Live migration of confidential guests

Ashish Kalra
SMTS Software System Design Engineer, AMD
September 2021

Live migration of confidential guests

- AMD SEV encrypts the memory of VMs , the hypervisor will not be able to simply copy ciphertext between machines to migrate a VM.
- Instead, AMD SEV key management API provides a set of functions which the hypervisor can use to package guest encrypted pages for migration, by maintaining the confidentiality provided by AMD SEV.
- Source VMs guest memory is decrypted with GCTX.VEK & then encrypted with GCTX.TEK with SEND_UPDATE_DATA command, and on the target VM, ciphertext data is decrypted with GCTX.TEK & then re-encrypted with GCTX.VEK of the target VM and written to guest memory with RECEIVE_UPDATE_DATA command.

Two proposals for Live Migration

- AMD Secure Processor (PSP) based migration, which uses AMD Secure Processor to export/import pages wrapped with a transport key. This is Slow Migration.

- In-guest migration or guest assisted migration. Fast Migration of confidential guests using an in-guest migration helper (MH) that is implemented as part of VM's firmware in OVMF. The MH runs in a separate mirror VM.

State of patches:

Mailing lists the live migration discussions/patches are posted on (with links and references to the latest versions of all the relevant patch-sets)

Hypervisor/host Linux Kernel patch merged in kernel v5.14

Guest kernel and guest API patches v6 posted upstream on 8/24/21

- https://lore.kernel.org/kvm/cover.1629726117.git.ashish.kalra@amd.com/

OVMF patches v7 posted upstream on 8/19/21

- https://edk2.groups.io/g/devel/message/79573?p=%2C%2C%2C20%2C0%2C0%2C0%3A%3Acreated%2C0%2Cashish+kalra%2C20%2C2%2C0%2C84997450

QEMU patches v4 posted upstream on 8/4/21

- https://lore.kernel.org/qemu-devel/cover.1628076205.git.ashish.kalra@amd.com/

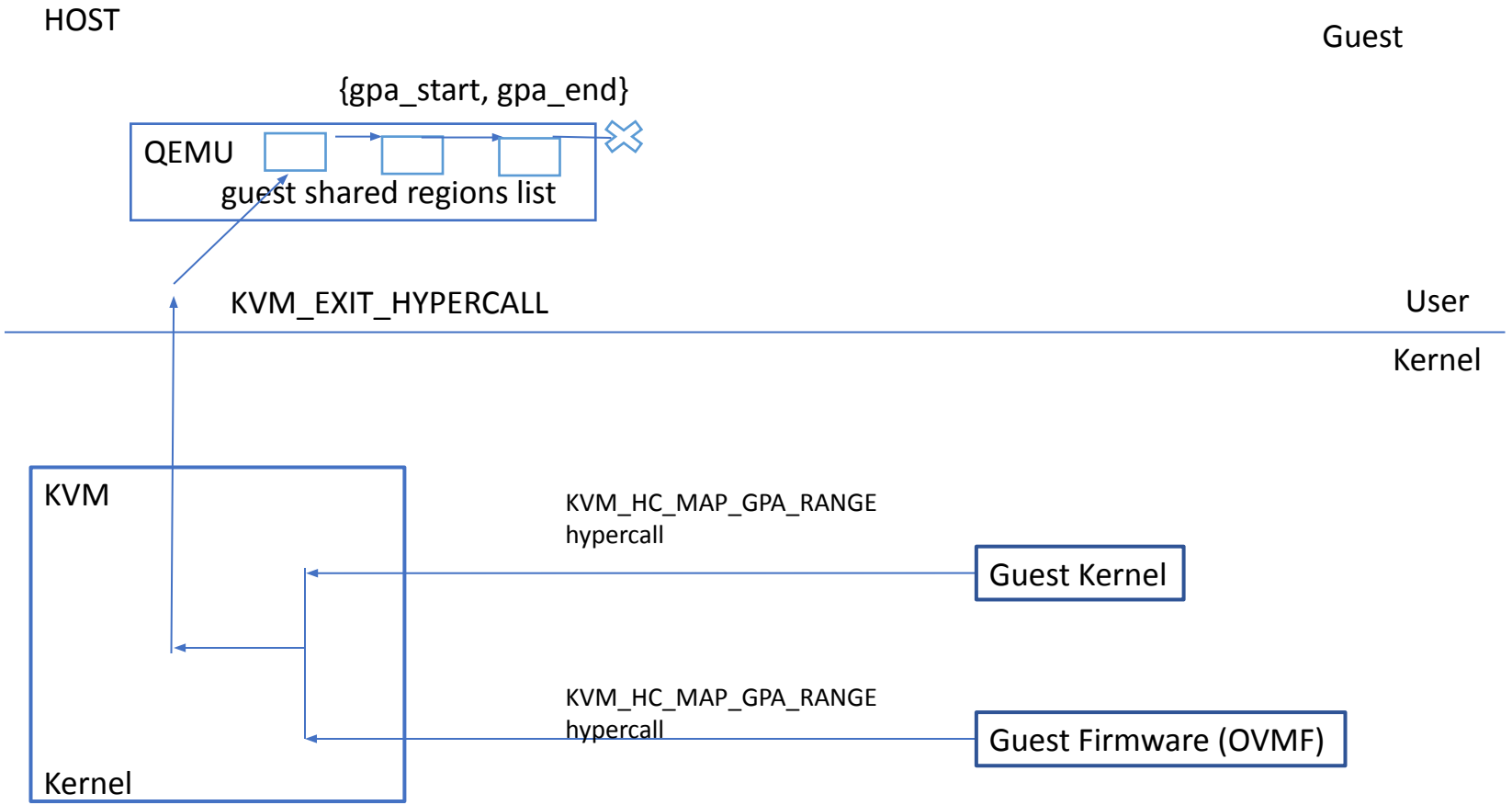
Guest assisted migration (Fast Migration) :
RFC patches for mirror VM posted upstream. Lot of related discussion going on KVM/QEMU mailing lists about alternative approaches, security issues, etc.

https://lore.kernel.org/qemu-devel/cover.1629118207.git.ashish.kalra@amd.com/

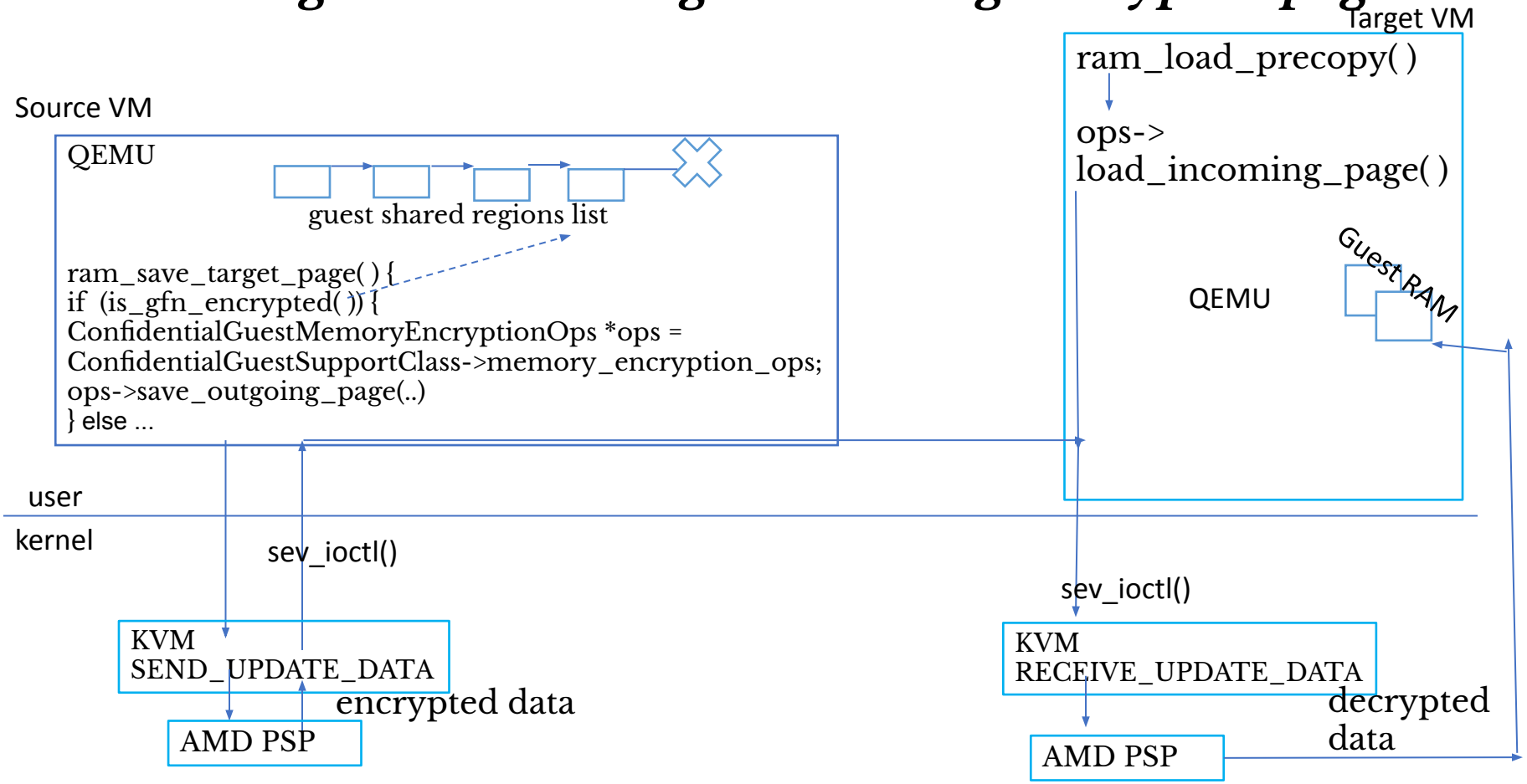
https://lore.kernel.org/qemu-devel/20210823141636.65975-1-dovmurik@linux.ibm.com/

https://edk2.groups.io/g/devel/message/79517?p=%2C%2C%2C20%2C0%2C0%2C0%3A%3Acreated%2C0%2Ctobin%2C20%2C2%2C0%2C84982978

Guest Page encryption status tracking



RAM Migration: Sending & receiving encrypted pages



Guest Kernel & guest API support

- Pre-alternatives hypercalls invoked to mark the "__bss_decrypted" section, per-cpu GHCB pages (SEV-ES) and per-cpu apf-reason, steal-time & kvm_apic_eoi as decrypted.

- apply-alternatives() called much later during setup_arch(), so we need of an early, pre-alternatives hypercall interface.

- All early hypercalls made via early_set_memory_decrypted() / encrypted() interfaces, which in turn invoke paravirt_ops (pv_ops).

- early_set_memory_XX()-> pv_ops.mmu.notify_page_enc_status_changed()

SEV/TDX specific hypercall

Continued….

- Guest support for detecting & enabling live migration feature vs. the following logic:

❑ kvm_init_platform() checks if it is booted under EFI

❑ If not EFI,

i) if kvm_para_has_feature(KVM_FEATURE_MIGRATION_CONTROL) issue a wrmsrl (MSR_KVM_MIGRATION_CONTROL) to enable SEV live migration support.

❑ If EFI,

i) If kvm_para_has_feature (KVM_FEATURE_MIGRATION_CONTROL) query “SevLiveMigrationEnabled” UEFI runtime variable.

ii) The variable indicates live migration support is enabled on Host & guest firmware, issue wrmsrl (MSR_KVM_MIGRATION_CONTROL) to indicate all three components support & have enabled live migration feature.

QEMU Support for Live Migration

- To protect confidentiality of data while in transit need to add platform specific hooks to save or migrate guest RAM.
- Introduce new “ConfidentialGuestMemoryEncryptionOps” which will be used during encrypted guest migration.

```
typedef struct ConfidentialGuestMemoryEncryptionOps {

    /* Initialize the platform platform specific state before staring migration */
    int (*save_setup) (MigrationParameters *p);

    /* Write the encrypted page and metadata associated with it */
    int (*save_outgoing_page) (QEMUFile *..., unit8_t* ptr);

    /* Check if gfn is in shared/unencrypted region */
    bool (*is_gfn_in_shared_region) (unsigned long gfn);

    /* Save the shared regions list */
    int (*save_outgoing_shared_regions_list) (QEMUFile *...);
```

Continued….

```
        /* Load the shared regions list */
        int (*load_incoming_shared_regions_list) (QEMUFile *...);

};

typedef struct ConfidentialGuestSupportClass {
   Object Class parent;

+ConfidentialGuestMemoryEncryptionOps *memory_encryption_ops);

}
```

edk2/ovmf support

- The patch-set detects if it is running under KVM hypervisor & then checks for SEV migration feature support via KVM_FEATURE_CPUID, if detected, it sets up a new UEFI Runtime variable to indicate OVMF support for SEV live migration.

- This is part of a 3-way negotiation of the live migration feature between hypervisor, guest firmware and guest kernel.