

Guidelines on debugging and introduction to Totalview

**Dominique Lucas
George Mozdzynski**

Outline

- Different approaches to debugging
 - Printing
 - Compiling environment aiding tools
 - Signal trapping
- Debugging with totalview

Don't Panic

- “Programming is an art” ... and so is debugging
 - “You don’t need a sledgehammer to crack a nut”
- Most problems are trivial and easy to fix.
- Look at stack trace/point of failure.
- Intuition, experience, luck all play a part.
- Check your code “again”.
- Explain/show your code to somebody else.
- Use make to build your libraries/models, to guarantee that the same options are used everywhere.

Some suggestions

- Is the problem reproducible on a rerun?
 - Does it fail in exactly the same way and place
- Try `OMP_NUM_THREADS=1`
- No stack trace?
 - suspect 'you' have trashed memory
- Try to reproduce problem at a lower resolution

The Universal Debug Tool

(the print/write statement)

- Will always be there for you!
- What to write out.
- Must be selective to keep file size(s) manageable.
- Can be activated with an DEBUG variable or Namelist entry.
- For “production” runs, try and switch off verbose output mode.

Memory Constraints

- Task stack limit 4 Gbytes
- Thread stack limit
 - Master thread 4 Gbytes
 - Other threads 256 Mbytes
- Large arrays (>1 Mbyte) should be declared allocatable
 - allocatable arrays use the (per task) heap
 - no practical limit on heap size
- Fortran 90/95 (xlf90/xlf95) put local data on stack.
- Fortran 77 (xlf) puts statically allocated local data on heap.

Eoj – check memory and CPU utilisation

```

eoj vers1.4 run at Wed Jun 11 17:35:46 GMT 2003 on hpca2501 for jobstep hpca2301.347796.0
Queued      : Wed Jun 11 17:11:09 GMT 2003 for      898 seconds
Dispatched  : Wed Jun 11 17:26:07 GMT 2003 for      579 seconds
Job Name    : edhm_model_fcgroup1
Step Name   : 0
Owner       : rdx
Unix Group  : rd
Account     : ecpreops
STDIN       : /dev/null
STDOUT      : /hpca/rdx_dir/log/mpm/edhm/fc/fcgroup1/model.1
STDERR      : /hpca/rdx_dir/log/mpm/edhm/fc/fcgroup1/model.1
Class       : np
Step Type   : General Parallel
Node Usage  : shared
Step Cpus   : 8
Total Tasks :
Blocking    :
Node actual : 1
Adapter Req. : (csss,MPI,shared,US)
Resources   : ConsumableCpus(1) ConsumableMemory(900.000 mb)
*** Next 3 times NOT up-to-date (TOTAL CPU TIME given later IS accurate)
Step User   Time : 00:13:41.940000
Step System Time : 00:00:33.760000
Step Total   Time : 00:14:15.700000 (855.7 secs)
*** Last 3 times NOT up-to-date (TOTAL CPU TIME given later IS accurate)
Context switches : involuntary =      28637, voluntary   =      12378
                  per second =         49                21
Page faults      : with I/O   =      13056, without I/O =     1203613
                  per second =         22                2078

<----- CPU -----> <----- MEM ----->
Node   ? #T #t secs/CPU (Eff%) (Now%) max/TSK mb (Eff%) (Now% - mb ) Task list
-----
hpca1503 M 8 1 120.33 ( 20%) ( 52%) 764.85 ( 84%) ( 88% - 7680) 0:1:2:3:4:5:6:7:
-----

Elapsed = 579 secs 900 mb = ConsumableMemory
CPU Tot = 962.64 ( 0+00:16:02) Average: 963 s/node, 120 s/task
System Billing Units used by this jobstep = 1.037

```

Debugging – compiler options

- checking:
 - argument checking: `-qextchk`
`$ xlf -qextchk prog.f -o prog`
Note that checking is done at compilation/linking
 - array bounds checking: `-C`
`$ xlf -C prog.f -o prog`
`$./prog`
Note that checking is done at runtime
 - undefined reference checking
`$ xlf -qinitauto=FF -qsigtrap \`
`-qfltrap=inv:over:nanq:zero:en prog.f -o prog`
Note that checking is done at runtime

Floating point exception

- In IEEE, a floating point exception sets status flag
- By default execution continues
- Trapping the exception requires software checking of status flags
- Utilities for enabling and checking
 - CALLs to fpgets and fpsets within program
- Automatic trapping by compiler (option `-qflttrap`)
 - high overhead to do precisely for whole program
 - "IMPrecise" option checks only at subprogram entry and exit
 - can apply flttrap to routines selectively

Floating point exception

- IEEE exception types
 - OVerflow, UNDerflow, ZERODivide, INValid, INEXact
- Other -qflttrap options
 - IMPrecise: check at routine exit and entry only
 - ENable
 - must specify in main program
 - may as well specify it everywhere

Floating point exception

- Examples

```
$ xlf -qflttrap=overflow:invalid:zerodivide:enable \  
    -qsigtrap prog.f -o prog  
$ ./prog
```

- One can also use other exception handlers
- Relatively expensive ... up to 20%

Debugging – core files

- Core files – how to get a traceback
 - `$ dbx ./prog core <<eof`
where
eof

ECMWF local signal trap - ECLIB

```
INTEGER*4 CORE_DUMP_FLAG, IRETURN, SIGNALS(1), SIGNAL_TRAP
REAL A
CORE_DUMP_FLAG = 0
SIGNALS(1) = 0
IRETURN = SIGNAL_TRAP(CORE_DUMP_FLAG, SIGNALS)
IF (IRETURN .LT. 0) THEN
  PRINT *, 'ERROR'
ELSE IF (IRETURN .EQ. 0) THEN
  PRINT *, 'FPE TRAPPING IS NOT SET'
ELSE
  PRINT *, 'FPE TRAPPING MODE =', IRETURN
ENDIF
call b(-2.)
end

subroutine b(a)
real a
write(*,*)sqrt(a)
return
END
```

- Link using \$ECLIB, e.g.
 - \$ xlf -c prog.f
 - \$ xlf prog.o -o prog \$ECLIB

Signal_trap - arguments

- CORE_DUMP_FLAG:
 - = 0: no core dumped
 - Not = 0: core dumped
- SIGNALS – integer array with signals to trap
 - Signals(1)=0 => SIGFPE, SIGILL, SIGBUS, SIGSEGV, SIGXCPU.
 - See “kill -l “ for list of signals.
- Sample traceback ...

Signal received: SIGTRAP - Trace trap

Signal generated for floating-point exception:

FP division by zero

Instruction that generated the exception:

`fdivs fr01,fr01,fr02`

Source Operand values:

`fr01 = 1.0000000000000000e+00`

`fr02 = 0.0000000000000000e+00`

Traceback:

Offset 0x00000040 in procedure sub_

Offset 0x00000048 in procedure ifs_model

--- End of call chain ---

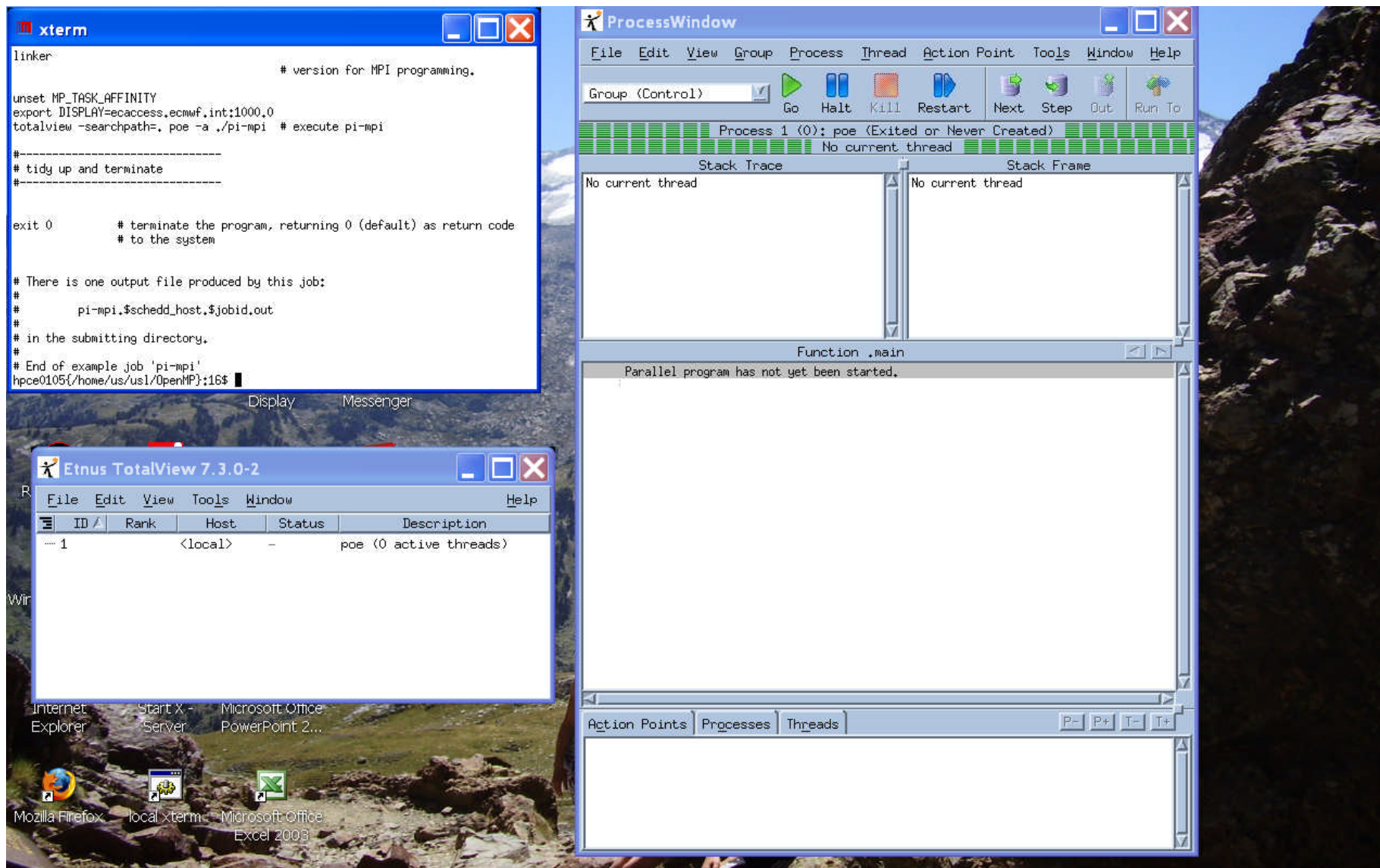
Totalview

- Recompile your application without optimization (and `-g`):
 - `-qnooptimize` for all routines
 - `-qsmp=noopt` for routines with OpenMP
 - Beware `-qsmp=omp` implies optimization
 - Best choice: `-g [-qoptdebug] -qfullpath`
- Command Line interface exists
 - `$ totalviewcli`
- Recommended use of totalview in batch mode, i.e. with the GUI version.

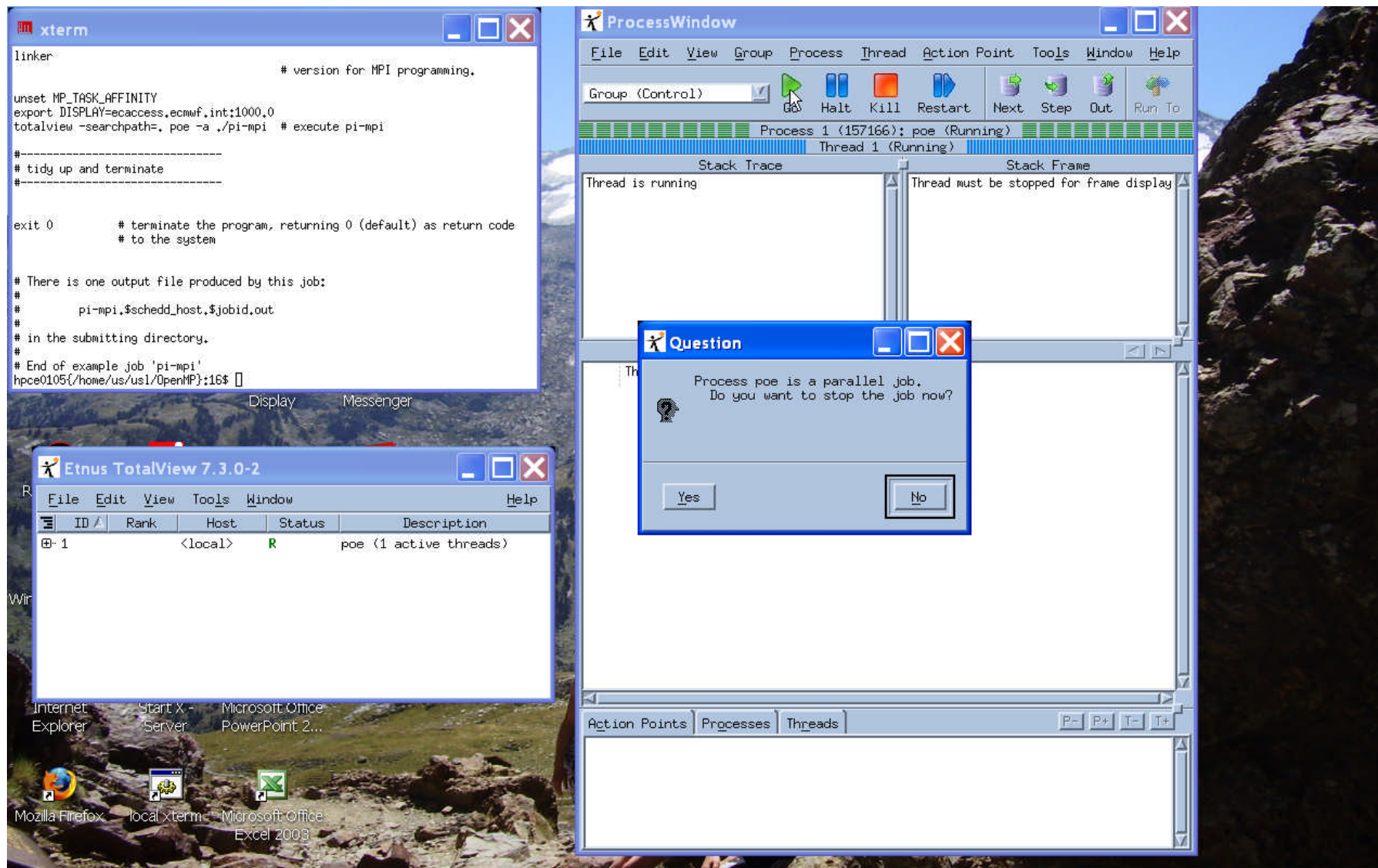
Totalview – GUI interface

- Before submitting batch job to launch totalview, request an X11 proxy at login time via ECaccess (ssh -X or NX).
- Include in your batch job the display:
`export DISPLAY=<your_ecaccess_display>`
- If needed, include source code searchpath, e.g.
`searchpath='dir_1/,dir_2/,...,dir_n/'`
- For MPI-parallel (load-leveller jobs):
`totalview -searchpath=$searchpath poe -a <executable> <args>`
- For serial/OpenMP only (interactive)
`totalview -searchpath=$searchpath <executable> -a <args>`

Totalview – GUI interface



Totalview – GUI interface



Totalview – GUI interface

The screenshot displays the Totalview GUI interface with three main components:

- Terminal Window (xterm):** Shows a script for running an MPI program. The script includes commands like `unset MP_TASK_AFFINITY`, `export DISPLAY=ecaccess.ecmwf.int:1000,0`, and `totalview -searchpath=. poe -a ./pi-mpi`. It also includes a `Display` command and a `Messenger` window.
- ProcessWindow:** Displays the execution state of the process. It shows the `Stack Trace` and `Stack Frame` for the function `pi` in `pi-mpi.f`. The `Stack Frame` section lists registers for the frame, including `R0`, `SP`, `RTOS`, `R3`, `R4`, `R5`, `R6`, and `R7`.
- Table of Running Processes:** A table with columns `ID`, `Rank`, `Host`, `Status`, and `Description`. It lists three processes:

ID	Rank	Host	Status	Description
1		<local>	B	poe (1 active threads)
2	0	<local>	T	poe<pi-mpi>.0 (1 active threads)
3	1	<local>	T	poe<pi-mpi>.1 (1 active threads)

The background of the desktop shows a rocky landscape. The taskbar at the bottom includes icons for Internet Explorer, Start X-Server, Microsoft Office PowerPoint 2003, Mozilla Firefox, local xterm, and Microsoft Office Excel 2003.

Totalview – GUI interface

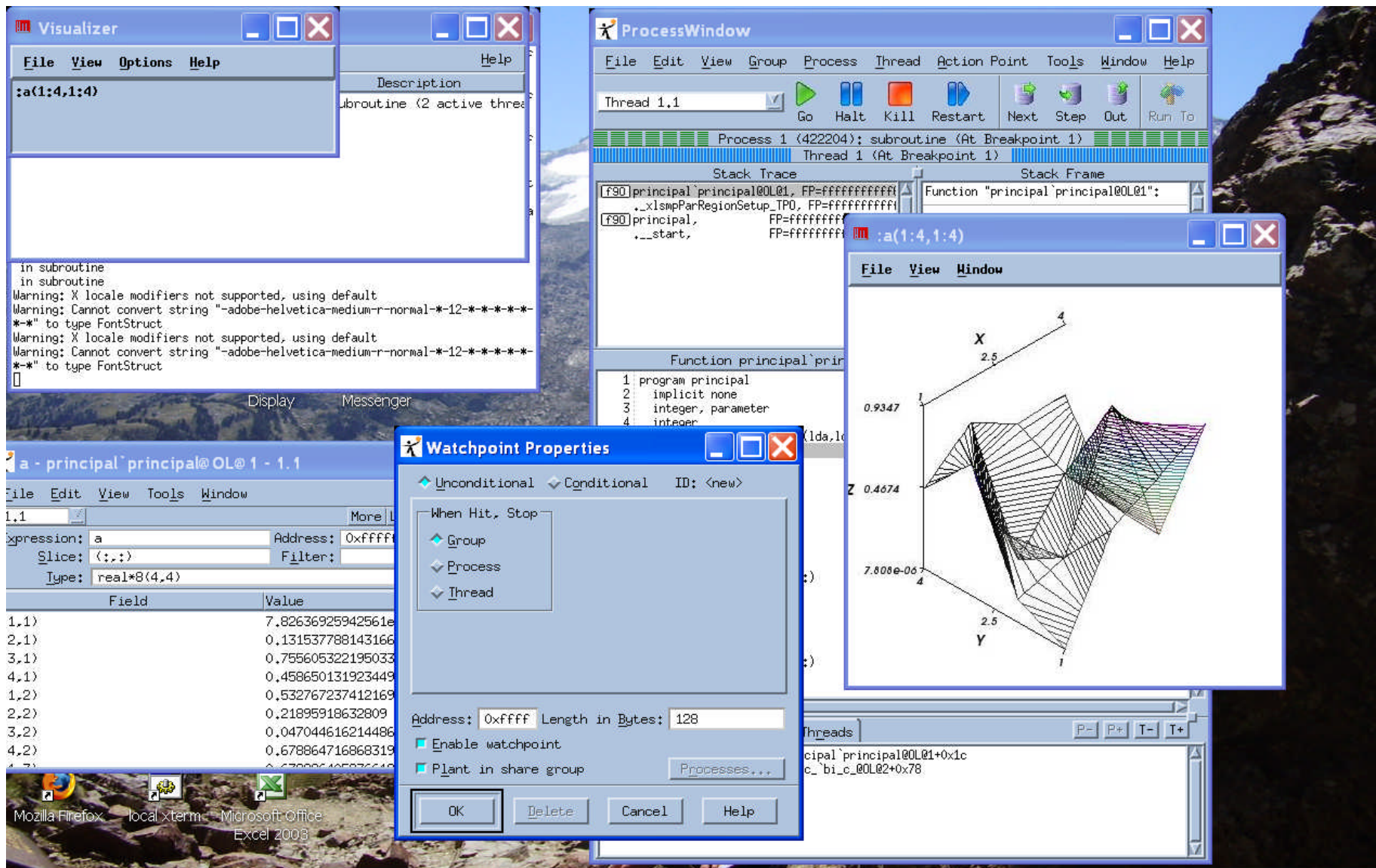
The screenshot displays the Etnus TotalView 7.3.0-2 GUI interface, which is used for debugging and performance analysis. The interface is divided into several windows:

- Main Window:** Shows a list of threads. The first thread is selected, showing its ID, Rank, Host, Status, and Description. Below the list, there is a warning message: "WARNING: icd_aix_symbol_info_t::lookup_symbols_in_ldinfo: '.__tls' was not preloaded. Performance may suffer." and a "Display" button.
- ProcessWindow:** Shows the stack trace and stack frame for the selected thread. The stack trace shows the function "principal" and its subroutines. The stack frame shows the registers for the frame, including R0, SP, RT0C, R3, R4, R5, and R6.
- Variable Window:** Shows the value of the variable "a" at the current execution point. The variable is of type "real*8(4,4)" and its value is displayed in a table.

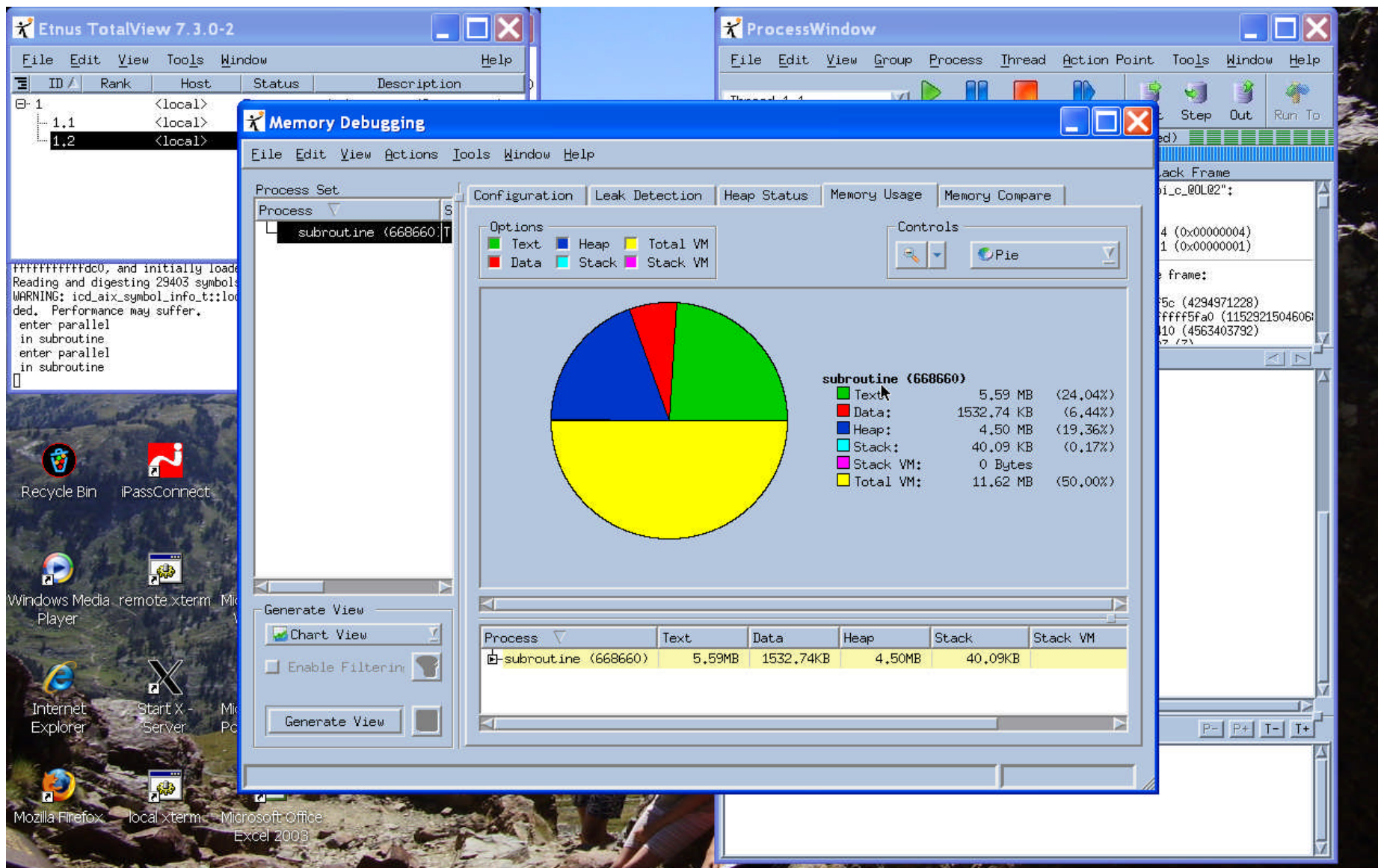
The variable window shows the following data:

Field	Value
1,1)	7.82636925942561e-06
2,1)	0.131537788143166
3,1)	0.755605322195033
4,1)	0.458650131923449
1,2)	0.532767237412169
2,2)	0.21895918632809
3,2)	0.0470446162144861
4,2)	0.678864716868319

Totalview – GUI interface



Totalview – GUI interface



The screenshot displays the Etnus TotalView 7.3.0-2 GUI interface. The main window shows a list of processes, with 'subroutine (668660)' selected. The 'Memory Debugging' window is open, showing a pie chart and a table of memory usage for the selected process.

Memory Usage Data:

Category	Size	Percentage
Text	5.59 MB	(24.04%)
Data	1532.74 KB	(6.44%)
Heap	4.50 MB	(19.36%)
Stack	40.09 KB	(0.17%)
Stack VM	0 Bytes	
Total VM	11.62 MB	(50.00%)

The 'Generate View' section shows the 'Chart View' selected, and the 'Enable Filtering' checkbox is unchecked. The 'Process' dropdown is set to 'subroutine (668660)'.

Case Study – “bug” in EC-EARTH

- EC-EARTH – international project for earth system modelling.
- Experimental runs on our HPC systems.
- Coupled model runs, using MPMD parallel modelling approach:
 - `poe -pgmmodel mpmd -cmdfile ${cmdfile} ...`

with cmdfile like:

oasis3.MPI1.x

ifsMASTER -v ecmwf -e ewhx

opa_exe.ORCA2_OASIS3.1.1

appl-tm5.x

Case Study – initial error

- Oasis compiled with `-O3`, TM5 with `-g`, IFS with `-g` and signal trap:

1:signal_drhook(SIGABRT=6): New handler installed at 0x100ac5b8; old preserved at 0xa062ea70

ERROR: 0031-250 task 3: Segmentation fault

1: Traceback:

1: Offset 0x0000086c in procedure pm_async_thread

1: Offset 0x000000dc in procedure _pthread_body

1: --- End of call chain ---

- Error (segmentation fault) in TM5 and IFS reports where it failed ...

Case Study – totalview (1)

- The job has been adapted to use totalview:

```
export DISPLAY=galahad:0.0
```

```
export MP_PGMMODEL='mpmd'
```

```
export MP_CMDFILE="${cmdfile}"
```

```
export MP_HOSTFILE="${hostfile}"
```

```
export MP_LABELIO='yes'
```

```
/usr/local/apps/toolworks/totalview.8.6.1-1/bin/totalview /usr/bin/poe
```

- Error in TM5 in a routine PRISM_Put_Proto (see following page).

Totalview – first run

The screenshot shows the Totalview 8.6.1-1 interface. The main window displays the source code of the function `tm5_tm5_prism_puts_` in `tm5.F90`. The code is being debugged, and the stack trace shows the current state of the program. The stack frame shows the local variables and their values.

The secondary window, titled "TotalView 8.6.1-1 <@hpce0506>", shows a list of active threads:

ID	Rank	Host	Status	Description
1		<local>	T	/usr/bin/poe (1 active threa
2	0	<local>	T	poe<oasis3,MPI1.x>.0 (8 acti
3	1	<local>	T	poe<ifsMASTER>.1 (8 active t
4	2	<local>	T	poe<opa_exe,ORCA2_OASIS3.1.1
5	3	<local>	E	poe<appl-tm5.x>.3 (8 active

The main window shows the source code of the function `tm5_tm5_prism_puts_` in `tm5.F90`. The code is being debugged, and the stack trace shows the current state of the program. The stack frame shows the local variables and their values.

The stack trace shows the following frames:

- `tm5_tm5_prism_puts_` (FP=fffffffffb280)
- `tm5_tm5_model_run_` (FP=fffffffffbaf0)
- `appl_tm5` (FP=fffffffffc420)
- `__start` (FP=fffffffffc6f0)

The stack frame for `tm5_tm5_prism_puts_` shows the following local variables:

- `ivar`: 14 (0x0000000e)
- `prism_t`: 0 (0x00000000)
- `ifs_ll`: (real*8, allocatable:: (60,45))
- `ilev`: 1 (0x00000001)
- `region`: 1 (0x00000001)
- `imr`: 60 (0x0000003c)
- `jmr`: 45 (0x0000002d)
- `l...`: 45 (0x0000002d)

The source code shows the function `tm5_tm5_prism_puts_` in `tm5.F90`. The code is being debugged, and the stack trace shows the current state of the program. The stack frame shows the local variables and their values.

Case Study – signal trap

- Activate signal trap in TM5:
 - 3: Signal received: SIGSEGV - Segmentation violation
 - 3:
 - 3: Traceback:
 - 3: Location 0x000000010072c7fc
 - 3: Offset 0x00001d08 in procedure __tm5_NMOD_tm5_prism_puts_, near line 1101 in file tm5.F90
 - 3: Offset 0x00001b50 in procedure __tm5_NMOD_tm5_model_run_, near line 842 in file tm5.F90
 - 3: Offset 0x000004d0 in procedure appl_tm5, near line 189 in file appl-tm5.F90
 - 3: --- End of call chain ---
- Not much more information yet, because oasis not compiled with “-g”.
- Recompiling oasis with “-g” and running totalview produces ...

Totalview oasis compiled with "-g"

Rank 3: poe<appl-tm5.x>.3 (Error)

Thread 1 (1): appl-tm5.x (Error) <Segmentation Violation>

Stack Trace		Stack Frame	
f90	mod_prism_put_proto`prism_put_proto_r28_, FP=	Function "mod_prism_put_proto`prism_put_proto_r28_":	
f90	tm5`tm5_prism_puts_, FP=fffffffffb270	id_port_id:	170 (0x000000aa)
f90	tm5`tm5_model_run_, FP=fffffffffbbae0	kstep:	0 (0x00000000)
f90	appl_tm5_, FP=fffffffffc410	rd_field_2d:	(real*8(60,45))
	__start, FP=fffffffffc6f0	kinfo:	0 (0x00000000)
		Local variables:	
		rd_field:	(real*8(2700))
		il_newtime:	0 (0x00000000)
		info:	150994944 (0x090000000)
		isend:	0 (0x00000000)
		ip:	0 (0x00000000)
		iport:	170 (0x000000aa)

Function mod_prism_put_proto`prism_put_proto_r28_ in mod_prism_put_proto.F90

```

1202 |
1203 |   rd_field(1:myport(4,id_port_id)) = RESHAPE (rd_field_2d(:, :), &
1204 |       (/myport(4,id_port_id)/))
1205 |
1206 |   If the user indicated in the namcouple that the field must be
1207 |   accumulated or averaged (keyword 'AVERAGE' or 'ACCUMUL' at the
1208 |   end of the field 2nd line), do the local transformations.
1209 |
1210 |   IF (ig_def_trans(iport) .EQ. ip_instant) THEN
1211 |     cl_str = 'inst(ident(X))'
1212 |
1213 |
1214 |     dg_field_trans(:, iport) = rd_field (:)
1215 |   ELSEIF (ig_def_trans(iport) .EQ. ip_average) THEN
1216 |     dg_field_trans(:, iport) = (real*8, allocatable)::(9312,249))
1217 |     IF (ig_def_trans(iport) .EQ. ip_min .OR. &
1218 |         ig_def_trans(iport) .EQ. ip_max) THEN
1219 |       IF (ig_number(iport) .EQ. 0) dg_field_trans(:, iport) = 0
1220 |       il_nbin = myport(4, iport)
1221 |       il_nbout = il_nbin
1222 |       cl_topps = 'ave, inst, t_sum, t_min, t_max'
1223 |       IF (ig_def_trans(iport) .EQ. ip_average) THEN
1224 |         cl_str = 'ave(ident(X))'

```

Action Points | Processes | Threads

P- P+ T- T+

dg_field_trans and rd_field are not conformant.

References

- Xlf Compiler Reference:

http://www.ecmwf.int/publications/manuals/hpcf_power6/xlf12.1_cr.pdf

- Xlf Optimisation and Programming Guide:

http://www.ecmwf.int/publications/manuals/hpcf_power6/proguide.pdf

- Totalview:

<http://www.roguewave.com/support/product-documentation/totalview-family.aspx#totalview>