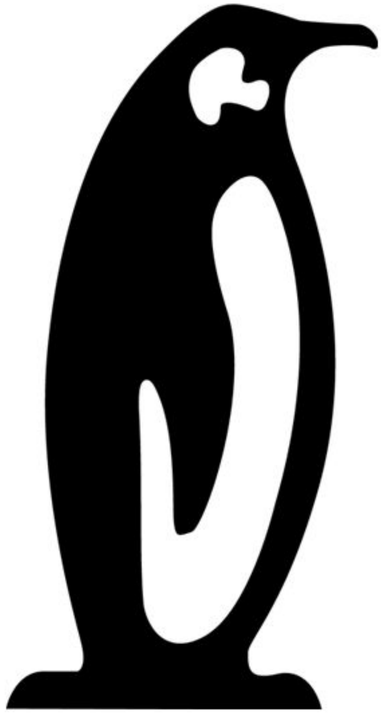

Linux Plumbers Conference

Vienna, Austria | September 18-20, 2024

Guest private memory for software-based hypervisors

Fuad Tabba (*tabba@google.com*)
Elliot Berman (*quic_eberman@quicinc.com*)

LINUX PLUMBERS CONFERENCE | Vienna, Austria
Sept. 18-20, 2024

Since LPC 2023…

Discussions on guest_memfd support for:

- Mapping into host userspace
- Huge pages
- Removing guest memory from the host kernel's direct map
- Page migration and compaction

Most of this was discussed in the Linux MM Alignment Session on July 10 2024.

Refactor guest_memfd as a library

Abstract core-mm decisions about managing folios associated with the file

Provide an easier way to reason about memory in guest_memfd:

- KVM supports multiple confidential computing implementations

Provide a common implementation for other hypervisors (e.g., Gunyah) to use

Patch series: [PATCH RFC v2 0/5] mm: Introduce guest_memfd library

Support `mmap()` of guest_memfd into host userspace

Gunyah and pKVM protect host/guests using Stage-2 (EPT) page tables (software-based)

Guest memory is not encrypted

In-place shared ⇔ private conversion is a requirement:

- Ability to mmap() and GUP *only* memory *shared* with the host
- Private memory should *never* have valid mapping at the host userspace

Patches:

- KVM: Restricted mapping of guest_memfd at the host and pKVM/arm64 support
- mm: guest_memfd: Add ability for userspace to mmap pages
- 1G page support for guest_memfd

Ongoing issues: Tracking host mappings

Does the host have any valid mappings of guest memory?

To maintain the invariant that private pages should not have valid mappings

- `folio_mapped() + folio_maybe_dma_pinned()` not sufficient
- Compare against a "safe" refcount of the folio
  ⇒ A "safe" refcounts accounts for the known references held.

Ongoing issues: Tracking if memory is shared with host

Is the host allowed to access guest memory?

To maintain the invariant that private pages should not be mappable

- The hypervisor (Stage-2/EPT) protects the guest, but errant accesses could crash the host
- Cannot trust the userspace-toggleable `PRIVATE` memory attribute

Dankeschön!

Ongoing issues:

How do know if the host has valid mappings of guest_memfd()?

Where to track if memory is shared with host, i.e., host can legally access it?