HLD: PMC pm8001 Forensic Debug

Primary Contact: <mark_salyzyn@us.xyratex.com>

References

- 2090556_spc_fw_forensic_appnote_063051.pdf
- 20800222_pm8001_pm_081181.pdf

Abstract

The pm8001 driver lacks detailed forensics of the underlying hardware to debug field reported issues. PMC has a list of details from the hardware to simplify debugging of sequencer issues and currently we have no means to provide the data. These details are part of the 2090556_spc_fw_forensic_appnote_063051.pdf or later in PMC's document list. In addition we need to acquire the SPC event log as described in section 3,17 of the Programmers Manual 20800222_pm8001_pm_081181.pdf.

Definitions

LSB

- Linux Standards Base
- A definition for the organization of the filesystem tree.

sysfs

- exportable kernel object (kobj) representation in user space
- Limited to pagesize of data (4KB).
- Complies with LSB and is a GPL ABI
- tree mounted to /sys
- each node is read-only, read-write or write-only one-simple-purpose one-utf8-line (human readable) representation
- not atomic, rarely binary, no user-context, asynchronous and typically non-blocking.
- preferred over legacy ioctl (IO Control), a binary or atomic ioctl is a last-resort if feature or function can not be implemented as a sysfs node.
- replaces/deprecates the legacy proc (/proc) representations

debugfs

- exported file_operation representation in user space.
- No limit to the data transfer size
- Complies with LSB
- tree mounted to /sys/kernel/debug/
- each node is read-only, read-write or write-only and a less strict format requirements than sysfs.
- preferred over sysfs for enduring non-standardized driver-specific operations that do not purely follow the sysfs guidelines.
- Extends the scope of sysfs.

Forensic handler mechanics

Requirements

- Must abide as closely as possible to the 2090556_spc_fw_forensic_appnote_063051.pdf or later as supplied by the PMC vendor.
- Must abide as closely as possible to the 20800222_pm8001_pm_081181.pdf or later as

supplied by the PMC vendor.

- All changes must built for CentOS 5.4 to latest, RHEL6 to latest, Scientific Linux to latest, Manufacturing kernels (2.6.26-2-686-bigmem-kci-1.0) and on kernel.org latest.
- It is expected that any dynamic content will be voided if no special arrangements are made to freeze the pm8001 system. Because Linux is a running system, data can and will be fleeting. The user is expected to understand this limitation. Thus the Forensic data could be considered a dull knife (better than no knife at all).
- The sequencer version(s) needs to be acquired to provide context.
- There should be no special tools required to retrieve the forensic data, collection must be possible with a simple combination of POSIX coretools. Some illustrative examples follow, the final approach is to be documented in the associated DLD.
    - eg: find | sort -u is one possible POSIX coretools command that could produce a list of nodes in the appnote order, possibly utilizing the section-numbers in the appnote as part of the naming.
    - Use of find or sort should not be necessary since POSIX Bourne Shell performs expansion in sorted order. find should only really be necessary to divine through a debugfs tree should it be necessary.
    - eg: cat `find | sort -u` is one possible POSIX coretools command that could produce a complete set of results that is in the format specified in the appnote and can be directly emailed to the vendor.
    - eg: (while true ; do cat <event-node> ; done | tee /dev/stderr | grep '<nul>' >/dev/null) 2>&1 to pull a complete list of logs.
- To ease collection of the data, recommend that the compartmentalized results are readable from a hierarchy of debugfs nodes.
    - Each compartmentalized result is somewhat limited to 4KB of human-readable utf8 data as a self-imposed limit, if one element or array takes up more that this amount of data, then the nodes may need to be divided up into smaller logical references or a binary representation should be considered.
    - A human familiar with the appnote can see the names and hierarchy of the nodes in the debugfs tree and understand which subset of data could be acquired.
    - The compartmentalization should be as a minimum one node per forensic dump requirement as itemized in the appnote.
    - We will break from the sysfs standard by producing more than one line of output per recognized entity if due to the volume of data required. This decision results in debugfs being a better home for these operations.
- The feature of presenting portions of the forensic hierarchy must be considered and if necessary turned off by default due to Heisenberg's uncertainty principal and fault isolation requirements in case the data is too dynamic requires any system resources (CPU or memory) or too dangerous. A debugfs write with a specific simple signature (eg echo 'on' > /sys/.../forensic/enable) could be used to turn the isolated feature on, and conversely off at runtime.
- Collection of the Forensics needs to be integrated with the Trinity Diagnostic Dump facility.
- Forensic and Debug add-ons for drivers are resisted by the kernel community, but are

generally accepted if they are by-default ifdefed out through a kernel CONFIG compile-time variable. This should be considered acceptable and utilized in tandem with the dynamic feature enablement. The lack of the CONFIG variable is considered a state.

Architectural Challenges & Risks

- The pm8001 driver is GPL, and carries with it some License, Expected and Implied responsibilities that must be followed.
- An audit of all data required needs to be performed to determine the boundaries for the compartmentalization.
- SPC event log has a possible size of 4MB of each logs(msgu & IOP) and a need to record at various severity levels (eg: severity=4). This may not be not conducive for use with sysfs/debugfs retrieval, and suggest that each atomic read retrieves one entry with a signature to indicate a <nul> entry to terminate.
- An audit of all data size is required as some dumps may exceed the 4KB limit and a logical compartmentalization may not be possible.
- The intent is to match the PMC appnote as much as possible, we will necessarily be required to update when if the appnote changed. The design must be flexible enough to endure these changes.
- The dump assumes that the card operations have quiesced. Duplication of the problem may natural halt the system, but there may be a subsidiary requirement to lock the pm8001 activities and immunize the driver and controller from operations and timeouts.
- All efforts must be made to prevent the introduction of harm to the runtime Linux environment by the forensic requirements.
- There is no way to acquire this data if the operating system is inoperational due to the issues being investigated. This will no doubt preclude diagnosis regardless of environment and this limit needs to be accepted. The sequencer's UART becomes the next avenue of attack for the person performing the problem triage.
- Functional with a near-late-model wide variety of kernel releases CentOS5.4, SL6, Ubuntu (manufacturing) and latest on kernel.org as a minimum. Current Version and one behind.
- Changes in patch-form need to be individually apply-able to Xyratex branch, or Distribution branches of the driver. Scaling the management of these could concern us if the number of kernels escalates (especially if pushing merges into multiple avenues).
- debugfs interfaces have changed somewhat over time, making sure the same code functions on CentOS5.4 and SL6 is of concern.
- SL6 does not, by default, mount debugfs. Any forensic scripting must check and mount debugfs before retrieving the content.

Future Directions

- Submit changes to the kernel community.