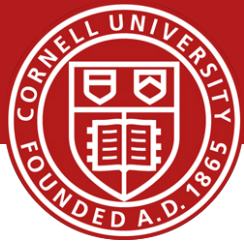


Data Center Virtualization: Xen and Xen-blanket

Hakim Weatherspoon

Assistant Professor, Dept of Computer Science

CS 5413: High Performance Systems and Networking

November 17, 2014

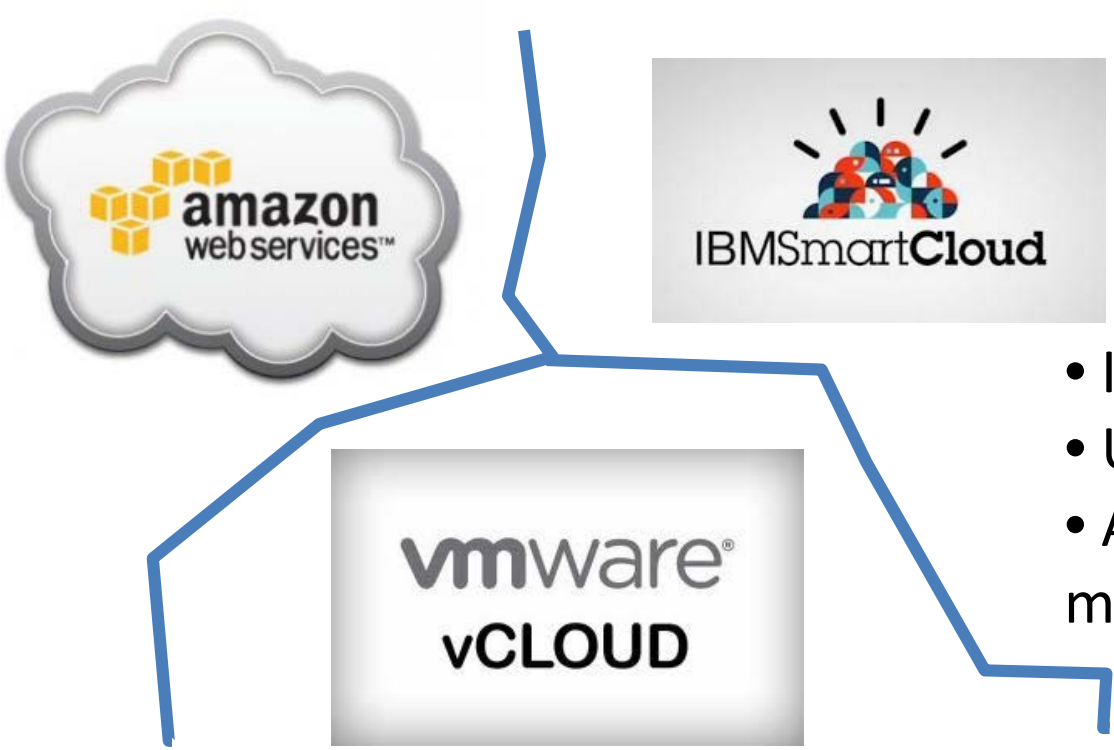
Slides from ACM European Conference on Computer Systems 2012 presentation of "The Xen-Blanket: Virtualize Once, Run Everywhere" and Dan Williams dissertation

Goals for Today

- The Xen-Blanket: Virtualize Once, Run Everywhere
  - D. Williams, H. Jamjoom, and H. Weatherspoon. ACM European Conference on Computer Systems (EuroSys), April 2012, pages 113-126..

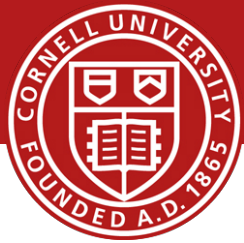
Background & motivation

- Infrastructure as a Service (IaaS) clouds



- Inter-cloud migration?
- Uniform VM image?
- Advanced hypervisor level management?

research challenges



- Lack of interoperability between clouds
  - How can cloud *user* homogenize clouds?
- Lack of control in cloud networks
  - What cloud network abstraction enables enterprise workload to run without modification?
- Lack of efficient cloud resource utilization
  - How can cloud users exploit oversubscription in the cloud while handling overload?

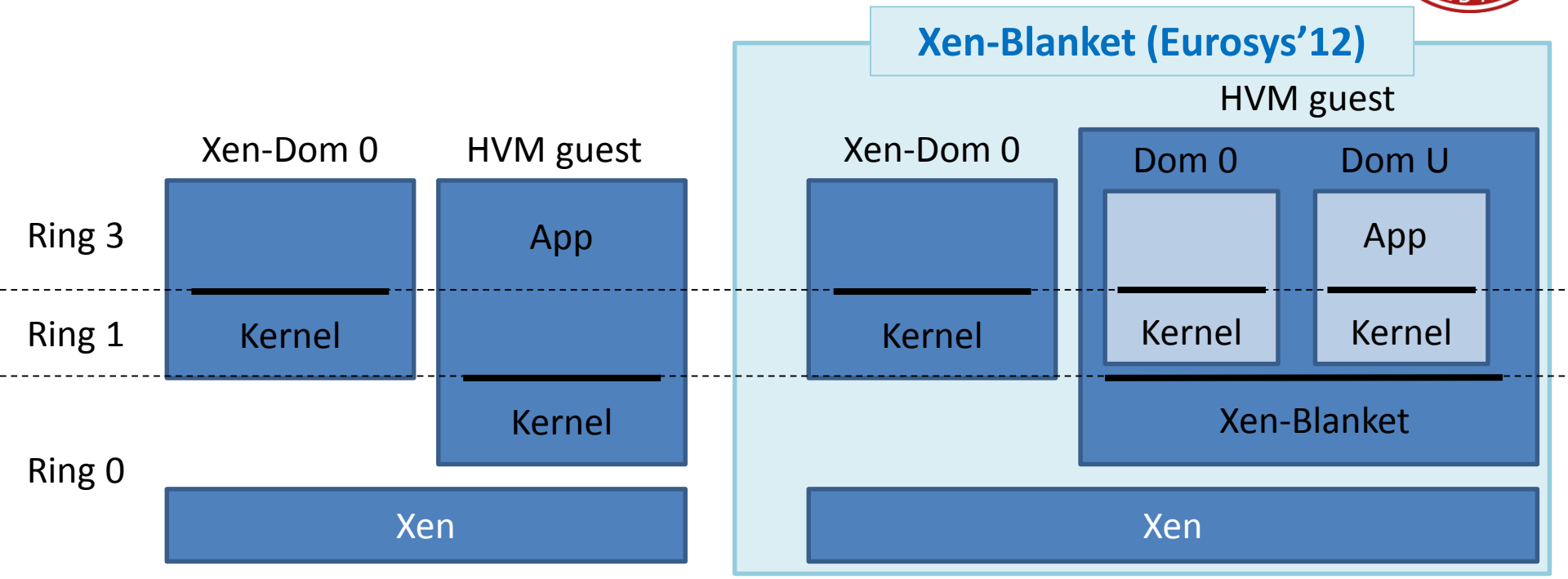
Xen-Blanket

- A second-layer hypervisor

- Inter-cloud migration?
- Uniform VM image?
- Advanced hypervisor level management?

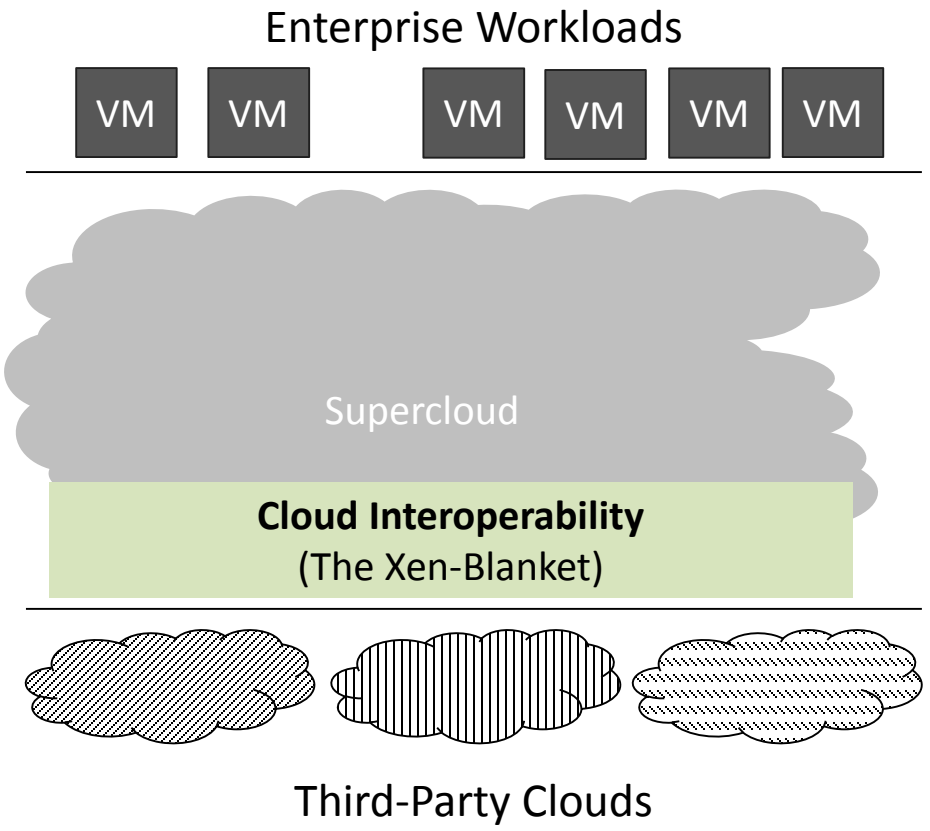
Xen-Blanket

Xen-Blanket


Xen-Blanket (Eurosys'12)
Xen-Dom 0
HVM guest
Ring 3
Ring 1
Ring 0
Kernel
App
Kernel
Xen
Xen-Dom 0
HVM guest
Dom 0
Dom U
Kernel
App
Kernel
Kernel
Xen-Blanket
Xen

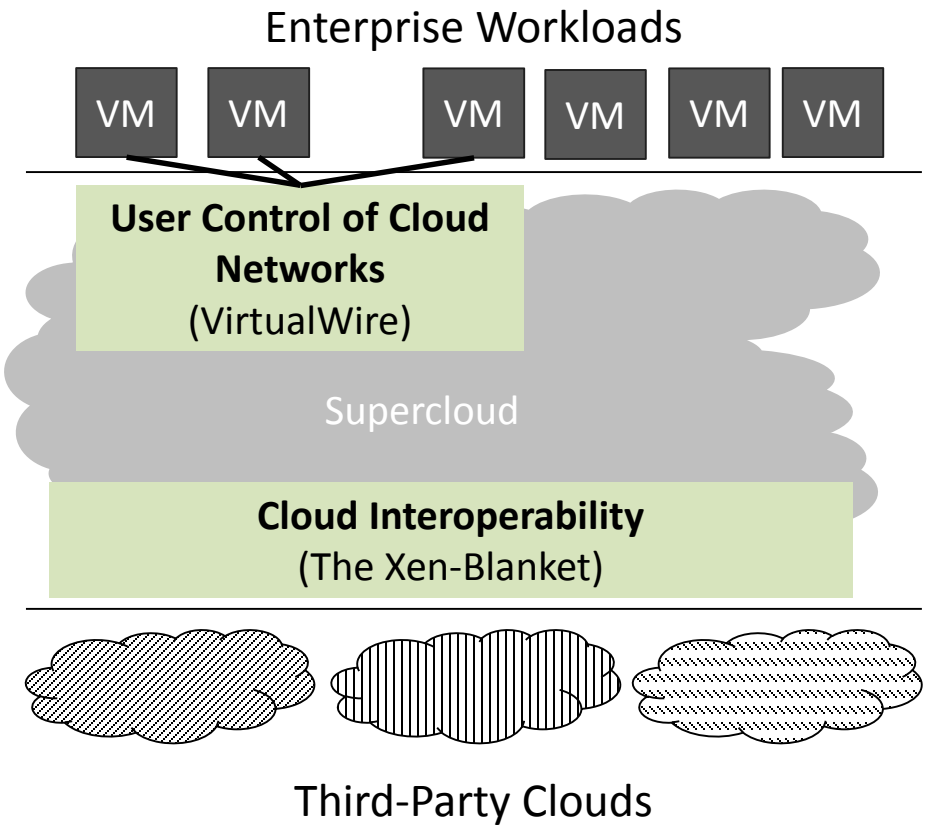
contributions towards superclouds

- Cloud interoperability
  - Enable cloud user to homogenize clouds
  - The Xen-Blanket



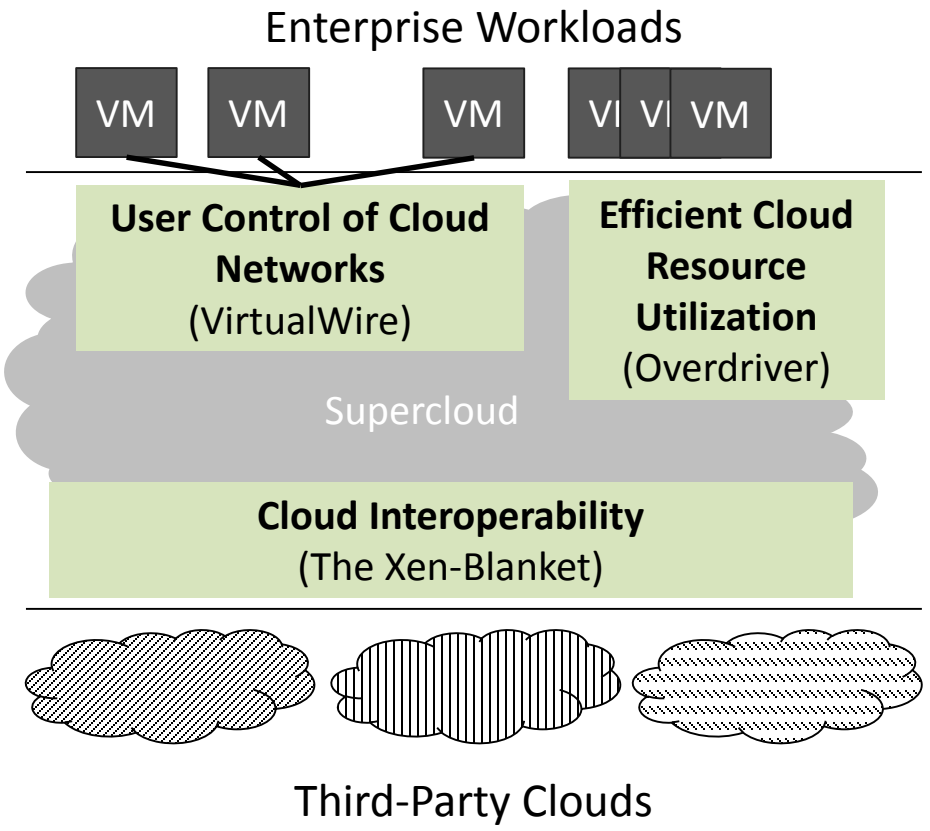
contributions towards superclouds

- Cloud interoperability
- User control of cloud networks
  - Enable cloud user to implement network control logic
  - VirtualWire



contributions towards superclouds

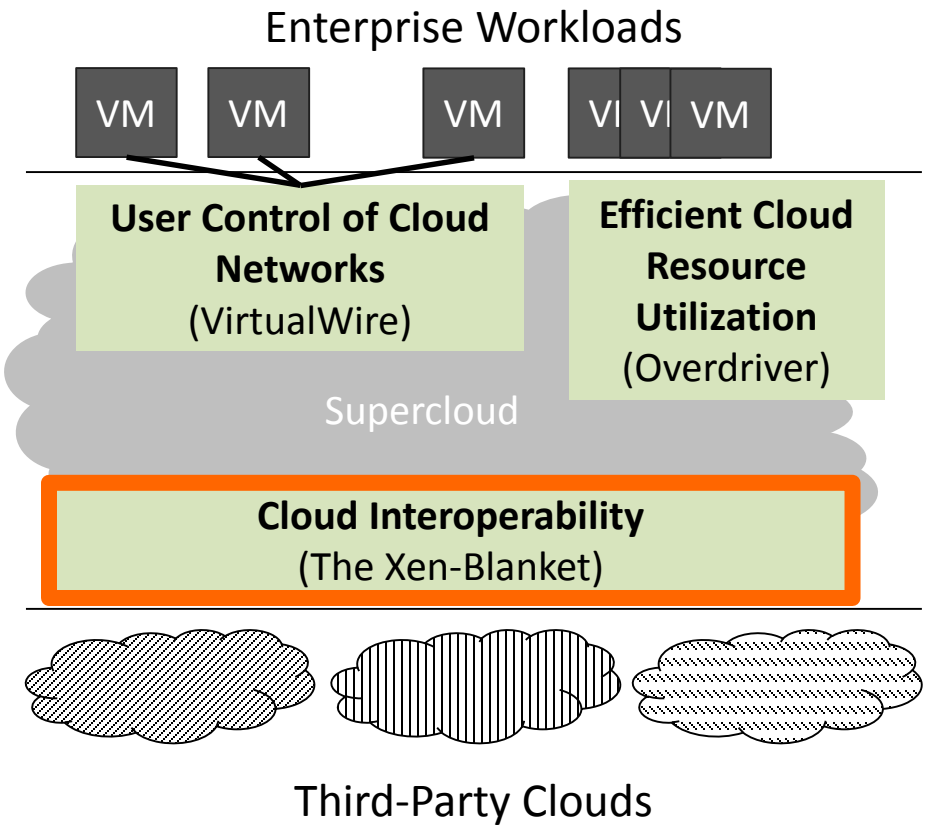
- Cloud interoperability
- User control of cloud networks
- Efficient cloud resource utilization
  - Enable cloud user to oversubscribe resources and handle overload
  - Overdriver



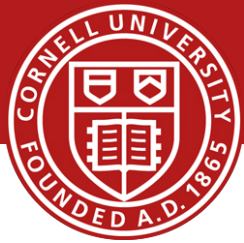
roadmap: towards superclouds

- Cloud interoperability
- User control of cloud networks
- Efficient cloud resource utilization

- Related work
- Future work
- Conclusion



Clouds are not interoperable

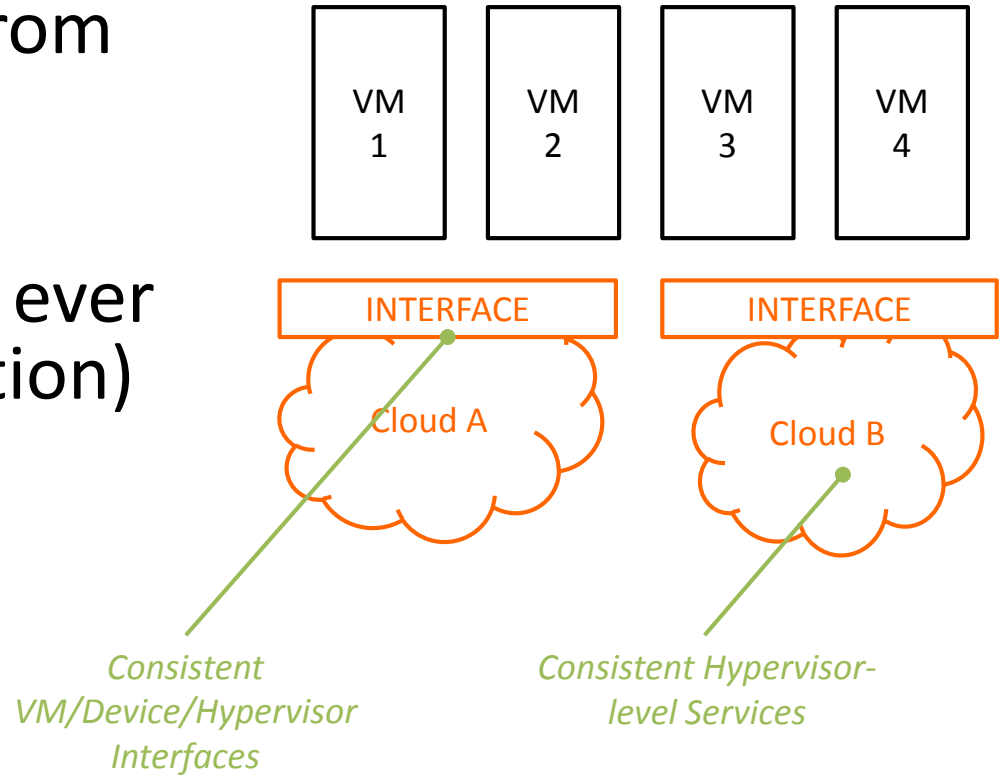


- Image format not yet standard
  - AMI, Open Virtualization Format (OVF)
- Paravirtualized device interfaces vary
  - **`virtio`**, Xen
- Hypervisor-level services not standard
  - Autoscale, VM migration, CPU bursting

Need *homogenization* (consistent interfaces, services across clouds)

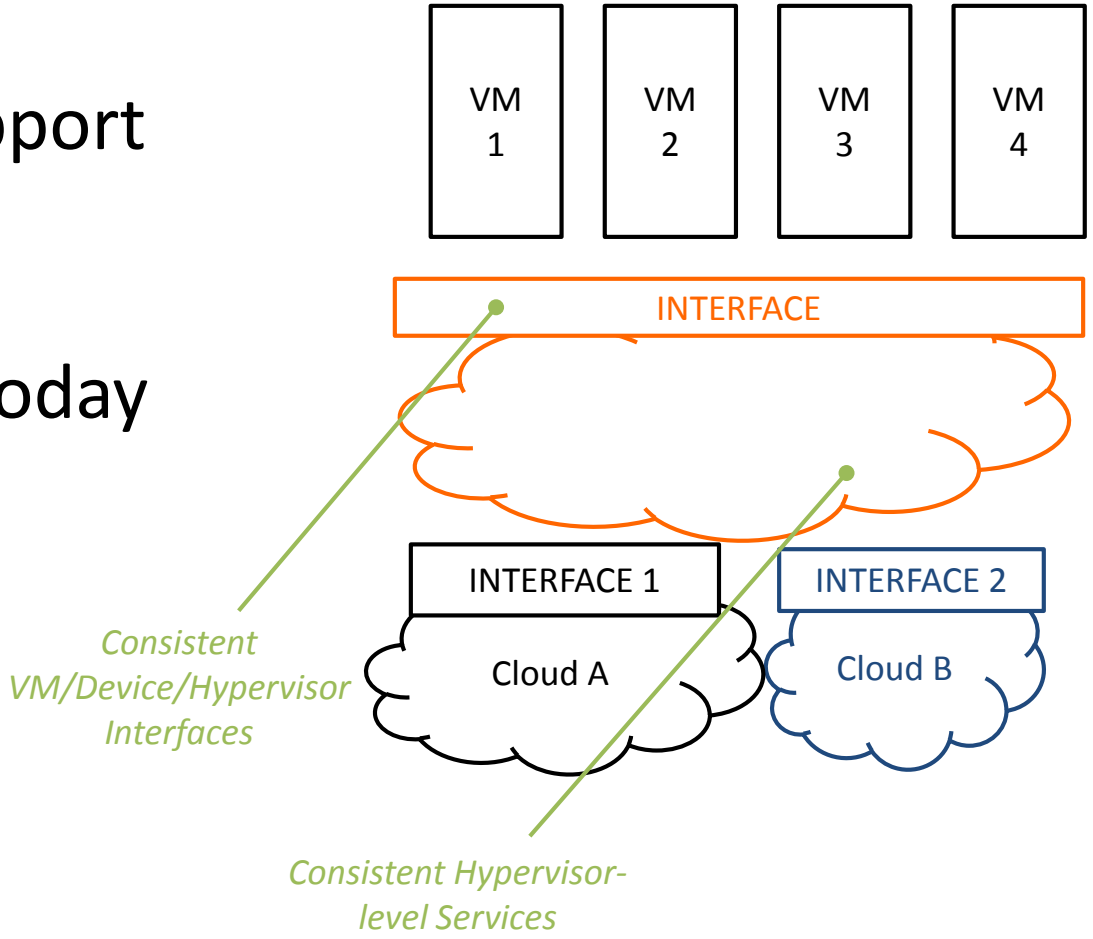
provider-centric homogenization

- Rely on support from provider
- May take years, if ever (e.g., standardization)
- "Least common denominator" functionality

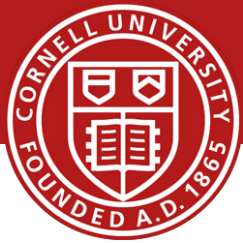


user-centric homogenization

- No special support from provider
- Can be done today
- Custom, user-specific functionality



nested virtualization approaches



- Require support by bottom level hypervisor

  No modifications to top-level hypervisor

  The Turtles Project (OSDI'10) (provider-centric)

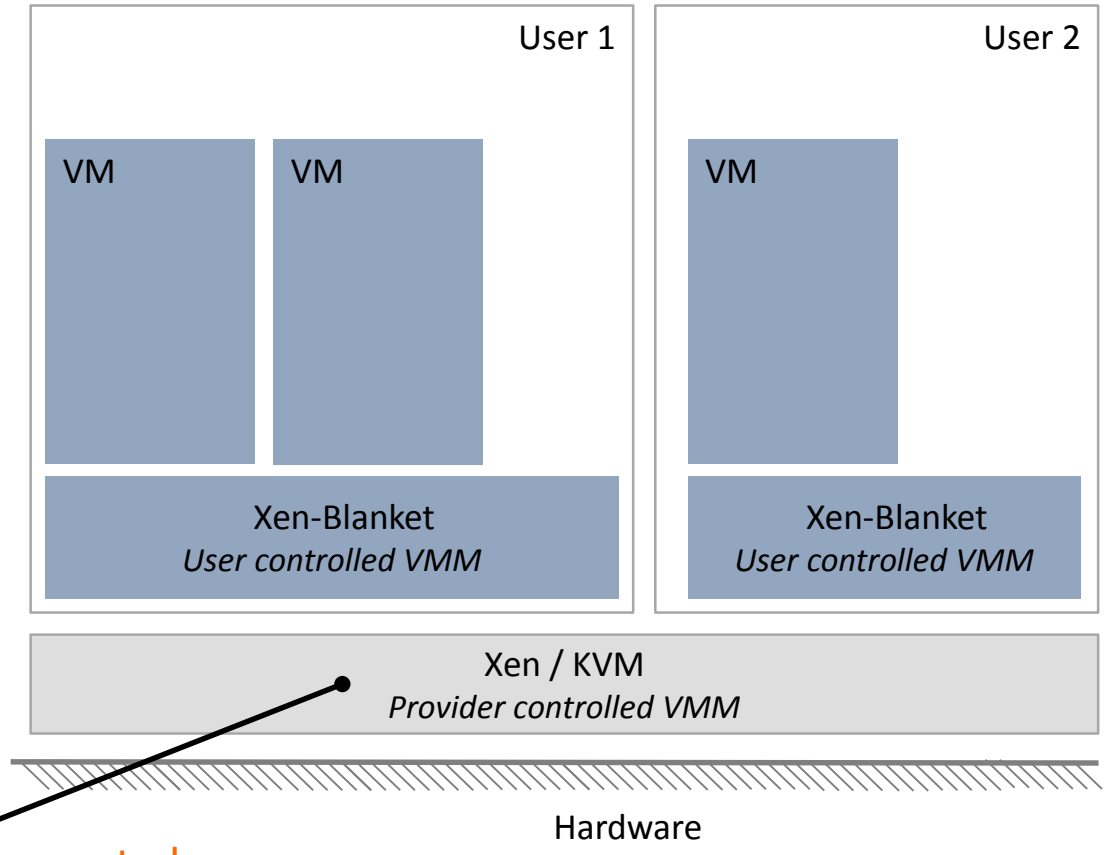
- No support from bottom level hypervisor

  Modify top-level hypervisor

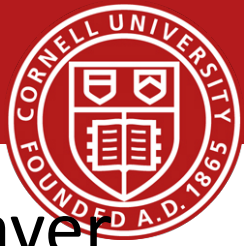
  The Xen-Blanket (user-centric)

the xen-blanket

- Assumption:
  - Existing clouds provide full virtualization (HVM)
- Future work:
  - Xen-Blanket in paravirtualized guest



without hypervisor support

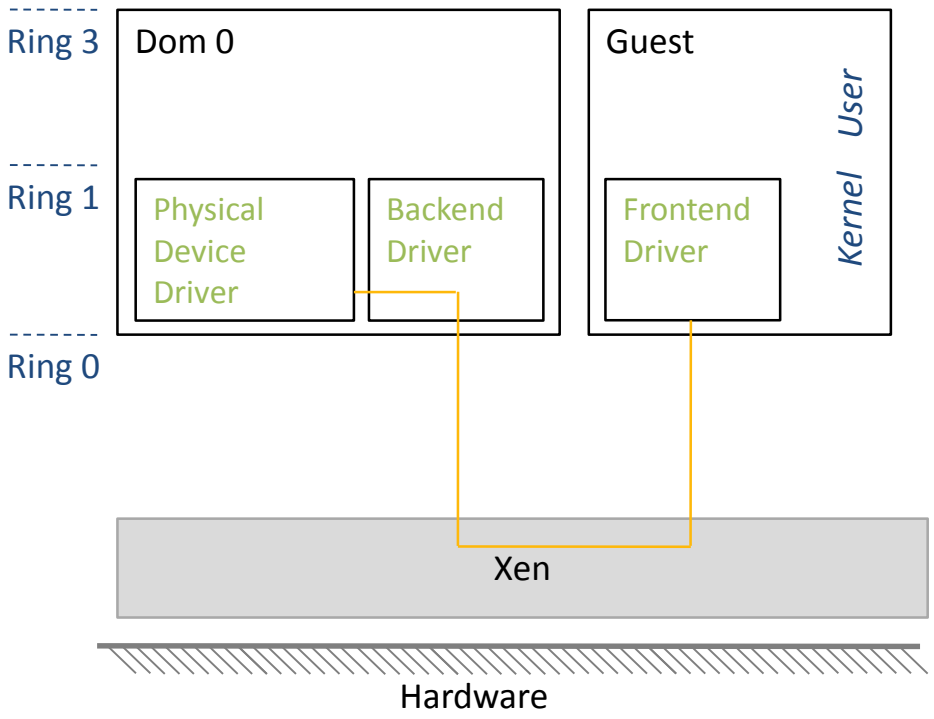


- No virtualization hardware exposed to second layer
  - Can use paravirtualization or binary translation
  - We use paravirtualization (Xen)

- Heterogeneous device interfaces
  - Create set of Blanket drivers for each interface
  - We have built drivers for Xen and KVM (**`virtio`**)

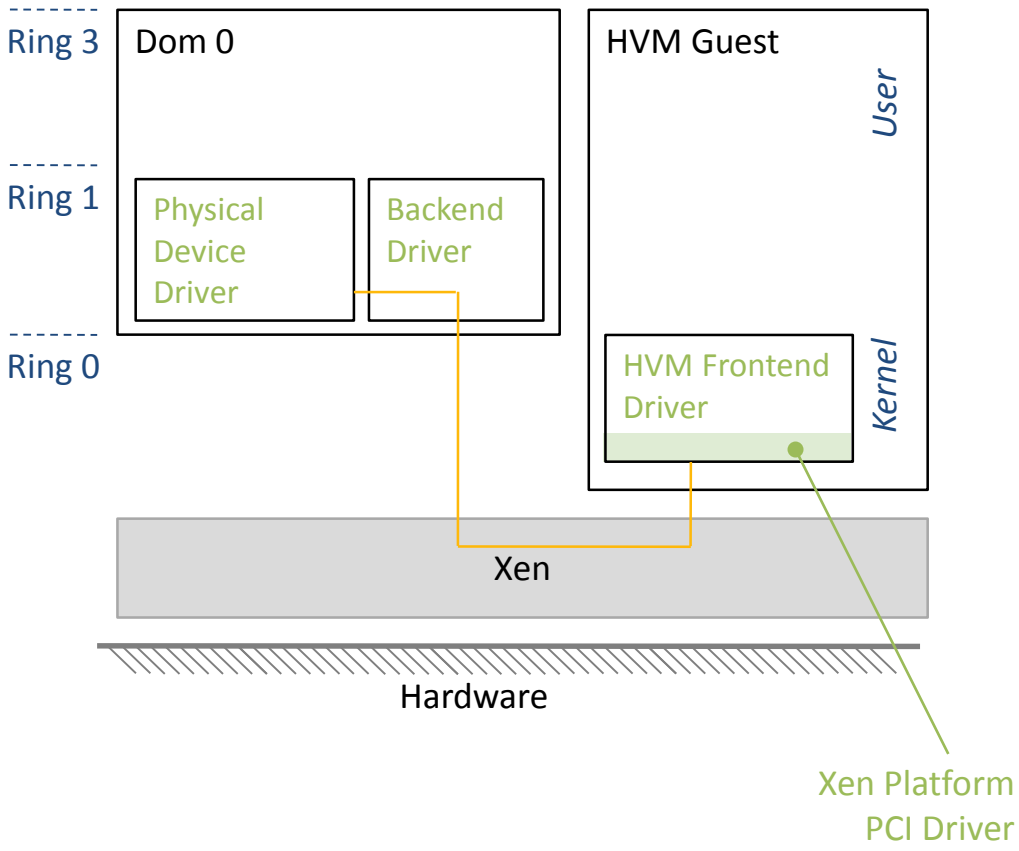
PV device I/O

- Paravirtualized device I/O essential for performance
- Domain 0 hides physical device details from guests



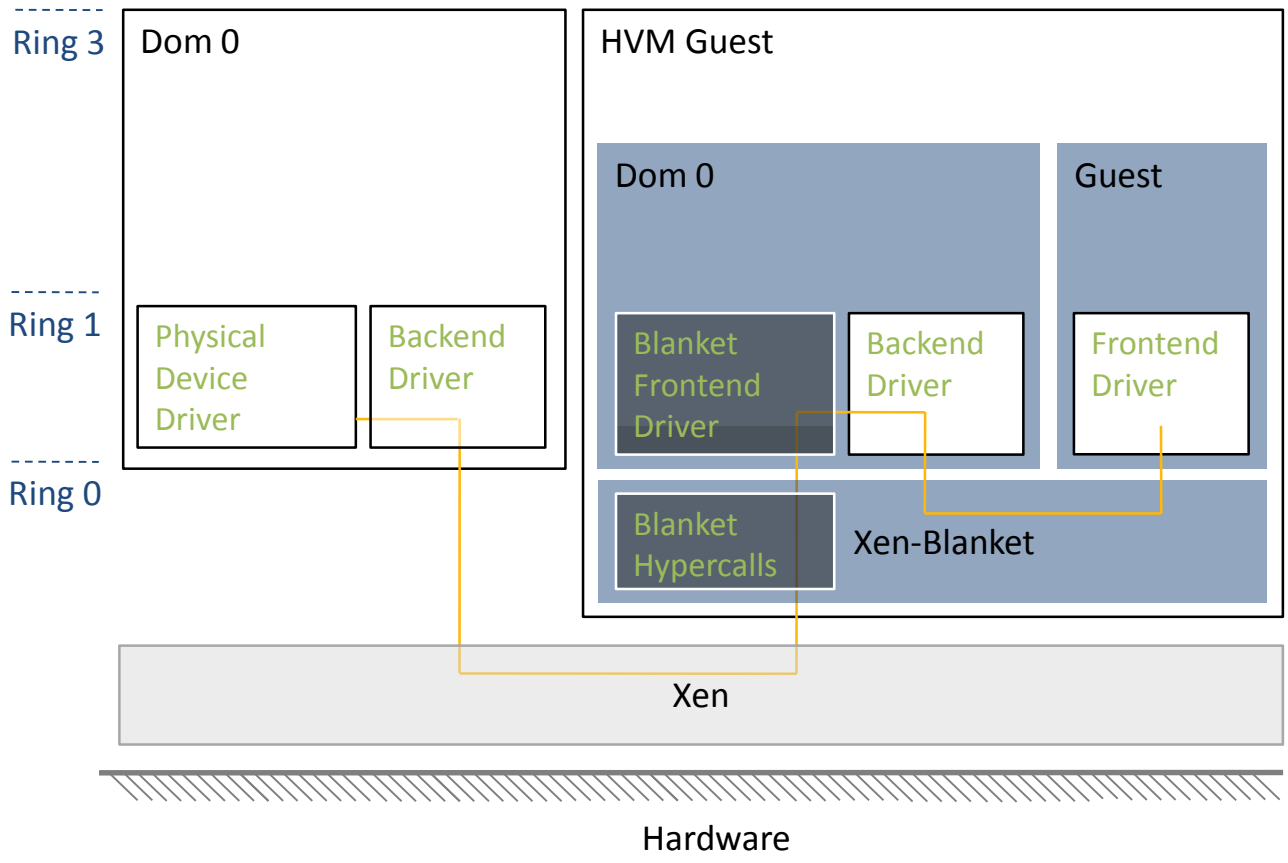
PV-on-HVM

- HVM guest still needs PV device I/O
- Platform PCI Driver makes Xen internals look like PCI device
- Physical device details still hidden from guests



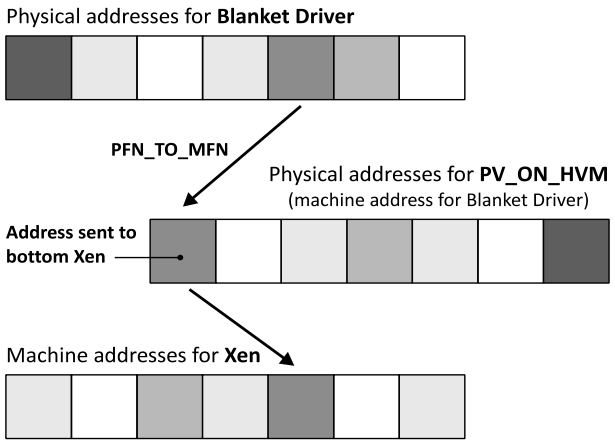
blanket drivers

- Physical device details are hidden from entire Xen-Blanket instance
- Blanket Frontend Driver interfaces with provider-specific device interface
  - like PV-on-HVM
- Provider-specific device interface details are hidden from second-layer guests



technical details

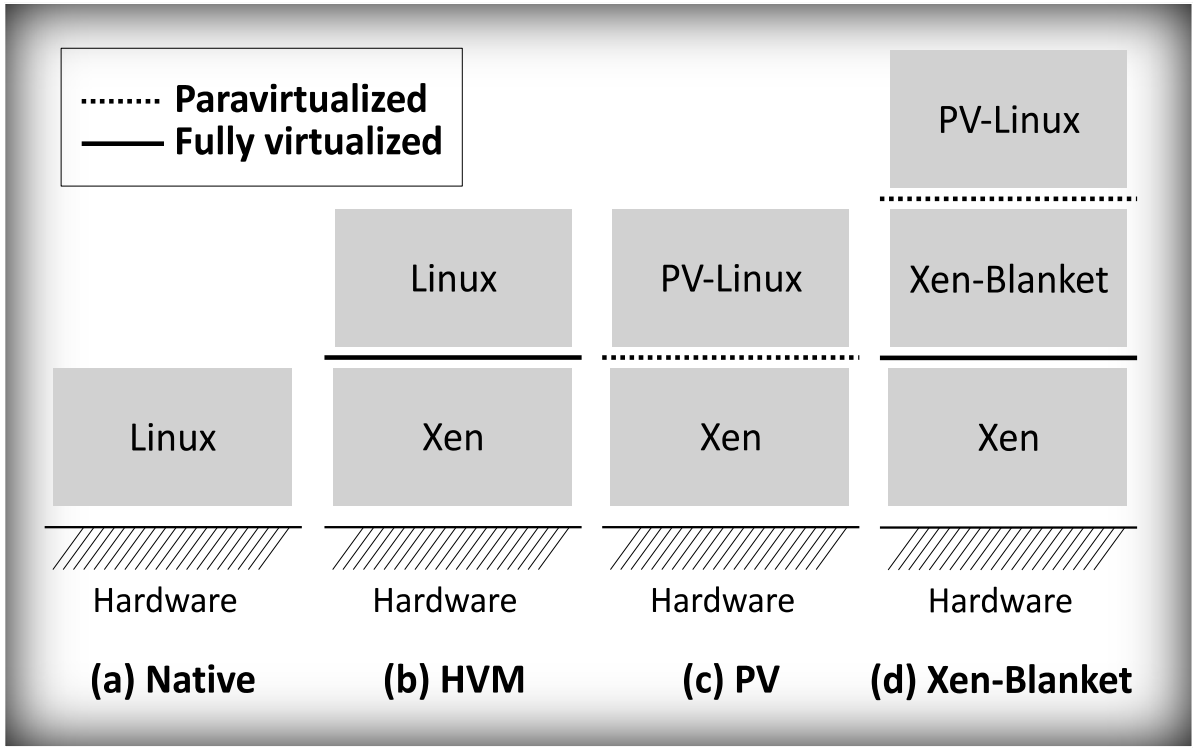
- Address translation
  - Virtual addresses are two translations from machine addresses (needed for DMA)
- Hypercall assistance
  - Communication between frontend blanket driver and backend driver
  - vmcall must be issued from ring 0
  - Most hypercalls are passthrough



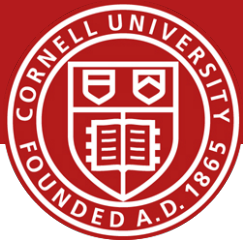
Many more details in thesis

overhead evaluation setup

Used up to 2 physical hosts (six-core 2.93 GHz Intel Xeon X5670 processors, 24 GB of memory, four 1 TB disks, and 1 Gbps link)



Imbench microbenchmarks

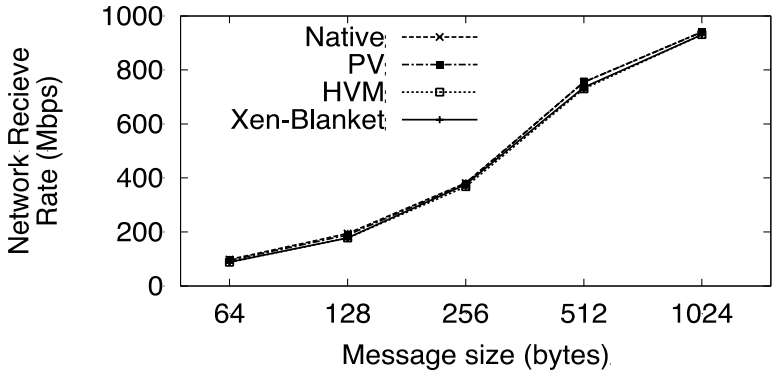


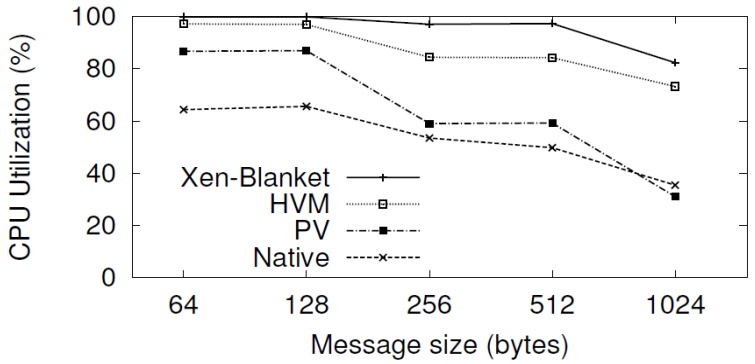
	Native *(µs)*	HVM *(µs)*	PV *(µs)*	Xen-Blanket *(µs)*
Null Call	0.19	0.21	0.36	0.36
Fork Proc	67	86	220	258
Ctxt switch (2p/64K)	0.45	0.66	3.18	3.46
Page fault	0.56	0.99	2.00	2.10

Compare Xen-Blanket to PV

blanket driver overhead

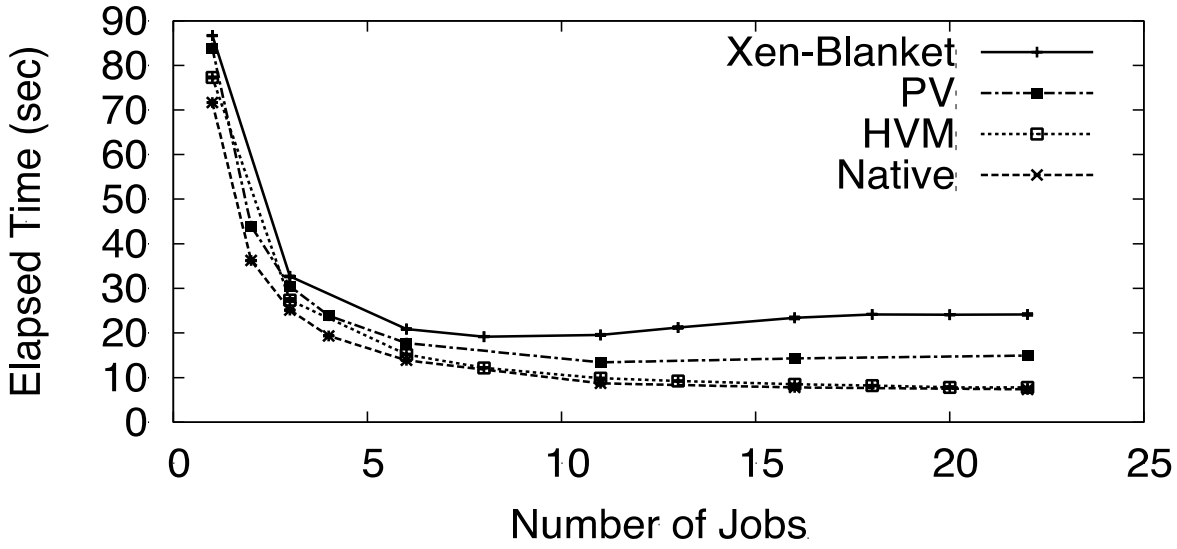
- Two VMs on two physical hosts using netperf
- Can receive at line speed on 1Gbps link
- Within 15% CPU utilization of single layer



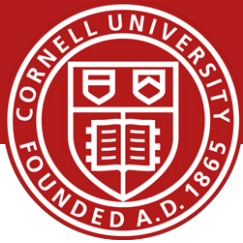


kernbench

- Up to 68% overhead on kernbench
  - APIC emulation causes many vmexits



user-defined oversubscription



Type	CPU (ECUs)	Memory (GB)	Disk (GB)	Price ($/hr)
Small	1	1.7	160	0.085
Cluster 4XL	33.5	23	1690	1.60
Factor	33.5x	13.5x	10x	18.8x

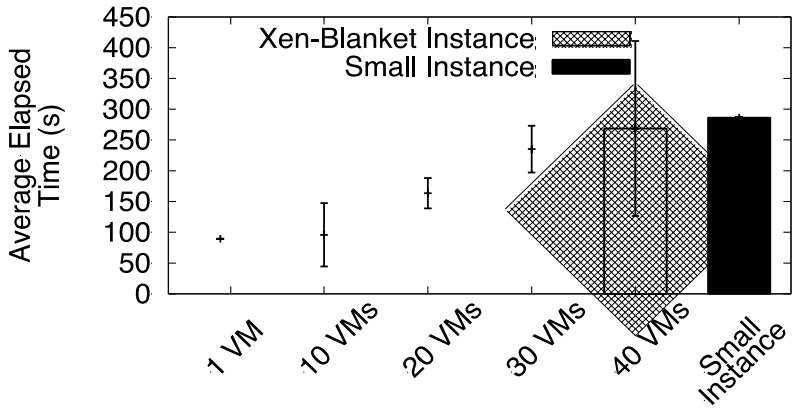
Resources do not all scale the same as price

Opportunity to exploit CPU scaling

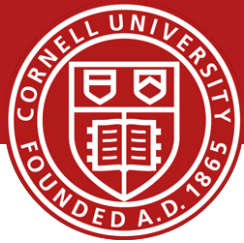
kernbench revisited

- kernbench kernel compile benchmark
- Rent one 4XL EC2 instance
- Use Xen-Blanket to partition it 40 ways
- All instances (on average) finished the same time as EC2 small instance

- 47% price reduction per VM per hour



cloud interoperability



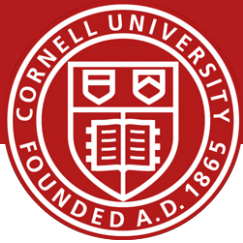
The Xen-Blanket

- User-centric homogenization
- Nested virtualization without support from underlying hypervisor
- Runs on today's clouds (e.g., Amazon EC2)
- Download the code:
  - `http://code.google.com/p/xen-blanket/`

New opportunities

- *performance*: user-defined oversubscription

Before Next time



- Project Interim report
  - **Due Monday, November 24.**
  - And meet with groups, TA, and professor
- Fractus Upgrade: Should be back online

- ***Required review and reading for Wednesday, November 19***
  - Extending networking into the virtualization layer, B. Pfaff, J. Pettit, T. Koponen, K. Amidon, M. Casado, S. Shenker. ACM SIGCOMM Workshop on Hot Topics in Networking (HotNets), October 2009.
  - http://conferences.sigcomm.org/hotnets/2009/papers/hotnets2009-final143.pdf

- Check piazza: http://piazza.com/cornell/fall2014/cs5413
- Check website for updated schedule