

VM Analysis

Execution Path Analysis of Virtualized Environments using Host Kernel Tracing

Hani Nemati

Dec 7, 2017

Polytechnique Montréal Laboratoire **DORSAL**

Agenda

Introduction

• Research update and research motivation

New Investigations

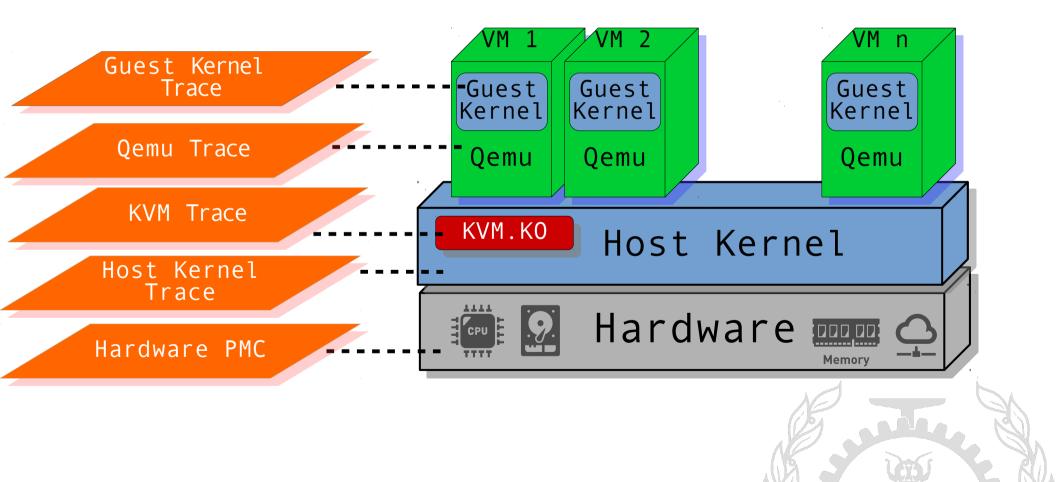
- Execution Path Analysis of virtualized environments using
- host kernel tracing
 - Sate of the art
 - Proposed Algorithm
 - Demo

Conclusion and in-progress



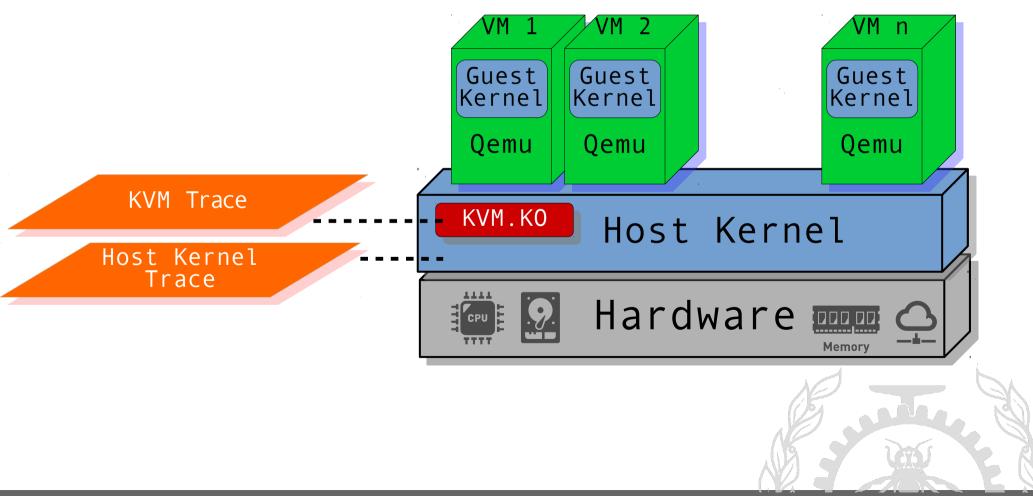
Previously on "VM Analysis"

Available Trace-Points in different layers



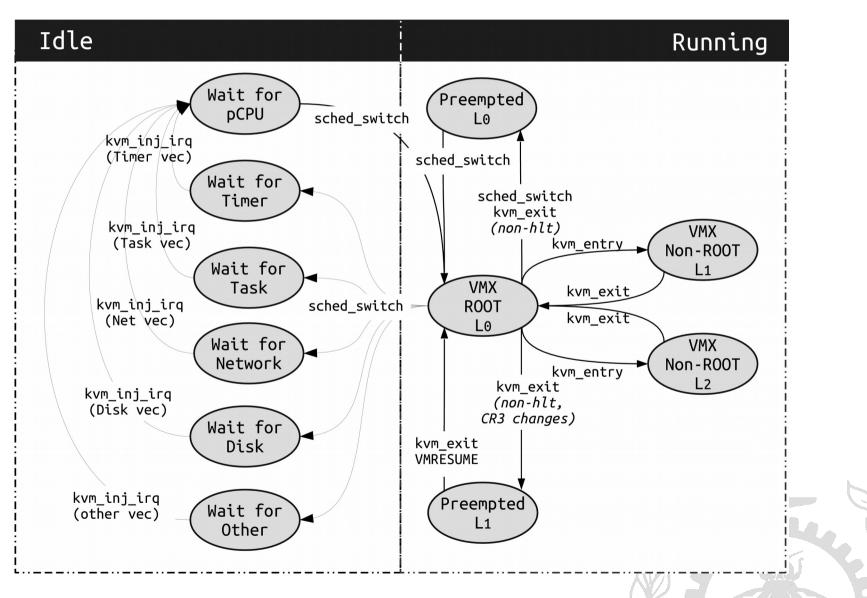
Previously on "VM Analysis"

Used Trace-Points in our approaches



Previously on "VM Analysis"

States for a Process inside VM



Motivation

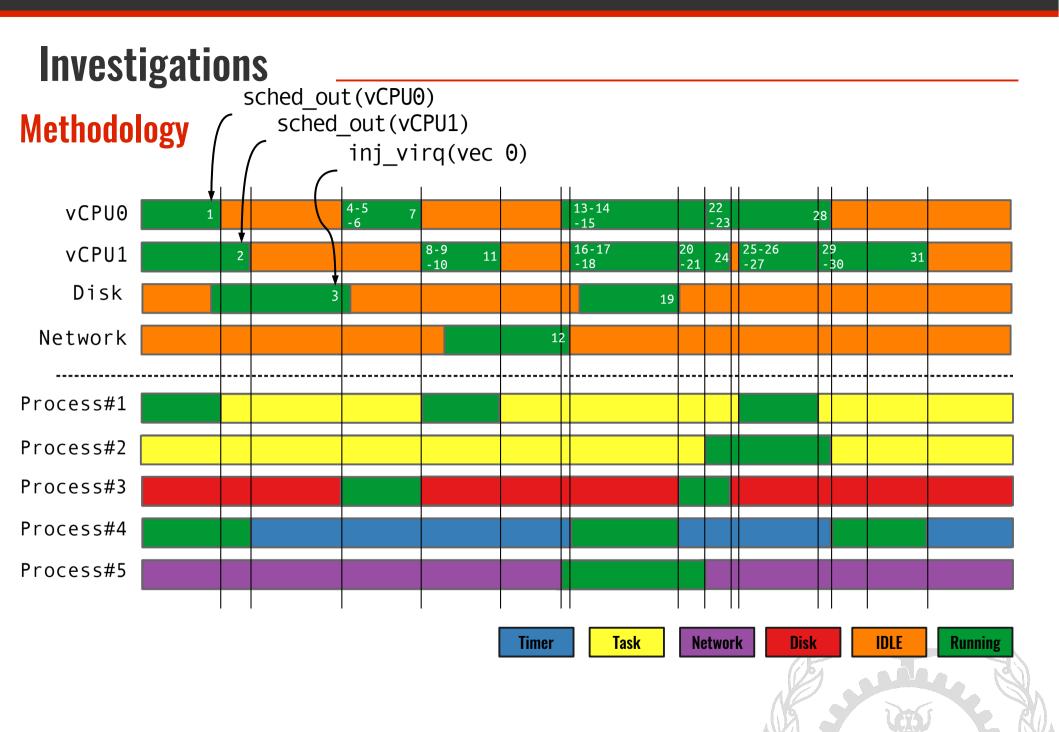
Why is the VM waiting?

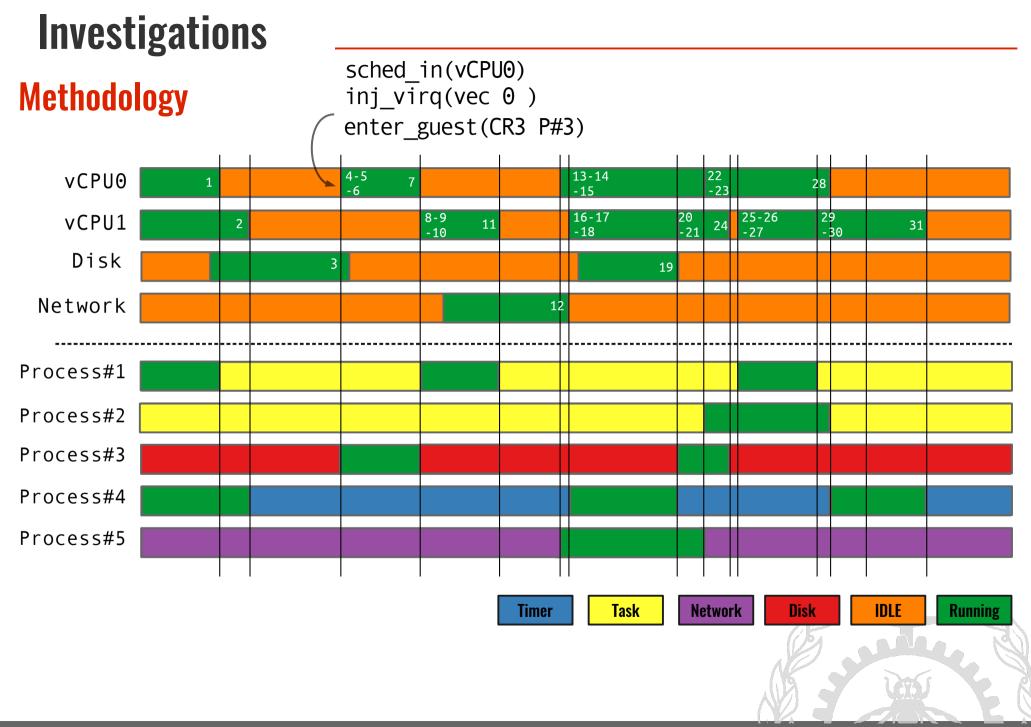
🖩 Resources 🖾	🗄 exp		ø	= 🙆 🗞 🖏	· 🕂 🤑 🔍 🔍 🗖	' 🗆
	2017 May 02	14:51:30.960	14:51:30.980	14:51:31.000	14:51:31.020	
⊞ Hypervisor Re □ VM	source View					
□ external1/te	stU1					
□ vCPU				i		
CPU 1		•				
vProcess						
19940597						
207841689 207775744		· · · ·			- I I I I I I I I I I I I I I I I I I I	
20564541						
20238704						
88752947	2					
		er 🖶 Qemu View 🛱 🔫	Progress ➡ Contr Ø ♣ ☷		ss View 🗖	' 🗆 • 💠
		er ➡ Qemu View ☎ ➡ 14:51:30.960				
🖻 Tasks E Sta	te System Explor		🧳 🆆 🖽	10 B B O	.0 € € 1	
 Tasks I Sta Process qemu-system qemu-system 	te System Explore TID n-x86 65 em-x86 65		🧳 🆆 🖽	10 B B O	.0 € € 1	
 Tasks I≡ Sta Process qemu-system qemu-system qemu-system 	te System Explore TID m-x86 65 em-x86 65 em-x86 65		🧳 🆆 🖽	10 B B O	.0 € € 1	
 Tasks I Sta Process qemu-systen qemu-systen qemu-systen qemu-systen 	te System Explore TID m-x86 65 em-x86 65 em-x86 65 em-x86 65		🧳 🆆 🖽	10 B B O	.0 € € 1	
Tasks E Sta Process qemu-system qemu-system qemu-system qemu-system qemu-system qemu-system	te System Explore TID m-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 65		🧳 🆆 🖽	10 B B O	.0 € € 1	
 Tasks I Sta Process qemu-syster qemu-syster qemu-syster qemu-syster qemu-syster qemu-syster 	te System Explore TID m-x86 6! em-x86 6! em-x86 6! em-x86 6! em-x86 6! em-x86 6!		🧳 🆆 🖽	10 B B O	.0 € € 1	
 Tasks I Sta Process qemu-syster qemu-syster qemu-syster qemu-syster qemu-syster 	te System Explore m-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 76		🧳 🆆 🖽	10 B B O	.0 € € 1	
 Tasks I Sta Process qemu-syster qemu-syster qemu-syster qemu-syster qemu-syster qemu-syster qemu-syster 	te System Explore m-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 65 em-x86 76		🧳 🆆 🖽	10 B B O	.0 € € 1	

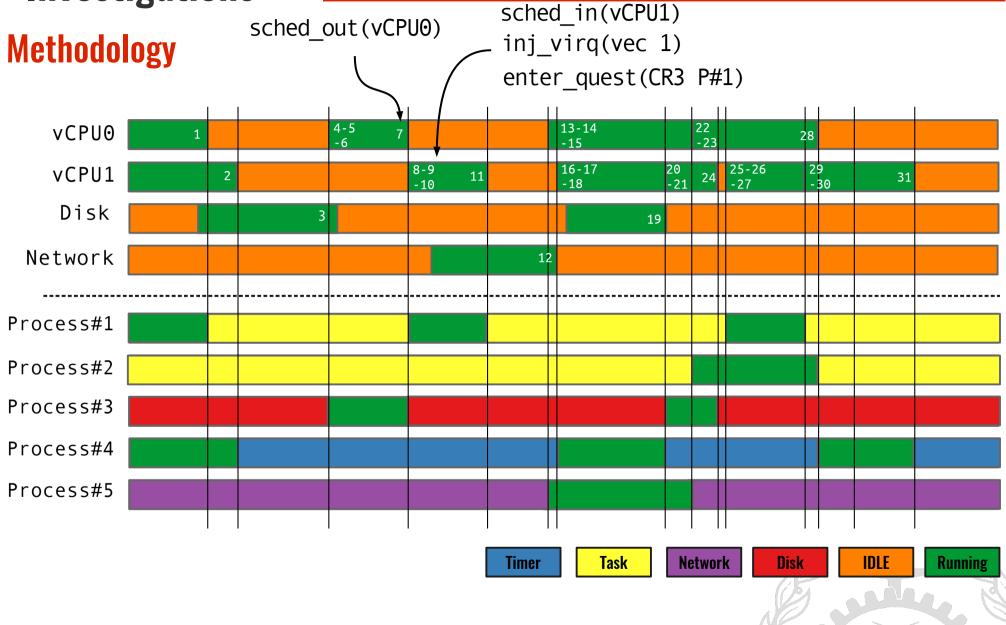
Motivation

Let's use the Critical Flow view of Trace Compass?

Process TID • kernel • systemd 1 • qemu-system-x86 6503 qemu-system-x86 6504 qemu-system-x86 6505 qemu-system-x86 6508 qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663 qemu-system-x86 7762	3 1 4 6503 5 6503 8 6503 9 6503	Birth time 14:51:30.028394188 14:51:30.028818415 14:51:30.028819071 14:51:30.028819571 14:51:30.028820255		14:51:31.420	š š ⊽ ∲ ∲ € 14:51:31.440	≷	(\$ \$)	▼
▼ kernel I ▼ systemd 1 ▼ qemu-system-x86 6503 qemu-system-x86 6504 qemu-system-x86 6505 qemu-system-x86 6508 qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663	3 1 4 6503 5 6503 8 6503 9 6503	14:51:30.028394188 14:51:30.028818415 14:51:30.028819071 14:51:30.028819571 14:51:30.028820255		14:51:31.420	14:51:31.440	14:51:31.460		_^
▼ systemd 1 ▼ qemu-system-x86 6503 qemu-system-x86 6505 qemu-system-x86 6508 qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663	4 6503 6503 8 6503 9 6503	14:51:30.028818415 14:51:30.028819071 14:51:30.028819571 14:51:30.028820255		Ï				
▼ qemu-system-x86 6503 qemu-system-x86 6504 qemu-system-x86 6505 qemu-system-x86 6508 qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663	4 6503 6503 8 6503 9 6503	14:51:30.028818415 14:51:30.028819071 14:51:30.028819571 14:51:30.028820255		i				
qemu-system-x86 6504 qemu-system-x86 6505 qemu-system-x86 6508 qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663	4 6503 6503 8 6503 9 6503	14:51:30.028819071 14:51:30.028819571 14:51:30.028820255						
qemu-system-x86 6505 qemu-system-x86 6508 qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663	5 6503 8 6503 9 6503	14:51:30.028819571 14:51:30.028820255						
qemu-system-x86 6508 qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663	8 6503 9 6503	14:51:30.028820255						
qemu-system-x86 6509 qemu-system-x86 6511 qemu-system-x86 7663	9 6503							
qemu-system-x86 6511 qemu-system-x86 7663							ii i	
qemu-system-x86 7663	1 6503	14:51:30.028820839						
		14:51:30.028821464						
qemu-system-x86 7762	3 6503	14:51:30.028822038						
	2 6503	14:51:30.028822602						
								Ē
)•)
🗄 kernel 🛛 🛱 Critical Flo	low Vie	w ¤		🥖 💷 🖹 🟠 🛡	5 😎 🖡 6 6 .	- 6 0 0	▽ □	
Process		Elapsed	Perce	14:51:31.420	14:51:31.440	14:51:31.460		Ê
▼ "5e0d50ad-28dd-4eb2-b3f0-6	-6cb3cf29	95a62" 2.360704919	64.26					
[qemu-system-x86,6508]		2.360704919	64.26	Ā			i A	
[qemu-system-x86,6503]		1.188970541	32.36				1	_
[qemu-system-x86,7762]		0.007940134	0.22	1				
[qemu-system-x86,6509]		0.088618179	2.41	<u> </u>			1	
[sshd,7760]		0.003013658	0.08	1				
[vhost-6503,6506]		0.017323691	0.47	Y			1	
[qemu-system-x86,7663]		0.002998406	0.08				1	
[scp,7761]		0.004311896	0.12					_
.(4(







POLYTECHNIQUE MONTREAL – Hani Nemati

Cont	rc		Flov	v View								
control Flow 🛱 🕼	Resou	rces	🗆 Statistics 😫	Critical Flow View								
									Ē	smart filter	i =	备
cess	TID	PTID	Birth time	14:19:12.180	14:19:12.200	14:19:12.220	14:19:12.240	14:19:12.260	14:19:1	12.280	14:19:	12.3
st-rpc-25-30												

Investigations

🖶 Control Flow 😫 🕯	Reso	urces	🗆 Statistics 😫	Critical Flow View									- 8
									🖆 smart filter 券	iii 👔 🦻 🤜 🌾	li l	. 🤍 🕅 🍖 🔶 🌾	• ⇒ ≝ • ▽
Process	TID	PTID	Birth time	14:19:12.180	14:19:12.200	14:19:12.220	14:19:12.240	14:19:12.260	14:19:12.280	14:19:12.300	14:19:12.320	14:19:12.340	14:19:12.360
▼ host-rpc-25-30													
▼ systemd	1		14:19:10.125187 -										
▼ qemu-system-x86	9162	1	14:19:10.125783										
qemu-system-x86	9163	9162	14:19:10.125784										
qemu-system-x86	9164	9162	14:19:10.125785										
qemu-system-x86	9167	9162	14:19:10.125785										I
qemu-system-x86	9168	9162	14:19:10.125786		1							++	I
qemu-system-x86	9170	9162	14:19:10.125787										
▼ qemu-system-x86	8917	1	14:19:10.125932										
qemu-system-x86	8929	8917	14:19:10.125933										
qemu-system-x86	8930	8917	14:19:10.125933										
qemu-system-x86	8947	8917	14:19:10.125934										I
qemu-system-x86	8953	8917	14:19:10.125935									+	
qemu-system-x86	9003	8917	14:19:10.125935										

vCPU View

🗄 vCPU v	/iew 🛱 🚟 VM Process View								🆆 🗄 🟠 🤜	№ 8 8 - 6 4 9 9	🤍 者 🔻 🖓 🗖 [
Name	ParentID	14:19:12.180	14:19:12.200	14:19:12.220	14:19:12.240	14:19:12.260	14:19:12.280	14:19:12.300	14:19:12.320	14:19:12.340	14:19:12.360
▼ rpc-25-3											
▼ 8917											
1				-							-
▼ 9162											
0					1					()	
					·						
									1		
											121
									115		
									(12)		(12)



8





Control Flow View





🕏 Control Flow 🛿 🕯	Reso	urces	🗆 Statistics 😫	Critical Flow View									
									🖆 smart filter 🇳	표 💧 🦁 🧕 🌾	k k - A € €	(Q 🕅 ሎ 🔶 🍋	⇒ ▼ ▽
Process	TID	PTID	Birth time	14:19:12.180	14:19:12.200	14:19:12.220	14:19:12.240	14:19:12.260	14:19:12.280	14:19:12.300	14:19:12.320	14:19:12.340	14:19:12.360
▼ host-rpc-25-30													
▼ systemd	1		14:19:10.125187										
▼ qemu-system-x86	9162	1	14:19:10.125783										
qemu-system-x86	9163	9162	14:19:10.125784										
qemu-system-x86	9164	9162	14:19:10.125785										
qemu-system-x86	9167	9162	14:19:10.125785		1								
qemu-system-x86	9168	9162	14:19:10.125786		1					(
qemu-system-x86	9170	9162	14:19:10.125787										
▼ qemu-system-x86	8917	1	14:19:10.125932										
qemu-system-x86	8929	8917	14:19:10.125933										
qemu-system-x86	8930	8917	14:19:10.125933										
qemu-system-x86	8947	8917	14:19:10.125934 -										
qemu-system-x86	8953	8917	14:19:10.125935		Ì				i i i i i i i i i i i i i i i i i i i		1	i-	
qemu-system-x86	9003	8917	14:19:10.125935										
													•

vProcess View

Task	Timer	Network	Running
------	-------	---------	---------

🕻 vCPU View 🗮 V	M Process	s View 🛙														⇒ =	🖞 🧟 😫	P. B. B	∂ ♣ € €	₫ - ▽	-
lame	TI I	12.180 1	4:19:12.190	14:19:12.200	14:19:12.210	14:19:12.220	14:19:12.230	14:19:12.240	14:19:12.250	14:19:12.260	14:19:12.270	14:19:12.280	14:19:12.290	14:19:12.300	14:19:12.310	14:19:12.320	14:19:12.330	14:19:12.340	14:19:12.350	14:19:12.360	
rpc-25-30-delay	- IÏ																				
▼ 8917																					
142794752																					
155238400																					+
156831744			1				-			1							1				+
159281152																	-				+
1690939392																	-				+
1820450816																	-				+
1896505344																	-				+
1898397696																					
2071756800																					
392470528																					+
443297792																					+
443887616			1														1				+
9162																					
1150103552																	-				+
1191071744																					+
1247629312																	-				+
1956872192																					-
1958658048																					+
1964007424			-													1				-	+
1971576832																					
743825408			-																		+
831987712																					+-
832233472																					+
864694272																					
															/	11					Ŵ

Investigations Execution Flow Analysis View

Running Block

- -

B Control Flow	E Resources	Statistics	🔄 Critical Flow View

								🖆 🛛 smart filter	🌦 🗉 🟠 🌄 🧏	% % - ∲ ∲ € €	▼ ▼ 🖺 🌾 📣 な ▼
Process	TID	PTID Birth	h time	16:07:58.000	16:07:58.500	16:07:59.000	16:07:59.500	16:08:00.000	16:08:00.500	16:08:01.000	16:08:01.500
▼ host-slow-disk											
▼ systemd	1	16:07	7:52.112301								
▼ qemu-system-x86	24500	1 16:07	7:52.113522 🛉	···· · · · · · · · · · · · · · · · · ·							
qemu-system-x86	24503	24500 16:07	7:52.113523								
qemu-system-x86	24506	24500 16:07	7:52.113523				· ·	·			
qemu-system-x86	24538	24500 16:07	7:52.113525 🗧							· · · · · · · · · · · · · · · · · · ·	
qemu-system-x86	24539	24500 16:07	7:52.113526								
qemu-system-x86	24541	24500 16:07	7:52.113527			and the second	and the second second	an aga a a an			
qemu-system-x86	24632	24500 16:07	7:52.113528 =								
qemu-system-x86	24633	24500 16:07	7:52.113529 =								
qemu-system-x86	24634	24500 16:07	7:52.113530 =								
qemu-system-x86	24635	24500 16:07	7:52.113531 =								
qemu-system-x86	24636	24500 16:07	7:52.113532 =								
qemu-system-x86	24637	24500 16:07	7:52.113533 =								
qemu-system-x86	24638	24500 16:07	7:52.113534 =						· · · · · · · ·		
qemu-system-x86	24639	24500 16:07	7:52.113535 =								
qemu-system-x86	24640	24500 16:07	7:52.113536 =								
qemu-system-x86	24641	24500 16:07	7:52.113537 =								
qemu-system-x86	24642	24500 16:07	7:52.113537 🗧								
qemu-system-x86	24643	24500 16:07	7:52.113539 =								
qemu-system-x86	24644	24500 16:07	7:52.113540 =								
qemu-system-x86	24645	24500 16:07	7:52.113540 =								

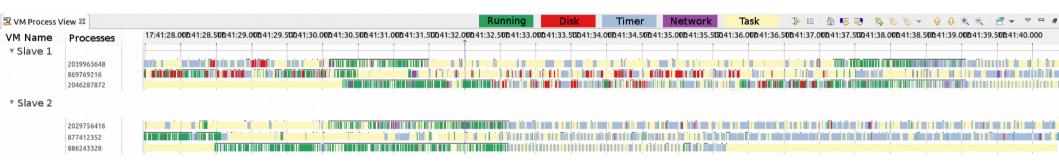


🚟 vCPU V	iew 🛱 🚟 VM Process View						+1. U) 💷 🟠 👼 🤜 🛝 🏷 🖏	- 0 0 0 0 0 1 1
Name	ParentID	16:07:58.000	16:07:58.500	16:07:59.000	16:07:59.500	16:08:00.000	16:08:00.500	16:08:01.000	16:08:01.500
▼ 24500									
▼ 24500									
0						IT.		· · · · · · · · · · · · · · · · · · ·	
1					İ	1 🖬 1			



B Control Flow	Resources	🗆 Statistics 🛱 Ci	ritical Flow View 🛙					i≡ 【) 🕏 💐 🖄 🖧 🖧 👻 🗘 -	ঢ় •、 ৹、 🔶 ৹^† 📑 ▾ ▽ 🗖 🗖
Process	Elapsed	Percent	16:07:58.000	16:07:58.500	16:07:59.000	16:07:59.500	16:08:00.000	16:08:00.500	16:08:01.000	16:08:01.500
♥ 24500 [VMkernel/26,26] [VMkernel/27,27] [VMkernel/27,27] [VMkernel/27,27] [VMkernel/27,27] [VMkernel/23,22] [VMkernel/33,33] [VMkernel/34,34] [VMkernel/34,34] [VMkernel/35,35] [VMkernel/36,36]	0.134587178 2.158543618 0.309552161 0.044711207 0.000359375 0.00031142 0.03002915 0.000171739 0.002177666 0.054021762 0.258265613 0.053352339	26.19 3.25 52.10 7.47 1.08 0.01 0.00 0.00 0.00 0.00 0.05 1.30 6.23 1.29								

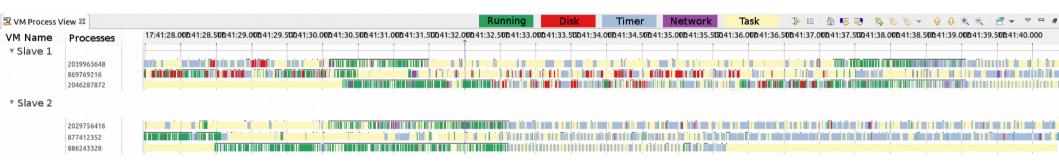
Slow Hadoop Slave



Processes	Root	non-Root	Task	Timer	Disk	Net
2039963648	0.004	1.499	54.040	43.512	0.198	0.217
869769216	0.001	0.979	39.938	56.978	1.251	0.700
2046287872	0.004	3.125	57.357	38.022	0.766	0.283
2029756416	0.002	1.898	36.508	60.883	0.098	0.340
877412352	0.001	0.970	24.947	72.767	0.029	1.130
886243328	0.005	8.350	85.240	5.588	0.003	0.258



Slow Hadoop Slave



Processes	Root	non-Root	Task	Timer	Disk	Net
2039963648	0.004	1.499	54.040	43.512	0.198	0.217
869769216	0.001	0.979	39.938	56.978	1.251	0.700
2046287872	0.004	3.125	57.357	38.022	0.766	0.283
2029756416	0.002	1.898	36.508	60.883	0.098	0.340
877412352	0.001	0.970	24.947	72.767	0.029	1.130
886243328	0.005	8.350	85.240	5.588	0.003	0.258

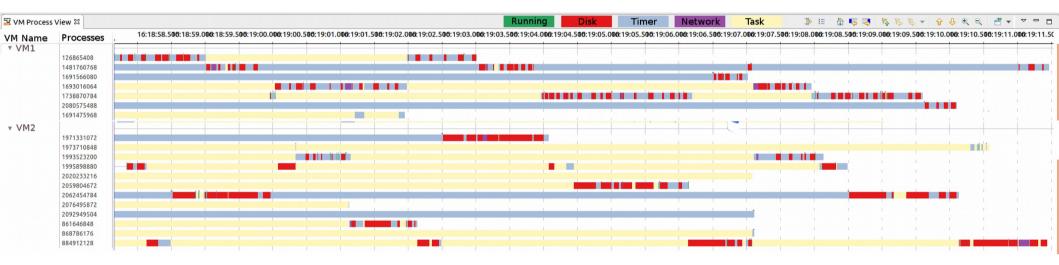
Cap on Network



Without Cap on Network

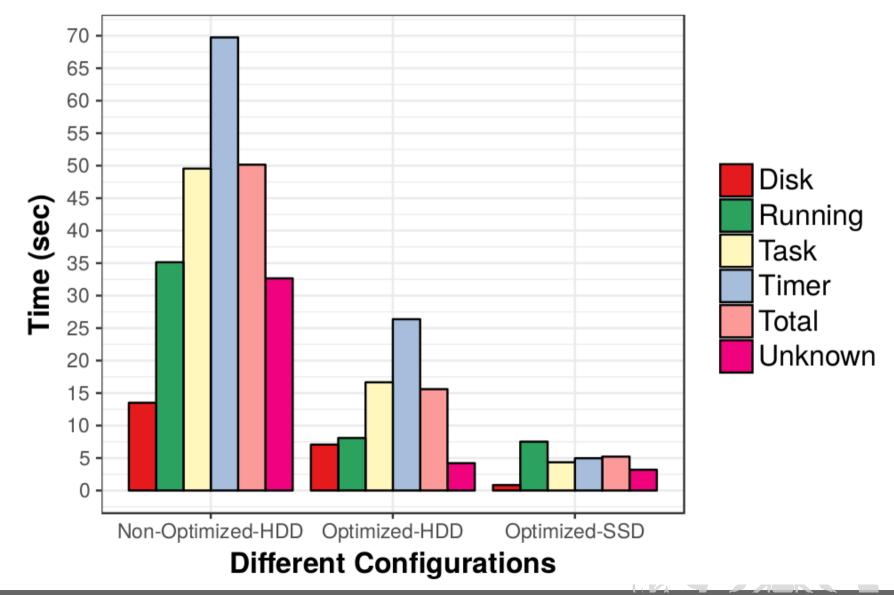
🗄 VM Process	View ន	R	unning)	Disk		Timer	N	letwork	Ta	sk		
VM Name	Processes	Pa		19:07:3	33.520	19:0	7:33.540	19:07	:33.560	19:07:3	3.580	19:07:33.	600
▼ VM1	894611456			Ī				i		Í	Ī		
▼ VM2	2035716096								Ī				
											and i		

Contention on Block Disk





VM Boot-up comparison



Demo



How to try these new features?

- Access to **Host** only
- Run **LTTng** on Host with my new added tracepoint (vcpu_enter_guest)
- Clone **TraceCompass** from github (incubator)
 - Open vCPU block View of TraceCompass (XML view)
 - Open vProcess block View of TraceCompass (XML view)
 - Use Execution Flow Analysis of TraceCompass





One More Thing ...



Let's make TraceCompass Intelligent

Demo



Conclusion and in-progress

Inferences

- Wait Analysis of process inside VM
 - A process is waiting for
 - A Disk Block request to finish
 - A Network packet to receive
 - Another process
 - A Timer to fire
 - Other devices
- Critical Path Analysis of process inside VM

In progress

• VM contention avoidance based on VM classification

[1] Hani Nemati, Michel R. Dagenais, "VM Processes State Detection by Hypervisor Tracing. "Submitted to IEEE System Conference 2018

[2] Hani Nemati, Genevieve Bastien, Michel Dagenais "Wait analysis of Virtual Machines using host kernel tracing", Accepted at IEEE Cloud Summit 2018



Questions?

Hani.nemati@polymtl.ca

https://github.com/Nemati

