Considerations for Benchmarking Network Performance in Containerized Infrastructure

draft-dcn-bmwg-containerized-infra-08

Minh-Ngoc Tran, Jangwon Lee, Younghan Kim (Soongsil University), Kyoungjae Sun (ETRI), Huynsik Yang (KT),

Draft purpose

- Distinguish benchmarking of containerized infrastructure from previous benchmarking methodology for VM-based NFV infrastructure
- Investigate different network models that support packet acceleration technologies based on vSwitch location considering vSwitch is the network principle of containerized infrastructure.
- Investigate different deployment configurations (resource isolation, hugepages, service function chaining,...) impact on containerized networking

Updates Summary (from v3 to v8)

- Sections 3.1 & 4 are now new section 3
- Section 3.2 is now part of new section 4
- Section 3.3 is now part of section 5
- New Appendix Sections shows our own benchmarking experiences through multiple Hackathons.

	<u>1</u> . Introduction	2
	2. Terminology	
	 Benchmarking Considerations Benchmarking Considerations 	
-	<u>3.1</u> . Comparison with the VM-based Infrastructure	
v3	<u>3.2</u> . Container Networking Classification	
	<u>3.3</u> . Resource Considerations	_
	4. Benchmarking Scenarios for the Containerized Infrastructure . 1	
	<u>5</u> . Additional Considerations	
	<u>6</u> . Security Considerations	<u>14</u>

			roduct																							
	<u>2</u> . 1	Term	ninol	ogy					•		•	•	•			•	•	•	•		•	•	•	•	•	4
	<u>3</u> . (Cont	:aine	rize	d In	fra	str	uct	une	e 0	ve	rv	ie	W												4
	<u>4</u> . N	Netw	orki	ng Ma	odel	s i	n C	ont	air	ner	iz	ed	I	nf	'na	st	nu	ct	ur	e						8
	4.1	1.	Kerne	el-s	pace	vS	wit	ch	Mod	del																9
	4.2	<u>2</u> .	User	-spa	ce v	Swi	tch	Мо	de]	1																<u>10</u>
	4.3	<u>3</u> .	eBPF	Acce	eler	ati	on	Mod	lel																	10
	4.4	4.	Smart	t-NI(C Ac	cel	era	tic	n I	1od	el															<u>12</u>
	4.5	<u>5</u> .	Mode:	l Cor	mbin	ati	on		•				•													<u>13</u>
	<u>5</u> . F	Perf	formai	nce 🛛	Impa	cts			•																	<u>14</u>
/8	5.1	<u>1</u> .	CPU :	Isola	atio	n /	NU	MA	Aft	fin	it	У	•													<u>14</u>
	5.2	2.	Huge	page	s.																					15
	5.3	<u>3</u> .	Serv:	ice H	Func	tio	n C	hai	inir	ng																15
	5.4	<u>4</u> .	Addi	tiona	al C	ons	ide	rat	ior	ns																<u>16</u>
	<u>6</u> . 9	Secu	urity	Cons	side	rat	ion	s.																		<u>16</u>
	<u>Z</u> . F	Refe	erence	es																						<u>16</u>
	7.1	1.	Info	rmat:	ive	Ref	ere	nce	es :																	<u>16</u>
	Apper	ndix	<u>A</u> .	Ben	chma	rki	ng	Exp	eri	ien	ce	(C	on	ti	v-	VP	P)									<u>18</u>
			Benc																							
			Troul																							
	Apper																									
			Benc																							
			Troul																							
	Apper																									
		-																								

Detailed Updates (1)

New Section 3

Containerized Infrastructure Overview

- Comparison with VM-based Infrastructure (old 3.1)
 - $\,\circ\,$ The lack of hypervisior
- Classifies different containerized deployment methods (new)
- Based on that, 4 Benchmarking scenarios for the Containerized Infrastructure (old 4)
 - \circ Container2Container
 - BMP2BMP
 - BMP2VMP
 - \circ VMP2VMP

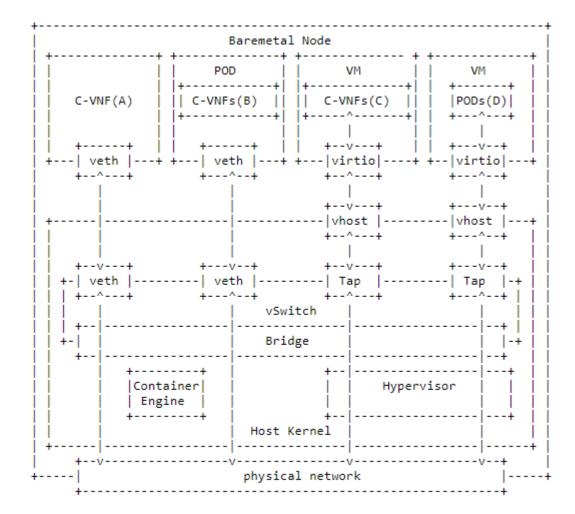


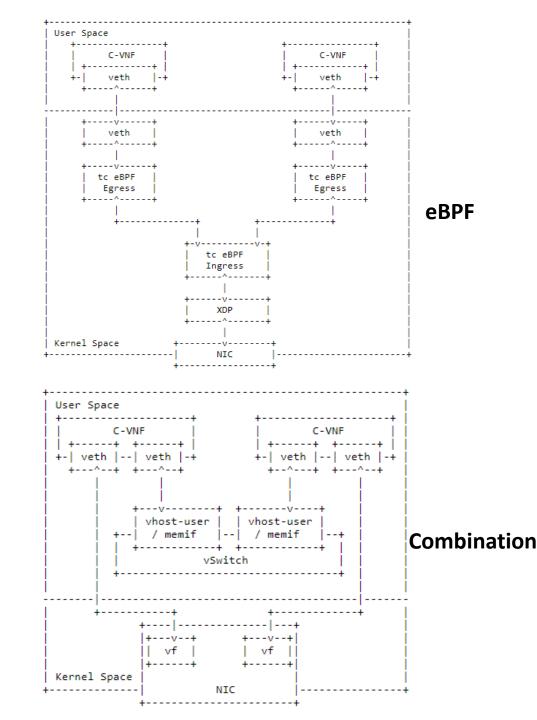
Figure 1: Examples of Networking Architecture based on Deployment Models - (A)C-VNF on Baremetal (B)Pod on Baremetal(BMP) (C)C-VNF on VM (D)Pod on VM(VMP)

Detailed Updates (2)

New Section 4

Networking Models in Containerized Infrastructure

- Container Networking Classification (old 3.2)
 - Kernel-space vSwitch model
 - \circ User-space vSwitch model
 - SR-IOV (Rename to Smart-NIC Acceleration Model)
- Adding 2 new categories
 - \circ eBPF Acceleration Model
 - o Notable used in Cillium, Calico CNI Plugin
 - $\circ\,$ Model Combination
 - $\circ~$ Notable used in Service Function Chaining
 - User-space vSwitch for East-West traffic
 - SR-IOV for North-South traffic



Detailed Updates (3)

New Section 5

Performance Impacts

- Different resource considerations (old 3.3)
 - Hugepages
 - NUMA & CPU Isolation

Adding 2 new impacts

- Service Function Chaining (new 5.3)
 - o In NFV environment, physical network port is commonly connected to multiple VNFs rather than single VNF
 - $\circ~$ Aspects needed to be considered when benchmarking service function changing
 - $\circ~$ Number of VNFs
 - Different network acceleration technologies (which provide VNF to VNF networking)
- Inter-node networking (as new additional consideration 5.4)
 - As defined in ETSI-NFV-IFA-038, different inter-node networking technologies may affect container network performance between nodes
 - Tunnel end point (VXLAN), Border Gateway Protocol (BGP), Layer 2 underlay, direct using dedicated NIC, load balancer.

Detailed Updates (4)

New Appendix Section

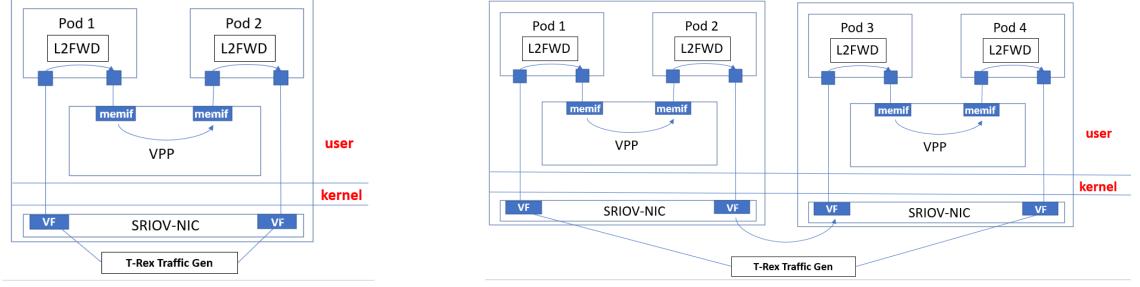
Our Hackathon benchmarking Collection

- Different network models benchmarking
 - VPP
 - SR-IOV
- Different performance impacts benchmarking
 - NUMA & CPU Isolation (included in Appendix A,B,C)
 - Service function chaining (Appendix C)

Appendix A. Benchmarking Experience(Contiv-VPP)			
A.1. Benchmarking Environment			<u>18</u>
A.2. Trouble shooting and Result			
Appendix B. Benchmarking Experience(SR-IOV with DPDK)			23
<u>B.1</u> . Benchmarking Environment			24
<u>B.2</u> . Trouble shooting and Results			27
Appendix C. Benchmarking Experience(Multi-pod Test) .			
<u>C.1</u> . Benchmarking Overview			27
C.2. Hardware Configurations			<u>28</u>
C.3. NUMA Allocation Scenario			30
C.4. Traffic Generator Configurations			
C.5. Benchmark Results and Trouble-shootings			30

- Service Function Chaining Benchmarking
 - Measure throughput when using SR-IOV and VPP combination

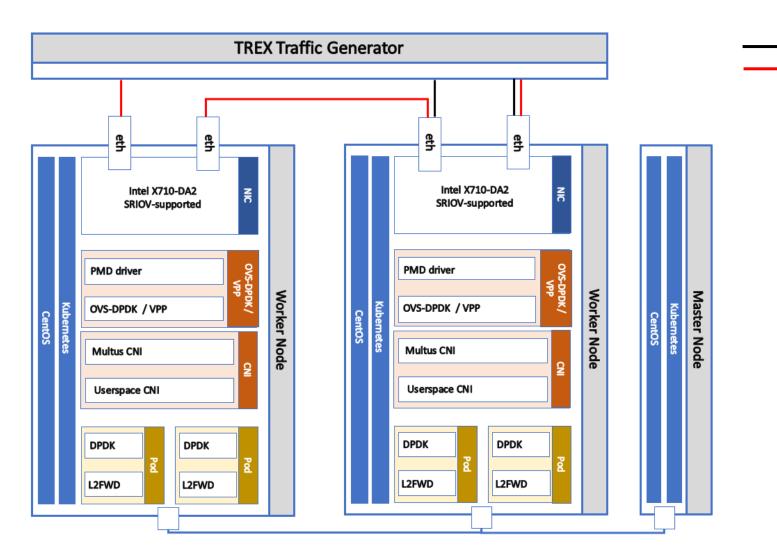
- Scenarios
 - Service Function chaining in single node
 - Service Function changing in multi-nodes (using L2 underlay as inter-node networking technique)
 - Test number of VNFs impacts (2,4,6 pods)



Single node scenario

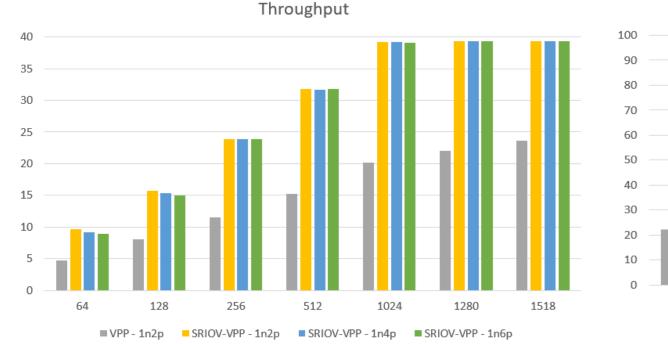
Multi nodes scenario

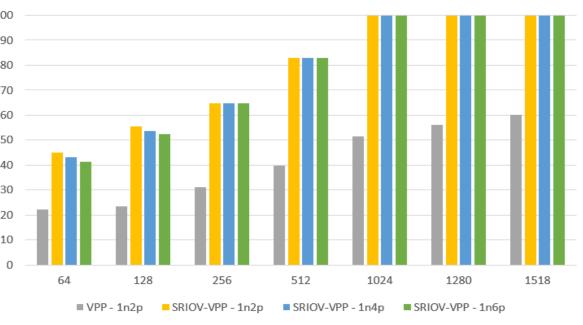




Single node scenario
 Multi nodes scenario

- Benchmarking Performance Results Single node SRIOV-VPP service chain
 - 1. SRIOV-VPP performed significantly better than VPP only (packets through VPP need go through vSwitch, no need with SRIOV)
 - 2. Increase number of pod slightly reduce throughput 2% at small packet size (64,128)

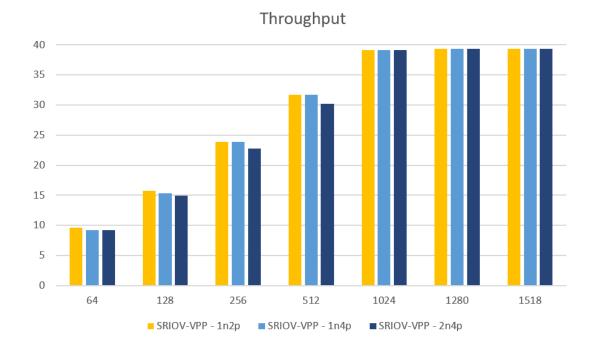


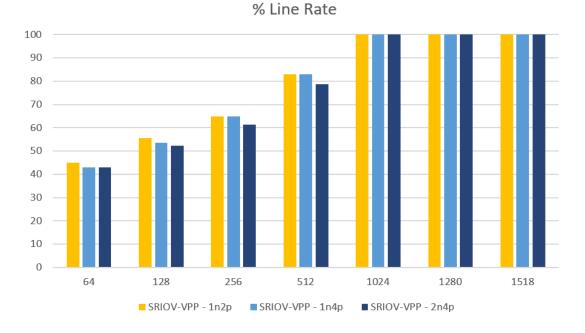


% Line Rate

• Benchmarking Performance Results – Multi-nodes SRIOV-VPP service chain

1. Throughput in multi-nodes scenario in slightly smaller than single-node with smaller packet size (<512) due to increasing in number of pod (4 pods total in multi-nodes > 2 pods in single node)





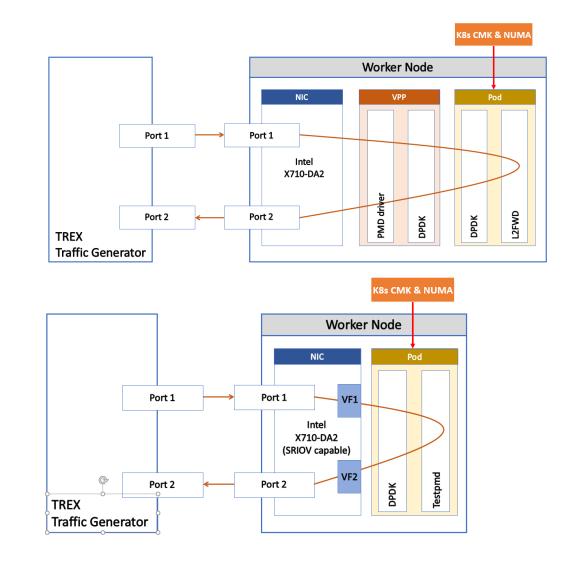
Next Steps

- Keep updating the drafts based on latest technologies
- Any comments or feedbacks are welcome
- IETF BMWG Hackathon
 - Test inter-node networking technique impacts on container network performance
 - Test performance of eBPF acceleration model (Cilium/Calico) with/without NIC offloading
 - Proof our draft scenarios and features
 - Sharing results to the BMWG

Backup Slides

Benchmarking Experiences (Contiv-VPP + SRIOV)

- Test performance of user-space based model and SmartNIC (VPP and SRIOV)
- Figure out impact of CPU isolation (using CMK – CPU Manager for Kubernetes) and NUMA to network performance
 - Without CMK
 - CMK-shared mode (2 pods share 2 CPUs)
 - CMK-exclusive mode (1 dedicated CPU/pod)



Benchmarking Experiences (Contiv-VPP + SRIOV)

What we learned

• VPP and SRIOV has nearly the same performance

CPU Isolation:

- CPU Isolation (CMK) significantly improves throughput
- Exclusive mode is better than Shared mode

NUMA alignment:

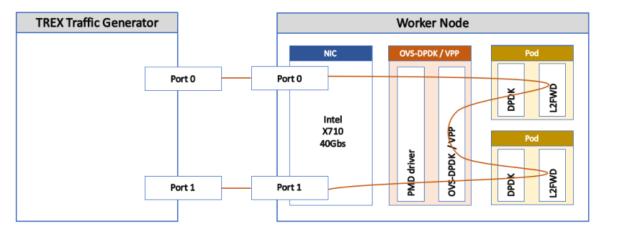
• Assigning CPU in the same NUMA node is better than in different NUMA nodes

Model	NUMA Mode (pinning)	Result(Gbps)
Maximum Line Rate	N/A	3.1
Maximum bine Rate	same NUMA	9.8
Without CMK	N/A	1.5
	same NUMA	4.7
		3.1
	same NUMA	3.5
CMK-shared Mode -	Different NUMA	2.3

CPU Isolation and NUMA location impact in VPP test with 10G Intel X710-DA2 NIC

Benchmarking Experiences (Multi-pods)

- Test performance of VPP in service function chain scenario (2 pods)
- Figure out impact of NUMA allocation over CNF, vSwitch, NIC
 - 6 scenarios
 - vSwitch same with NIC
 - vSwitch same with input CNF and vice versa
 - vSwitch different with NIC
 - vSwitch same with input CNF and vice versa



different

with NIC

NUMA node

—	NUM			~							—	NUN			0		—						
<u> </u>				-	-	NUMA node 1					L-i			_	-		· ·		IA n		3 24 8 29 3 34 8 39		
0	1	2	3	4	-				23	24	0	1	2	3	4		_	itch		23			
5	6	7	8	9	_		NF1	27	28	29	5	6	7		F 2		CN		27	28			
10	11	12	13	14	4	C	NF 2	32	33	34	10	11	12	13	14		30	31	32	33	34		
15	16	17	18	19	9	35	36	37	38	39	15	16	17	18	19		35	36	37	38	39		
	м	emo	ry				Ν	lemo	ry			м	emo	ory			Memory						
	Syst	tem (Bus				Sys	tem	Bus			Sys	tem	Bus			System Bus						
Sce	ena	rio	1 (s1	L)			NIC			Sc	ena	ario	5 2	(s2))			NIC	:			
											_					,	_						
												NUN	/A n	ode	0	1		NUM	IA no	ode 1	1		
V	sw	ito	ch								0	NUN 1	/An 2	ode 3	0		-	NUM itch		23	24		
•			ch								-	-		3	-		-	itch	_	_	_		
•	sw am		ch								0	1	2	3	4		vSw	itch	22	23	24		
sa		e		nc	bd	e					0	1	2 7 12	3 CN	4 F 1		vSw CN	<mark>itch</mark> F 2	22 27	23 28	24 29		
sa N	am	ne MA	٩r		bd	e					0 5 10	1 6 11 16	2 7 12	3 CN 13 18	4 F 1 14		VSW CN 30	F 2 31 36	22 27 32	23 28 33 38	24 29 34		
sa N	am UN	ne MA	٩r		bd	e					0 5 10	1 6 11 16	2 7 12 17	3 CN 13 18 0ry	4 F 1 14		VSW CN 30	itch F 2 31 36 M	22 27 32 37	23 28 33 38	24 29 34		
sa N	am UN	ne MA	٩r		bd	e					0 5 10	1 6 11 16 N Sys	2 7 12 17 17	3 CN 13 18 0ry Bus	4 F 1 14 19		VSW CN 30	itch F 2 31 36 M Sys	22 27 32 37 emo	23 28 33 38 ry Bus	24 29 34		

NUMA node 0 NUMA node 1										NUMA node 0						NUMA node 1					
vSwitch		3	4	20	21	22	23	24	vSv	vitch	2	3	4		20	21	22	23	24		
CNF1	7	8	9	25	26	27	28	29	CN	F1	7	8	9		CN	IF 2	27	28	2		
CNF 2	12	13	14	30	31	32	33	34	10	11	12	13	14		30	31	32	33	3		
15 16	17	18	19	35	36	37	38	39	15	16	17	18	19		35	36	37	38	3		
N	lemo	ory			N	1emo	ory					М	emo	ory							
Sys	tem	Bus			Sy	stem	Bus			Sys	tem	Bus				Sys	tem	Bus			
Scena	Ē		NIC	:		Sc	ena	rio	5 (s	:5)				NIC							
										NUN	1A n	ode	0	1			A no	do 1			
Vsw	itc	h								ital		2	Ŭ.		NUMA node 1						

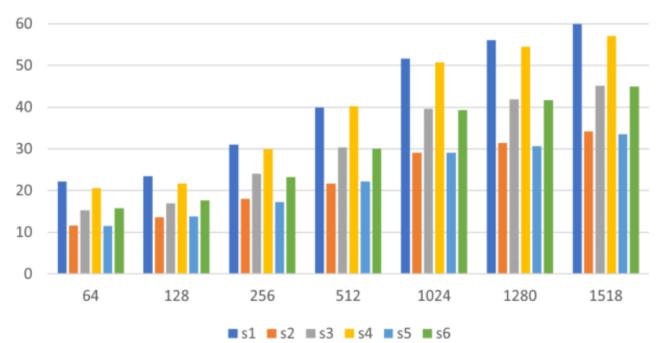
	NUM	IA no	ode ()	NUMA node 1						
vSv	vitch	2	3	4	20	21	22	23	24		
CN	F 2	7	7 8		CNF 1		27	28	29		
10	11	12	13	14	30	31	32	33	34		
15	16	17	18	19	35	36	37	38	39		
	М	emo	iry		Memory						
	Sys	tem	Bus		System Bus						
So	en	ario	o6 (s6)			NIC				

Benchmarking Experiences (Multi-pods)

What we learned

NUMA alignment:

- vSwitch and NIC in different nodes slightly degrade performance in 1024+ packet size
- **CNFs and vSwitch** in different nodes degrade performance by 10-15%
- Input CNF and vSwitch in different node has better performance



VPP (%line rate)