s390 (IBM Z) Boot/IPL of Protected VMs

Summary

The memory of Protected Virtual Machines (PVMs) is not accessible to I/O or the hypervisor. In those cases where the hypervisor needs to access the memory of a PVM, that memory must be made accessible. Memory made accessible to the hypervisor will be encrypted. See Documentation/virt/kvm/s390-pv.rst for details."

On IPL (boot) a small plaintext bootloader is started, which provides information about the encrypted components and necessary metadata to KVM to decrypt the protected virtual machine.

Based on this data, KVM will make the protected virtual machine known to the Ultravisor (UV) and instruct it to secure the memory of the PVM, decrypt the components and verify the data and address list hashes, to ensure integrity. Afterwards KVM can run the PVM via the SIE instruction which the UV will intercept and execute on KVM's behalf.

As the guest image is just like an opaque kernel image that does the switch into PV mode itself, the user can load encrypted guest executables and data via every available method (network, dasd, scsi, direct kernel, ...) without the need to change the boot process.

Diag308

This diagnose instruction is the basic mechanism to handle IPL and related operations for virtual machines. The VM can set and retrieve IPL information blocks, that specify the IPL method/devices and request VM memory and subsystem resets, as well as IPLs.

For PVMs this concept has been extended with new subcodes:

Subcode 8: Set an IPL Information Block of type 5 (information block for PVMs) Subcode 9: Store the saved block in guest memory Subcode 10: Move into Protected Virtualization mode

The new PV load-device-specific-parameters field specifies all data that is necessary to move into PV mode.

- PV Header origin
- PV Header length
- List of Components composed of
    - AES-XTS Tweak prefix
    - Origin
    - Size

The PV header contains the keys and hashes, which the UV will use to decrypt and verify the PV, as well as control flags and a start PSW.

The components are for instance an encrypted kernel, kernel parameters and initrd. The components are decrypted by the UV.

After the initial import of the encrypted data, all defined pages will contain the guest content. All non-specified pages will start out as zero pages on first access.

When running in protected virtualization mode, some subcodes will result in exceptions or return error codes.

Subcodes 4 and 7, which specify operations that do not clear the guest memory, will result in specification exceptions. This is because the UV will clear all memory when a secure VM is removed, and therefore non-clearing IPL subcodes are not allowed.

Subcodes 8, 9, 10 will result in specification exceptions. Re-IPL into a protected mode is only possible via a detour into non protected mode.

Keys

Every CEC will have a unique public key to enable tooling to build encrypted images. See s390-tools for the tooling.