NUMASFP: NUMA-Aware Dynamic Service Function Chain Placement in Multi-Core Servers

Venkatarami Reddy Chintapalli, Sai Balaram Korrapati, Bheemarjuna Reddy Tamma and Antony Franklin A

Networked Wireless Systems (NeWS) Laboratory
Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad, India

14th International Conference on COMmunication Systems & NETworkS

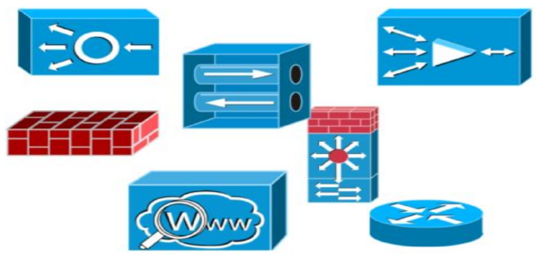
5 January 2022

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

COMSNETS

Network Function Virtualization (NFV)

- Hardware (middlebox) ➡ Software (VNFs)



Classic approach

NFV approach

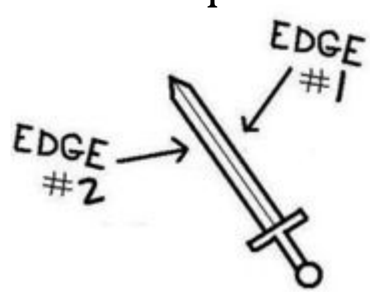
Low Cost

Flexibility

Scalability

......

Virtualization Techniques

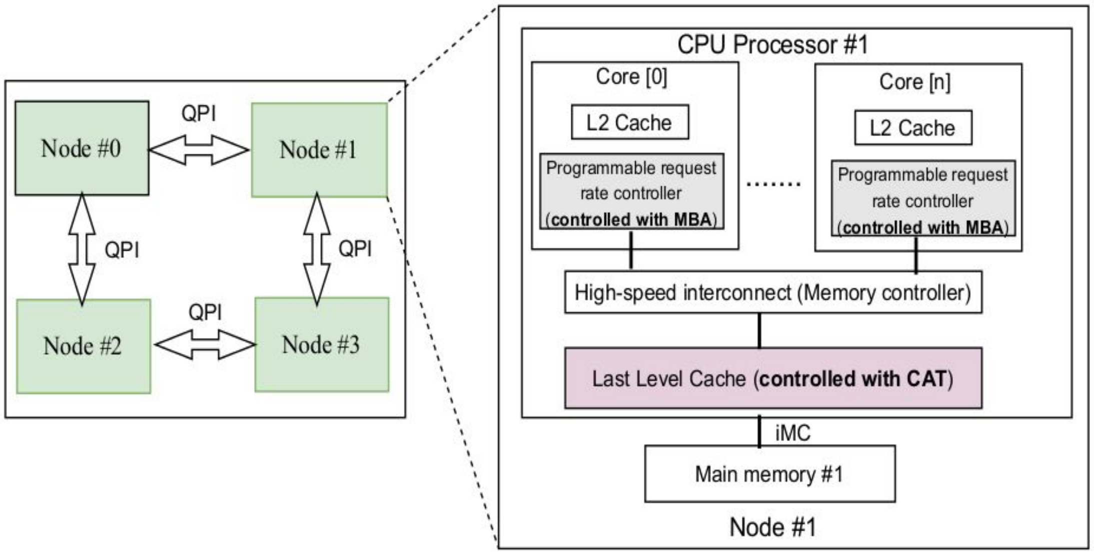


Low Throughput

High Latency

Multi-core System Architecture

- CPU cores are grouped into NUMA nodes
- Each core in a node may contend for shared resources
  - Intra-node contention
    - Last Level Cache (LLC)
    - integrated Memory Controller (iMC)
  - Inter-node contention
    - Quick Path Interconnect (QPI)



An example of a modern multi-core architecture

- VNF consolidation causes throughput degradation ranges from 12% to 50% as more VNFs are consolidated on the same server[1][2]

[1] Zeng C, Liu F, Chen S, Jiang W, Li M. Demystifying the performance interference of co-located virtual network functions, In Proc. of *IEEE INFOCOM*, 2018

[2] Manousis A, Sharma RA, Sekar V, Sherry J. Contention-Aware Performance Prediction For Virtualized Network Functions, In Proc. of *ACM SIGCOMM,* 2020

Recent Studies on NUMA Impact

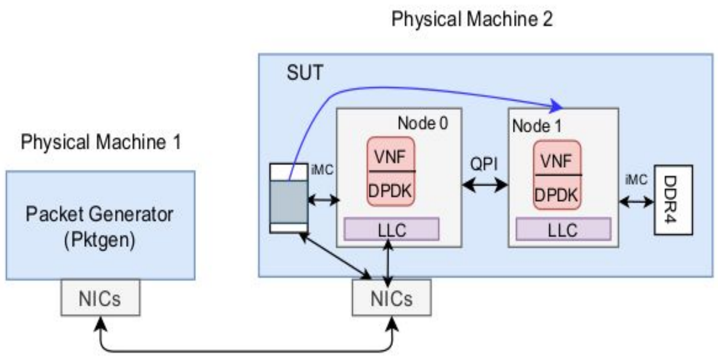
- Service Function Chain (SFC)

- Recent studies[1][2] demonstrated randomly placing VNFs of an SFC has a significant impact on the performance degradation
  - Due to inter-node resource contention
- But these works overlooked
  - Intra-node contention
  - Performance guarantee between running SFCs
  - Impact of dynamic variation in SFC traffic

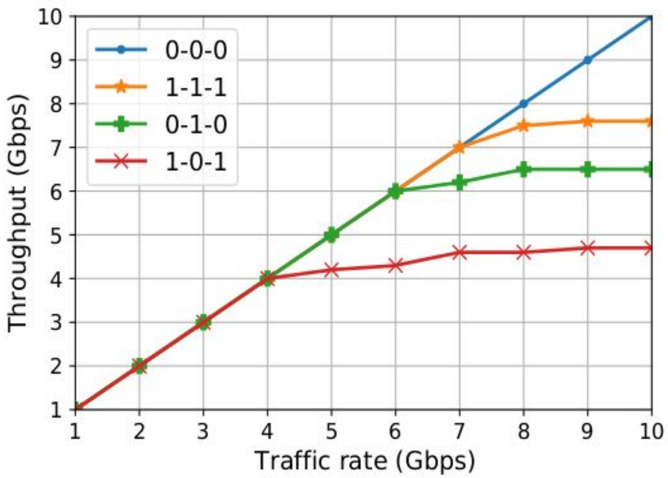
[1] Zheng Z, Bi J, Yu H, Wang H, Sun C, Hu H, Wu J. Octans: Optimal placement of service function chains in many-core systems, in Proc. of *IEEE TPDS,* 2021

[2] Sieber C, Durner R, Ehm M, Kellerer W, Sharma P. Towards optimal adaptation of nfv packet processing to modern cpu memory architectures, In Proc. of *ACM CAN,* 2017

Motivation (NUMA Impact- SFC)



Experimental setup

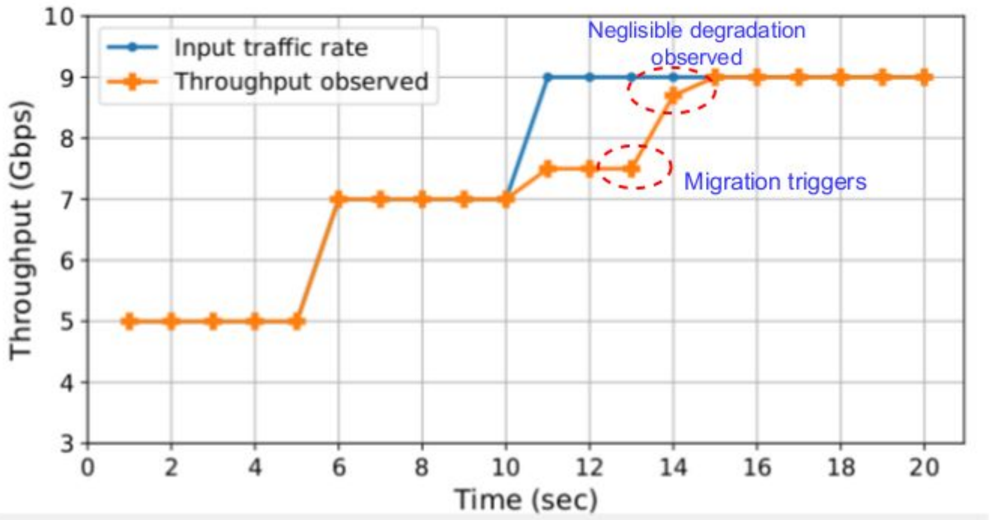


Variation in SFC throughput vs. input traffic rate for running VNFs in different combinations of *Node 0* and *Node 1*.
VNF1: Basic Monitor; VNF2: Router; VNF3:Simple Forward

- When SFC is placed in remote node $(1-1-1)$, the performance degrades by *22%* due to remote memory access overhead
- Randomly selecting cores to place VNFs in an SFC (1-0-1) can result in *52%* lower throughput compared to an optimal placement solution $(0-0-0)$

Recommends to place all VNFs of an SFC in the same node. Migrate SFCs/VNFs between nodes based on traffic rates to increase their throughput and thereby meeting the SLAs

VNF Migration between NUMA Nodes



The effect of VNF migration between NUMA nodes on VNF performance in a server

- Initially, VNF is running in *Node 1*
- Migrate VNF to *Node 0* at *13th sec* and notice seamless migration with a minimal performance impact in the immediate second
- Docker feature *cpuset* used to migrate VNF between cores in a server

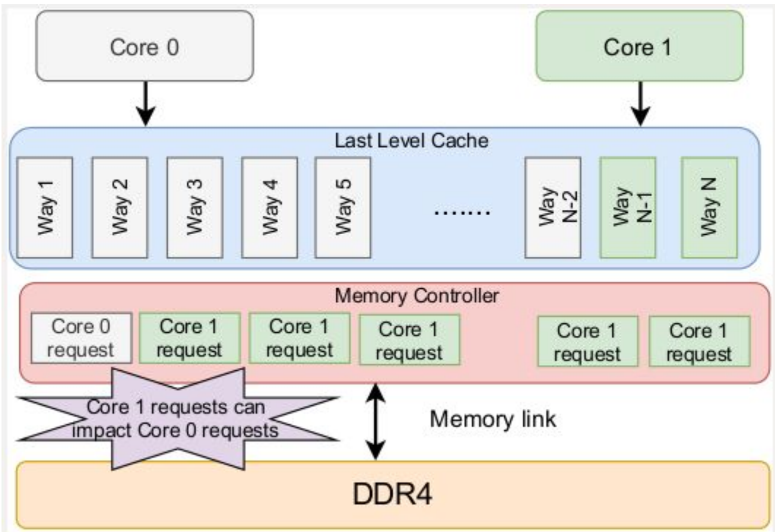
What about Performance Guarantee?

- Recent works[1],[2],[3] identified that contention at LLC is one of the primary causes for performance degradation
- Addressed it by LLC resource partitioning using Intel's Cache Allocation Technology (CAT) mechanism[4]

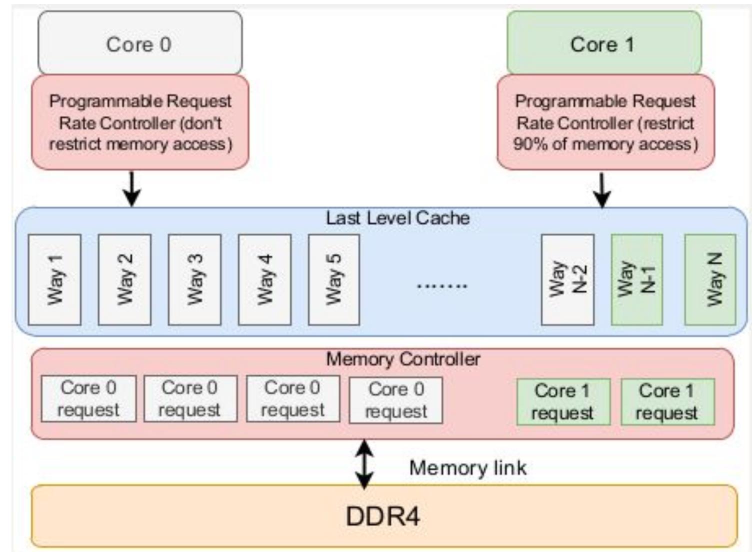
Is LLC isolation sufficient to ensure performance isolation?

[1] Veitch P, Curley E, Kantecki T. Performance evaluation of cache allocation technology for nfv noisy neighbor mitigation, In Proc. of I*EEE NetSoft,* 2017
[2] Tootoonchian A, Panda A, Lan C, Walls M, Argyraki K, Ratnasamy S, Shenker S. Resq: Enabling slos in network function virtualization, In Proc. of *ACM NSDI,* 2018
[3] Li B, Wang Y, Wang R, Tai C, Iyer R, Zhou Z, Herdrich A, Zhang T, Haj-Ali A, Stoica I, Asanovic K.RLDRM: closed loop dynamic cache allocation with deep reinforcement learning for network function virtualization, In Proc. of *IEEE Netsoft,* 2020
[4] Andrew Herdrich, Edwin Verplanke, Priya Autee, Ramesh Illikkal, Chris Gianos, Ronak Singhal, and Ravi Iyer. Cache qos: From concept to reality in the intel® xeon® processor e5-2600 v3 product family, In Proc. of *IEEE HPCA*, 2016

Impact of LLC allocations on Memory Bandwidth (MB)



Controlling LLC only (using CAT)

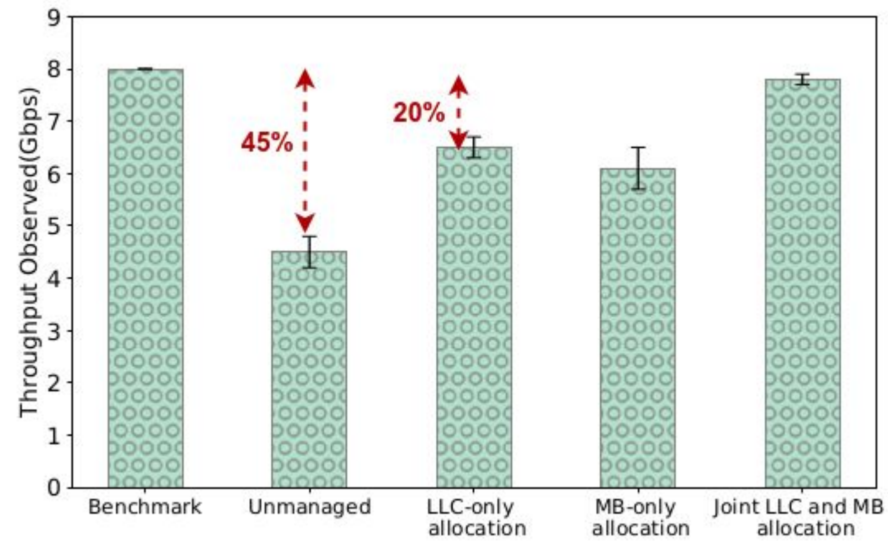


Controlling LLC as well as memory bandwidth

- Intel RDT resource partitioning technologies:
  - LLC partitioning: CAT
  - MB partitioning: Memory Bandwidth Allocation (MBA)

Motivation (Importance of MB along with LLC)

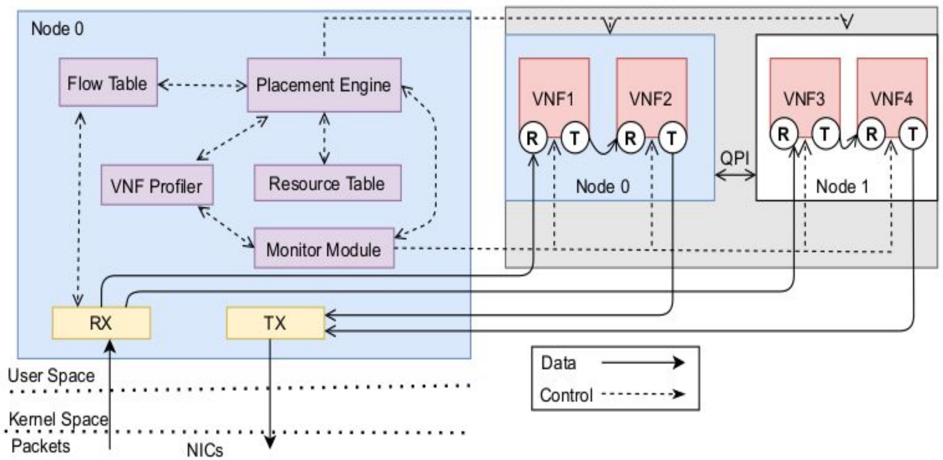
- Running *Firewall* VNF and *stress-ng* (noisy neighbour) together
- Dedicated core is pinned to VNF
- When resources are not allocated to VNF (*Unmanaged*), performance drops by *45%* due to resource contentions compared to when VNF runs alone (*Benchmark*)
- Allocating *LLC only* does not always result in performance isolation due to bandwidth saturation and results in *20%* reduction in throughput



Throughput observed for Firewall VNF for various experimental scenarios (VNF is running in Node 0)

Performance of a VNF is highly dependent on both LLC and MB allocations

NUMA-aware Dynamic SFC Placement (NUMASFP)

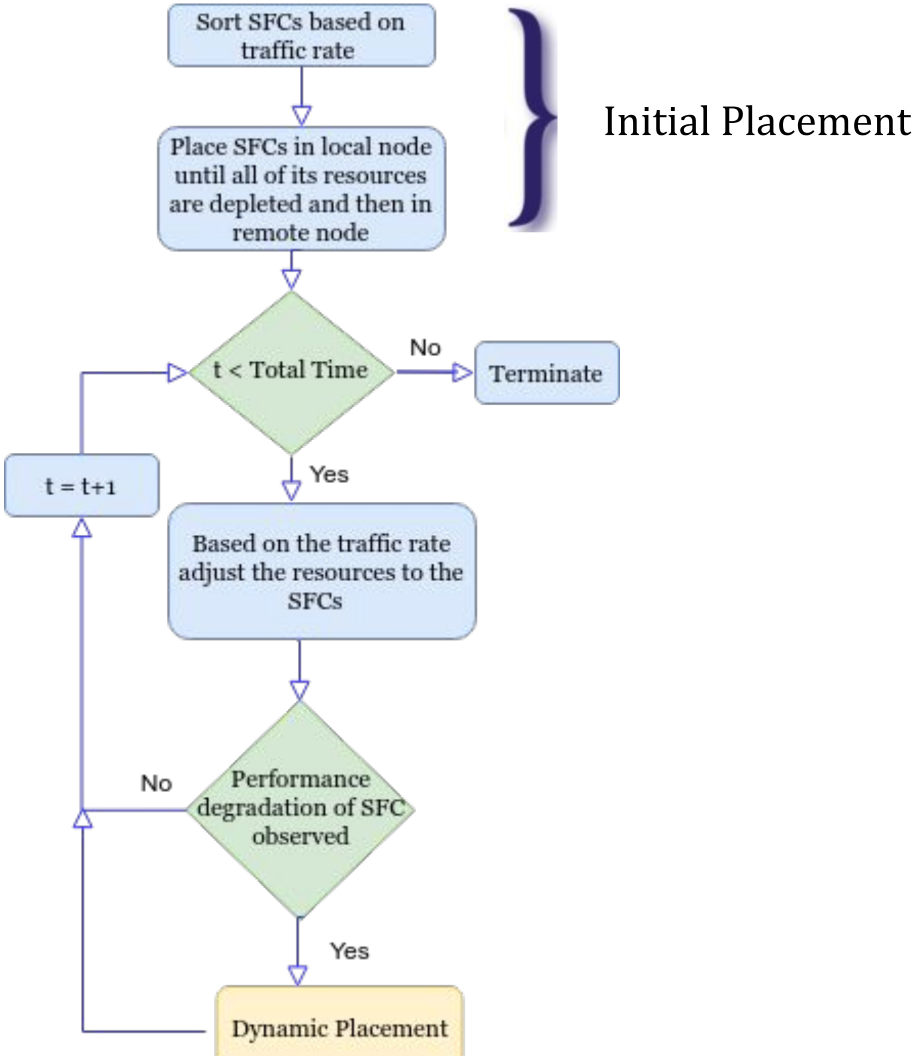


NUMASFP Architecture

- Considers the impact of NUMA and traffic rate and figure out which SFC need to be migrated in order to improve aggregate throughput of all running SFCs
- Built on OpenNetVM[1], a high performance NFV platform
- *Placement engine* allocates resources and migrates SFCs among NUMA nodes
- *Resource table* maintains the allocated resources information (in terms of cores, number of LLC-ways, and percentage of MB) of each SFC
- *VNF Profiler* generates a lookup table which maps resource requirement of each VNF based on traffic rate

[1] Wei Zhang, Guyue Liu, Wenhui Zhang, Neel Shah, Phillip Lopreiato, Gregoire Todeschi, KK Ramakrishnan, and Timothy Wood. Opennetvm: A platform for high performance network service chains, In Proc. of *ACM HotMiddlebox*, 2016

NUMASFP Procedure



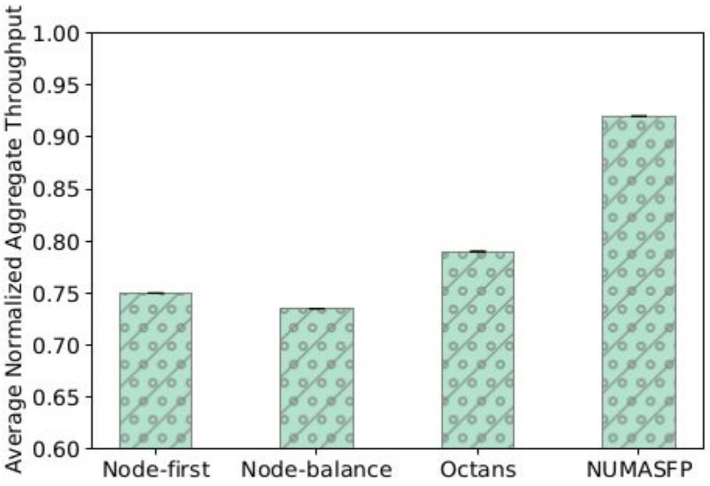
Performance Evaluation

- Alternative placement mechanisms
  - ***Node-balancing:*** Places SFCs by dividing them into all nodes to balance each node's core utilization
  - ***Node-first:*** Place SFCs on the local node first until all resources are consumed and then places the SFCs on the remote node
  - ***Octans[1]:*** Place SFCs with the high traffic rates in the local node until all of its resources are depleted
- All of these approaches assume that the placement of SFCs is static

[1] Zheng Z, Bi J, Yu H, Wang H, Sun C, Hu H, Wu J. Octans: Optimal placement of service function chains in many-core systems, in Proc. of *IEEE TPDS,* 2021

Simulation Results

- Considered 20 different SFCs of length 3 each and VNFs are randomly picked from profiled VNFs[1]
- Randomly select five SFCs and place them in each server and each SFC receives traffic for 120 Secs
- ***Normalized Aggregate Throughput (NAT):*** Ratio of aggregate throughput achieved for all SFCs over total offered load
- ***Average Normalized Aggregate Throughput (ANAT):*** Average of NAT for all time instances

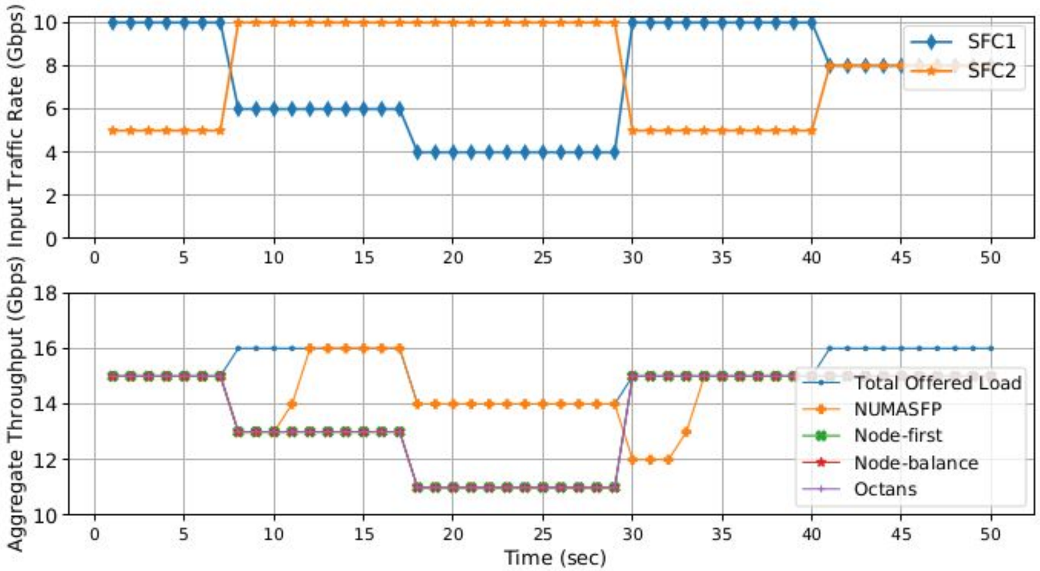


ANAT of all servers

NUMASFP achieves 16%, 25%, and 23% more ANAT than *Octans*, *Node-balance*, and *Node-first* approaches, respectively

[1] Wei Zhang, Guyue Liu, Wenhui Zhang, Neel Shah, Phillip Lopreiato, Gregoire Todeschi, KK Ramakrishnan, and Timothy Wood. Opennetvm: A platform for high performance network service chains, In Proc. of *ACM HotMiddlebox*, 2016

Testbed Results



Aggregate throughput for various methods over time

- Considered 2 SFCs of length 3 each and transmitted traffic ranging from 1 Gbps to 10 Gbps independently
- At time instance 11, NUMASFP migrates SFC1 to remote and SFC2 to local node
- NUMASFP outperformance other mechanisms most of the times

Conclusions and Future Directions

- NUMASFP places SFCs in a many-core NVF server and dynamically migrates SFCs among NUMA nodes based on their traffic rates while maximizing the aggregate throughput of all SFCs
- Built a prototype of NUMASFP on OpenNetVM
- The evaluations of NUMASFP on testbed and simulation reveal that it provides performance guarantee while significantly improving aggregate system performance
- Future work
  - ML-based solution
    - Predicting the future traffic rate and migrate accordingly to avoid ping-pong effects

THANK YOU!

email: cs17resch01007@iith.ac.in

Visit NeWS Lab @ IITH for more related research: https://newslab.iith.ac.in/

Queries