



# Profiling in Xen

J. Renato Santos  
G. (John) Janakiraman  
Yoshio Turner  
Aravind Menon

HP Labs

Xen Summit

September 7-8, 2006



- OProfile introduction
- XenOprofile overview
- Current status
- Further work
- Tutorial

# OProfile – System Profiling in Linux

- Profile code execution (application & kernel)
- For selected hardware events, such as:
  - a) unhalted clock cycles: (most common use)
    - % of time spent on each user/kernel function
  - b) intructions:
    - % of instructions executed by each user/kernel function
  - c) L2 cache misses:
    - % of L2 misses in each user/kernel functionetc...
- Statistical profiling: continually sample currently executing code at every N hardware events. Large set of samples approximates the real hardware event distribution.

# OProfile – example



CPU: P4 / Xeon, speed 2794.74 MHz (estimated)

Counted GLOBAL\_POWER\_EVENTS events (time during which processor is not stopped) with a unit mask of 0x01 (mandatory) count 1000000

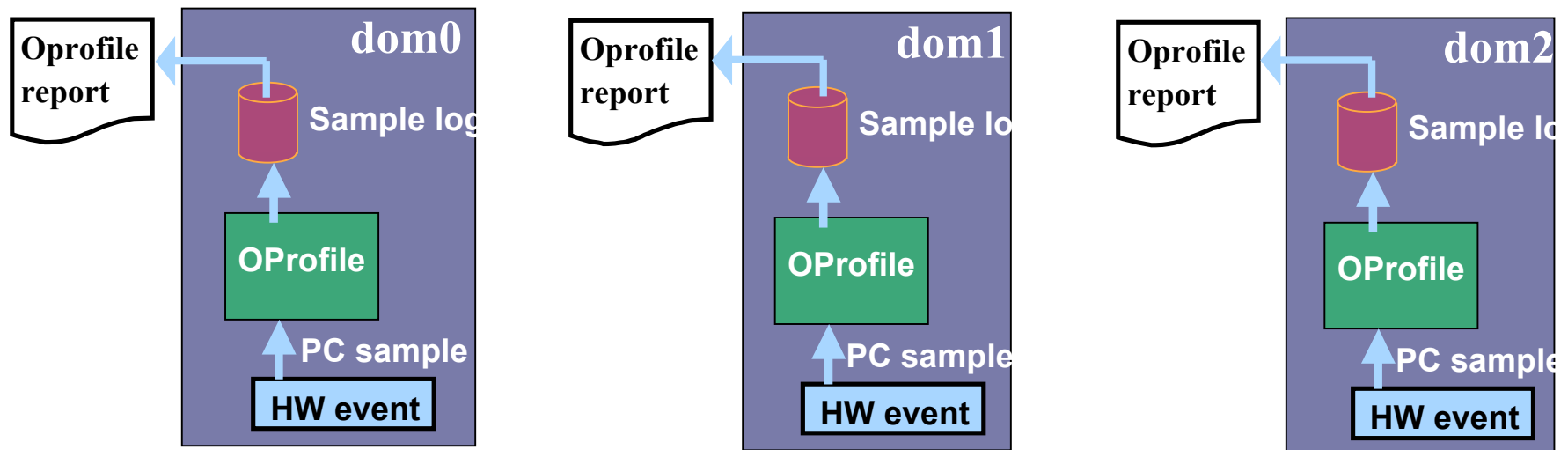
samples	%	app name	symbol name
9652	60.7694	libc-2.3.6.so	fputs_unlocked
3368	21.2051	libc-2.3.6.so	_IO_file_xsputn@@GLIBC_2.1
1829	11.5155	yes	(no symbols)
448	2.8206	libc-2.3.6.so	__i686.get_pc_thunk.bx
85	0.5352	vmlinux-syms-2.6.16.13	get_offset_pmtmr
51	0.3211	vmlinux-syms-2.6.16.13	delay_pmtmr
37	0.2330	vmlinux-syms-2.6.16.13	mark_offset_pmtmr
31	0.1952	bash	(no symbols)
28	0.1763	vmlinux-syms-2.6.16.13	timer_interrupt
25	0.1574	vmlinux-syms-2.6.16.13	page_fault
23	0.1448	vmlinux-syms-2.6.16.13	default_idle
21	0.1322	vmlinux-syms-2.6.16.13	e1000_update_stats
15	0.0944	libc-2.3.6.so	__gconv_transform_utf8_internal
12	0.0756	libc-2.3.6.so	mbrtowc
11	0.0693	vmlinux-syms-2.6.16.13	ide_inb
11	0.0693	vmlinux-syms-2.6.16.13	sysenter_past_esp
9	0.0567	libcrypto.so.0.9.7f	(no symbols)
6	0.0378	vmlinux-syms-2.6.16.13	do_wp_page
5	0.0315	libc-2.3.6.so	_int_malloc
5	0.0315	vmlinux-syms-2.6.16.13	apic_timer_interrupt
5	0.0315	vmlinux-syms-2.6.16.13	irq_entries_start
4	0.0252	libc-2.3.6.so	_IO_default_xsputn
4	0.0252	libc-2.3.6.so	_IO_file_overflow@@GLIBC_2.1
4	0.0252	vmlinux-syms-2.6.16.13	__copy_to_user_ll
4	0.0252	vmlinux-syms-2.6.16.13	__d_lookup
4	0.0252	vmlinux-syms-2.6.16.13	i8042_interrupt
4	0.0252	vmlinux-syms-2.6.16.13	vfs_write
3	0.0189	libc-2.3.6.so	_dl_mcount_wrapper_check
3	0.0189	libc-2.3.6.so	malloc

# XenOprofile – System Profiling in Xen



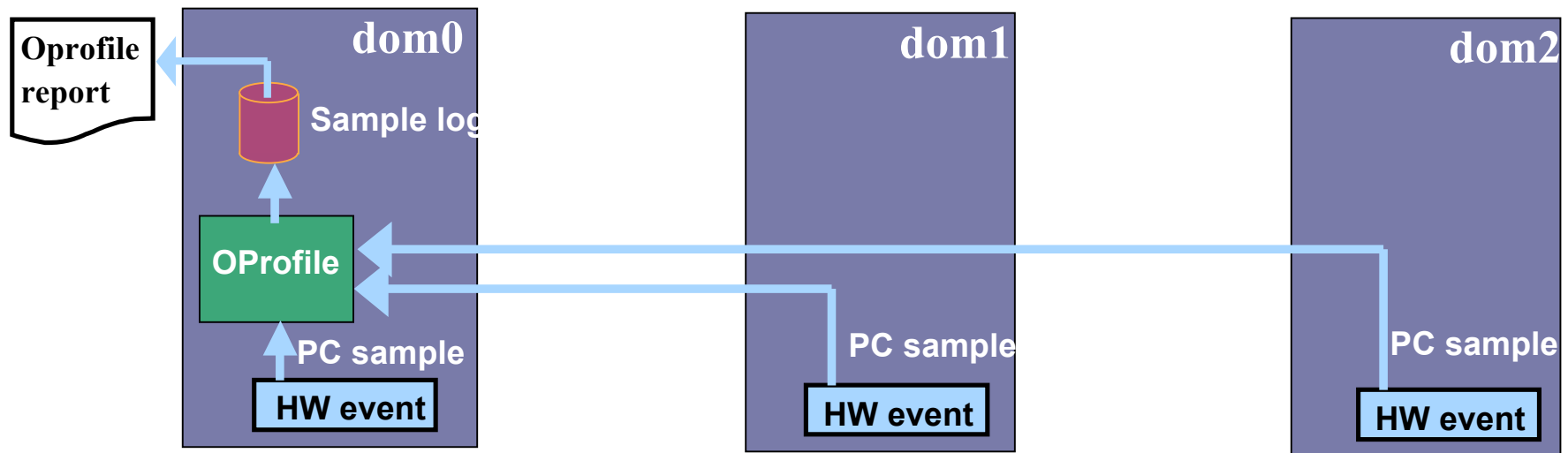
- Extensions to Xen and OProfile for enabling profiling in Xen
- System wide profile:
  - Full profile across multiple domains
  - Covering code in user processes, kernel & Xen (including interrupts, etc.).
- Domain roles in a profiling session
  - Initiator (dom0): Coordinates the session
  - Active domain:
    - Run an instance of Oprofile
    - Process and store its own OProfile samples
  - Passive domain:
    - No OProfile instance running on domain
    - Its samples are processed by initiator (dom0)

# Active Domain Profiling



- Complete detailed profile for each guest (xen, kernel & user)
- Requires domain coordination.
  - OProfile commands in dom0 and active domain must follow a given sequence
- Each domain generate its own profiling report.

# Passive Domain Profiling



- Does not require OProfile in guest
  - Useful if using other OS'es, e.g. Windows
- Function level profiling only for Xen & kernel (Linux)
  - No detailed profile (functions) for user processes and kernel modules
- Easier to use (No domain coordination)
- Single aggregate OProfile report

# Current Status



- Integrated in current unstable tree (first stable release: Xen 3.0.3)
- Current supported architectures:
  - X86 and X86-64 (both Intel and AMD cpu models)
- HVM guest support
  - Final stages of tests for AMD SVM (thanks to help from Tom Woller and Ray Bryant from AMD).
  - Need small fix on user level tools to deal with address overlap in Xen and kernel
- Passive domain support added (thanks to Xiaowei Yang, Intel)
- OProfile integration
  - First version accepted & merged into OProfile CVS tree
  - New version (for passive domain) needs some clean up before submission
  - Date for next OProfile release(including Xen support) is not determined yet
  - Xen patch for OProfile 0.9.1 available in <http://xenoprof.sourceforge.net>

# Further Work (short term)



- Fix known limitations/bugs
  - Cannot profile domain in active mode after profiling it in passive mode and vice-versa.
  - Passive buffer flushed only on dom0 sample (possible overflow/unlikely)
  - Export statistics to user level (buffer overflow, total samples, etc.)
  - Identify source of anonymous samples for user level applications
- OProfile modifications cleanup and merge
  - Current representation of passive domain in sample files not ideal.
    - needs to create multiple symbolic links representing Xen/kernel samples
    - Better approach would use OProfile “–separate” feature.
- Support for other architectures
  - IA64, PPC. (simple for someone familiar with architecture - port from linux)
- Support for Intel VT
  - Need validation of fixes for SVM on VT.
- Support for call-trace profile
  - Work in progress by Intel.

# Further Work (long term)



- Performance counter virtualization
  - Enable guests to run any tool that access performance counters (including OProfile) in guest context
  - Performance counter read/write is expensive (~1000 cycles)
    - Need some form of lazy save/restore on context switch
- Performance counter access from privileged domain
  - System wide instrumentation for other tools in addition to OProfile
- Physical CPU awareness in OProfile
  - OProfile can provide individual profiles for each CPU (VCPU in Xen)
  - capability to separate profiles based on physical CPU may also be useful
  - Combining physical and virtual CPU profile could give insights into effects of Xen CPU scheduling.



i n v e n t