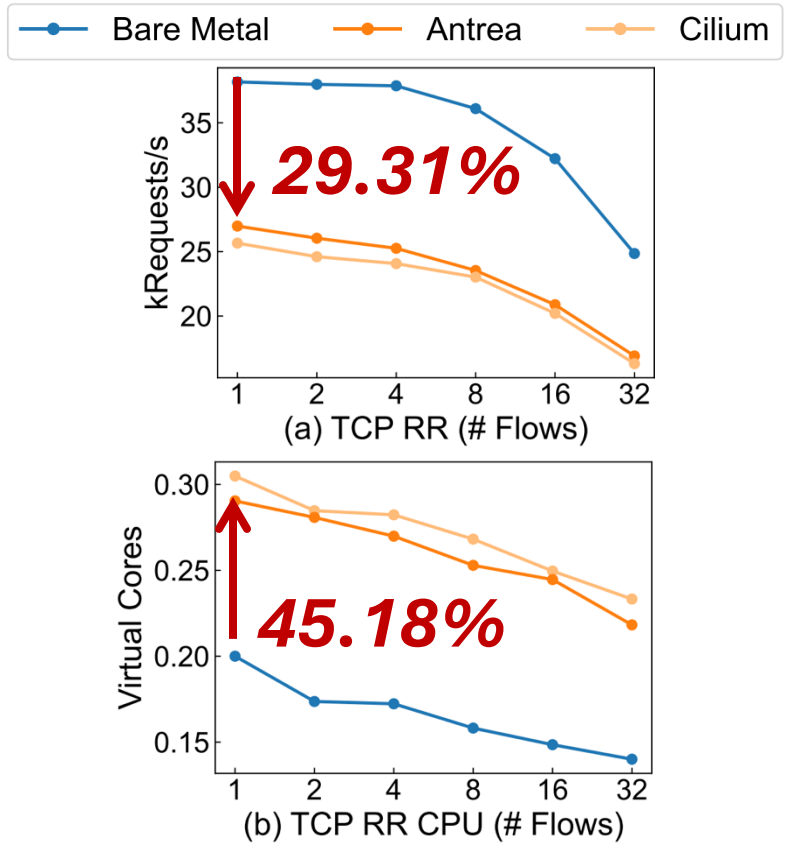
flannel

ANTREA

cilium



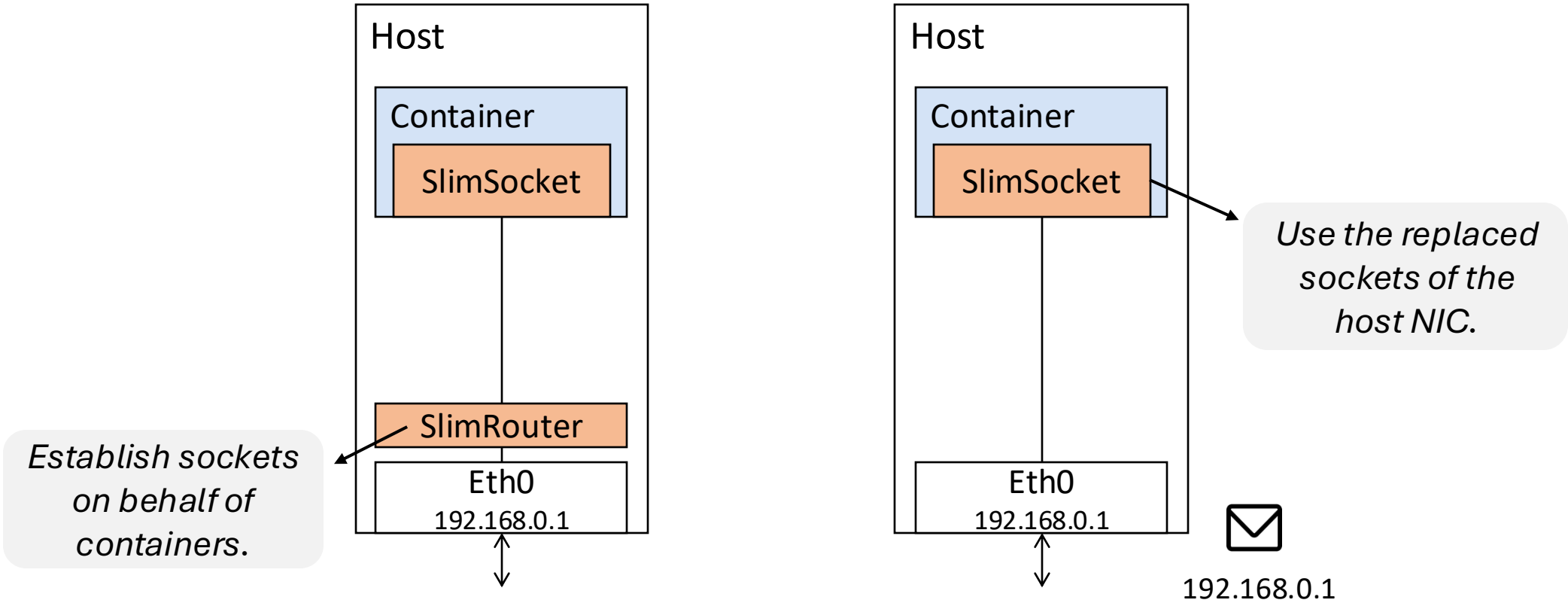
Incur significant overhead.



👍 **High flexibility, Good compatibility**

👎 **High overhead**: More processing on container packets.

Poorer compatibility: socket replacement



High flexibility, Low overhead

Poor compatibility: Poor support for UDP, ICMP, etc.

[1] Zhuo, Danyang, et al. "Slim:{OS} Kernel Support for a {Low-Overhead} Container Overlay Network." 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19). 2019.

None of current solutions are perfect

Technology	Performance	Flexibility	Compatibility
Host	✓	✗	✓
Bridge	✓	✗	✓
Macvlan	✓	✗	✓
IPvlan	✓	✗	✓
SR-IOV [64]	✓	✗	✓
Overlay	✗	✓	✓
Falcon [37]	✗	✓	✓
Slim [74]	✓	✓	✗
ONCache	✓	✓	✓

[1] Jiaxin Lei, Manish Munikar, Kun Suo, Hui Lu, and Jia Rao. Parallelizing packet processing in container overlay networks. EuroSys 2021, 2021.
[2] Zhuo, Danyang, et al. "Slim:{OS} Kernel Support for a {Low-Overhead} Container Overlay Network." 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19). 2019.

In this talk…

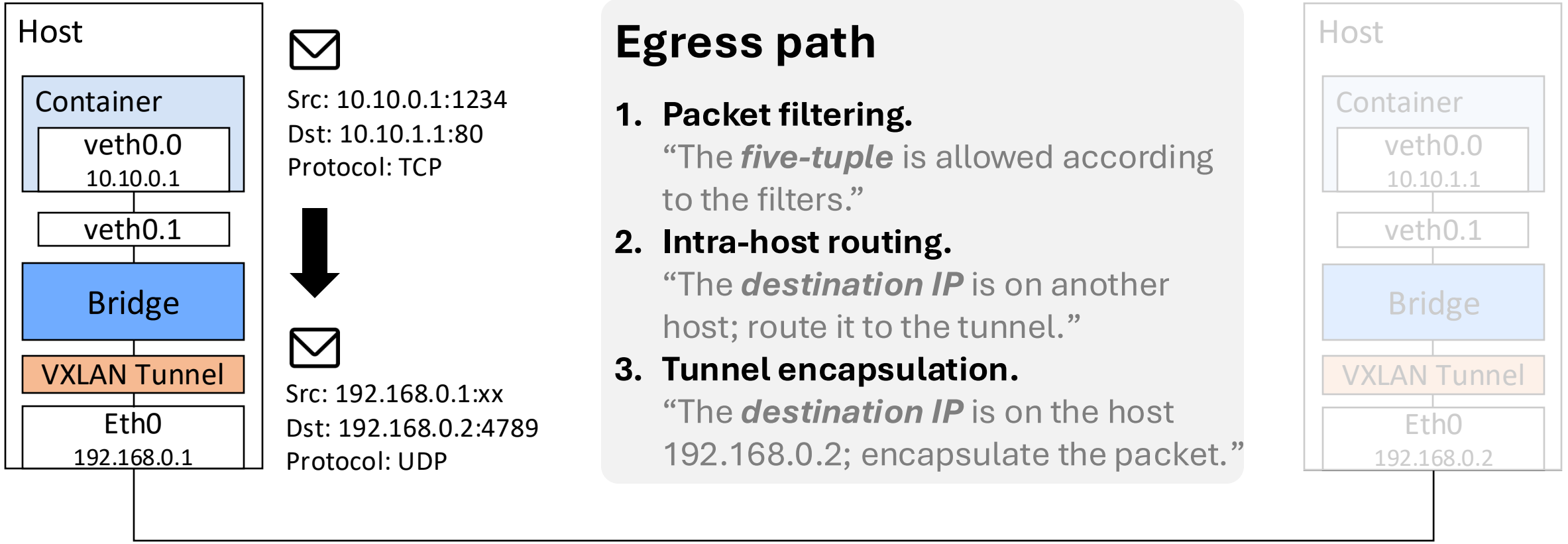
- Revisiting container networks: None achieves low overhead, high flexibility, and high compatibility at the same time.
- ONCache: A Cache-Based Low-Overhead Container Overlay Network.
- ONCache saves CPU cycles and improves performance.

Dive into the overhead of overlay networks

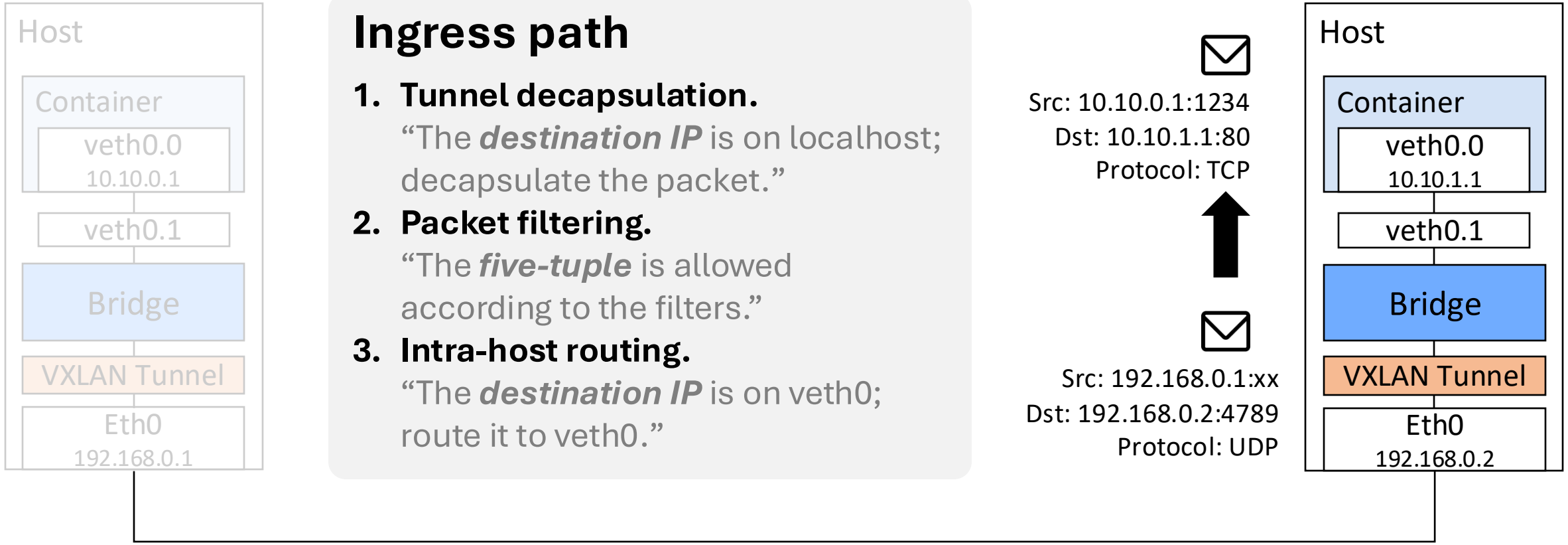
	Egress				Ingress			
Data path	**Overhead type**	**Antrea**	**Cilium**	**Bare metal**	**Overhead type**	**Antrea**	**Cilium**	**Bare metal**
Application network stack	skb allocation	1505	1566	1461	skb releasing	715	818	780
	Conntrack	778	0	788	Conntrack	616	0	600
	Netfilter	0	0	305	Netfilter	0	0	173
	Others	423	560	547	Others	838	1016	979
Veth pair	**NS traversing**	**562**	**594**		**NS traversing**	**400**		
eBPF	**eBPF**		**1513**		**eBPF**		**1429**	
Open vSwitch	**Conntrack**	**872**			**Conntrack**	**758**		
	Flow matching	**354**			**Flow matching**	**308**		
	Action execution	**92**			**Action execution**	**66**		
VXLAN network stack	**Conntrack**	**0**	**471**		**Conntrack**	**0**	**271**	
	Netfilter	**667**	**421**		**Netfilter**	**466**	**303**	
	Routing	**50**	**468**		**Routing**	**294**	**554**	
	Others	**319**	**127**		**Others**	**619**	**444**	
Link Layer	Link Layer	1858	1763	1799	Link Layer	2790	2848	2800

Overlay networks take up near **half of CPU time** on egress & ingress.

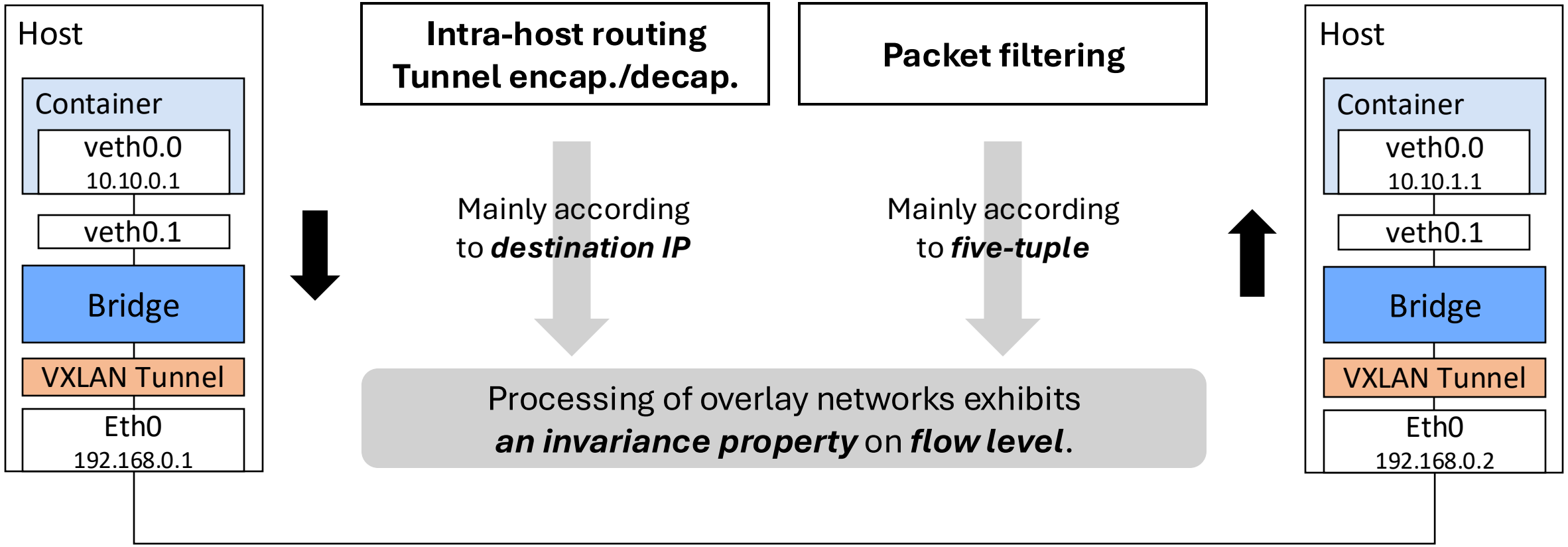
Dive into the overhead of overlay networks



Dive into the overhead of overlay networks

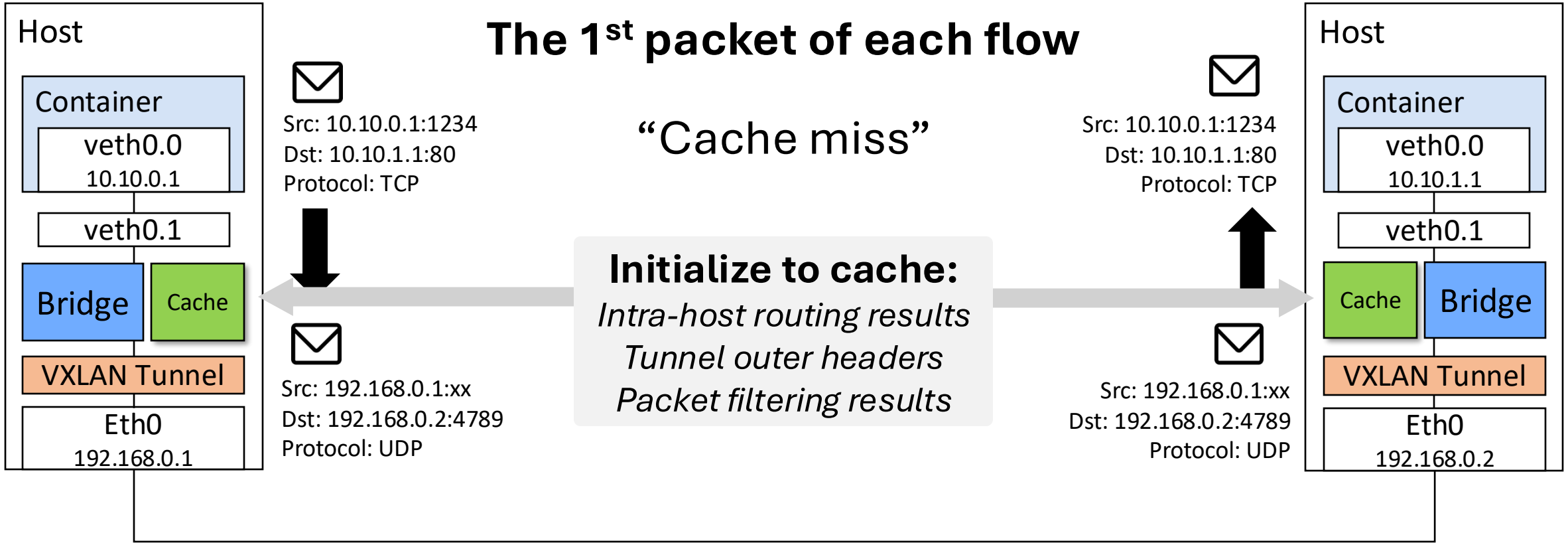


Opportunity: *the invariance property*



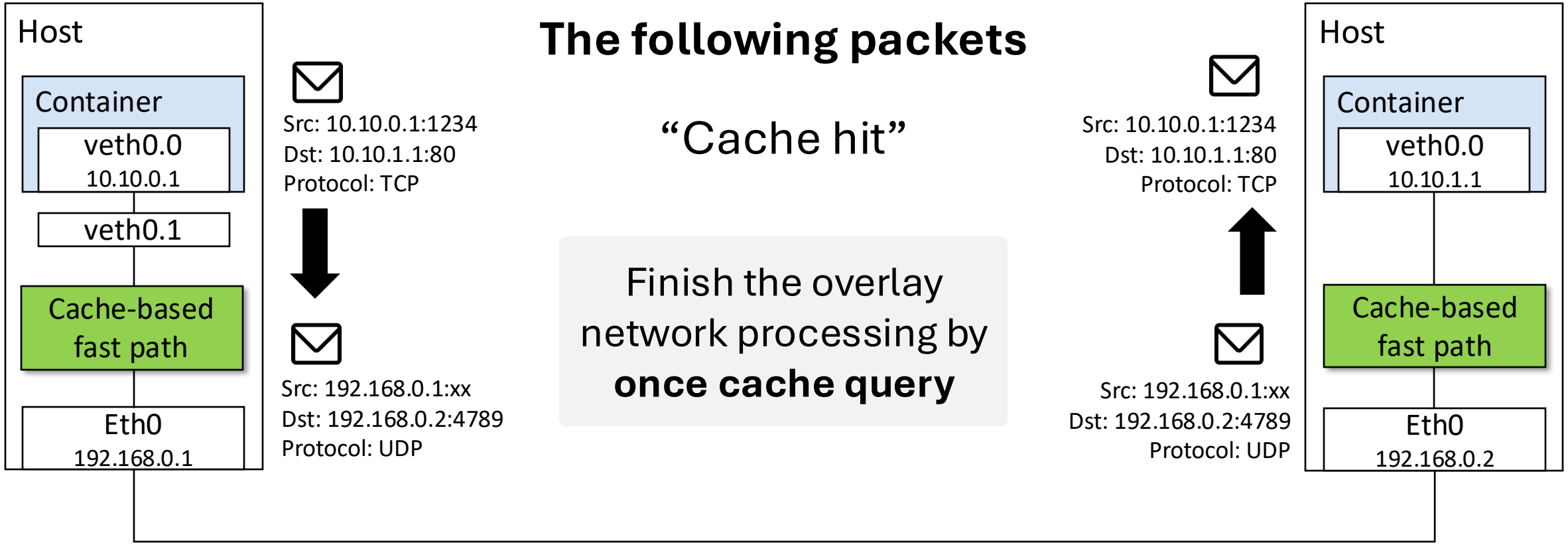
ONCache: bring caches into overlay networks

ONCache = **O**verlay **N**etwork **Cache**



ONCache: bring caches into overlay networks

ONCache = **O**verlay **N**etwork **Cache**



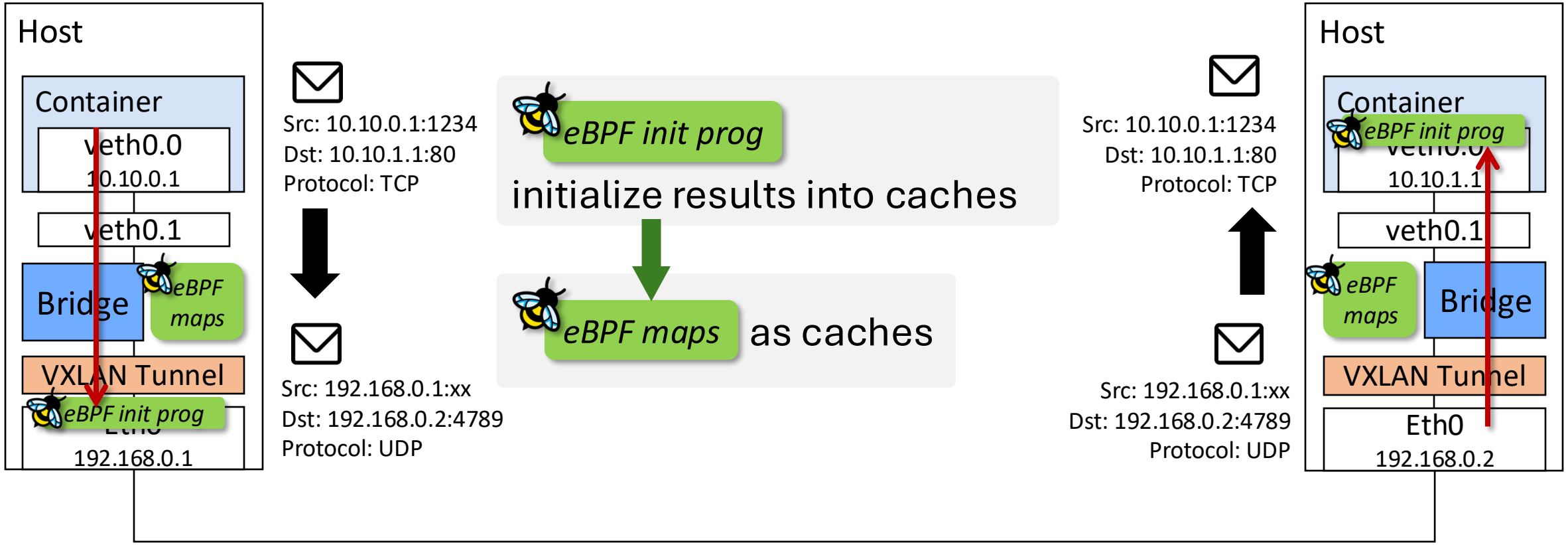
ONCache: bring caches into overlay networks

ONCache = **O**verlay **N**etwork **Cache**

How to implement such design in the network stack?

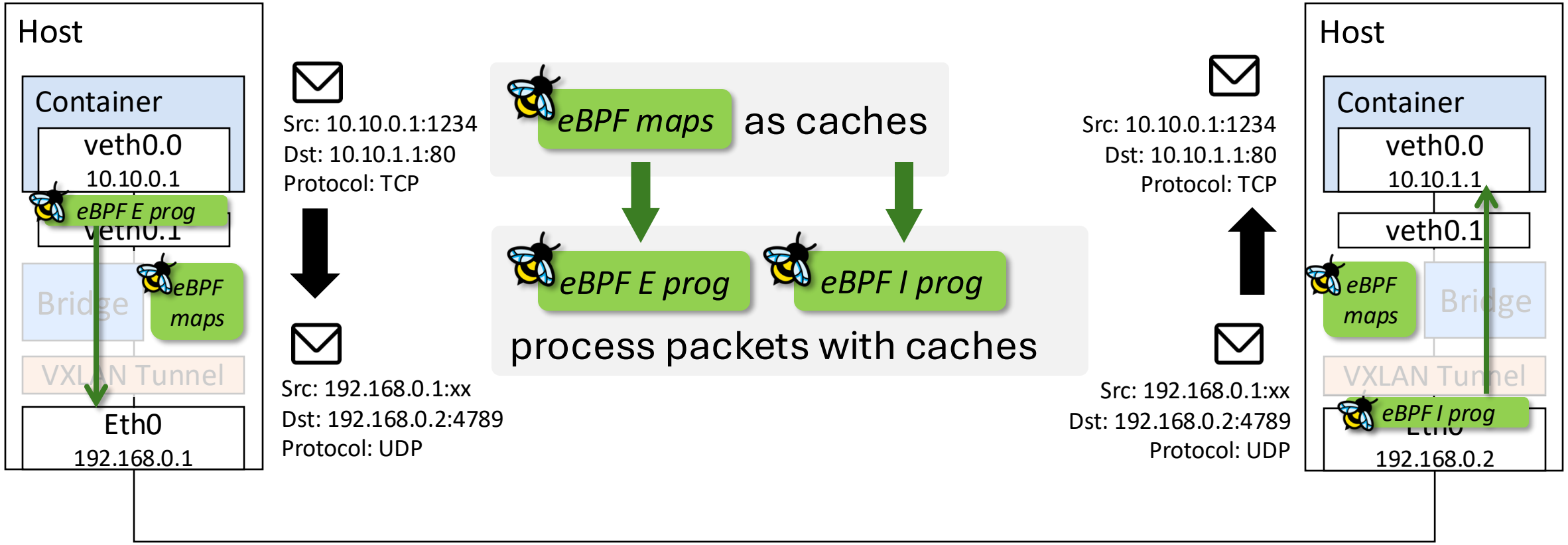
ONCache: eBPF-based design

eBPF (extended Berkeley Packet Filter) can safely run programs in kernel at runtime.



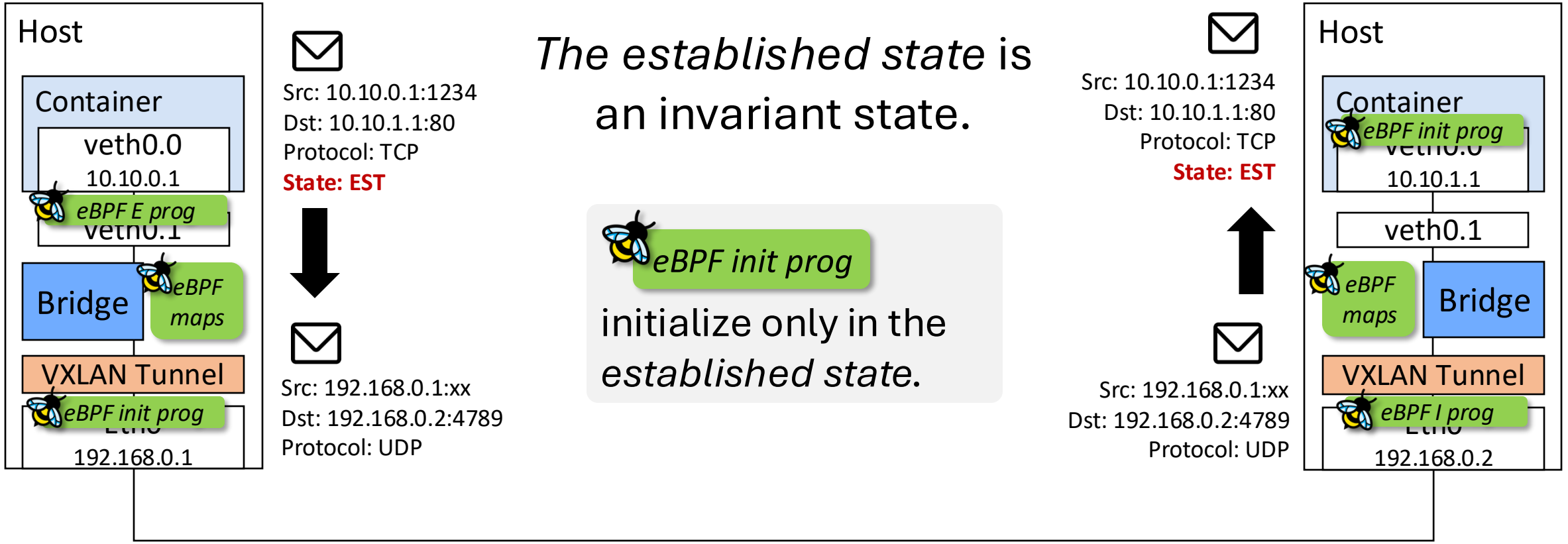
ONCache: eBPF-based design

eBPF (extended Berkeley Packet Filter) can safely run programs in kernel at runtime.

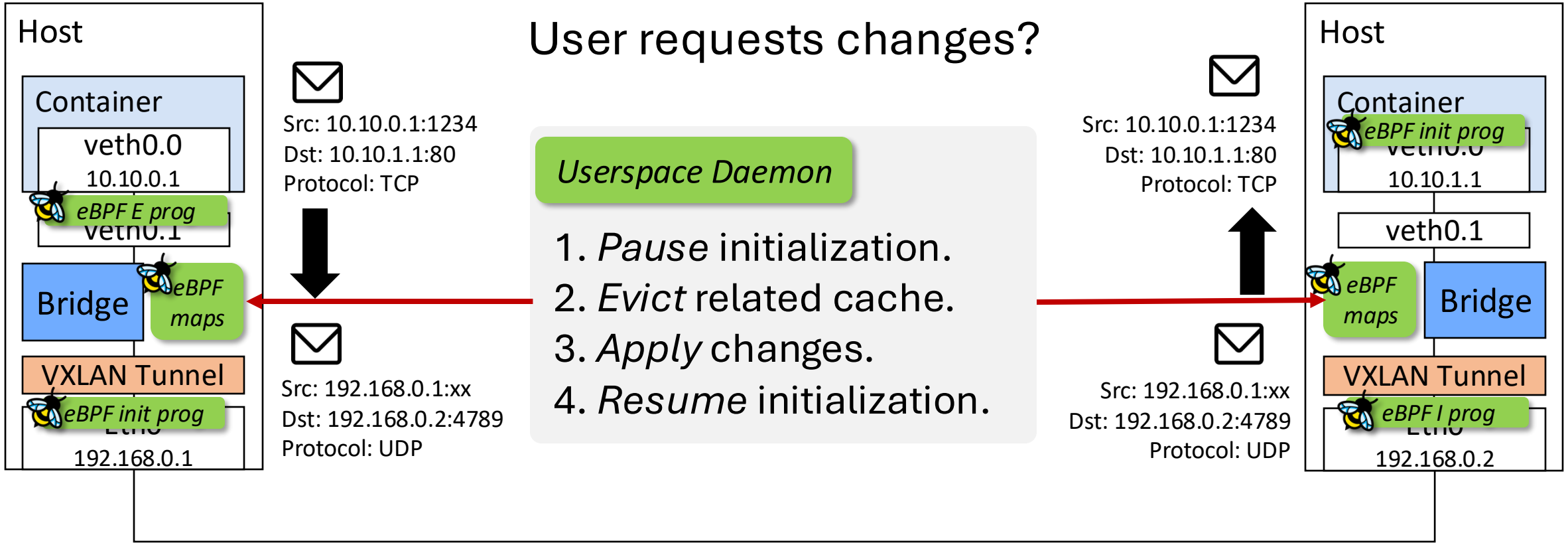


ONCache: handling stateful policies

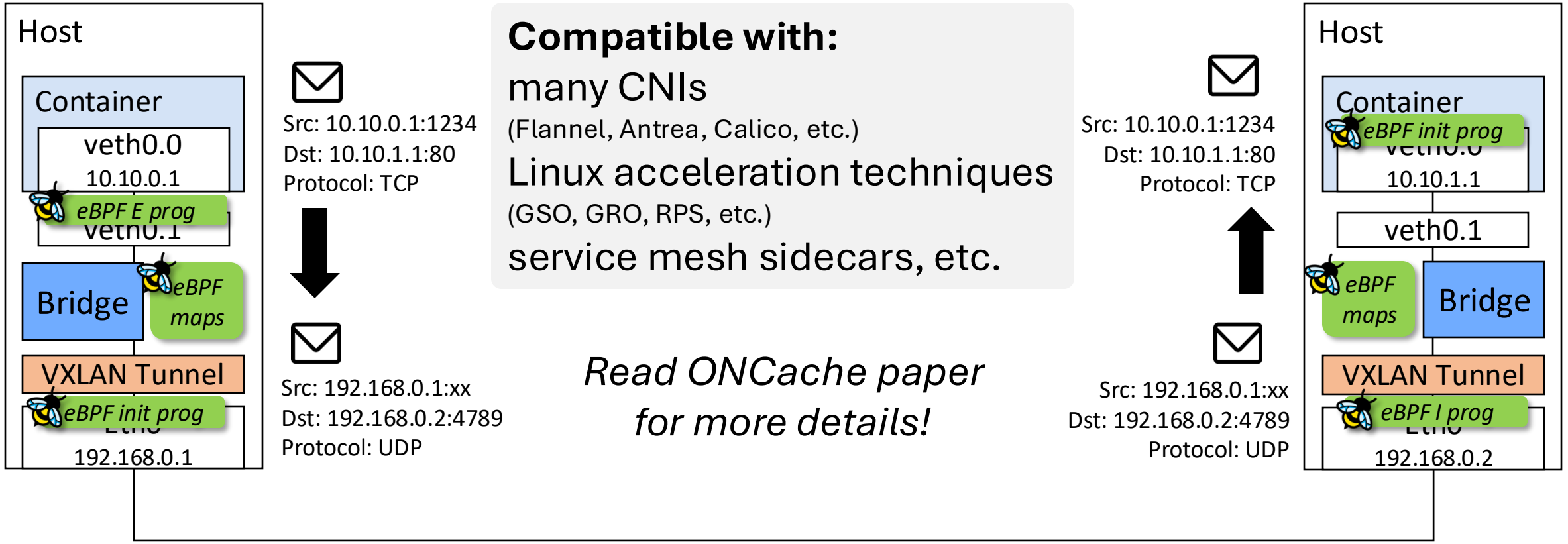
Stateful policies also match on the connection states.



ONCache: cache coherency



ONCache: compatibility



In this talk…

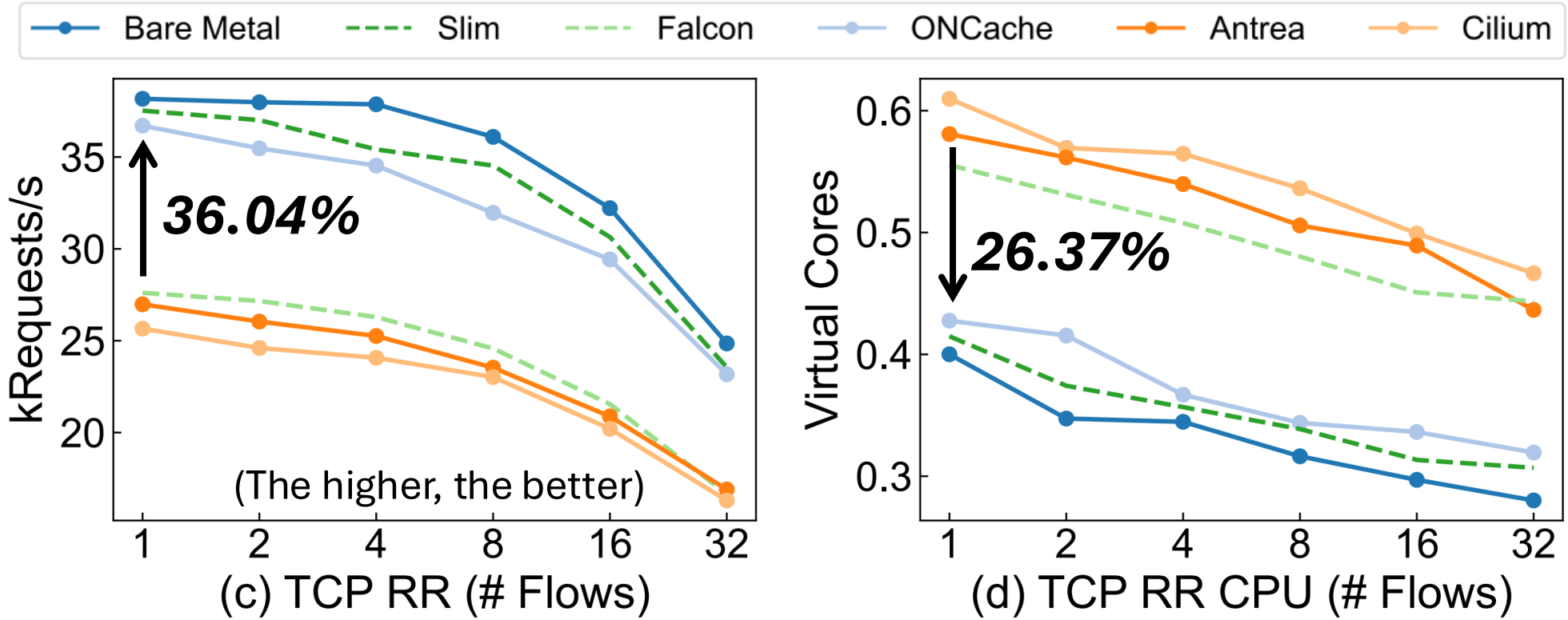
- Revisiting container networks: None achieves low overhead, high flexibility, and high compatibility at the same time.
- ONCache: A Cache-Based Low-Overhead Container Overlay Network.
- ONCache saves CPU cycles and improves performance.

CPU overhead of ONCache

	Egress				Ingress			
Data path	**Overhead type**	**Antrea**	**ONCache**	**Bare metal**	**Overhead type**	**Antrea**	**ONCache**	**Bare metal**
Application network stack	skb allocation	1505	1509	1461	skb releasing	715	714	780
	Conntrack	778	763	788	Conntrack	616	592	600
	Netfilter	0	0	305	Netfilter	0	0	173
	Others	423	519	547	Others	838	982	979
Veth pair	**NS traversing**	**562**	**489**		**NS traversing**	**400**		
eBPF	**eBPF**		**511**		**eBPF**		**289**	
Open vSwitch	**Conntrack**	**872**			**Conntrack**	**758**		
	Flow matching	**354**			**Flow matching**	**308**		
	Action execution	**92**			**Action execution**	**66**		
VXLAN network stack	**Conntrack**	**0**			**Conntrack**	**0**		
	Netfilter	**667**			**Netfilter**	**466**		
	Routing	**50**			**Routing**	**294**		
	Others	**319**			**Others**	**619**		
Link Layer	Link Layer	1858	1700	1799	Link Layer	2790	2737	2800

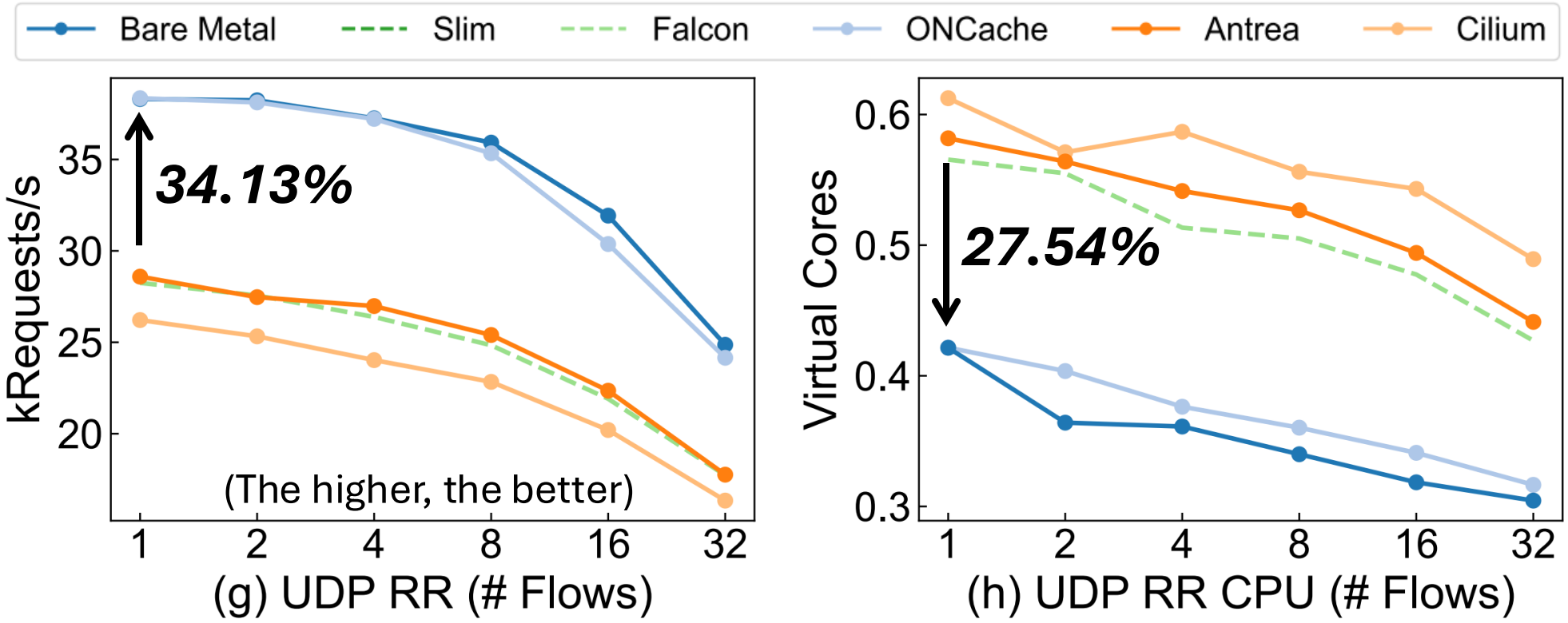
ONCache significantly eliminates the overlay networks' overhead.

TCP 1-byte Request-Response transaction



RR increased by 35.81% ~ 40.91%
CPU utilization reduced by 26.02% ~ 32.03% per byte

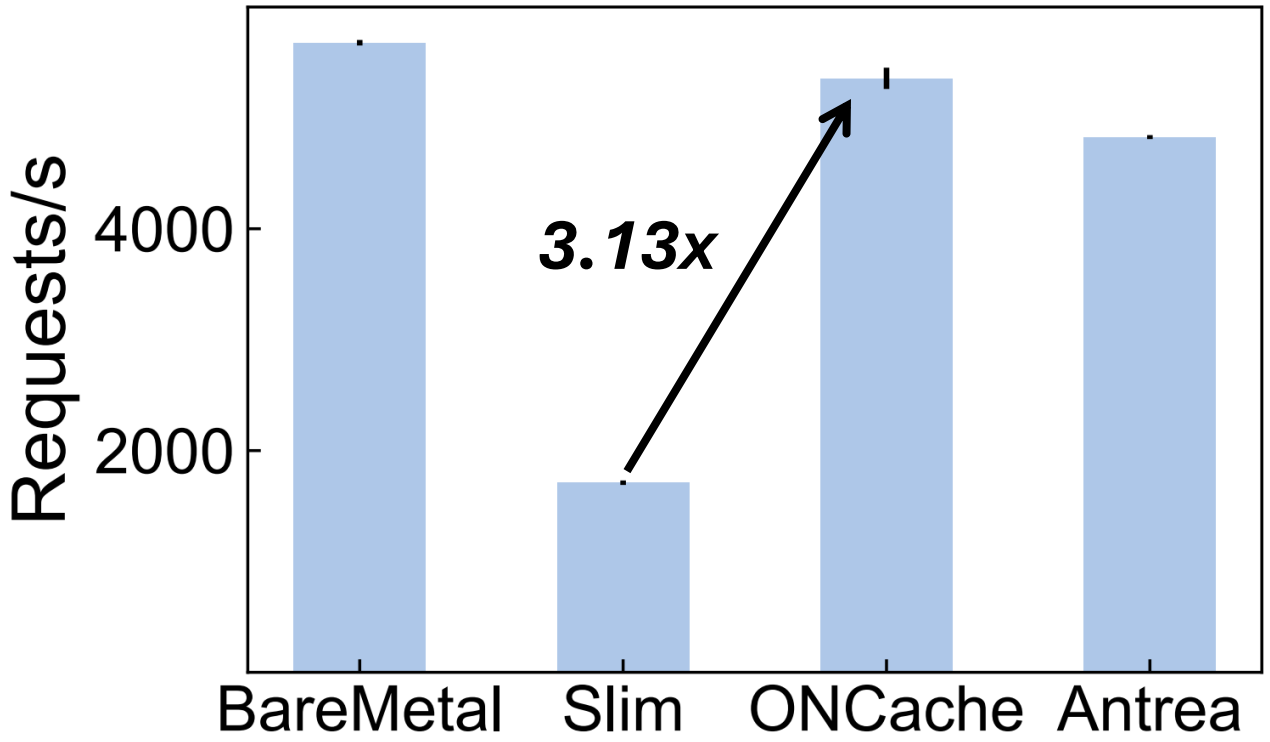
UDP 1-byte Request-Response transaction



RR increased by 34.13% ~ 39.12%
CPU utilization reduced by 27.54% ~ 31.59% per byte

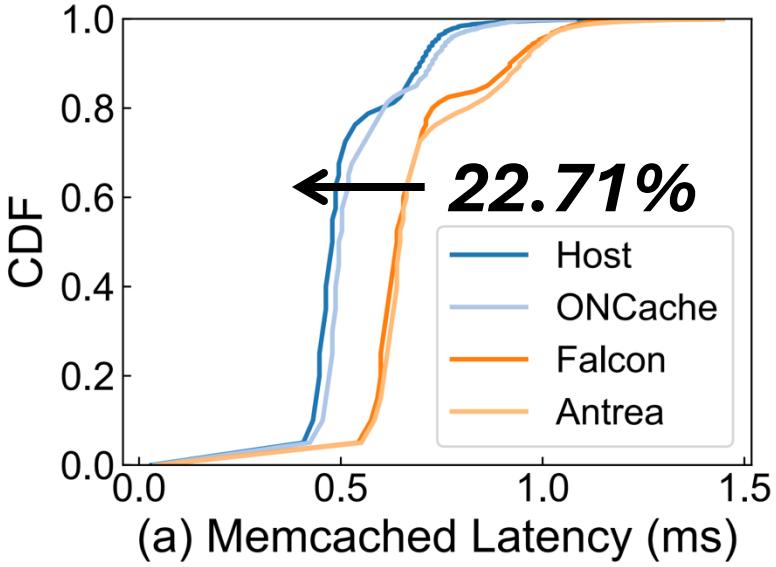
Connect-Request-Response (CRR) rate

(The higher, the better)

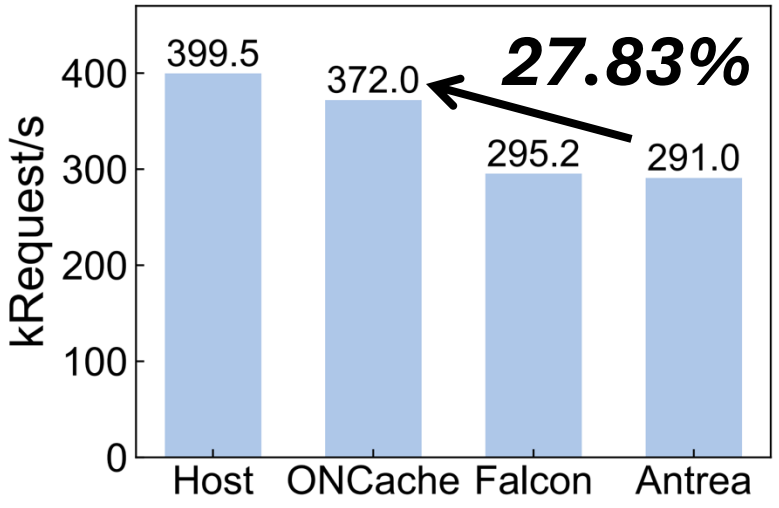


Connect-Request-Response (CRR) rate is
3.13x of that of Slim.

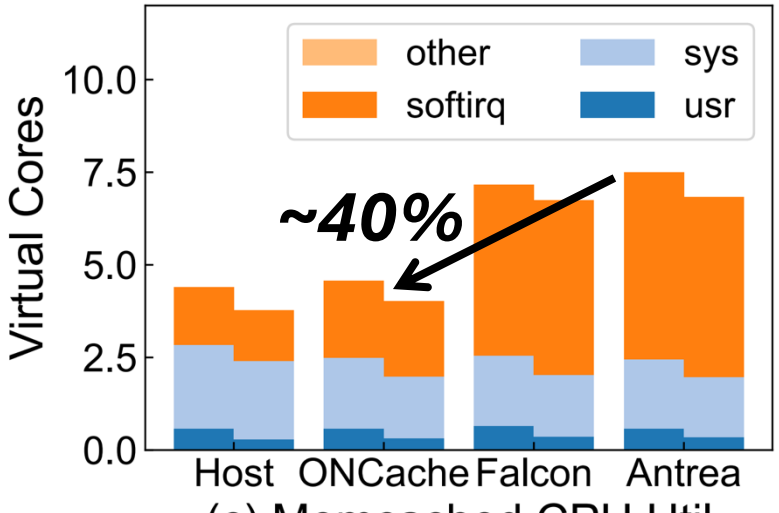
Memcached benchmarks



(a) Memcached Latency (ms)



(b) Memcached TPS



(c) Memcached CPU Util

Beneficial for latency, transaction rate, and CPU utilization of popular distributed applications.

ONCache conclusion

ONCache = **O**verlay **N**etwork **Cache**

- ONCache brings **cache** to container overlay networks.
- Significantly reduces overhead while maintaining flexibility and compatibility.
- Easy-to-deployment eBPF-based design.

Try ONCache at https://github.com/shengkai16/ONCache

Welcome to raise issues and pull requests!

Email: jefflin@sjtu.edu.cn ; shizhenzhao@sjtu.edu.cn

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

BROADCOM®