IPADS INSTITUTE OF PARALLEL AND DISTRIBUTED SYSTEMS

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

(Mostly) Exitless VM Protection from Untrusted Hypervisor through Disaggregated Nested Virtualization

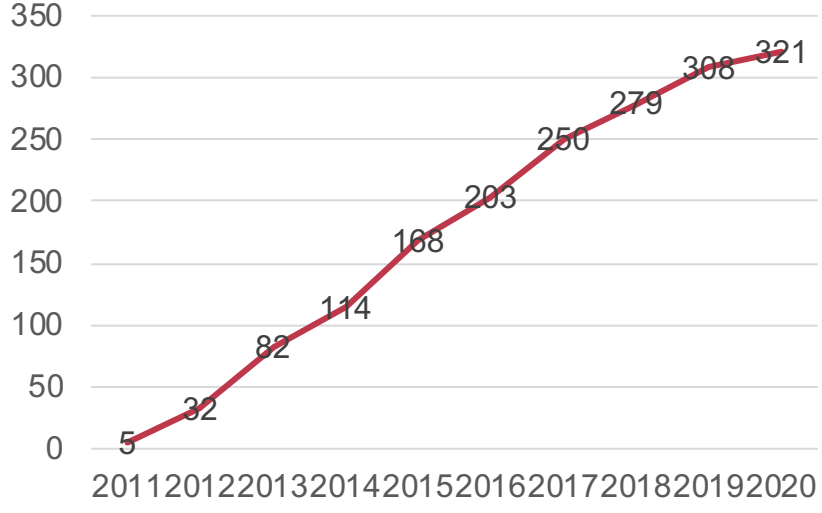
Zeyu Mi, Dingji Li, Haibo Chen, Binyu Zang, Haibing Guan

Shanghai Jiao Tong University

https://ipads.se.sjtu.edu.cn

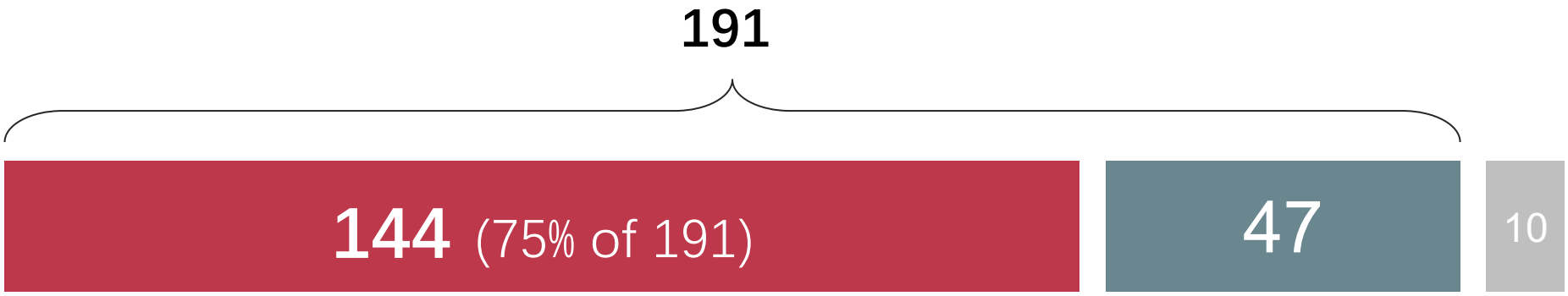
Vulnerable Commercial Hypervisors

- **Xen CVE is growing**
  - LoC: from 45K (v2.0) to 2,649K (v4.14.0)
  - 321 XSA
- **KVM and VMware**
  - KVM: 110+ CVE
  - VMware: 140+



https://xenbits.xen.org/xsa/

Analyzing 201 of Xen's Vulnerabilities (XSA)



144 are in the hypervisor

- E.g., Host DoS, privilege escalation, etc.
- Use hypervisor to attack VM

47 are not in hypervisor

- Some are in Domain-0
- Some are in Qemu

10 are ignored

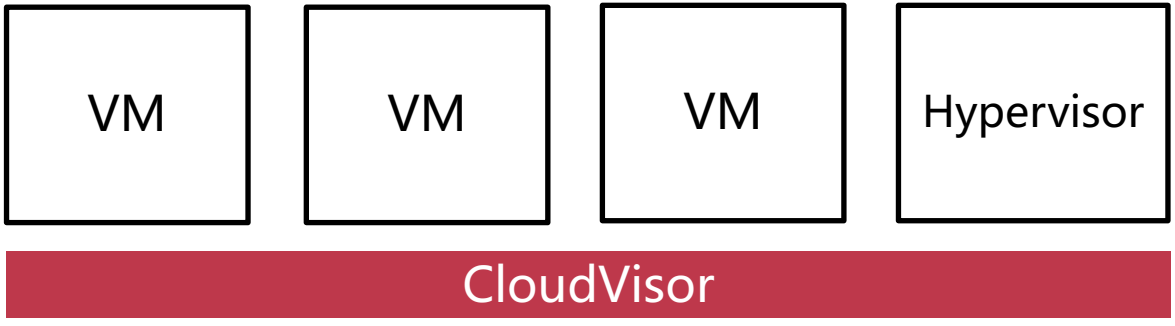
Shi L, Wu Y, Xia Y, et al. Deconstructing Xen. NDSS. 2017.

Existing Approaches

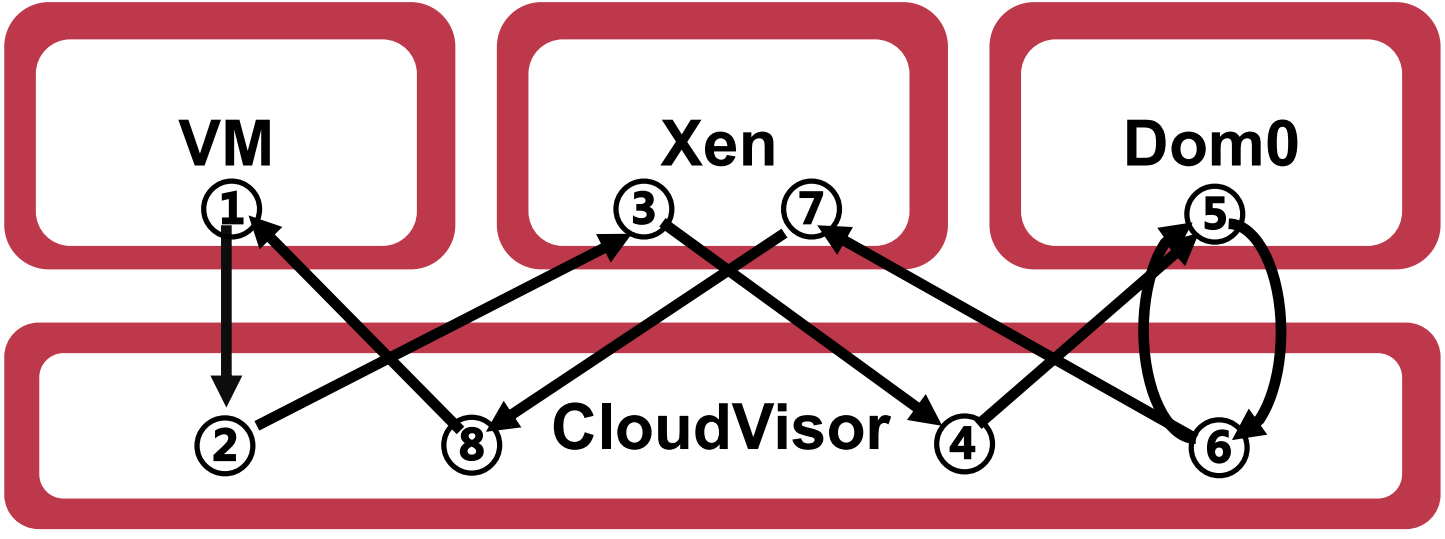
- **Software Method**
  - In-the-box: harden the hypervisor layer
    - Cannot eliminate the risks of exploiting hypervisor vulnerabilities
  - Out-of-the-box: nested virtualization
    - Numerous VM exits bring performance overhead
- **Hardware Method (Intel SGX)**
  - Only available to run in user mode
  - Limited EPC memory incurs significant performance overhead

CloudVisor (SOSP' 11)

- **Observation: protection logics for VMs are mostly fixed**
- **Idea: Separate management from protection**
  - Deprivilege the commercial hypervisor to non-root mode
- **Result: Minimized TCB**
  - VMM and CloudVisor separately designed and evolved



The Cost of Protection: Excessive VM Exits



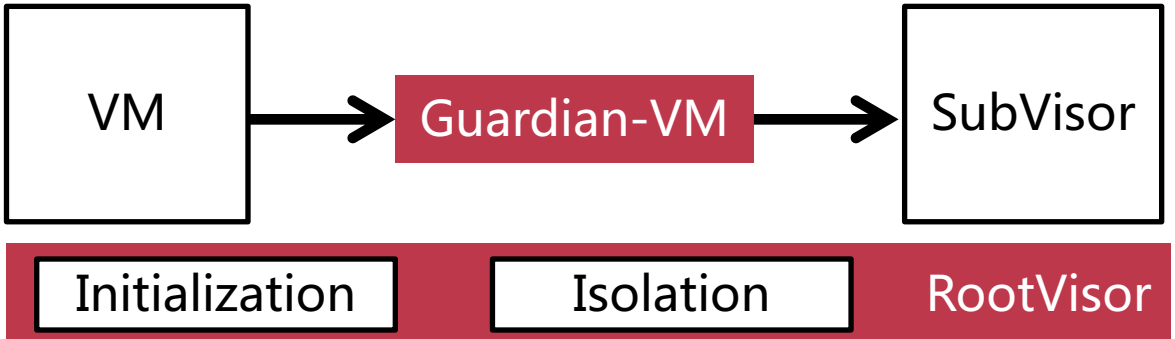
Operation	Times
Hypercall	>= 2X
EPT Violation	2 – 6 X
DMA Operation	>= 2X

CloudVisor-D: No Compromise for Security & Performance

- **A secure and efficient design to shield VM in untrusted clouds**
  - Do not trust the commercial hypervisor
  - Introduce negligible overheads compared to the Xen hypervisor
- **Disaggregated nested virtualization**
  - Deprivilege the hypervisor through **nested virtualization**
  - Disaggregate the nested hypervisor
    - **Offload** VM operations and their protection work to the non-privileged mode

Architecture of CloudVisor-D

- **A tiny nested hypervisor in root mode**
- **A Guardian-VM for each VM in non-root mode**
- **Most VM ops offloaded to Guardian-VM**
  - Hypercalls
  - Memory virtualization
  - I/O operations

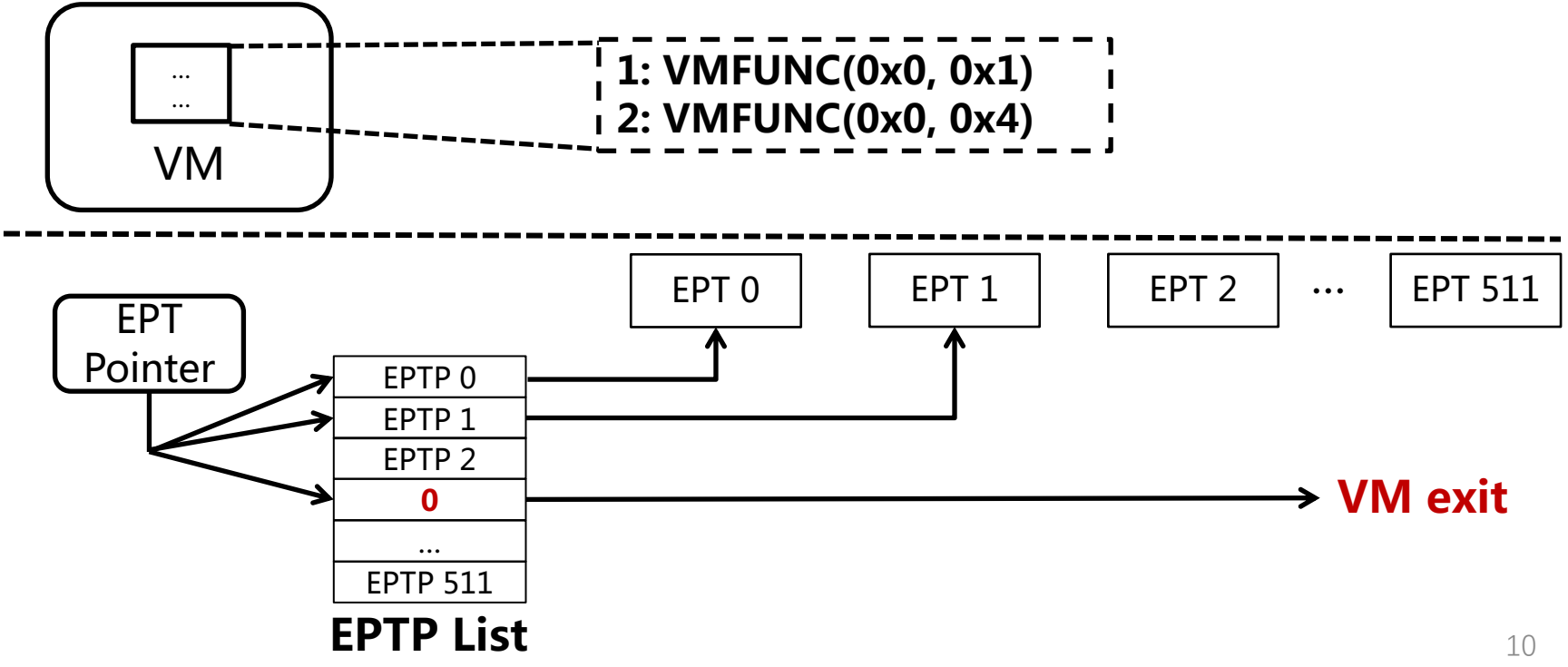


Threat Model

- **TCB: RootVisor and each Guardian-VM**
- **Distrusing: SubVisor and all guest VMs**
- **Out of scope**
  - physical attack
  - Side-channel attacks
  - DoS attacks

Key Secrets: VMFUNC

- Switch EPT efficiently without VM Exits
- Faster than VM exit (134 vs. 301 cycles on Intel Skylake)



CloudVisor-D as Reference Monitor

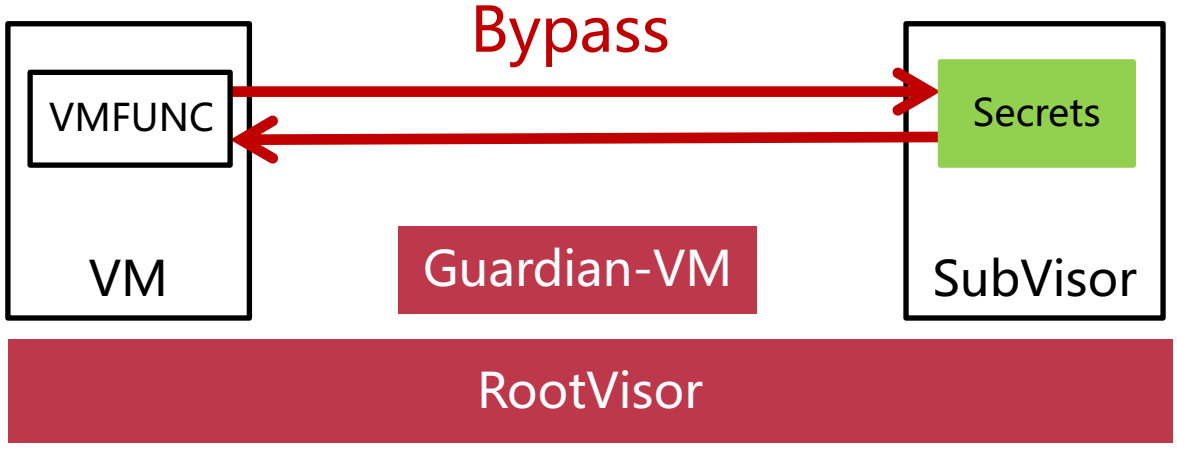
- **CloudVisor-D satisfies two properties**
  - **Tamperproof:** protect RootVisor and Guardian-VM from compromising
  - **Complete Mediation:** interpose on all communication paths between SubVisor and VMs

Complete Mediation

- **Two paths**
  - RootVisor Path: VM->RootVisor->SubVisor
  - Guardian-VM Path: VM->Guardian-VM->SubVisor

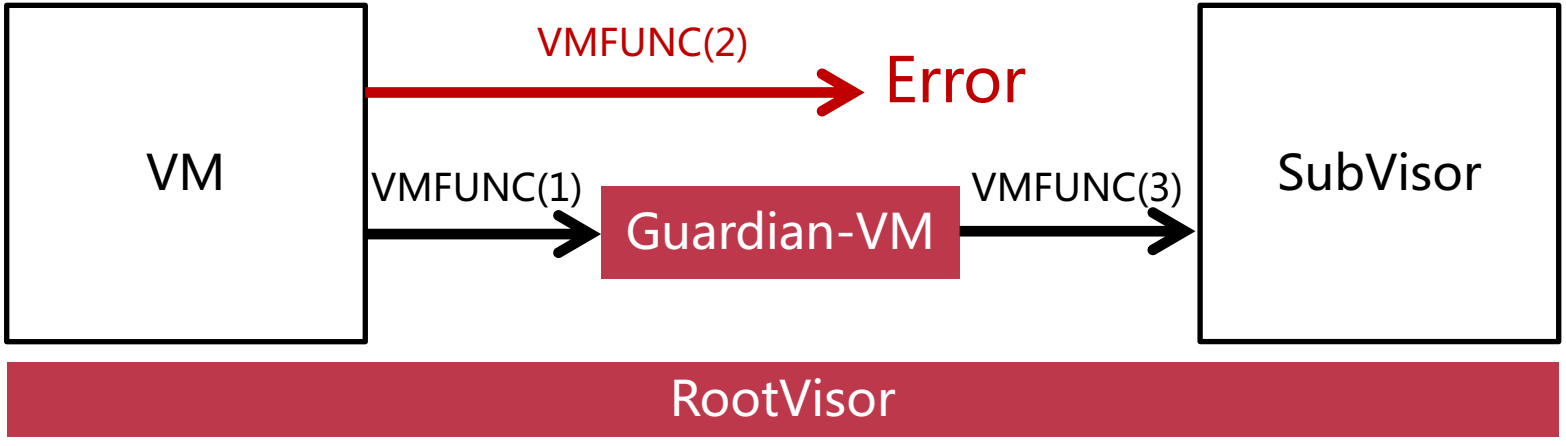
Faking VMFUNC Attacks

- **Type-1: Bypass Guardian-VM**
  - Access arbitrary memory region in VM or SubVisor
- **Type-2: Attack Guardian-VM**



Dynamic EPTP List Manipulation

- **An invalid EPTP entry triggers a VM exit**



0	Guest-EPT
1	Guardian-EPT
2	**0**
…	…
511	0

EPTP List

0	0
1	Guardian-EPT
2	SubVisor-EPT
…	…
511	0

EPTP List

Other Techniques

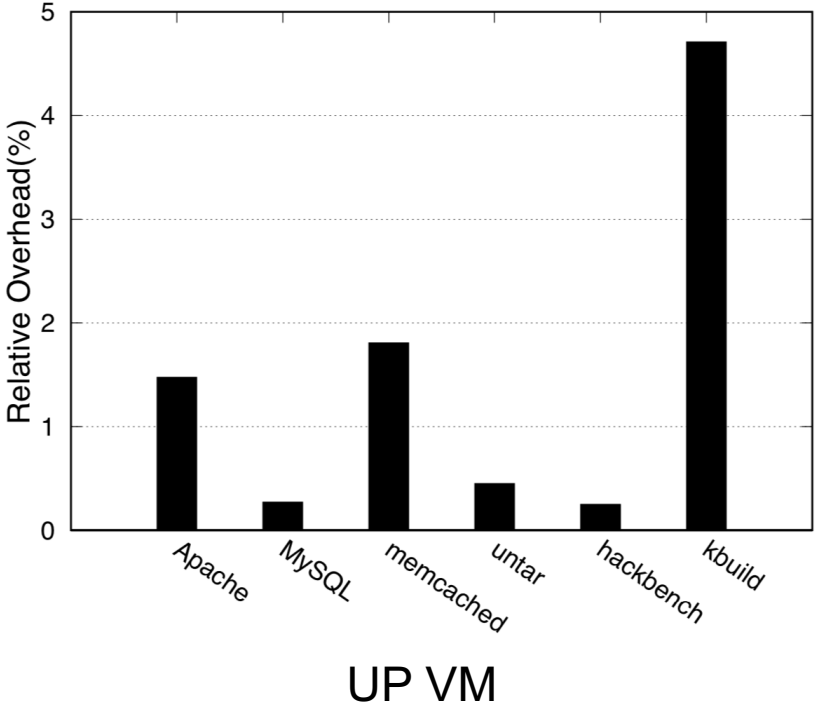
- **Isolated Guardian-VM Page Table**
- **Jump table**
- **Memory virtualization**
  - Shadow EPT and virtualization exception
- **I/O protection**
  - Compatible with PV I/O model
  - Encryption and integrity guarantee

Microbenchmark

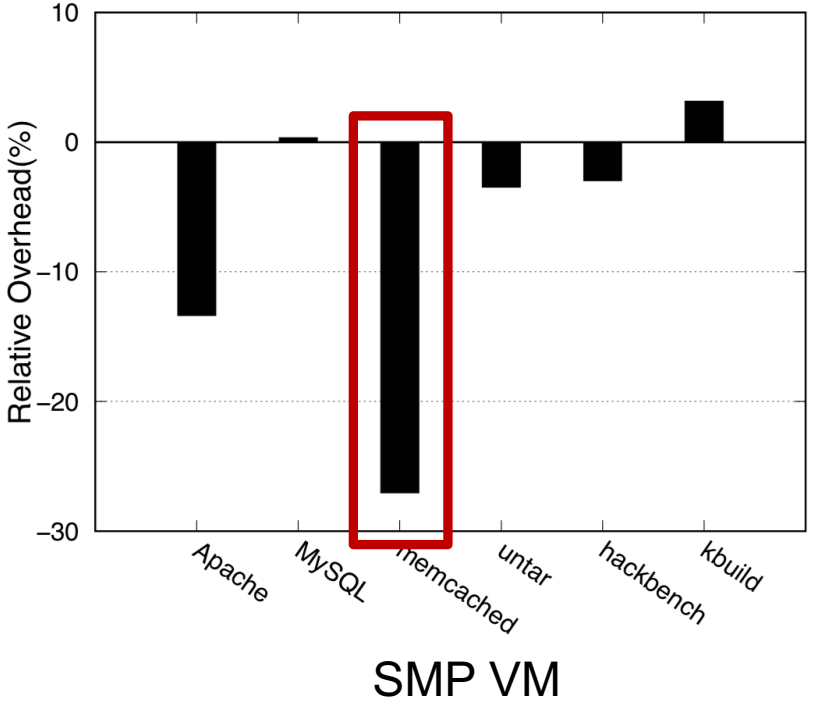
Operation	Xen	CloudVisor	CloudVisor-D	Speedup
Hypercall	1758	4681	1810	61.3%
EPT violation handling	5374	66301	9929	85.0%
Virtual IPI	11214	21344	13331	37.5%

Applications

#VM exits: 1,691,758 -> 63,909

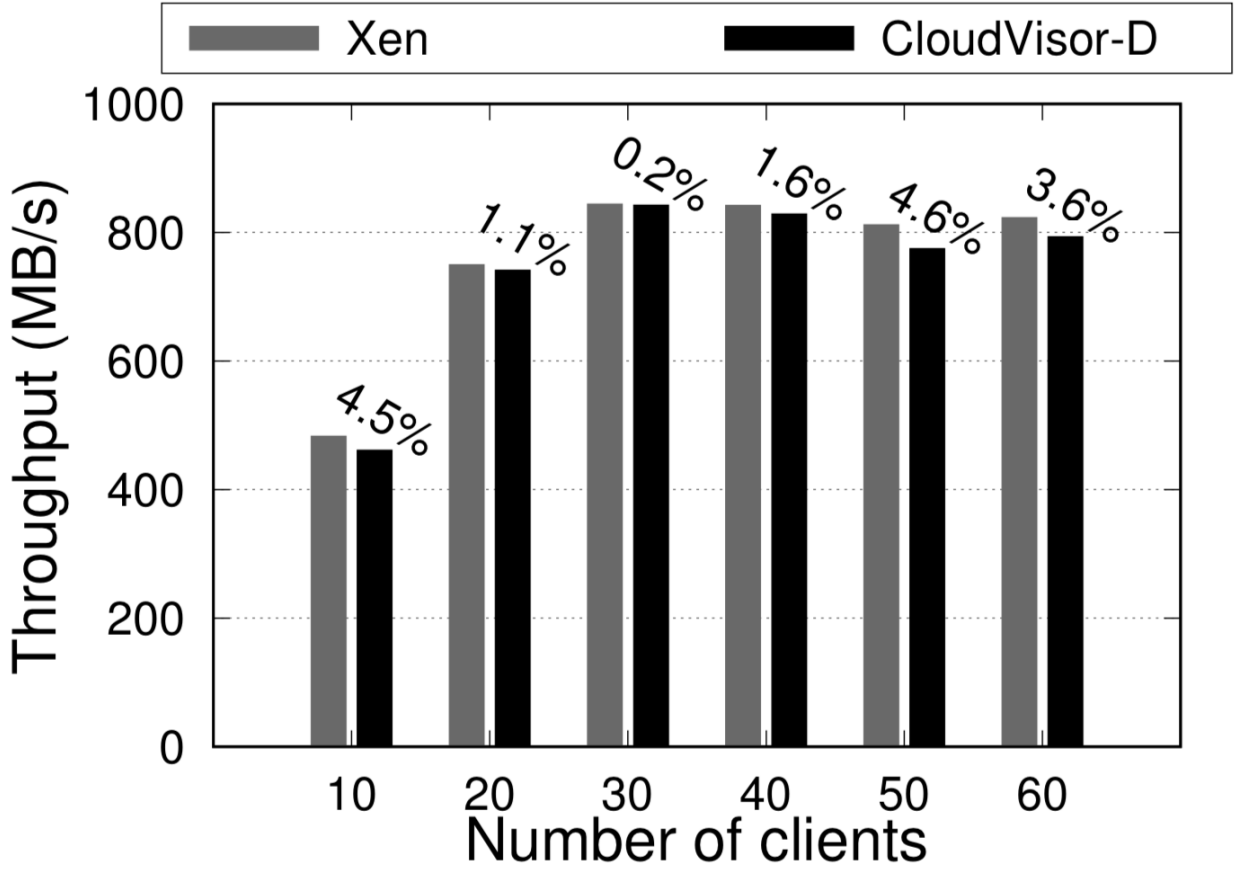


UP VM



SMP VM

dbench: I/O Performance



Conclusion

- **Today's cloud tenants are facing severe security threats**
- **A secure and efficient system to shield VM in untrusted clouds**
  - Disaggregated nested virtualization
  - Same level of security guarantee as nested virtualization
  - Introduce negligible overhead compared with the vanilla Xen

Q&A: yzmizeyu@sjtu.edu.cn

Thanks!