



Center for Information Services and High Performance Computing (ZIH)

Interactive Trace Analysis with Vampir

ORNL – Tools Workshop 19-20 Jan 2022



Outline

Welcome to the Vampir Tool Suite

- Mission
- Event Trace Visualization
- Vampir & VampirServer
- The Vampir Displays
- What's New
- Showcases







Mission

Visualization of dynamics of complex parallel processes Requires two components

- Monitor/Collector (Score-P)
- Charts/Browser (Vampir)

Typical questions that Vampir helps to answer:

- What happens in my application execution during a given time in a given process or thread?
- How do the communication patterns of my application execute on a real system?
- Are there any imbalances in computation, I/O or memory usage and how do they affect the parallel execution of my application?







Event Trace Visualization with Vampir

- Alternative and supplement to automatic analysis
- Show dynamic run-time behavior graphically at any level of detail
- Provide statistics and performance metrics

Timeline charts

- Show application activities and communication along a time axis, which can be zoomed and scrolled
- Master timeline showing all parallel processes/threads
- Process timeline focusing on a single process/thread

Summary charts

 Provide quantitative results for the currently selected time interval (e.g., Message Summary)





Vampir – Visualization Modes (1)

Directly on front end or local machine

% vampir &







Vampir – Visualization Modes (2)

On local machine with remote VampirServer



Vampir on OLCF Resources

Please follow instructions on

— <u>https://docs.olcf.ornl.gov/software/profiling/Vampir.html</u>

— Note that a running VampirServer is protected by a password

\$ vampirserver start -- -P <project>
Launching VampirServer...
Submitting LSF batch job (this might take a while)...
Warning: more than 1 task/rank assigned to a core
VampirServer 9.11.1 OLCF (4626dba5)
Licensed to ORNL
Running 4 analysis processes... (abort with vampirserver stop 29200)
User: <user>
Password: <password>
VampirServer <29200> listens on: h19n03:30035





Vampir Hands on

Connect Client to Vampir Server:







Vampir Hands on

Connect Client to Vampir Server:







The main displays of Vampir

Timeline Charts:

- Master Timeline
- Process Timeline
- Counter Data Timeline
- Performance Radar

Summary Charts:

- Function Summary
- Message Summary
- Process Summary
- Communication Matrix View





Vampir: Displays for a WRF Trace with 64 Processes



Master Timeline Vampir - [Trace View - /home/dolescha/tracefiles/feature-traces/wrf-p6

 \sim

🥁 File 🗸 View Help View Chart Filter

Detailed information about functions, communication and synchronization events for collection of processes.

	Timel	ne			Func	tion Legend
26.86	0 s 26.865 s 26.870 s 26.87	5s 26.880s 26	6.885 s 26.890 s	-	DYN	
	MDI Wait	MDI Wait	i i i		1/0	
ocess 0	MOL Wait	MPI_Wait	solve_em_		MEM	
ocess 2	MPI Wait	Cut-ware	solve_em_		MPI	
ocess 3	MPL Wait	MPI Wai	solve_em		PHYS	
ocess 4	odule em mp rk step prep		it solve em		WRF	
ocess 5	MPEWait		solve em		_	
ocess 6 m	odule em mp rk step prep		solve em			
ocess 7	MPI Wait		olve em			
ocess 8 m	odule_em_mp_rk_step_prepMRI	Wait	solve_em_		Co	ontext View
ocess 9	MPI_wait		solve_em_	-	Master Time	eline 🔀 🛛
ocess 10	MPI_Wait		solve_em_	Pr	operty	Value
ocess 11	MPLwait		solve_em_	Di	splay	Master Timeline
ocess 12	MPh_Wait		solve_em_	Typ	be	Function
ocess 13	MPT_Wait		solve_em_	Fu	nction Name	MPI_Wait MPI
ocess 14 mo	odule_em_mp_rk_step_prep_		solve_em_	Int	erval Begin	26.872264 s
ocess 15 🛛 📩	odule_em_mp_rk_step_prep_		solve_em_	Int	erval End	26.885236 s
ocess 16 🗖	odule_em_mp_rk_step_prep_		/solve_em_	Du	ration	0.012972 s
ocess 17 🛛 📩	odule_em_mp_rk_step_prep_		solve_em_	50	urce File	
ocess 18	MPI_Wait	MIR_Wait	solve_em_		arce Line	
ocess 19 🔤	odule_em_mp_rk_step_prep		solve_em_			
ocess 20 mo	odule_em_mp_rk_step_prep		solve_em_			
ocess 21	MPI_Wait	MRI Wait	solve_em_			
ocess 22	MPI_Wait		solve_em_			
ocess 23 mo	odule_em_mp_rk_step_prep	MPI_Wait	solve_em_			
ocess 24	MPI_	Wait	MPI_Wait			

Process and Counter Timeline Vampir - [Trace View - /home/dolescha/tracefiles/feature-traces/wrf-pt

 \mathbf{v}

¥ File View Help

Detailed information about different levels of function calls in a stacked bar chart for an individual process.



Function Summary Vampir - [Trace View - /home/dolescha/tracefiles/feature-traces/wrf-p64-i

V

Overview of the accumulated information across all functions and for a collection of processes.

🥁 File View Help ⊻iew <u>C</u>hart <u>F</u>ilter 🗮 📷 🛄 🌑 🔄 🏭 🎆 🖏 😘 🕓 Function Summary Function Summary All Processes. Accumulated Exclusive Time per Function Group 30 s 20 s 10 s 0 s 39.185731 s Sum 33.298935 s MPI Wait 3.591342 s MPI Isend 2.295455 s MPI Irecv MPI [39.185731 s] DYN [135.259444 s] PHYS [73.474569 s]

Process Summary Vampir - [Trace View - /home/dolescha/tracefiles/feature-traces/wrf-p64-io-mem-r

 \sim

Find groups of similar processes and threads by using summarized function information.

¥	File \	/iew Help					function	informatio
⊻i	ew <u>C</u> ł	hart <u>F</u> ilter						
1	2 H	i 🚻 🍋 🗮	i 🔚 📷 🖏 🤹 🕓			n Thilling Thilling		LIATING SITERONY SUITE
Π.				Pr	rocess Summary			
		0 s						
	—) s	5s 1	0s 15s	s 20)s 2	5s 30s	35 s
3		MPI_Bcast	solve_em_	moduleriver_ MPLV	Vait mor_	mod_mo	pen m	
12	┝┯┹┩	MPI_Bcast	solve_em_	moduliver_ MP		pder_1mod_ n	n Topen _m	
2	╓┷╥	MPI_Bcast	solve_em_ m	poduleriver_MPI_wai	t mor n	nodmop_dop	en m	and an load
<i>'</i>	┡╼╫╫	MPI_Bcast	solve_em_	module_river_LMPLW	Value and a	modimer mod	foren m	ext5e1eau
5	┢╺╄╨╇╢	MPI_Bcast	Isolve em	module_river	Vait mod e	mod m	open m	
9	┢╈╧╫	MPL Bcast	solve_em_	moduleriver MPI	Wait moder	modld Im	open Im	
4	H	MPI Bcast	solve em	moduleriver MPI W	/ait mod.	er mod mo	p open 1m	
3	HΨ	MPI Bcast	solve er	n moduleriv	er MPI Wait	moder mo	.d m m	
11	ΠΠ	MPI_Bcast	solve_em	moduleriver	PI_Waitmod.	.er _ mod_ m	open m	
		-			-			-
-								
		1	1					8



Message Summary

Vampir - [Tr	ace View - /home/d	iolescha/trace	files/feature-t	rac	es/wrf-p64-io-mem-r	usage/wrf.1h.otf] 🛛 🔤 🔤
File View Help						_8
ew <u>C</u> hart <u>F</u> ilter						
= 53 🔟 🌰 🧮 📰 🛊	i 🖏 🚯 🔘	*			I TUILINA TUINANNI A	
	Message Summary					Context View
7500 6000 45	500 3000	1500	0		Message Summar	y 🛛 📔 📲
93			15.039062 KiB	-	Property	Value
49	28		10.664062 KiB		Display	Message Summary
49	11		10.9375 KiB		Message Size	105.820312 KiB
49	00		23.203125 KiB		Message Transfer Rate	561.904092 MiB/s
	2800		16.875 KiB			Min: 21.282674 MiB/s, Max: 1.872321 GiB/
	2800		16.453125 KiB			
	22	54	48.398438 KiB			
	2	156	145.12 KiB			
	2	156	75.839844 KiB			
	2	2071	60.800781 KiB			
		1470	218.45 KiB			
		1288	36.09375 KiB			
		1288	35.273438 KiB			
1 1	Massaga Summan/	I	1	–		
1600 MiB/s 1280 MiB/s 96	0 MiB/s 640 MiB/s 3	20 MiB/s 0 M	iB/s			
	577.6	538488 MiB/s	212.68 KiB	-		
	562.	663161 MiB/s	217.59375 KiB			
	561.	904092 MiB/s	105.82 KiB			
	543	.888495 MiB/s	108.28125 KiB			
	493.213465 MiB/s		296.71875 KiB			
	477.257369 MiB/s		159.25 KiB			
	469.472582 MiB/s		162.9375 KiB			
	467.421276 MiB/s		35.273438 KiB			
	461.590886 MiB/s		36.09375 KiB			
	457.004166 MiB/s		325.875 KiB			
	451.100737 MiB/s		318.46875 KiB			
	432.708147 MiB/s		147.65625 KiB			
	404.654303 MiB/s		48.398438 KiB			
				•		

















Center for Information Services and High Performance Computing (ZIH)

Vampir: What's New



(released in Nov '21)

What's New Dialog when starting new version the first time New Chart Icons

😣 🖲 🔳 Trac	e View -	Default:/Trace	es/Large_Sco	re-P/traces.ol	:f2 * - Vampir	
<u>F</u> ile <u>E</u> dit <u>C</u>	hart F <u>i</u> lt	er <u>W</u> indow	<u>H</u> elp			
: 📰 👯 🚮	<u>M</u>	. 🖭 🔍 🔇	। 🖙 🐴 🕅	5 III i a	£ 🗄 🗡 🗐 🎦	83s
		Timeline			Context	t View 📃 🗷 🗷
	0 <u></u> s	25s	50s	75s	📰 🕕 Trace Info 🗱	⊼ +
Thread:0				anno	Property	Value
Thread:1	000	\mathbf{x}	20100000	anna	File	Default:/1
					Creator	Score-P 1.







(released in Nov '21)

Support new OTF2 3.0 features

- Accelerator devices and contexts
- Non-blocking collective operations

No Score-P release yet







(released in Nov '21)

Collective Operation Bursts

- Collated into bursts, when too many of them occur in a specific interval
- Collectives also visible if beginning is not in the zoom interval







(released in Nov '21)

Summarizations in the Performance Overlay

— Values are summarized into their parent groups











Center for Information Services and High Performance Computing (ZIH)

Vampir Case Study: Analyzing Load Imbalance in COSMO-SPECS



E IX	🔛 🔐	16 m	6 🔄		101	# 4		1 83	1	Q	s Ing pang				578 1	01 s			5	578 5
					T	imeline		1000					××			F	unction Summ	nary		
	0 s 50 s	s 100 s	150 s	200 s	25	0 s	300 s	350 s	400 :	s 450 s	500 s	550 s		All Process	es, Accumu	alated E	xclusive Tim	e per Function	on Group	
Drocess O					-	-	-	The second	-		TTTT OTT	-	s lix	20,000	s 15,00	20 s	10,000 s	5,000 s	0 s	
Process 0									144				18	23,104.03		_		_	MP	
Process 7		THE REAL													18,783.74	15			MPI	-
Process 2																	11,577,183	1 550 84	MP_U	JTIL.
Process 4																		1,290,50	Apple Apple	DIE
Process 5																		993.1	46 s METE	EO
Process 6																		4	4.366 s VT_AF	PI
Process 7	0313131																			
Process 8	0313131								100001											
Process 9																	Context Vie	N		
Process 10														=		× II	Trace			1 F
Process 11	0303031				1111	0000	1000							Property		1	Value			
Process 12	0313131													File			Leone les sur	has/Till/Ten	an Alamaia T. d	staalal
Process 13	0313131													File			Users/mawe	Der/ZiPl/Trac	oes/vampir-1ut	torial
Process 14	0303031				1999.0	88388								Creator		'	/ampirTrace	5.11.2		
Process 15	0303031													Version		1	1.11 alpha			
Process 16	0303031				12222									Number of	Processes	1	00			
Process 17	0303031				12 2 2 2 2									Timer Res	olution	Ę	50 ms			
Process 18	0303031												2	Trace Time	15					
Process 19	-30303I												2	Start		1	Thu Nov 3 12	:36:57 2011	(13203202177	7616
Process 20	000000				10000								2	Stop		1	Thu Nov 3 12	:46:30 2011	(1320320790)	09206
Process 21	000000				11000									Flansed			0:09:33 (57	31589891		
Process 22	0999999													VamairTra	aa Environa	nont '	10.00.00 (0)	01000001		
Process 23				14.00				444						vampirma	ce Environn	nem,				
Process 24		ADD DO DO						****						VI_MOD			RAGE			
Process 20 Process 28	CRIMINS.						-							VT_BUF	FER_SIZE	-	200M			
Process 20	· ·	ARRENT OF	STATE OF			1111								VT_SYN	C_FLUSH	r	10			
Process 28							11111							VT_SYN	C_FLUSH_L	EVEL 8	30			
Process 29						-								VT_MAX	FLUSHES	1				
Process 30													6	VT_MET	RICS	F	PAPLEP_OPS	6		
														VT_RUS	AGE		cnot set>			

Weather forecast code COSMO-SPECS with 100 processes COSMO: weather model (METEO group) SPECS: microphysics for accurate cloud calculation (MP and MP_UTIL group) Coupling of both models is done in COUPLE group







Compared to METEO, MP and MP_UTIL are very compute intensive, however this is due to more complex calculations and no performance issue

Problem: >32% of time spent in MPI

MPI runtime share increases throughout the application run







Zoom into the first three iterations MP/MP_UTIL perform four sub-steps in one iteration Low MPI time share Everything is balanced and looks okay







Zoom into the last three iterations

Very high MPI time share (>50%)

Large load imbalance caused by MP functions around **Process 54** and **Process 64**









PAPI_FP_OPS counter showing higher FLOPs rates on processes causing the imbalance

Reason for imbalance: Static grid used for distribution of processes. Depending on the weather, expensive cloud computations (**MP** group) may be only necessary on some processes







Process Summary helps finding outliers Groups processes by their behavior (similar call/duration profile) Number of expected groups is variable In this case 4 yields the best results











Center for Information Services and High Performance Computing (ZIH)

Vampir Showcase: Analyzing CUDA Applications



CUDA Application



Material science code LSMS CUDA is utilized for heavy computations

CUDA streams are children of the owning process

Allows an in-depth analysis of host-device communication





CUDA Application



Communication Matrix best for analyzing the general communication pattern Expectation: balanced communication, represented by a symmetric matrix Problem: communication with stream 7 is different





CUDA Application



Shared Resource Timeline offers a per device view on kernel executions Best suited for analyzing multi-GPUs per node scenarios Allows a dedicated analysis of kernel execution patterns Yields insights of the actual hardware usage









Center for Information Services and High Performance Computing (ZIH)

Vampir Showcase: Analyzing Multilayer File I/O Applications



File I/O recording

Omnipresent in todays HPC applications

Record interaction between multiple layers

- MPI I/O (MPI_File_open)
- ISO C I/O (fopen)
- POSIX I/O (open, interface to OS)

System tree information determine whether file resides in a shared filesystem

High level of detail

=> Trace data might increase dramatically





NetCDF & HDF5 not supported yet







MPI I/O benchmark

fs.hlrs.de/projects/par/mpi//b_eff_io/

MPI I/O instrumentation is enabled by default by Score-P

```
% scorep-mpicc -o b_eff_io b_eff_io.c
% export SCOREP_EXPERIMENT_DIRECTORY=scorep-b_eff_io-4-profile
% mpirun -n 4 -c 6 ./b_eff_io -MB 2048 -MT 98304 -rewrite -N 4 -T 60
% scorep-scorep -g scorep-b_eff_io-4-profile/profile.cubex
% export SCOREP_EXPERIMENT_DIRECTORY=scorep-b_eff_io-4-tracing
% export SCOREP_FILTERING_FILE=initial_scorep.filter
% export SCOREP_ENABLE_TRACING=true
% export SCOREP_TOTAL_MEMORY=31MB
% mpirun -n 4 -c 6 ./b_eff_io -MB 2048 -MT 98304 -rewrite -N 4 -T 60
```





Result visualization









Enabling ISO C and POSIX I/O instrumentation

Instrumentation might require threading support

















Center for Information Services and High Performance Computing (ZIH)

Summary & Conclusion

ORNL/OLCF Vampir and Score-P Training 2022 Jan 19



Summary

Vampir & VampirServer

- Interactive trace visualization and analysis
- Intuitive browsing and zooming
- Scalable to large trace data sizes (20 TByte)
- Scalable to high parallelism (200000 processes)

Vampir for Linux, Windows and Mac OS X

Note: Vampir neither solves your problems automatically nor points you directly at them. It does, however, give you FULL insight into the execution of your application.





Conclusion

Performance analysis very important in HPC

Use performance analysis tools for profiling and tracing

Use tracing tools with some precautions

- Overhead
- Data volume

Contact the Vampir Team for help on usage: service@vampir.eu







https://www.vampir.eu





Interactive Trace Analysis with Vampir

Slide 49