NAME

tnf_probes – TNF kernel probes

DESCRIPTION

The set of probes (trace instrumentation points) available in the standard kernel. The probes log trace data to a kernel trace buffer in Trace Normal Form (TNF). Kernel probes are controlled by **prex**(1). A snapshot of the kernel trace buffer can be made using **tnfxtract**(1) and examined using **tnfdump**(1).

Each probe has a *name* and is associated with a set of symbolic *keys*, or *categories*. These are used to select and control probes from **prex**(1). A probe that is enabled for tracing generates a TNF record, called an *event record*. An event record contains two common members and may contain other probe-specific data members.

Common Members

tnf type name	**member name**
tnf_probe_event	*tag*
tnf_time_delta	*time_delta*

tag encodes TNF references to two other records:

- *tag* describes the layout of the event record
- *schedule* identifies the writing thread and also contains a 64-bit base time in nanoseconds.

time_delta a 32-bit time offset from the base time; the sum of the two times is the actual time of the event.

Threads

thread_create

tnf_kthread_id	*tid*
tnf_pid	*pid*
tnf_symbol	*start_pc*

Thread creation event.

tid the thread identifier for the new thread
pid the process identifier for the new thread
start_pc the kernel address of its start routine.

thread_state

tnf_kthread_id	*tid*
tnf_microstate	*state*

Thread microstate transition events.

tid optional; if it is absent, the event is for the writing thread, otherwise the event is for the specified thread.

state indicates the thread state:
- running in user mode
- running in system mode
- asleep waiting for a user-mode lock
- asleep on a kernel object,
- runnable (waiting for a cpu)

- stopped.

The values of this member are defined in **<sys/msacct.h>**. Note that to reduce trace output, transitions between the *system* and *user* microstates that are induced by system calls are not traced. This information is implicit in the system call entry and exit events.

thread_exit

Thread termination event for writing thread. This probe has no data members other than the common members.

Scheduling

thread_queue

tnf_kthread_id	*tid*
tnf_cpuid	*cpuid*
tnf_long	*priority*
tnf_ulong	*queue_length*

Thread scheduling events. These are triggered when a runnable thread is placed on a dispatch queue.

cpuid	specifies the cpu to which the queue is attached.
priority	the (global) dispatch priority of the thread.
queue_length	the current length of the cpu's dispatch queue.

Blocking

thread_block

tnf_opaque	*reason*
tnf_symbols	*stack*

Thread blockage event. This probe captures a partial stack backtrace when the current thread blocks.

reason	the address of the object on which the thread is blocking.
symbols	references a TNF array of kernel addresses representing the PCs on the stack at the time the thread blocks.

System Calls

syscall_start

tnf_sysnum	*sysnum*

System call entry event.

sysnum	the system call number. The writing thread implicitly enters the *system* microstate with this event.

syscall_end

tnf_long	*rval1*
tnf_long	*rval2*
tnf_long	*errno*

System call exit event.

rval1
rval2 the two return values of the system call

errno the error return.

The writing thread implicitly enters the *user* microstate with this event.

Page Faults

address_fault

tnf_opaque *address*
tnf_fault_type *fault_type*
tnf_seg_access *access*

Address-space fault event.

address gives the faulting virtual address.

fault_type gives the fault type: invalid page, protection fault, software requested locking or unlocking.

access gives the desired access protection: read, write, execute or create.

The values for these two members are defined in **<vm/seg_enum.h>**.

major_fault

tnf_opaque *vnode*
tnf_offset *offset*

Major page fault event. The faulting page is mapped to the file given by the *vnode* member, at the given *offset* into the file. (The faulting virtual address is in the most recent **address_fault** event for the writing thread.)

anon_private

tnf_opaque *address*

Copy-on-write page fault event.

address the virtual address at which the new page is mapped.

anon_zero

tnf_opaque *address*

Zero-fill page fault event.

address the virtual address at which the new page is mapped.

page_unmap

tnf_opaque *vnode*
tnf_offset *offset*

Page unmapping event. This probe marks the unmapping of a file system page from the system.

vnode and *offset*
identify the file and offset of the page being unmapped.

Pageins and Pageouts

pagein

tnf_opaque *vnode*
tnf_offset *offset*
tnf_size *size*

Pagein start event. This event signals the initiation of pagein I/O.

vnode and *offset*
identify the file and offset to be paged in.
size specifies the number of bytes to be paged in.

pageout

tnf_opaque *vnode*
tnf_ulong *pages_pageout*
tnf_ulong *pages_freed*
tnf_ulong *pages_reclaimed*

Pageout completion event. This event signals the completion of pageout I/O.

vnode identifies the file of the pageout request.
pages_pageout the number of pages written out.
pages_freed the number of pages freed after being written out.
pages_reclaimed the number of pages reclaimed after being written out.

Page Daemon (Page Stealer)

pageout_scan_start

tnf_ulong *pages_free*
tnf_ulong *pages_needed*

Page daemon scan start event. This event signals the beginning of one iteration of the page daemon.

pages_free the number of free pages in the system.
pages_needed the number of pages desired free.

pageout_scan_end

tnf_ulong *pages_free*
tnf_ulong *pages_scanned*

Page daemon scan end event. This event signals the end of one iteration of the page daemon.

pages_free the number of free pages in the system.

pages_scanned the number of pages examined by the page daemon. (Potentially more pages will be freed when any queued pageout requests complete.)

Swapper

swapout_process

tnf_pid *pid*
tnf_ulong *page_count*

Address space swapout event. This event marks the swapping out of a process address space.

pid identifies the process.
page_count reports the number of pages either freed or queued for pageout.

swapout_lwp

tnf_pid *pid*
tnf_lwpid *lwpid*
tnf_kthread_id *tid*
tnf_ulong *page_count*

Light-weight process swapout event. This event marks the swapping out of an LWP and its stack.

pid the LWP's process identifier
lwpid the LWP identifier
tid member the LWP's kernel thread identifier.
page_count the number of pages swapped out.

swapin_lwp

tnf_pid *pid*
tnf_lwpid *lwpid*
tnf_kthread_id *tid*
tnf_ulong *page_count*

Light-weight process swapin event. This event marks the swapping in of an LWP and its stack.

pid the LWP's process identifier
lwpid the LWP identifier
tid the LWP's kernel thread identifier.
page_count the number of pages swapped in.

Local I/O

strategy

tnf_device *device*
tnf_diskaddr *block*
tnf_size *size*
tnf_opaque *buf*
tnf_bioflags *flags*

Block I/O strategy event. This event marks a call to the **strategy**(9E) routine of a block device driver.

device contains the major and minor numbers of the device.
block the logical block number to be accessed on the device.
size the size of the I/O request.
buf the kernel address of the **buf**(9S) structure associated with the transfer.
flags the **buf**(9S) flags associated with the transfer.

biodone

tnf_device *device*
tnf_diskaddr *block*
tnf_opaque *buf*

Buffered I/O completion event. This event marks calls to the **biodone**(9F) routine.

device	contains the major and minor numbers of the device.
block	the logical block number accessed on the device.
buf	the kernel address of the **buf**(9S) structure associated with the transfer.

physio_start

tnf_device	*device*
tnf_offset	*offset*
tnf_size	*size*
tnf_bioflags	*rw*

Raw I/O start event. This event marks entry into the **physio**(9F) routine which performs unbuffered I/O.

device	contains the major and minor numbers of the device of the transfer.
offset	the logical offset on the device for the transfer.
size	the number of bytes to be transferred.
rw	the direction of the transfer: read or write (see **buf**(9S)).

physio_end

tnf_device *device*

Raw I/O end event. This event marks exit from the **physio**(9F) routine.

device the major and minor numbers of the device of the transfer.

SEE ALSO

prex(1), **tnfxtract**(1), **tnfdump**(1), **TNF_PROBE**(3X), **biodone**(9F), **buf**(9S)