

# Linux Plumbers Conference

Richmond, Virginia | November 13-15, 2023



Linux  
Plumbers  
Conference | Richmond, VA | Nov. 13-15, 2023

# Taming the Incoherent Cache Issue in Confidential VMs

Jacky Li <jackyli@google.com>

Mingwei Zhang <mizhang@google.com>





- Problem Statement
  - Incoherent cache lines
  - Performance degradation
- Solution: Selective Cache Flushing
  - MMU Notifier
    - Introduction
    - filtering the reason



## Incoherent cache lines

- **C-bit**: mark whether a memory page is encrypted.

C-bit

0	0	0	...	010	00010
---	---	---	-----	-----	-------

1	0	0	...	010	00010
---	---	---	-----	-----	-------

|                      Cache Tag                      | Set Index |      Offset      |

- Memory Management (kernel) **recognizes** the C-bit
- Cache (hardware) **doesn't know** about C-bit







## Incoherent cache lines

	Tag						
T0: Cache Line 1	1	0	0	...	0	0	Data Block
T1: Cache Line 2	0	0	0	...	0	0	Data Block

- CVM releases the page -> non-CVM gets the same page
  - 2 **conflicting** cache lines => Data Corruption
  - Solution (2017): flush cache[1]



## Performance degradation

- SME\_COHERENT (2020): CPU cache recognized the C-bit so no need to flush. [2]
  - Vulnerability CVE-2022-0171.
    -  CPU => CPU
    -  CPU => DMA devices
  - Solution: cache flush in mmu notifier when the page leaves CVM [3]
    - Perf impact [4]

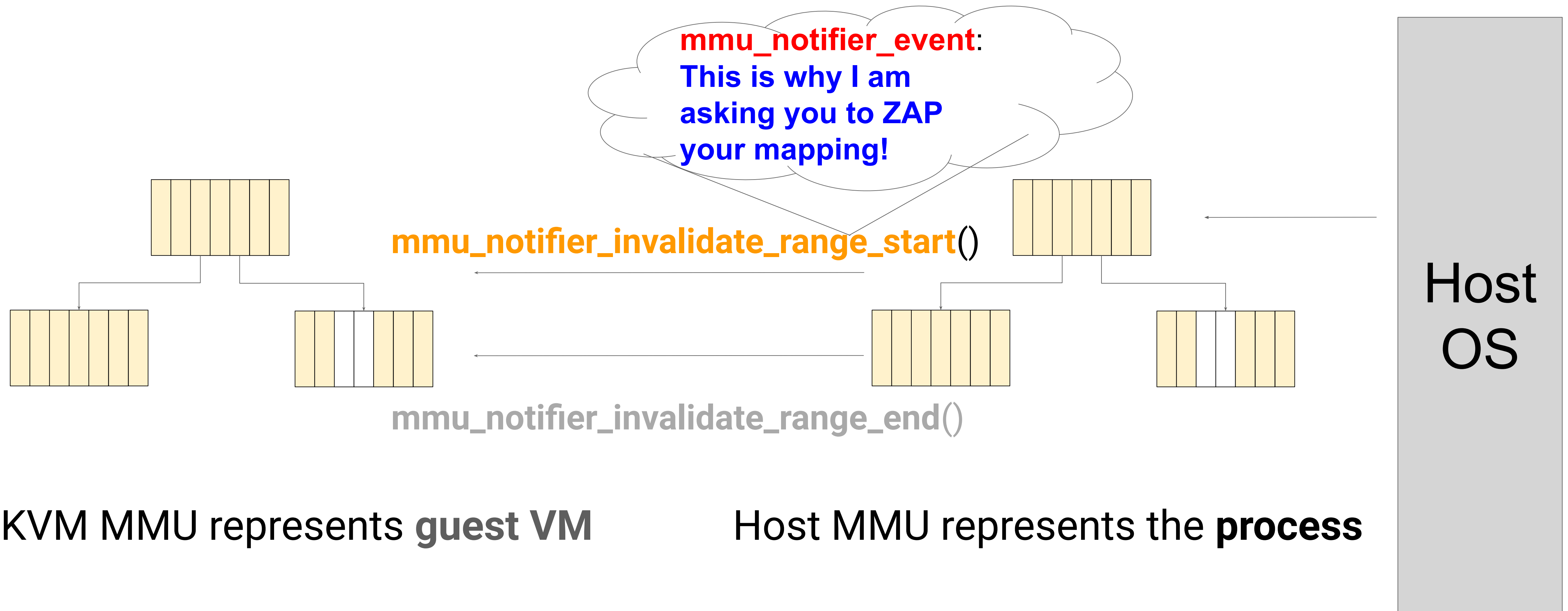


## Solution: Selective Cache Flushing

- Cache Flush **Only When** VM deallocate memory?
  - We are trying to...
  - KVM MMU does not manage memory...
  - We have to do it at MMU\_NOTIFIER
- Cache Flush in **SMALLER** Granularity?
  - We wish...
  - Non-trivial changes on KVM.
  - Non-trivial changes on MM.

**Note: this problem will only affect SEV/SEV-ES**

# MMU Notifiers: reasons



KVM MMU represents **guest VM**

Host MMU represents the **process**

**MMU notifier invalidation does contain a reason parameter which is unused currently.**





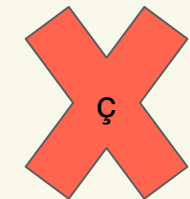
## MMU Notifier: filtering the reason



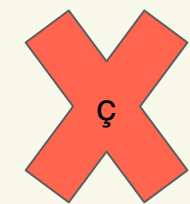
- **MMU\_NOTIFY\_UNMAP,**



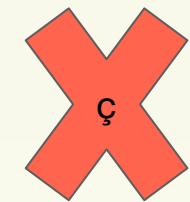
- **MMU\_NOTIFY\_CLEAR,**



- MMU\_NOTIFY\_PROTECTION\_VMA,



- MMU\_NOTIFY\_PROTECTION\_PAGE,



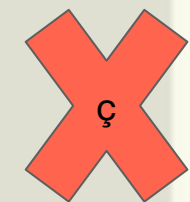
- MMU\_NOTIFY\_SOFT\_DIRTY,



- **MMU\_NOTIFY\_RELEASE,**



- **MMU\_NOTIFY\_MIGRATE,**



- MMU\_NOTIFY\_EXCLUSIVE,

munmap?

madvise?  
migrate?

process  
died?

NUMA  
balancing?



## MMU Notifier: filtering the reason

- Flush cache selectively on mmu\_notifier is **the most cost effective approach** with minimum changes to KVM
- We have had discussions with AMD about addressing this issue in future HW



Linux  
Plumbers  
Conference | Richmond, VA | Nov. 13-15, 2023

# Thank You! Q&A

Jacky Li <jackyli@google.com>

Mingwei Zhang <mizhang@google.com>





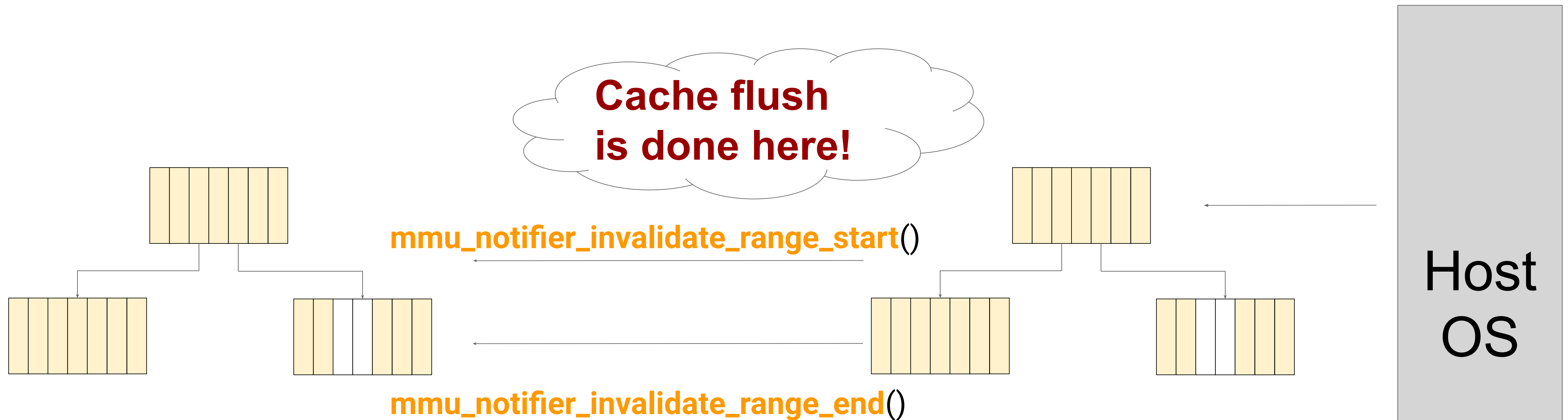
## Appendix:

- [1] 89c505809052 (“KVM: SVM: Add support for KVM\_SEV\_LAUNCH\_UPDATE\_DATA command”)
- [2] e1ebb2b49048 (“KVM: SVM: Don't flush cache if hardware enforces cache coherency across encryption domains”)
- [3] 683412ccf612 (“KVM: SEV: add cache flush to solve SEV cache incoherency issues”)
- [4]  
<https://lore.kernel.org/kvm/YzJFvWPb1syXcVQm@google.com/T/#mb79712b3d141cabb166b504984f6058b01e30c63>





# MMU Notifiers: reasons



KVM MMU represents **guest VM**

Host MMU represents the **process**

**MMU notifier is the memory reclaim interface between KVM and host MM**



## VM\_PAGEFLUSH: limited functionality

- MSR\_AMD64\_VM\_PAGE\_FLUSH (0xc001011e)
  - CPUID level 0x80000001f (EAX), bit 2
  - X86\_FEATURE\_VM\_PAGE\_FLUSH
  - Available on AMD EPYC v1 and later
- **VM\_PAGEFLUSH MSR does not work on user addresses**
  - Even if we disable SMAP (EFLAGS.AC)
  - **AMD APM updated on this at the end of 2022**