GPU Virtualization Methods

By Erik Gustafson and Drew Tosaya
A Quick History.....

....of Virtualization

....Of GPUs/Graphic Cards
What is Virtualization?

What are Hypervisors?

Hypervisor types

• Type 1
• Type 2

Examples

• Virtual Box
• VMware
• Xen
How does a GPU work?

Why do GPUs exist?

What resources does it have?

CPU and GPU interaction
The Virtualization Spectrum

- Direct Pass Through
- Mediated Pass Through
- API Forwarding
- Full Emulation
Direct Passthrough

Fastest Of All Options

Spotty support, some additional support required

Highly Secure

Most Expensive - by far
Mediated Pass Through

2nd Fastest, about \( \frac{1}{2} \) of baremetal, \( \frac{2}{3} \) of direct pass-through

Tons of back end code required for implementation

Highly Secure **IF** done correctly

Much cheaper the direct pass through, about the same cost as rest
API Forwarding

3rd Fastest, but you are looking at a fraction of bare metal

Back end code required for implementation

Once it’s written it’s mostly plug and play.

Generally Secure - depends on implementation

Much cheaper the direct pass through, about the same cost as rest
Full Emulation

4th Fastest, but you are looking at a (smaller) fraction of bare metal

Tons back end code required for implementation

Once it’s written it’s mostly plug and play.

Generally Secure - depends on implementation

Much cheaper the direct pass through, about the same cost as rest
Example Uses/Projects

- Intel-GVT-g (XenGT and KVMGT)
- VirGL
- VMWare 3D acceleration
- Virtual Box 3D acceleration+
What is The End Game?

Nvidia Grid

SecureView

Other consumer grade GPU virtualization
Nvidia Grid
My Personal Favorite Projects

XenGT

VirGL
References


https://www.slideshare.net/guestb3fc97/gpu-virtualization-on-vmwares-hosted-io-architecture-presentation

https://www.kraxel.org/slides/qemu-gfx/#gl-rendering


Questions?