

---

# MSE ClComp SEC 2 – Security for Clouds

Martin Gwerder

Marcel Graf

# Table of Content

---

- Introduction
- Recaps
- Known classes of attacks to virtualized infrastructures
  - Denial of Service
  - Privilege Escalation / Boundary Crossing
- Generic good practices to counter attacks to virtualisation

# Important note

---

- The focus of this lesson is not on security in general. All virtual machines remain as single entities as vulnerable as in the physical world. It focusses on attacks which are either specific to virtualisation or have “normal” attacks having a different (higher) impact on virtualisation systems.
- This does not mean that virtualisation is by design bad or vulnerable! Virtualisation is just complex.

# Some terms

---

- Hypervisor  
“hypervisor layer” of a host itself and the hypervisor management.
- Host  
A physical computer system accommodating a hypervisor.
- Guest  
A virtual machine residing on a host
- Hypervisor management  
Management software controlling the hypervisor layer locally. This is done either in a privileged processor mode, or in a privileged domain of the hypervisor (e.g. dom-0).
- Cluster management  
Management controlling the cross hypervisor functionality (e.g., vSphere or VMM). We always assume a cluster to automatically move VMs (due to rules or node failures). We assume no redundant running of VMs (FT in VMware terms). All virtual machines are freshly started on a new host if a host failure occurs.  
We do assume in this lecture that a failing cluster management leaves the hypervisors as they are (meaning: Virtual machines remain fully functional)

# Known classes of threats specific to virtualisation

## Denial of Service

---

- To achieve a successful Denial of Service an attacker has two main possibilities:
  - Deplete a resource of a system
  - Trap a system in a non-functional state (either temporarily or permanently)
- Virtualisation systems are a very interesting target as there is a high impact if successful.
- Virtualisation systems are very vulnerable due to their complexity (especially when used in clusters). Great efforts are taken to mitigate this.
- Note: Successful DoS attacks on virtualised infrastructures by exhausting resources are relatively rare as the sizing of such platforms normally outperforms their surroundings.

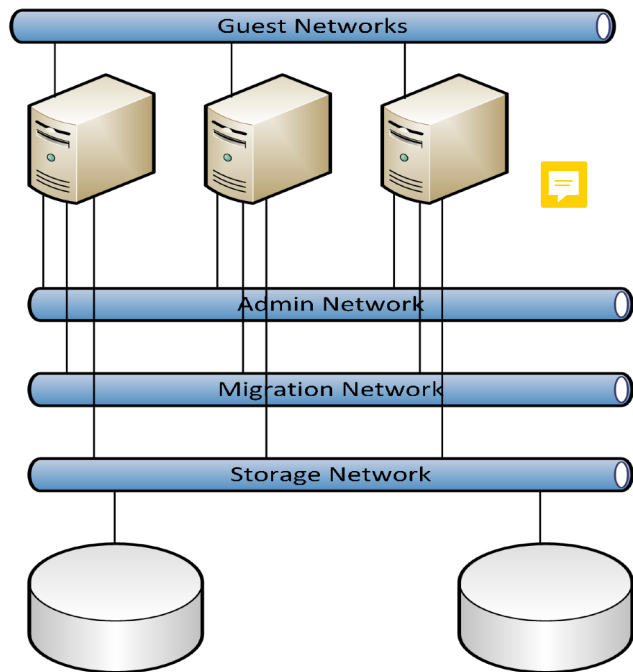
## Denial of Service: Exhausting host resources (e.g., CPU)

---

- Attack: Run one or more virtual guests at “full speed” thus occupying the CPU the whole time.
- Impact: Normally very limited impact as we do have good scheduling on a per-host basis.
- Mitigation:
  - Control over commitment/in-VM-elasticity (virtual machines should not be able to extend drastically if not needed; e.g., do not assign two or four vCPUS/vCore to a simple DNS-Server)
  - Limit impact with (resource-) reservations for vital virtual systems. Please note fat provisioning (as opposed to thin provisioning) of disks is also a reservation in this context.
  - Use a sensible resource allocation algorithm if given a choice. (e.g., VMDq)

# Known classes of threats specific to virtualisation


## Denial of Service: Exhausting cluster resources (e.g., network)



- Important note: A single host is due to the bandwidth-limitations of its internal busses normally unable to fill a 10G+ network (today ~2-4 hosts are usually required to fill 10GB network).
- Host Internal (virtual) networks are normally not bandwidth limiting (although we do have 1GB or 10GB adapter drivers in the VM)
- The easiest way to fill a physical net is triggering multiple live migration while creating high memory loads (assuming load balancing systems).
- Filling guest networks with broadcast traffic is another good way of overwhelming virtual/physical boundaries. This can be exploited by creating a high cross-host traffic (initiated my multiple nodes) congesting the physical network.

## Boundary Crossing / Privilege Escalation: General

---

- All normal attacks crossing boundaries such as buffer overflows belong in this category.
  - They are usually product-specific attacks depending on the guests weaknesses and not specific to virtualisation. We skip them here although they do exist.
- We do outline here attacks capable of crossing VM boundaries.  
You may cross boundaries by exploiting...
  - Hardware (most effective so far but inevitably hardware-dependent)
  - The virtualisation interfaces within a VM (remember full virtualisation vs. paravirtualisation )
  - The administration level of a hypervisor or a virtualisation cluster (whoever controls the host has full control over the guest)

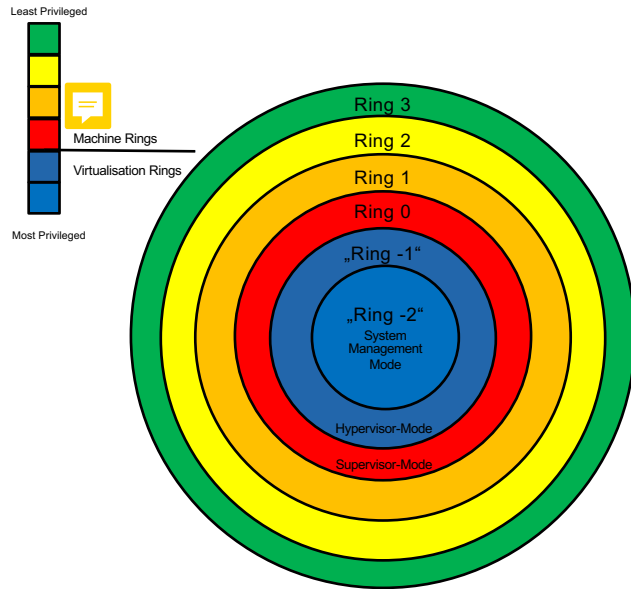


# Recap threats

---

- A system may be compromised at any level. Most interesting spots are out of reach for an OS such as System Management Mode (SMM) or the hardware as it is very hard for a system to detect an attack on this level.
- One of the main goals of the Trusted Platform Module (TPM) is to protect a machine and the OS from malicious software or hardware. Although it is often only seen as brilliant/evil mean to enforce DRM.

# Recap CPU-Rings



- Rings:
  - Ring 0-3 are traditional Machine Rings for Privilege encapsulation of Hardware
    - Ring 3 (may be higher or lower on non-x86-HW) is the Application Mode
    - Ring 0 is called “Supervisory Mode”
  - “Ring -1” is “Hypervisor Mode” (HVM)
  - “Ring -2” is “System Management Mode”
- The placement of the rings below zero is implementation dependent and not on all platforms true.

# Known classes of threats specific to virtualisation


## Boundary Crossing / Privilege Escalation: Management Attacks

---

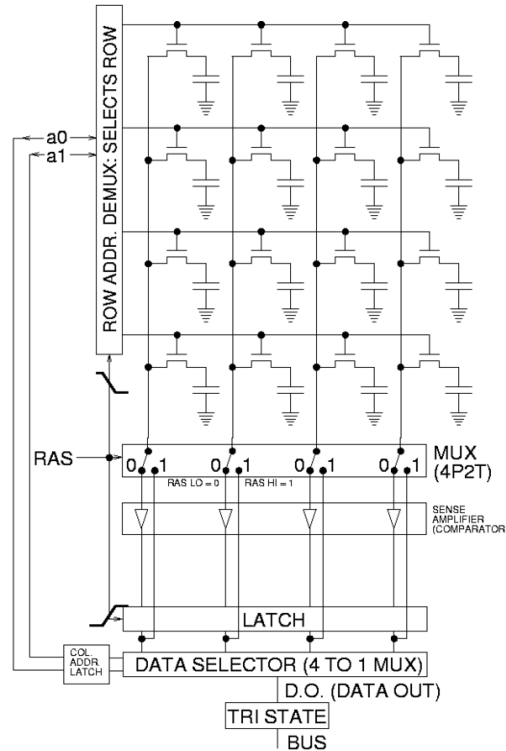
- Attack: Most of the attacks targeting hosts do not attack the hypervisor but its management. Simple attacks include weak passwords or badly secured APIs on the management level of the hypervisor
- Impact: Control over all aspects of the hypervisor including underlying infrastructure such as centralized storage. May pDoS (permanent DoS) all guests and the hypervisor.
- Mitigation:
  - Isolate management networks.
  - Prohibit general access to management networks.
  - Install patches.
  - Use strong creds (preferably certificates)

## Boundary Crossing / Privilege Escalation: Rootkit attacks

---

- Attack: Access virtual machines from a Ring “below ground” by using malicious software. There are known rootkits (e.g. evil maid, ACPI rootkits and Blue Pill) which are attacking on this layer 
- Impact: Free access to data and control (including subsequent systems of the hypervisor)
- Mitigation:
  - Enable trusted platform features for hypervisor (if available)
  - Enable trusted platform features for UEFI/BIOS >(e.g., Secure Boot)
  - Maintain whitelist process list (very weak mitigation)
- These weaknesses are known to be actively exploited by advanced tools.

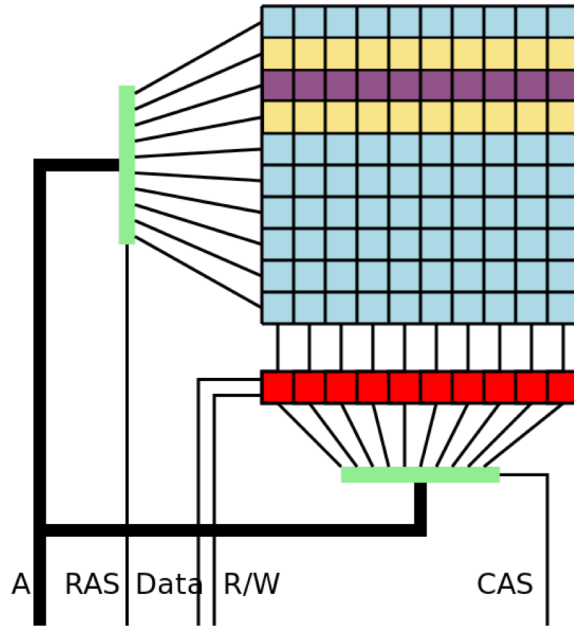
# Special knowledge: DRAM



- DRAMs store bits in very small and densely organised capacitors.
- Load of a capacitor is degraded over time.
- Reading a value normally destroys/degrades it (at first) this is why the read value is fed back to the capacitors to refresh charges while reading. (missing part on image)
- Bits are organized in rows and refreshed on a regular base (JEDEC at least every 64ms).

Source: by [Gloogier via Wikipedia](#); CC BY-SA 3.0

# Special knowledge: Rowhammer



Source: by [Dajmic via Wikipedia](#), CC BY-SA 3.0

- Very old attack based on work in early 1970s actively exploited in 2010+.
- By repeatedly writing adjacent memory rows a „victim row“ may be altered (disturbance errors).
- Approximately 1 of 1700 Bits are flippable.
- ECC-RAM is susceptible as well.
- DRAM3 is often susceptible (80% of tested DIMM modules). DRAM4 is normally not susceptible as they were developed after the first rowhammer attack came up.

## Boundary Crossing / Privilege Escalation: Flip Feng Shui (aka. Rowhammer crossing VM-Borders)

---

- Attack: Applying rowhammer attack to a virtual machine guest or host system to alter one bit of the public key (simplifies factorisation).
- Impact: Controlled RAM writing over VM boundaries possible.
- Mitigation:
  - Use ECC-Ram (proven as \*not\* effective for all cases)
  - Decrease maximum refresh time (increases power consumption and heat; currently recommended is currently at least all 52ms)
  - Enable pTRR (bandwidth reduction 2-4%) or TRR (no bandwidth effects)
  - Use DDR4
- Test code for “rowhammer” available under <https://github.com/google/rowhammer-test>.
- **IMPORTANT: Do not use this on third party hosts! This attack crosses all known CPU/Chipset boundaries if successful!**
- Note: This attack has been shown as academic research but not been proven as actively exploited (yet) in the virtual world. Rowhammer is actively being used in exploits already.

## Boundary Crossing / Privilege Escalation: Flip Feng Shui (how it works)

---

- How the attack works:
  - Search for two adjacent segments which are susceptible for rowhammer in own memory. (Result two segments with rowhammer weakness).
  - Connect to an openSSH server on the same host system. This creates a memory block containing the public of a user (for verification of the user).
  - Create the same block in the own memory in the previously found area.
  - Wait for the memory deduplication to pick the block up. There is a 50% chance that your block is chosen as the master copy. (effectively: sit and wait for a couple of minutes; You may try as many times as you want)
  - Make a rowhammer attack to the public key to flip one bit. This results in a public key which is no longer built from two primes. (you just altered the public key on the remote server)
  - Factorise the new public key and thus create a private key.
  - Login!
- Movie of the attack itself: <https://www.youtube.com/watch?v=TqWmP2owbdo>



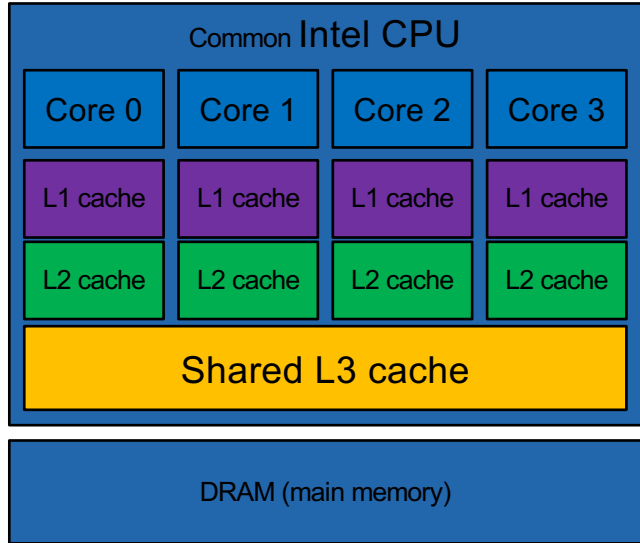
# Known classes of threats specific to virtualisation

## Boundary Crossing / Privilege Escalation: Flush & Reload

---

- Attack: Access read protected memories by exploiting timing of of L3-Cache
- Impact: Memory in other virtual machines can be read and information may be stolen.
- Mitigation:
  - Disable memory deduplication technologies on all layers (e.g. Kernel Samepage Merging).
- This attack is already further developed (e.g., S\$A)
- Test code available under <https://github.com/DanGe42/flush-reload>. This test code only works on HW with memory deduplication enabled.

## Boundary Crossing / Privilege Escalation: Flush & Reload (how it works)



- Cores, L1 and L2 caches are never concurrently shared across VMs
- Every layer is roughly 3-4 times faster than the layer below it
  - CPU cycle ~0.3 ns
  - Access time L1 cache ~4 cycles
  - Access time L2 cache ~10 cycles
  - Access time L3 cache ~70 cycles
  - Memory requires ~200 cycles
- Last level cache is normally shared
- Cached memory may be flushed / invalidated by code (may be required due to DMA access, clflush instruction)





# Generic hardening of a virtualisation Cluster

---

- We must take several measures to harden a virtualisation cluster. The first locations to look are usually:
  - Physical setup
  - Virtual network setup
  - Hypervisor (actually primarily the hypervisor management)
  - Cluster manager

- Separate network categories physically
  - At least separate admin, migration, storage and guest networks (possibly heartbeat as well).
  - No VLANs to combine multiple categories.
- Isolate admin networks from guest networks with advanced methods (such as a jump host to an admin zone. Do not realise the jump host as virtual machine on the same cluster).



- Limit bandwidth available to a single guest (not perfect but in reality helps a lot)
- Do not make admin networks available in guest space 
- Avoid cross-host traffic by defining affinities and anti-affinities in the guest (where possible). 
- Always offer tag-free networks to guests. Never connect them to trunks.

# Generic hardening of a virtualisation Cluster

## Hypervisor management

---

- Encrypt all traffic (includes M2M traffic)
- Make sure that always mutual authentication is applied
- Make sure that there is no connection downgrading possible (e.g. no algorithms/key sizes in cypher list)
- Make sure that there are no default passwords set.
- Make sure that you use certificates instead of passwords wherever possible
- Enable password changes by either syncing local accounts or introduce a centralized authentication and authorisation provider.
- Patch, patch, patch, and if it does not help: patch

# Generic hardening of a virtualisation Cluster Cluster Manager

---

- Same rules as for the hypervisor do apply here
- Do not virtualise cluster managers (and never ever do it on the virtualised platform you are managing)