

XICS interrupt controller

Device type supported: KVM_DEV_TYPE_XICS

Groups:

1. KVM_DEV_XICS_GRP_SOURCES

Attributes:

One per interrupt source, indexed by the source number.

2. KVM_DEV_XICS_GRP_CTRL

Attributes:

2.1 KVM_DEV_XICS_NR_SERVERS (write only)

The `kvm_device_attr.addr` points to a `__u32` value which is the number of interrupt server numbers (ie, highest possible vcpu id plus one).

Errors:

-EINVAL	Value greater than KVM_MAX_VCPU_IDS.
-EFAULT	Invalid user pointer for attr->addr.
-EBUSY	A vcpu is already connected to the device.

This device emulates the XICS (eXternal Interrupt Controller Specification) defined in PAPR. The XICS has a set of interrupt sources, each identified by a 20-bit source number, and a set of Interrupt Control Presentation (ICP) entities, also called "servers", each associated with a virtual CPU.

The ICP entities are created by enabling the KVM_CAP_IRQ_ARCH capability for each vcpu, specifying KVM_CAP_IRQ_XICS in `args[0]` and the interrupt server number (i.e. the vcpu number from the XICS's point of view) in `args[1]` of the `kvm_enable_cap` struct. Each ICP has 64 bits of state which can be read and written using the KVM_GET_ONE_REG and KVM_SET_ONE_REG ioctls on the vcpu. The 64 bit state word has the following bitfields, starting at the least-significant end of the word:

- Unused, 16 bits
- Pending interrupt priority, 8 bits Zero is the highest priority, 255 means no interrupt is pending.
- Pending IPI (inter-processor interrupt) priority, 8 bits Zero is the highest priority, 255 means no IPI is pending.
- Pending interrupt source number, 24 bits Zero means no interrupt pending, 2 means an IPI is pending
- Current processor priority, 8 bits Zero is the highest priority, meaning no interrupts can be delivered, and 255 is the lowest priority.

Each source has 64 bits of state that can be read and written using the KVM_GET_DEVICE_ATTR and KVM_SET_DEVICE_ATTR ioctls, specifying the KVM_DEV_XICS_GRP_SOURCES attribute group, with the attribute number being the interrupt source number. The 64 bit state word has the following bitfields, starting from the least-significant end of the word:

- Destination (server number), 32 bits

This specifies where the interrupt should be sent, and is the interrupt server number specified for the destination vcpu.

- Priority, 8 bits

This is the priority specified for this interrupt source, where 0 is the highest priority and 255 is the lowest. An interrupt with a priority of 255 will never be delivered.

- Level sensitive flag, 1 bit

This bit is 1 for a level-sensitive interrupt source, or 0 for edge-sensitive (or MSI).

- Masked flag, 1 bit

This bit is set to 1 if the interrupt is masked (cannot be delivered regardless of its priority), for example by the `ibm,int-off` RTAS call, or 0 if it is not masked.

- Pending flag, 1 bit

This bit is 1 if the source has a pending interrupt, otherwise 0.

Only one XICS instance may be created per VM.