CloudVisor: Retrofitting Protection of Virtual Machines in Multi-tenant Cloud with Nested Virtualization

Fengzhe Zhang, Jin Chen, Haibo Chen, Binyu Zang

System Research Group
Parallel Processing Institute
Fudan University
http://ppi.fudan.edu.cn/system_research_group

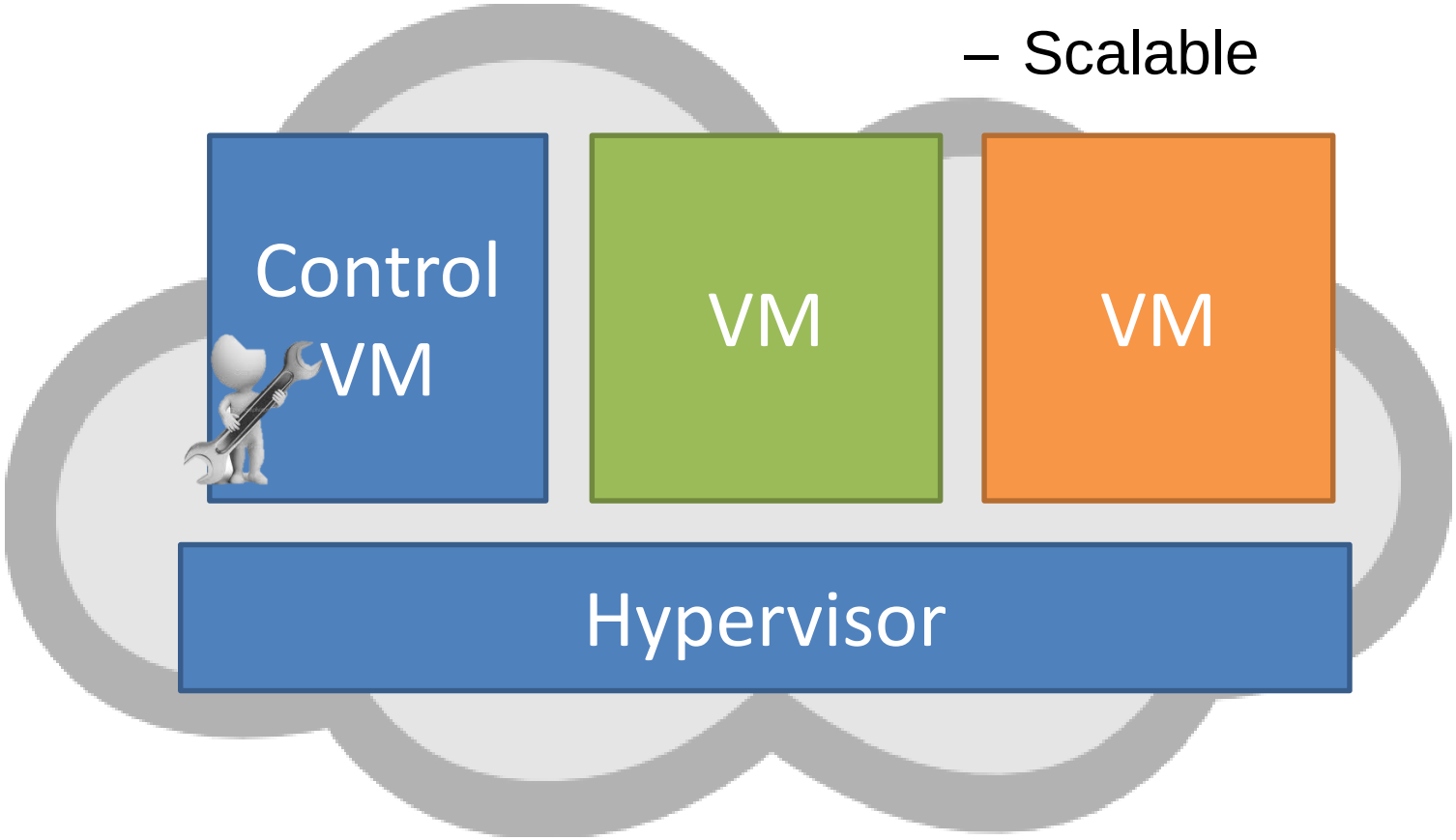
2011-10-24

Multi-tenant Cloud

- Widely available public cloud
  - Amazon EC2, RackSpace, GoGrid
- Infrastructure as a Service
  - Computation resources are rented as ***Virtual Machines***
- To save cost, VMs from different users may run side-by-side on the same platform

Multi-tenant Cloud Software Stack

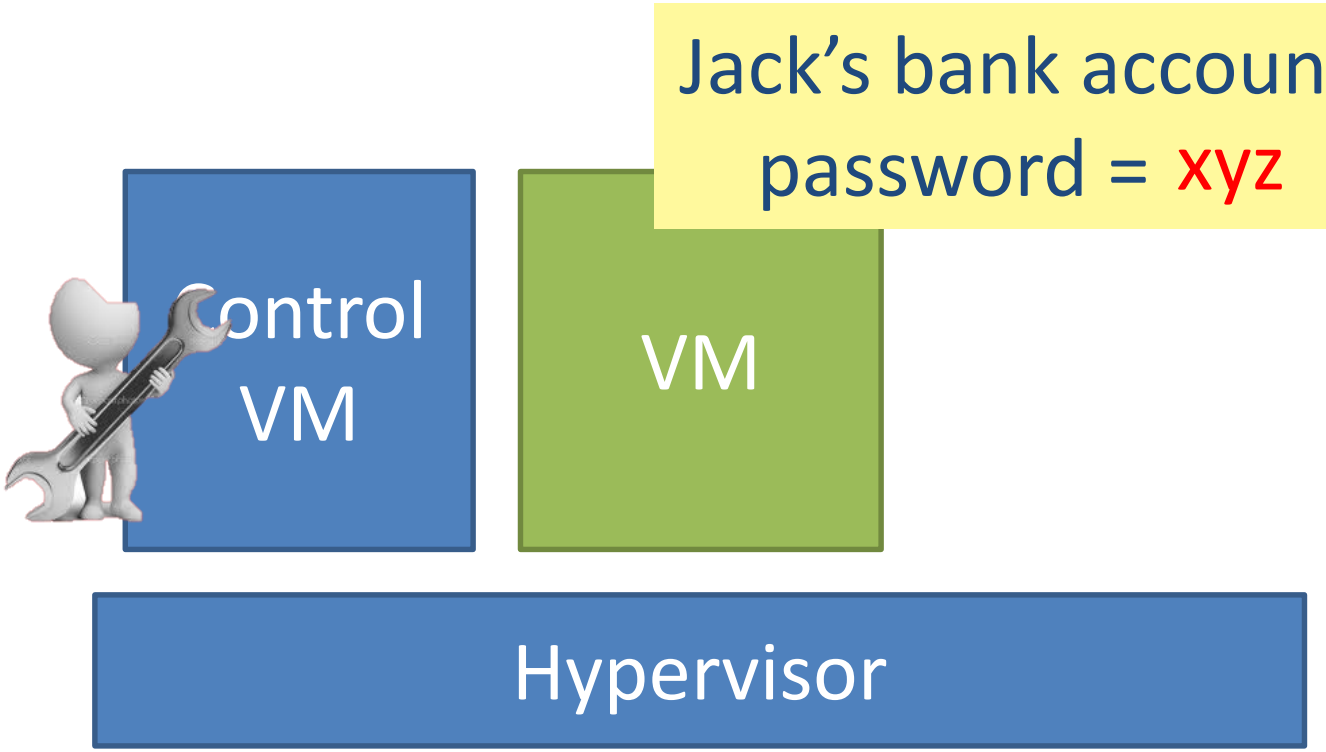
- Pay-as-you-go
  - Flexible
  - Scalable



Can we simply trust public cloud?

Probably Not !

Problem #1: Curious/malicious Administrator



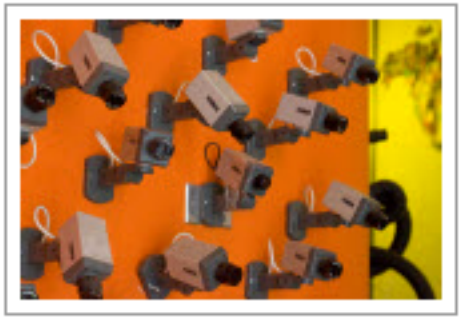
most concerned issue:

"*invisibly access* unencrypted data in its facility"- Gartner, 2008

Problem #1: Curious/malicious Administrator

Google Fires Employee Accused Of Spying On Kids

By Phil Villarreal on September 16, 2010 9:15 AM



(RAWRZ!)

For a Google engineer who was fired in July, it apparently wasn't enough just to Google people in order to stalk them. Instead, he allegedly abused his access and violated the company's privacy policies to snoop on users.

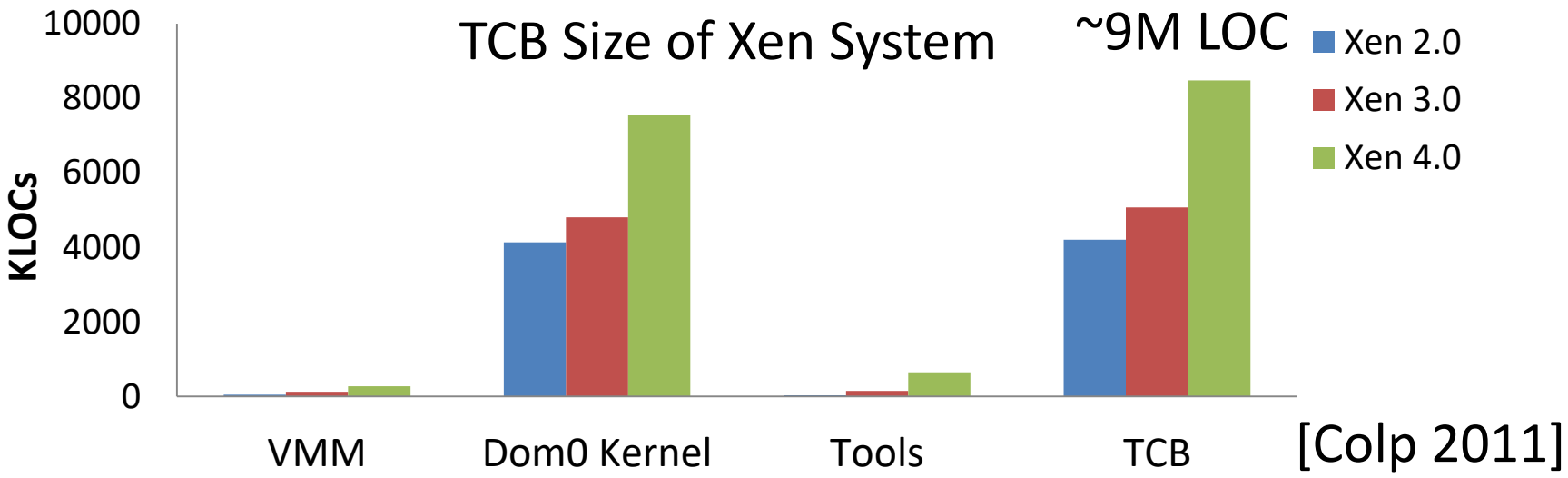
Valleywag reports the man spied on four teenagers, peeking in on emails, chats and Google Talk call logs for several months before the company discovered what was going on.

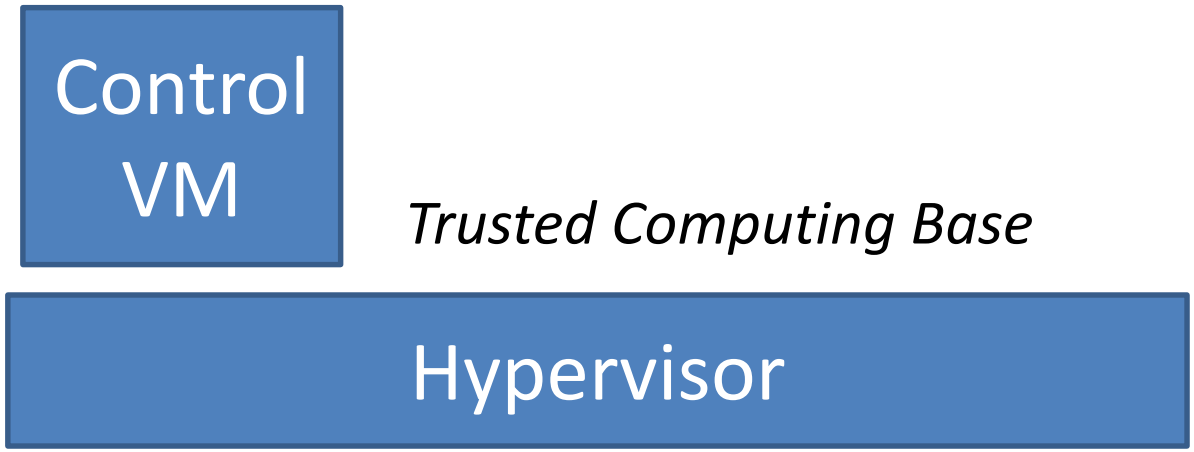
A Google s

peeking in on emails, chats and Google Talk call logs for several months before the company discovered...

"We
num
example, we are significantly increasing the amount of time we spend auditing our logs to ensure those controls are effective. That said, a limited number of people will always need to access these systems if we are to operate them properly-which is why we take any breach so seriously."

Problem #2: Large TCB for Cloud





monolithic virtualization stack

one point of penetration leads to full compromise

Result: Limited Security Guarantees in Public Cloud

Amazon AWS User Agreement, 2010

7.2. [illegible] but cannot guarantee that we will be successful at doing so [illegible] given [illegible]

acknowledge that you bear sole responsibility for adequate security, protection and backup of Your Content and Applications. We strongly encourage you, where available and appropriate, to (a) use encryption technology to protect Your Content from unauthorized access, (b) routinely archive Your Content, and (c) keep your Applications or any software that you use or run with our Services current with the latest security patches or updates. We will have no liability to you for any unauthorized access or use, corruption, deletion, destruction or loss of any of Your Content or Applications.

Microsoft Windows® Azure™ Platform Privacy Statement, Mar 2011

Security of Your Information

Microsoft is committed to protecting the security of your information. We maintain technical and organizational measures designed to provide and enable security for the Service. This includes a variety of security technologies and procedures to help protect your information from unauthorized access, use, or disclosure. For example, we store the information you provide on computer systems with limited access, which are located in controlled facilities.

Some personal information may be particularly sensitive to you or your organization, and hence may require a level of security that we do not provide. You or your organization is responsible for determining whether our security meets your requirements.

For more information, please see this Security Overview. If you have specific questions, please contact Support as described below under the Support Services section.

Data Encryption is not Enough

- Encryption is only good for *static data storage*
  - Data never decrypted in the cloud
  - Cloud is just used as online storage space

- As for computation cloud
  - Data are involved in computation, such as *web services*
  - Data should be decrypted during computation
  - Encryption is ***not enough*** in this case
  - Note, computation cloud is more widely desired

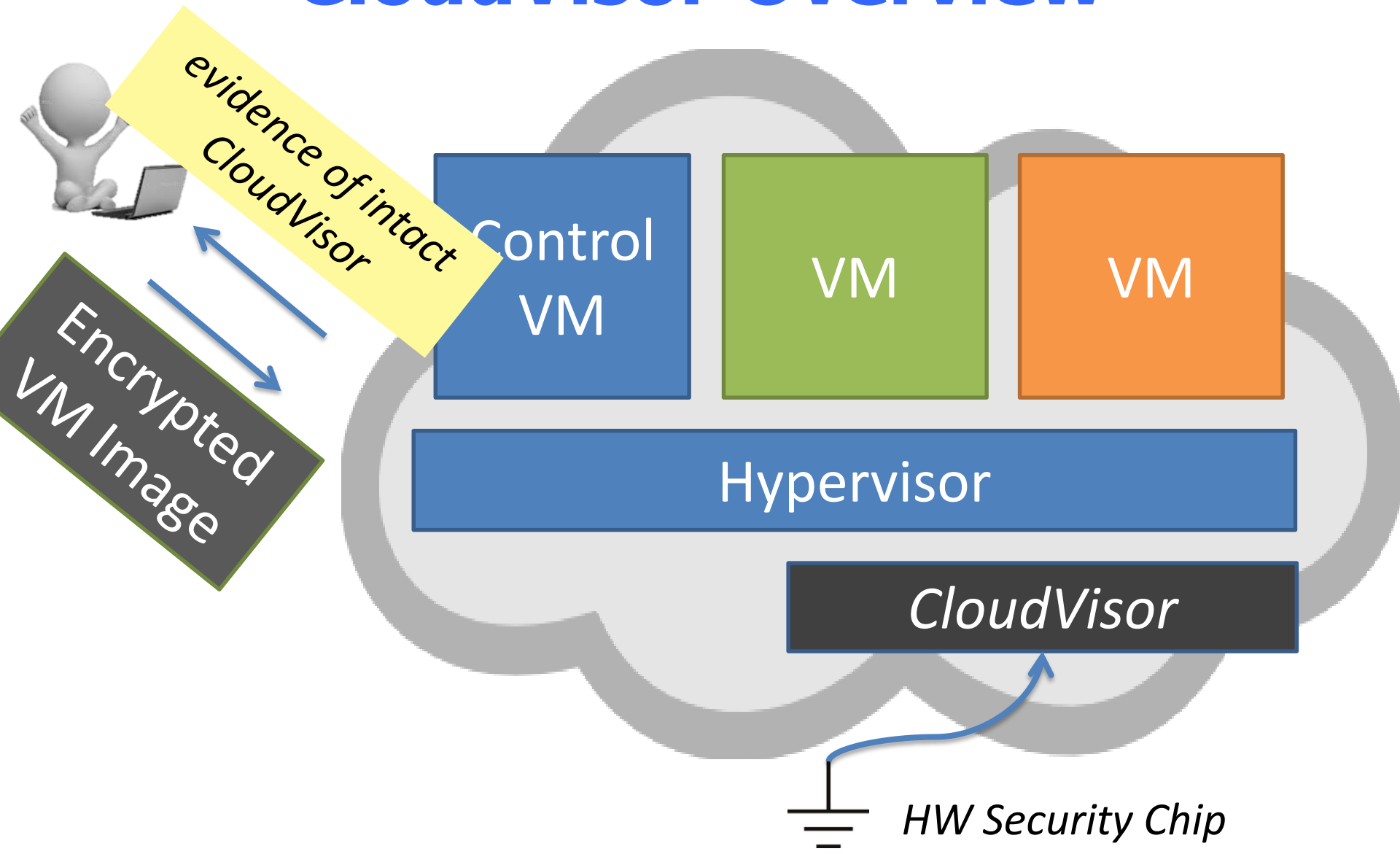
Goal of CloudVisor

- Defend again curious or malicious cloud operators
  - To ensure privacy and integrity of a user VM
- Be transparent to existing cloud infrastructure
  - No or little modifications to virtualization stack (OS, Hypervisor)
- Minimized TCB
  - Easy to verify correctness (e.g., formal verification)
- Non-goals
  - DOS
  - Side-channel attacks
  - Semantic attacks to VM services from network

Observation and Idea

- Key observation
  - Live with a *compromised* virtualization stack
- Idea: separate security protection from VM hosting
  - CloudVisor: another layer of indirection
    - In charge of security protection of VMs
    - Interposes between VMs and hypervisor
  - Hypervisor (unmodified)
    - VM multiplexing and management
- This separation results in
  - Minimized TCB
  - Hypervisor and CloudVisor separately designed and evolved

CloudVisor Overview

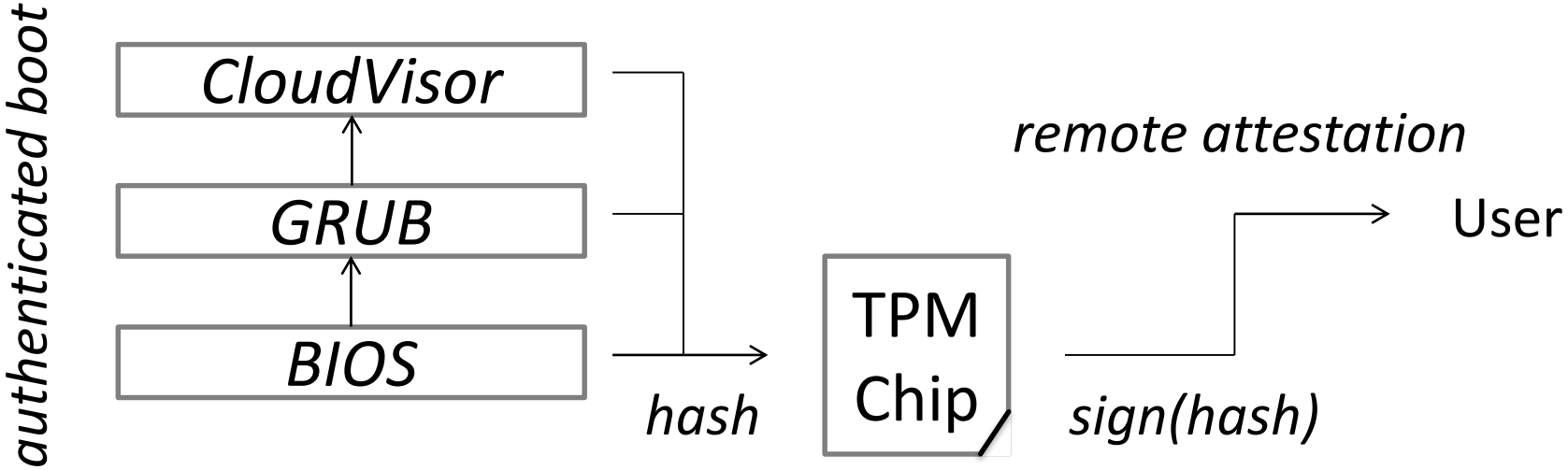


VM Protection Approach

Bootstrap	Uses *Trusted Computing* technology
Memory Pages	Interpose address translation from guest physical address to host physical address, disallow illegal mapping to VM memory
I/O data	Whole VM image encryption Transparent decrypt I/O data in CloudVisor Network I/O not encrypted
CPU states (in paper)	Interpose control switches between hypervisor and VM (i.e., VMexit), hides CPU register states from the hypervisor

Bootstrapping Trust

- 2 basic *Trusted Computing* techniques
  - Authenticated boot
  - Remote attestation



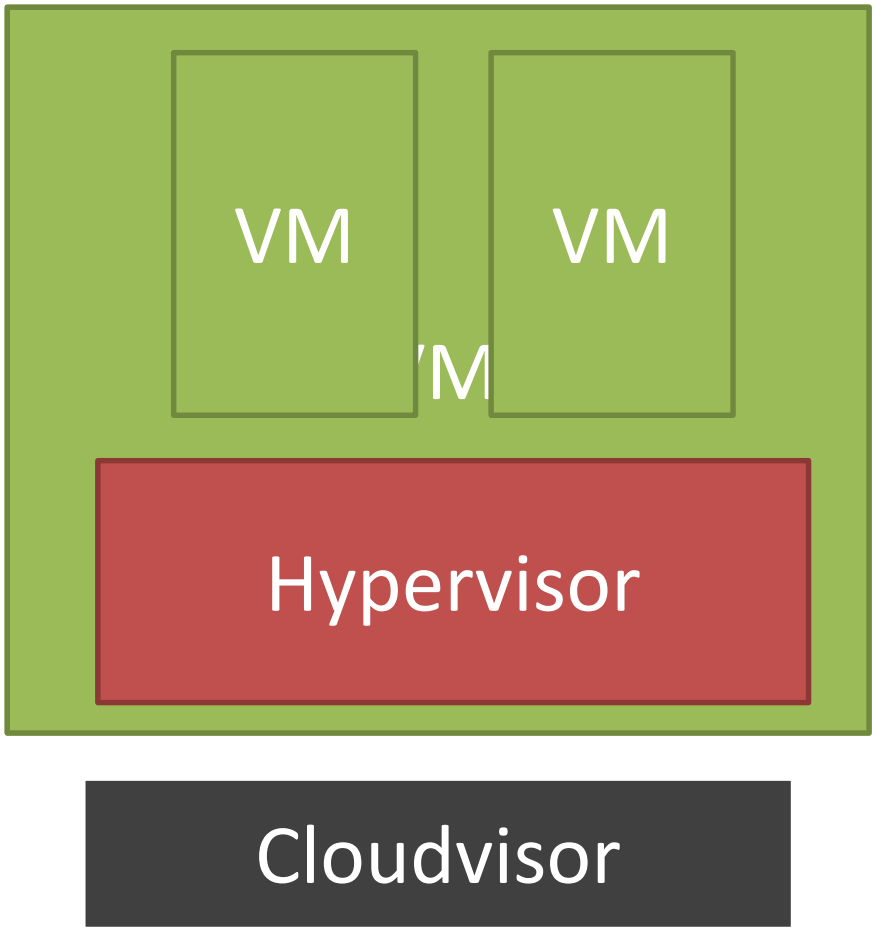
User can ensure a correct version of CloudVisor is running

Interposition with Nested Virtualization

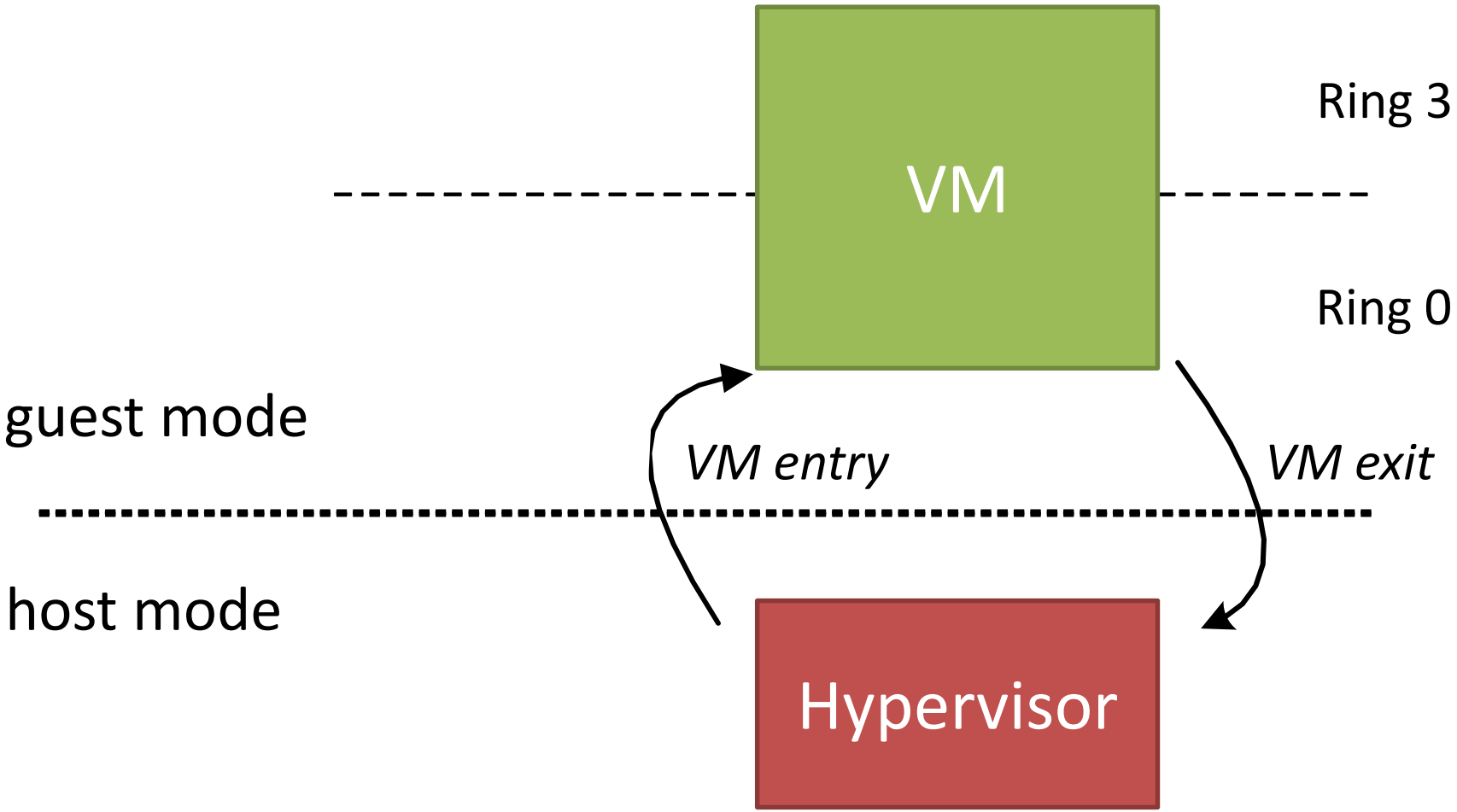
- CloudVisor is based on standard hardware support for virtualization like *VT-x*, *VT-d*
  - It can host only *1* hypervisor

- Hypervisor runs in un-privileged mode
- CloudVisor runs in most privileged mode

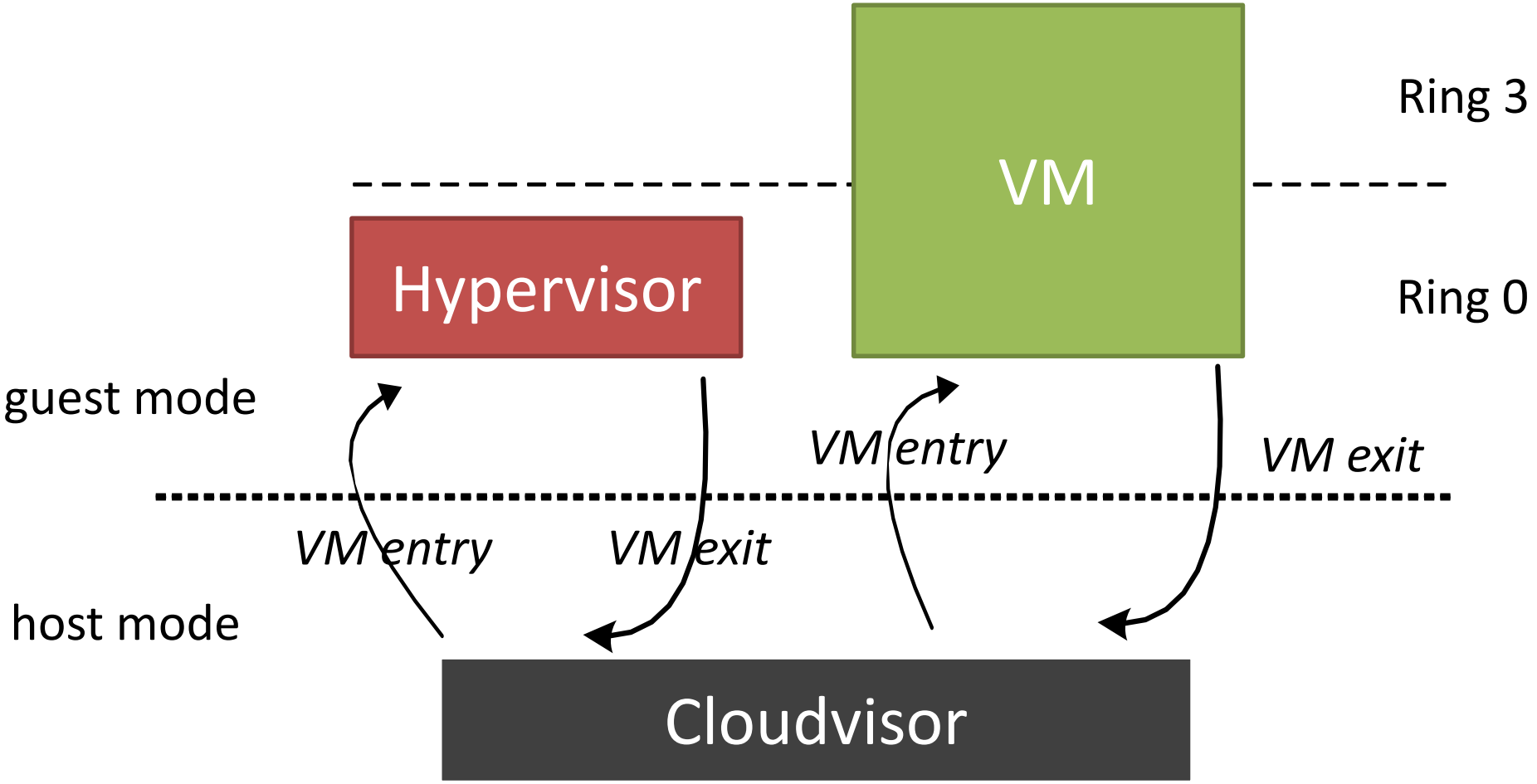
1-on-1 Nested Virtualization (Turtles, 2010)



Virtualization Preliminary: VT-x



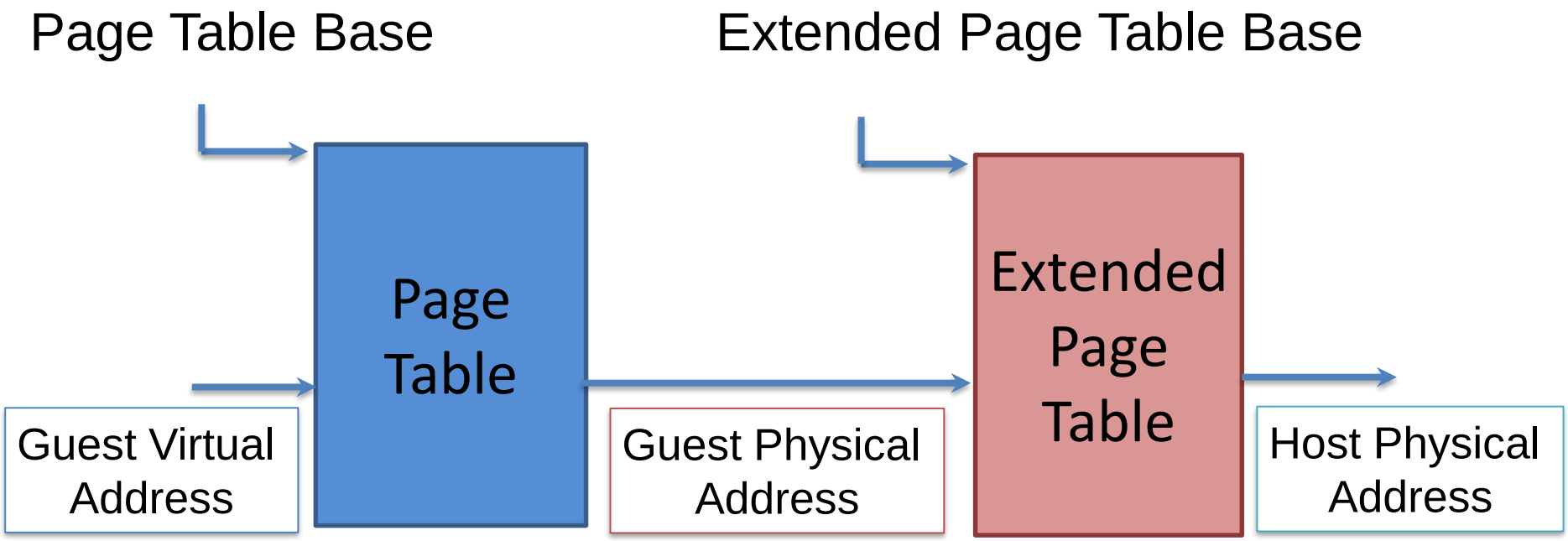
Interposition with CloudVisor



VM Memory Isolation

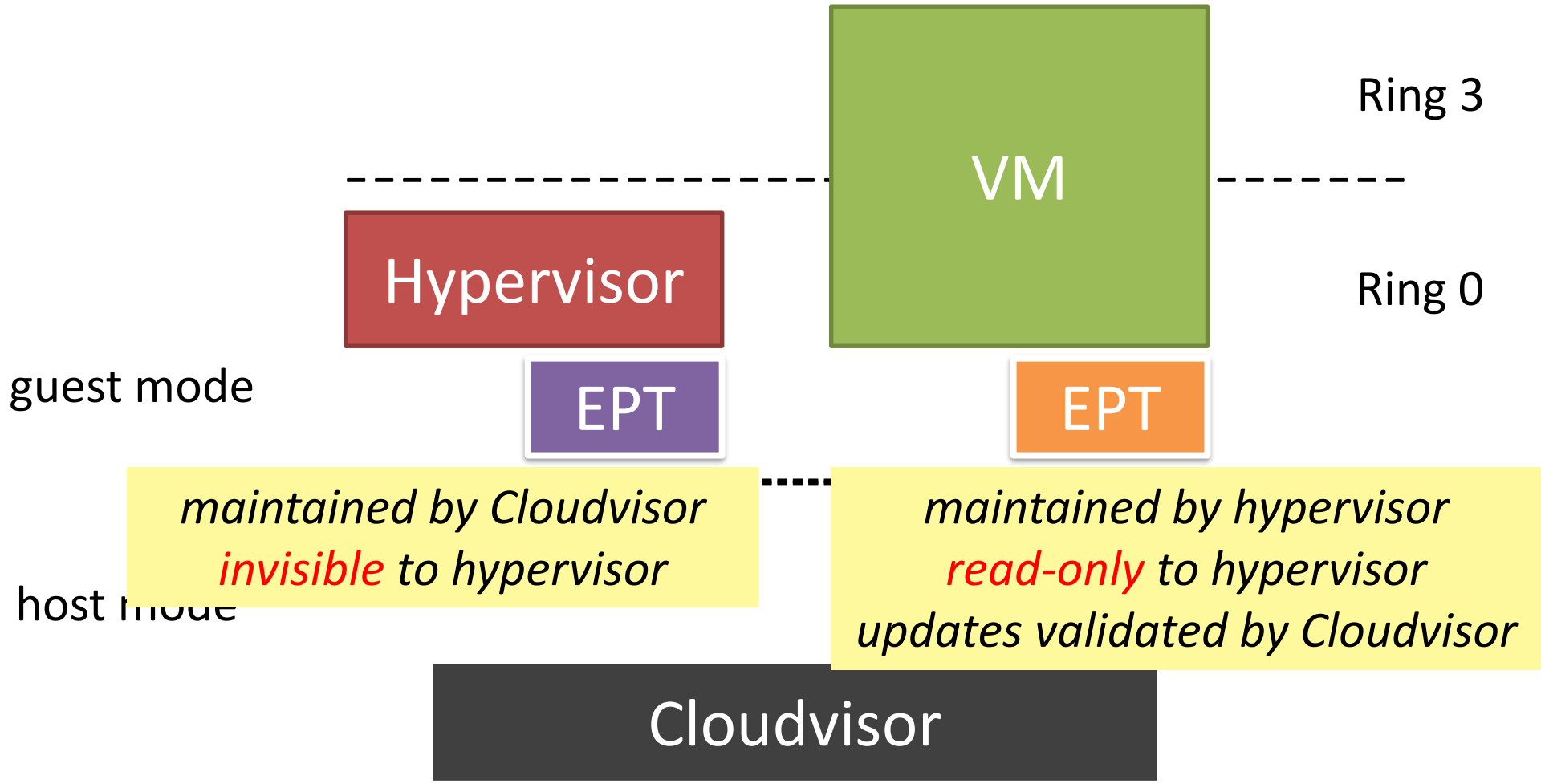
- Goal: forbid hypervisor access to VM memory
- Rules:
  - When a page is assigned to a VM, CloudVisor changes the ownership of the page
  - A memory page is only accessible to its owner

Memory Translation with EPT



Memory access initiated from
CPU: address translated by MMU (Page Table and EPT)
Devices: address translated by IOMMU

Memory Isolation with EPT



Memory Isolation with EPT

- In EPT maintained by CloudVisor
  - There's no mapping to VM memory
  - This guarantees a page is either mapped by hypervisor or a VM, not both
- CloudVisor tracks the ownership of every page
  - Encrypt unauthorized pages and store its hash

Implementing I/O Protection

- CloudVisor intercepts and parses disk I/O request
  - Programmed I/O, DMA
  - Encrypt/decrypt data transparent to VM and hypervisor
  - Calculate hash to verify the integrity of the data (in paper)

- Network I/O are not encrypted
  - User VM should protect the transferred data by itself

Disk Read: Transparent Decryption

- 1. *encrypted data* loaded from disk to hypervisor memory
- 2. hypervisor tries to copy data to I/O buffer in VM memory, fails because EPT fault
- 3. traps into CloudVisor, CloudVisor decrypts the data and copies it to corresponding I/O buffer in VM memory

Impact on VM Operations

CloudVisor works with Save/Restore/Migration

- VM save: transparently encrypted and hashed
- VM restore: transparently decrypted and verified
- Require key exchanges between two machines during migration (Mao et al. 2006)

Transparent memory sharing (not supported)

- Problem: each VM has different keys
- Sol#1: use a common key for page sharing
- Sol#2: provide only integrity protection for shared pages

Implementation

- Xen hypervisor
  - Run unmodified Windows, Linux Virtual Machine
  - ~200 LOC patch to Xen to reduce VMexit (Intel platform only, **Optional**)
- Run on SMP and support SMP VMs
- 5.5K LOCs
  - Intel *TXT* is used to further decrease code size

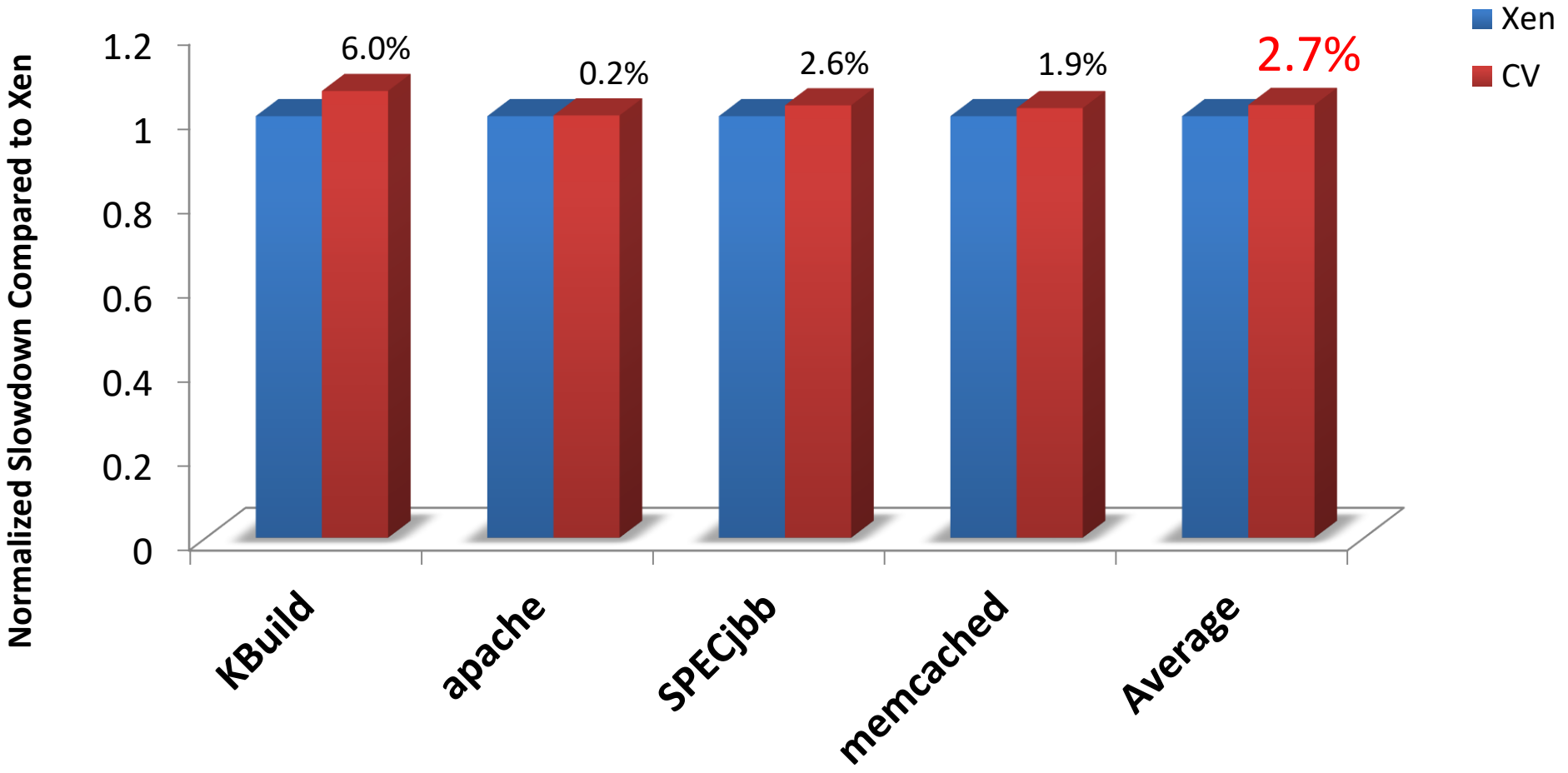
Performance Evaluation

- How much overhead does CloudVisor incur?
- What's the source of overhead?
- Is CloudVisor scalable on multicore?

Test Environment

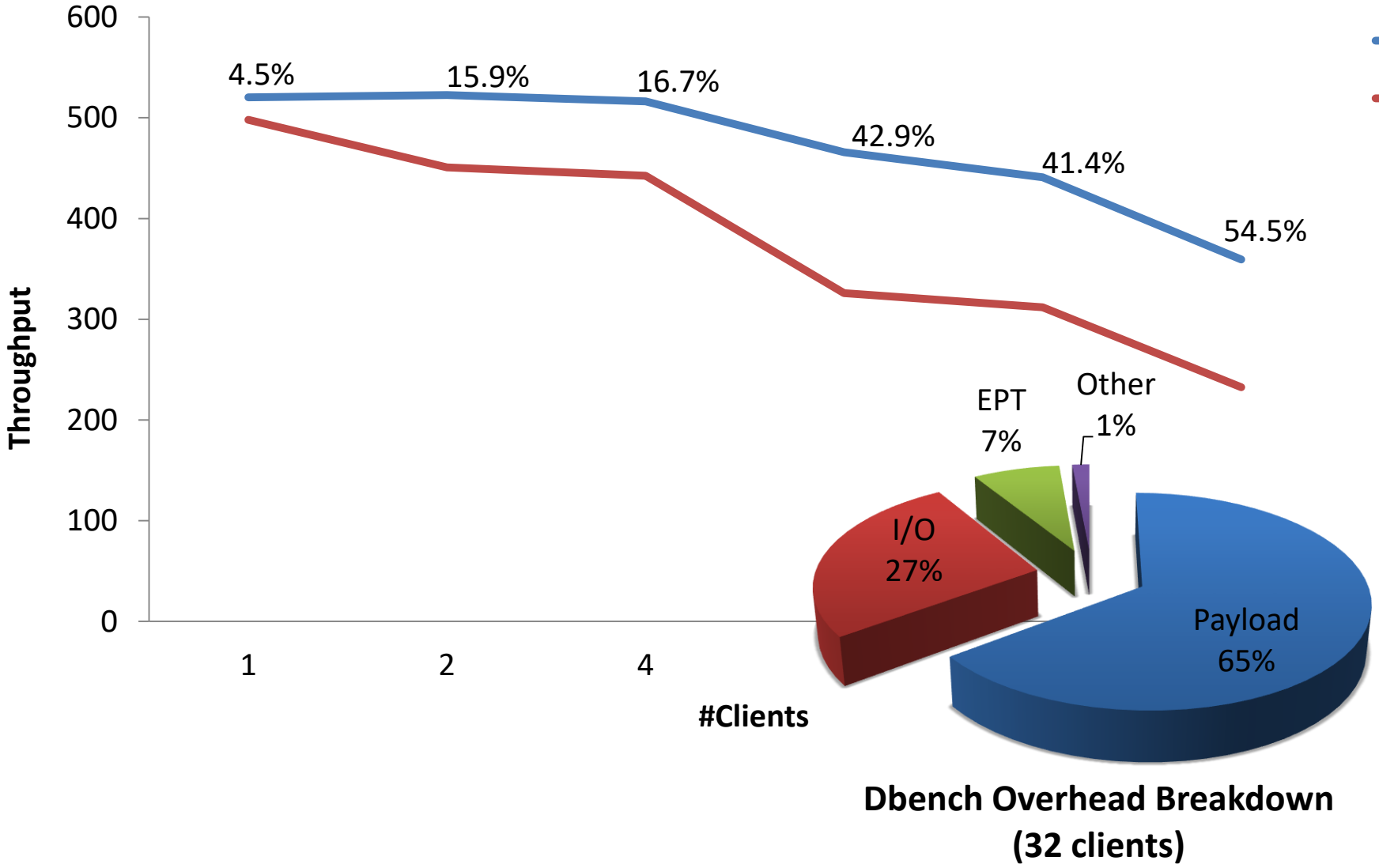
- Hardware: Dell R810
  - 1.8 GHz 8-core Intel processor with *VT-x*, *VT-d*, *IOMMU*, *EPT*, *AES-NI* and *SR-IOV* support
  - 32 Gbyte memory
- Software:
  - Xen-4.0.0 and XenLinux-2.6.31.13 as Domain0 kernel
  - Debian-Linux with kernel 2.6.31 and Windows XP with SP2, both are 64-bit version

Uniprocessor Performance



Average slowdown 2.7%

I/O Intensive Workload

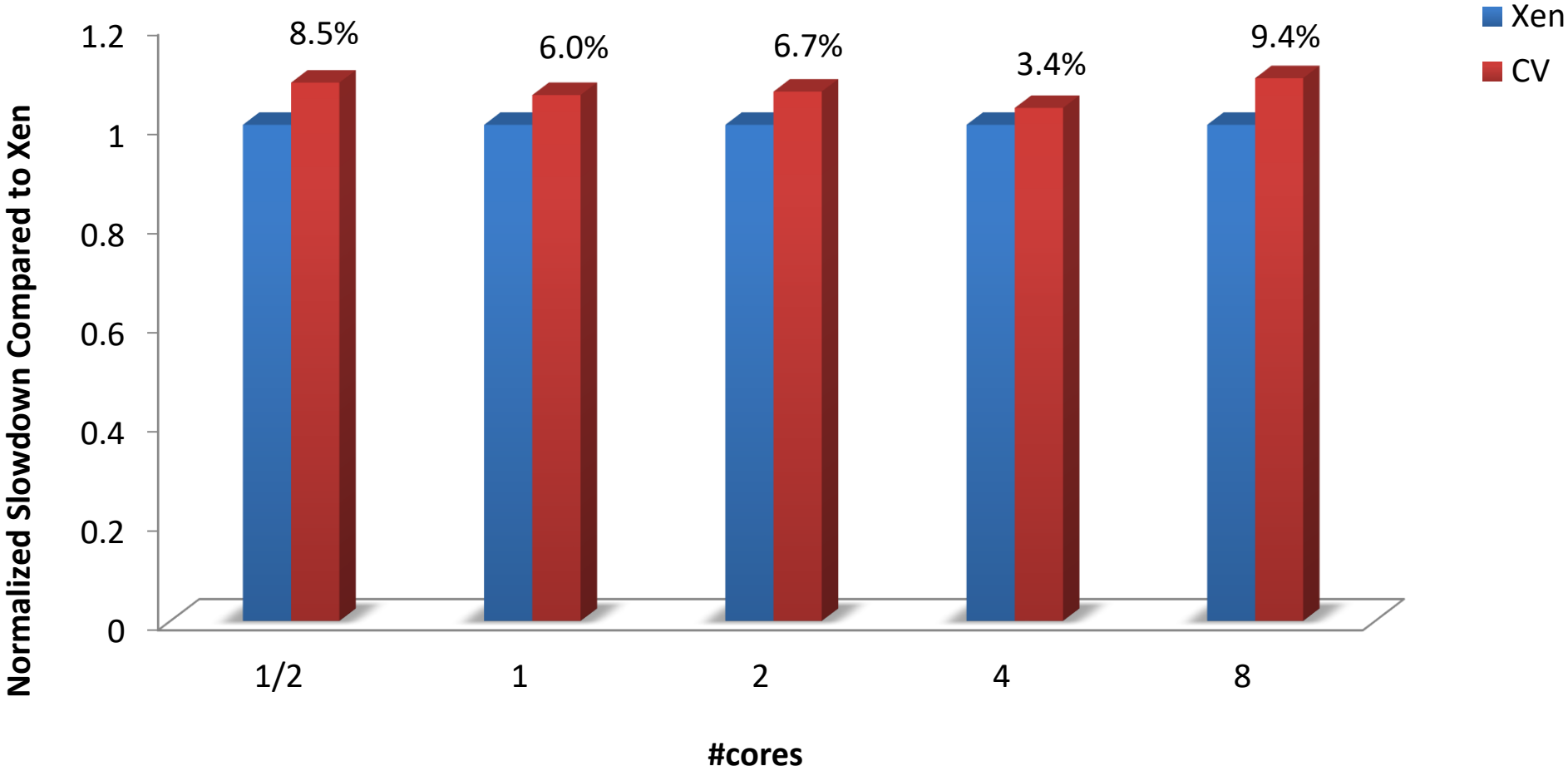


Dbench Overhead Breakdown (32 clients)

Source of Overhead

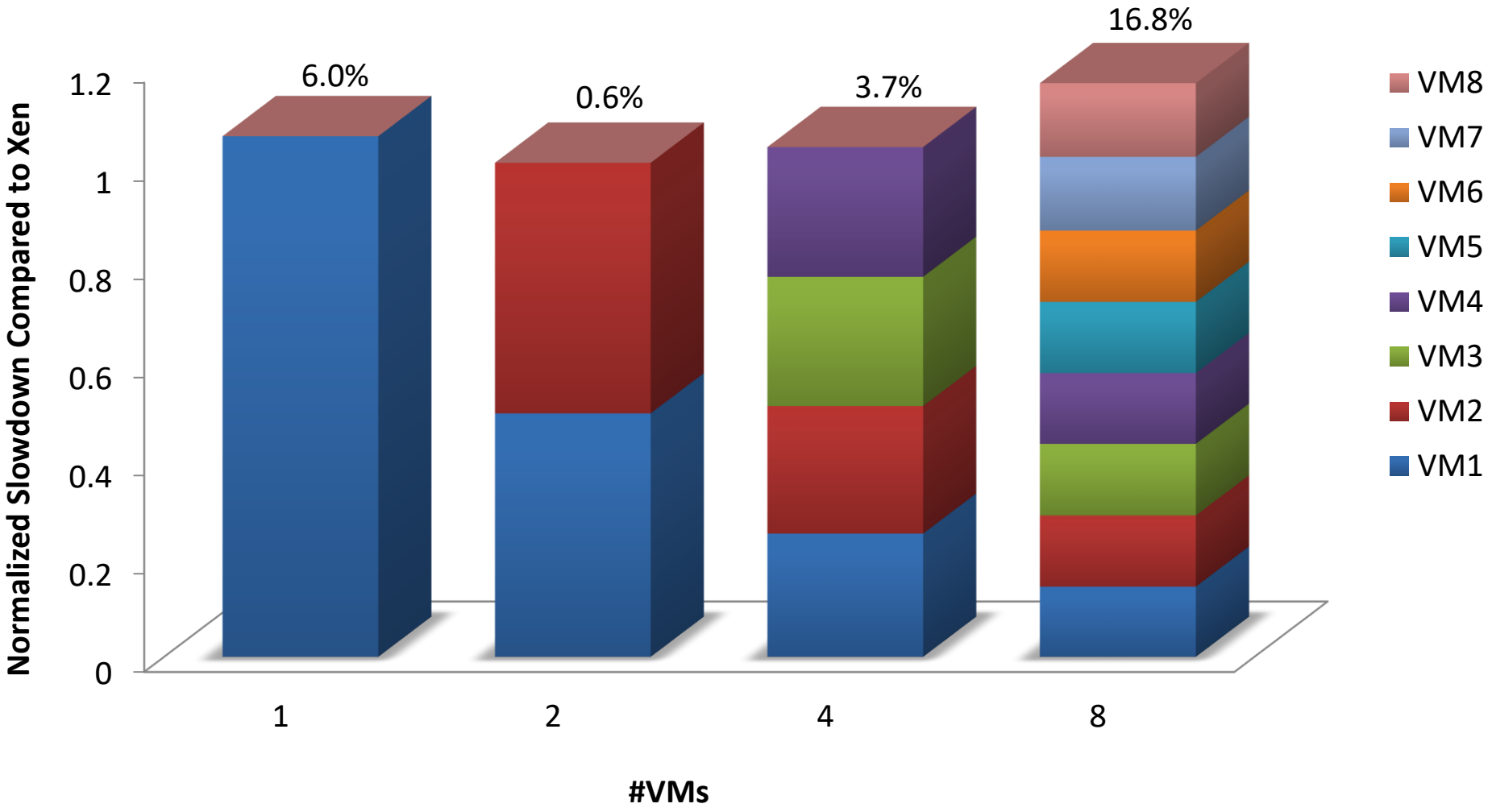
- Additional VMexits due to CloudVisor
  - Although CloudVisor only intercepts a small set of architectural events, VMexits caused by I/O buffer copying is inevitable
- Cryptographic operations
  - Encryption and hash

Multi-core scalability: KBuild



1/2 core means two processes on a core

Performance of Multiple VMs



Related Work

- Nested Virtualization (Turtles, 2010)
  - Support two layers of virtualization, no security protection
  - Result in an even larger TCB
- Virtualization-based rootkits
  - Bluepill, Subvirt
- VMM-based process protection
  - CHAOS, Overshadow
- Efforts in improving or reducing virtualization layer
  - NoHype: removal of virtualization layer
  - NOVA: microkernel based VMM
- Virtualization-based attacks and defenses

Conclusion and Future Work

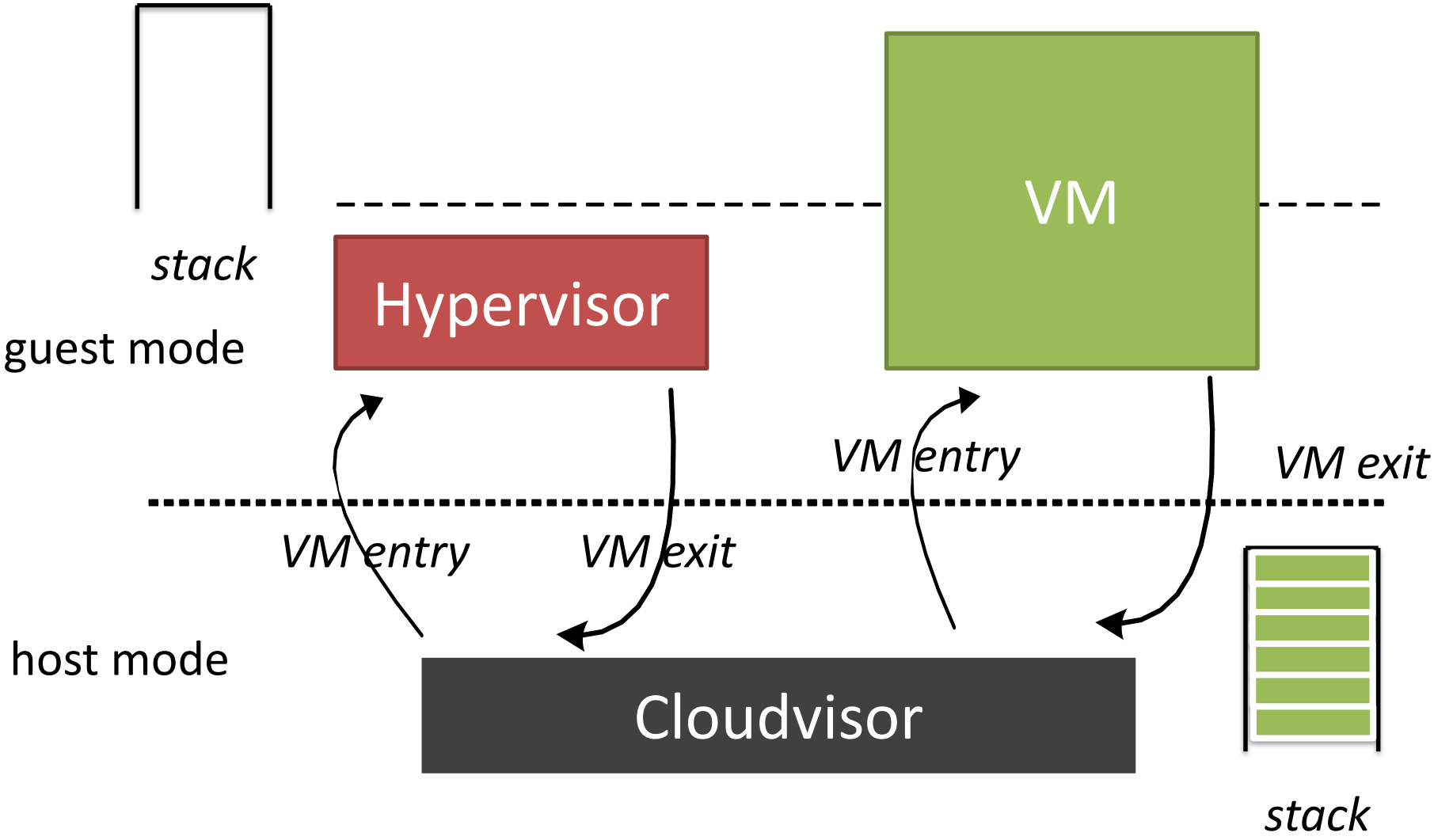
- Hypervisor can host VMs without knowing what's inside
  - That means: hypervisor can provide services without being trusted
- Hiding VM resources from the hypervisor can be done with a small code base (~*5.5 KLOC*)

- Future: HW support of CloudVisor
  - Reduce overhead and complexity

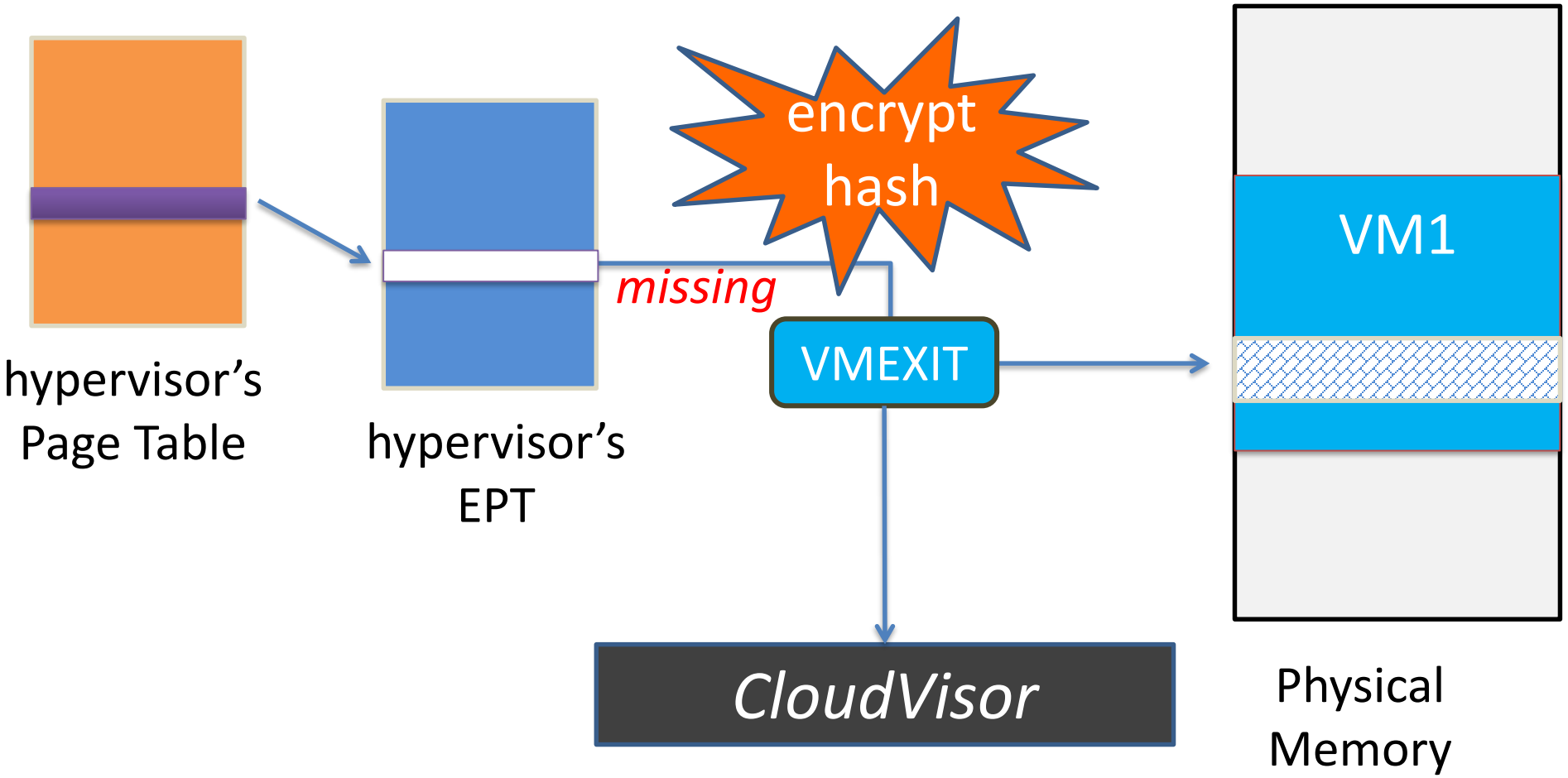
Thanks

Backup

Interposition with CloudVisor



Prevent Unauthorized Access



It is supposed that hypervisor will not use VM memory this way just in rare cases

Para-virtualization Support

- No visible architectural events, no interposition, not supported

- PV drivers
  - Memory sharing and event channel
  - Not supported now, maybe doable

Optimization

- Network benchmarks are beneficial from directly assigned network card
  - Apache, memcached
- I/O data encryption/decryption uses hardware crypto instructions
  - Intel AES-NI