tprof on NetBSD

Masanobu SAITOH (msaitoh@n.o)
Ryo SHIMIZU (ryo@n.o)
Who are we?

• We are working for IIJ to develop NetBSD based routers.
• The number of NetBSD developers in our team is 8.
pmc and tprof

• NetBSD had two different performance counter interfaces and commands.
• pmc(1)
  • Monitor PMCs.
  • Simple.
  • Specific to x86
• tprof(8) and tpfmt(1)
  • See the next page
What’s tprof?

• Sampling based profiler
  • Each time a performance counter overflows, the value of the PC at that time is recorded.
• Inspired by IBM AIX’s tprof?

History on NetBSD:
• First written by YAMAMOTO Takashi in 2008 (NetBSD 5.0).
  • global_power_events is used for the sampling event and can’t be changed to other event. It’s hardcoded in the backend.
  • x86 only
• Revamped by Maxime Villard in 2018 (NetBSD 9.0).
  • It allows users to choose which event to count.
  • Generic(MI) PMC interface is implemented.
  • And then, Jared McNeill wrote the code for ARM.
Removal of pmc(1) stuff

- problem
  - pmc(1) had not been maintained for years.
  - The code is duplicated between pmc(1) stuff and tprof stuff. Not shared at all. There are two different kernel interfaces.
  - tprof(8) can use only one performance counter. Supporting multiple counters is in the TODO list.

- So, pmc(1) stuff was removed. NetBSD 9.0 had no pmc(1).
At that time

- We were using `pmc(1)` at that time.
- We were **not** using `tprof` at that time.
- We thought
  - if `tprof` functionality includes `pmc` functionality, then removing `pmc` is OK.
- We did not oppose the proposal to remove `pmc(1)` stuff.
- Another reason why NetBSD developers working for IIJ didn’t oppose the proposal is that our routers are based on `netbsd-8` or prior.
Problems

- We, out team, develop new functionality and improvement on NetBSD-current first and then backport them to netbsd-8.
- Sometimes we used `pmc(1)` to see some performance counters.
- `tprof(8)` can use only one counter.
  - It’s important to monitor more than one counter at a time.
  - For example, we can’t calculate the last level cache’s hit ratio from llc-references and llc-misses.
- NetBSD-current has no `pmc(1)` anymore.
What’s new in NetBSD 10.0’s tprof?

• Support multiple counters at once.
• Subcommands:
  • list
  • monitor
  • analyze
  • count (<= new)
    • does not do any profiling, only outputs counters every interval.
  • top (<= new)
    • displays profiling results in real-time.
• (It’s not as feature-rich as FreeBSD’s pmcstat or Linux’s perf...)
TODO

• x86:
  • Add counter definitions that newer chips have.
  • Add support fixed PMC (Intel).
    • Currently support general counter only.
  • Get performance counter structure from CPUID 0x0a(Intel) and 0x80000022(AMD).
  • Use Intel PEBS (Processor Event-Based Sampling)
    • to reduce profiling overhead
    • to improve accuracy
  • AMD’s IBS is complexed. Need some investigation to use it.

• Collect events per LWP.
  • Our old pmc had the feature.

• Take some idea from FreeBSD and Linux’s PMC stuff.
Any questions?