Recent trend of OSS Virtualization development

Isaku Yamahata <yamahata@private.email.ne.jp>

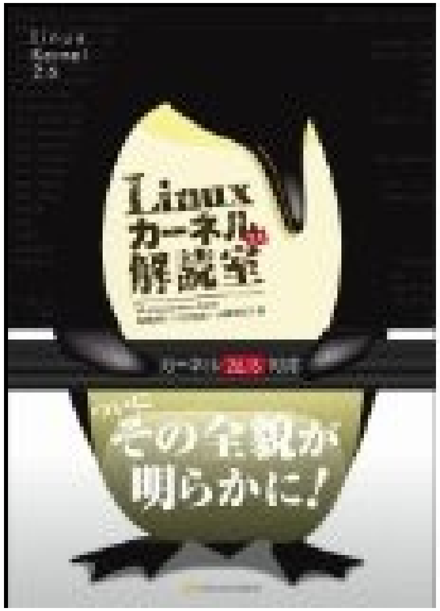
North Asia OSS promotion forum training camp 2012
November 14, 2012

Agenda

- Who am I?
- Technology trend
- Developing areas
- Summary

Who am I?

- Software engineer
- Has been contributed to OSS OS/virtualization related technologies for 7+ years
- Publications
  - Linux in details (Linux Kaidokushitu)
- Magazine
  - Xen in details (Xen Kaidokushitu)
  - Latest KVM virtualization technology(KVM no saishin kasouka gijutu)

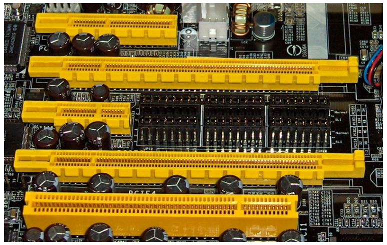


Contributed Projects

- Xen
- KVM/QEMU
- Open vSwitch
- Ryu
- OpenStack
  - Nova
  - Quantum

KVM/QEMU

- New chipset and PCI express support
  - Didn't complete it unfortunately
  - However, other developers started to revitalize
    - Seems making it into the upstream for 1.4
    - Good OSS community collaboration



From wikipedia

- Yabusame: postcopy live-migration
  - On-going work
  - Others also working on another implementation with RDMA

Yabusame is a joint project with Takahiro Hirofuchi, AIST and Satoshi Itoh, AIST.

This work is partly supported by JST/CREST ULP and KAKENHI (23700048).
The development of Yabusame was partly funded by METI (Minister of Economy, Trade and Industry) and supported by NTT Communications Corporation.

From wikipedia

Ryu and Open vSwitch

- Network virtualization
- SDN(Software Defined Network)
  - Openflow protocol
  - Make network programmable
- Ryu
  - Network Operating System
  - Openflow Controller
  - Integration with OpenStack
  - Multi tenant support
    - Mac-based L2 segregation
    - GRE tunneling
- Open vSwitch
  - Various contribution

Ryu

OPEN VSWITCH

An Open Virtual Switch

Ryu: Network Operating System

流

Flow

龍

Oriental Dragon

Open-sourced network operating system

Network operating system

- Logically centralized controller for managing thousands of network switches
- A platform for building network applications to manage switches

Open source software (Apache v2)

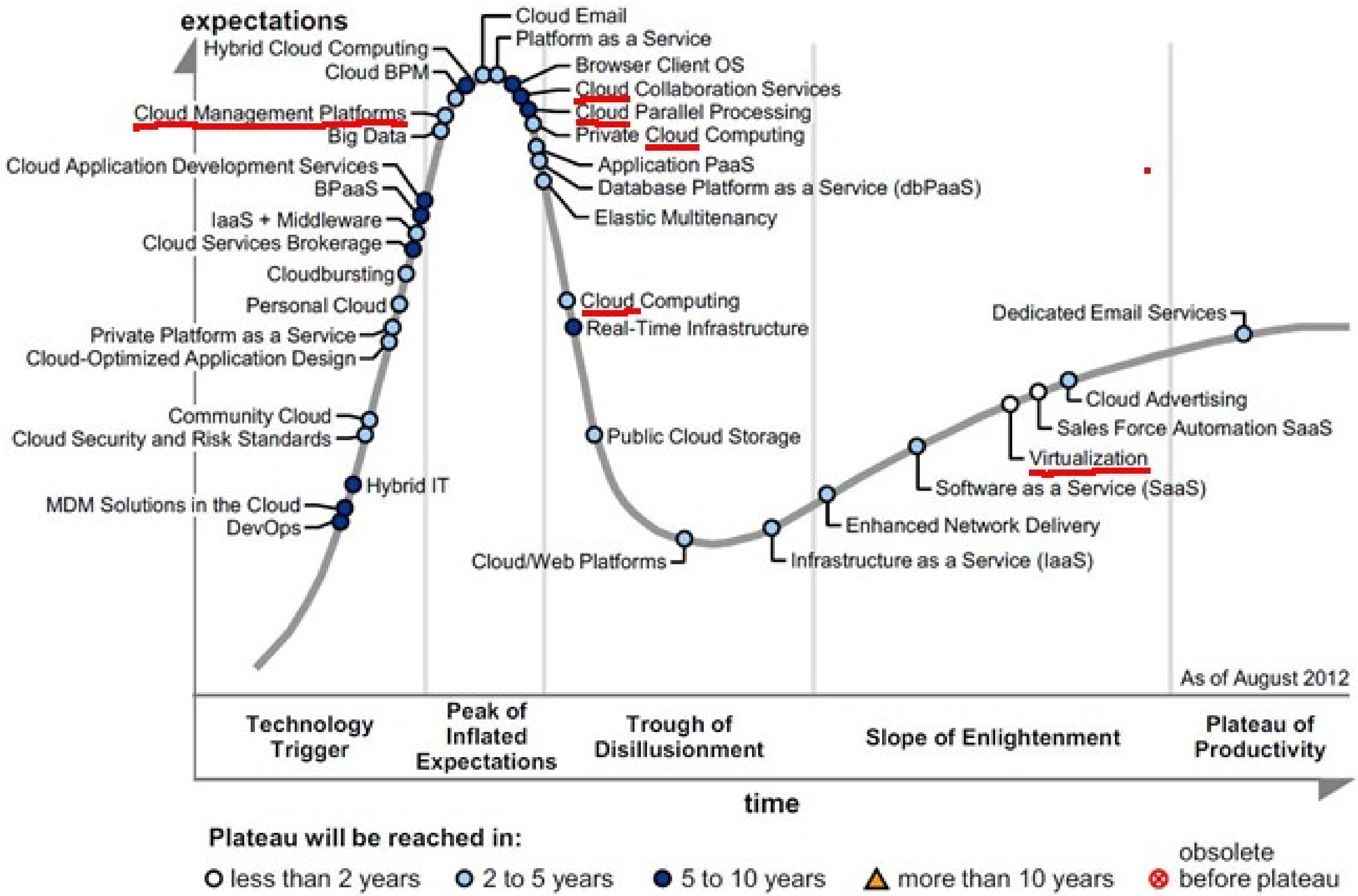
- Fully written in Python
- Project site: http://osrg.github.com/ryu/

OpenStack

- Cloud Management System
- Nova compute:
  - Boot-from-volume
  - Corresponds to AWS EBS boot
- Quantum: network
  - Ryu-plugin

Technology trend

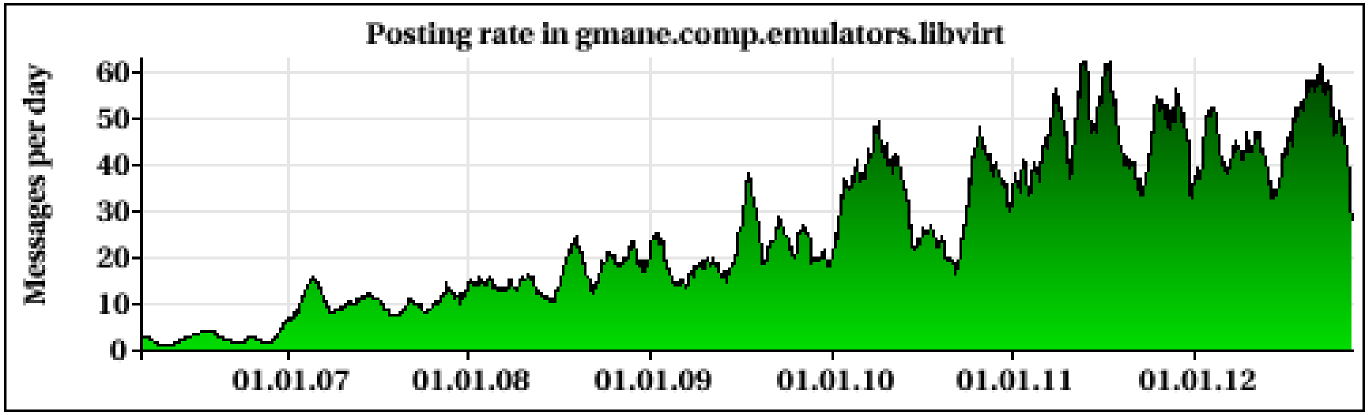
Garnter Hype-cycle



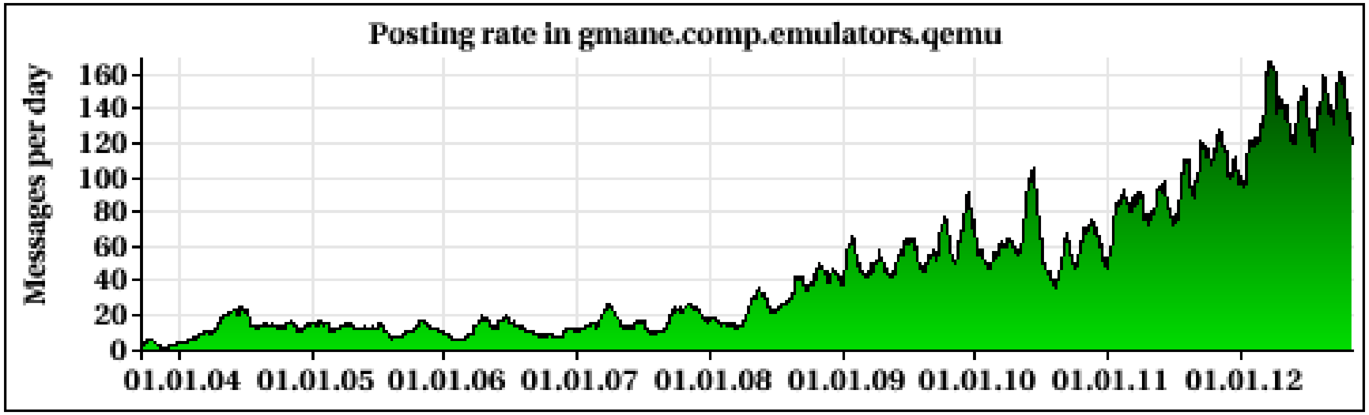
Source: Gartner (August 2012)

Development activity

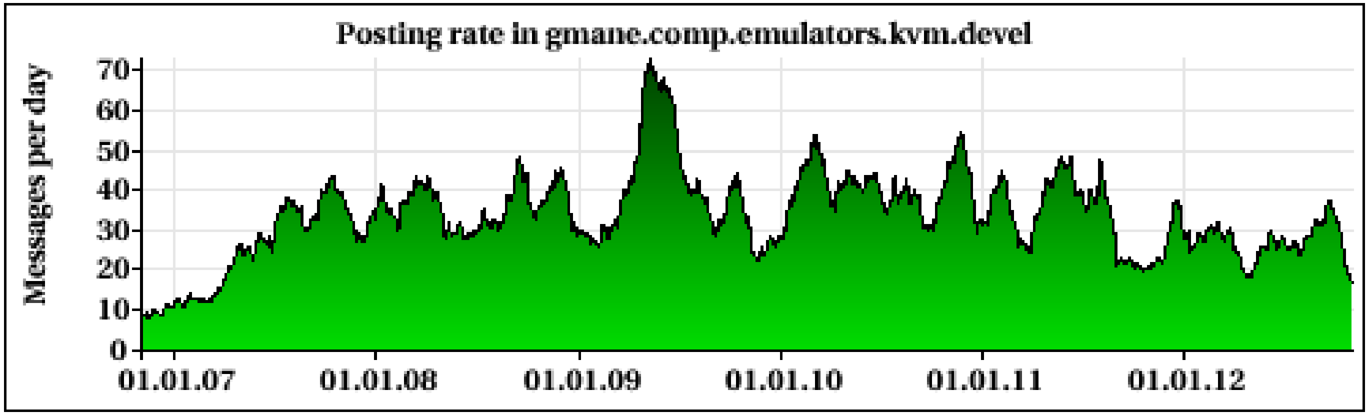
libvirt



qemu



kvm



NOTE: scale is different

From: gmane.org

Another facts

- VMWare (Virtualization Giant) joined Open Stack foundation
    - Even they have their own products
    - Committed to contribute
- Microsoft supports various guest OSes
    - Non-windows OS

Development Trend

- KVM: virtualization core technology
  - Cpu virtualization
- QEMU: virtualization technology that covers wider area
- Libvirt: management of virtualization technology
- Its forcus has shifted to surrounded area
  - Focus of core virtualization has move into scalability/usability
  - How to use virtualization technology better

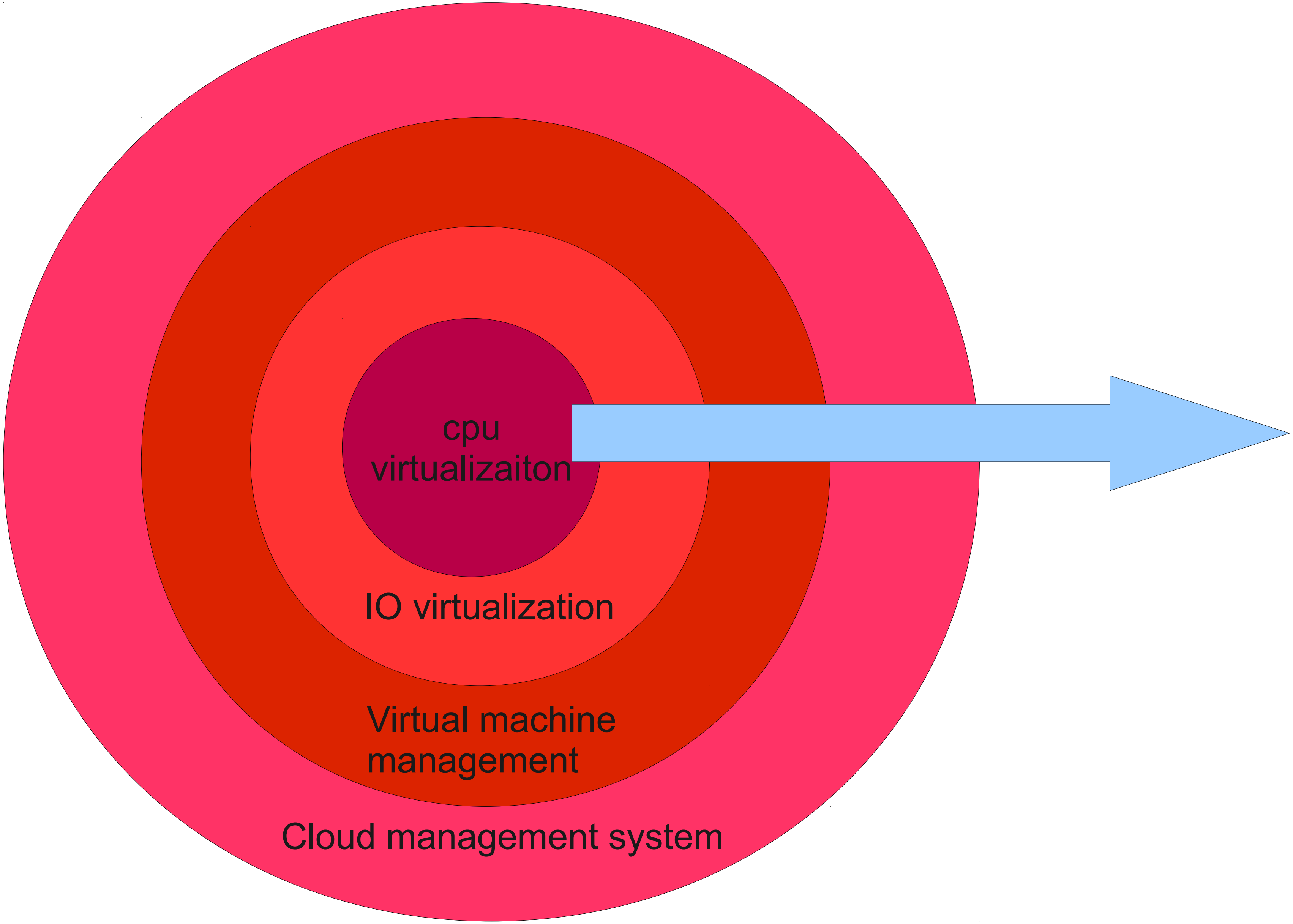
Software → hardware

- Software solution will become needless when hardware supports it
- If something is optimized by software, then hardware supports it directly.
- We are working to make our software achievement needless.
  - Software optimization proves that it's worthwhile for hardware optimization
- Difficult to differentiation
- Example
  - Any kinds of paravirtualization
    - VMX, SVM
    - Pause loop
    - Apic/interrupt virtualization
    - SR-IOV

New hardware feature

- Accessed/dirty flags for EPT
- VMFUNC
  - Vmfunction 0: EPTP switching
    - allows
    - Loads EPTP from EPTP list
- Interrupt/APIC virtualization
  - APIC-register virtualization
  - Virtual-interrupt delivery

Virtualizatin: core to surrounding



cpu virtualizaiton

IO virtualization

Virtual machine management

Cloud management system

Technology drivers

- Usage model is driving virtualization technology
- Cloud computing
  - Green
  - Memory: density
  - Power consumption
- Security
- Bigdata
- Mobile/embedded
  - ARM
  - Realtime

Developing areas

Existing technology

- Hypervisor
  - Bhyve
  - bitvisor
- Container,, OS virtualization
  - Linux Virtual Server
  - Cgroup, namespace
  - LXC
  - OpenVZ
  - LVS(Linux Virtual Server)
- BIOS
  - Seabios
  - Tiano core
- Libvirt
- Virt-manager
- oVirt
- Cloud management software
  - Openstack, cloud stack

Hardware emulation

- Kvm-tools: simple, easy to understand
- Qemu
- Threading
  - Removing
- Device modeling
  - Qapi
  - Live-migration
- New hardware
  - IOMMU

Hardware emulation(cont)

- Correct hardware emulation is difficult
  - Functionality emulation for virtualization
    - Not cycle accurate emulation(signal emulation)
  - With reasonable performance
  - But reasonable hardware modeling is required
- Hardware is
  - Asynchronous
  - Executes independently

BIOS

- Classic PC BIOS
- EFI
  - X86, Arm
  - Tiano core
  - EFI-application
  - Drivers
- ACPI

Scalability/stability

- Scalability
  - vcpu
  - Memory
  - Devices
- Stability under load
  - Live-migration
    - RDMA

Memory

- Memory aggregation
- Memory compression
- Transendent memory
  - Cleancache, frontcache
  - Zram, zcache
  - Ramster

Hot plug/unplug

- Cpu
- Memory
  - Dimm modeling
- Device
  - PCI/PCIe device
  - Serial ATA
  - USB
  - SCSI
  - ...
- ACPI support

ARM virtualization

- ARM introduced virtualization extension
- KVM/ARM, XEN/ARM is under heavy development
- For
    - Embedded
    - ARM server
- Would follow similar path of x86
- But with ARM own requirement

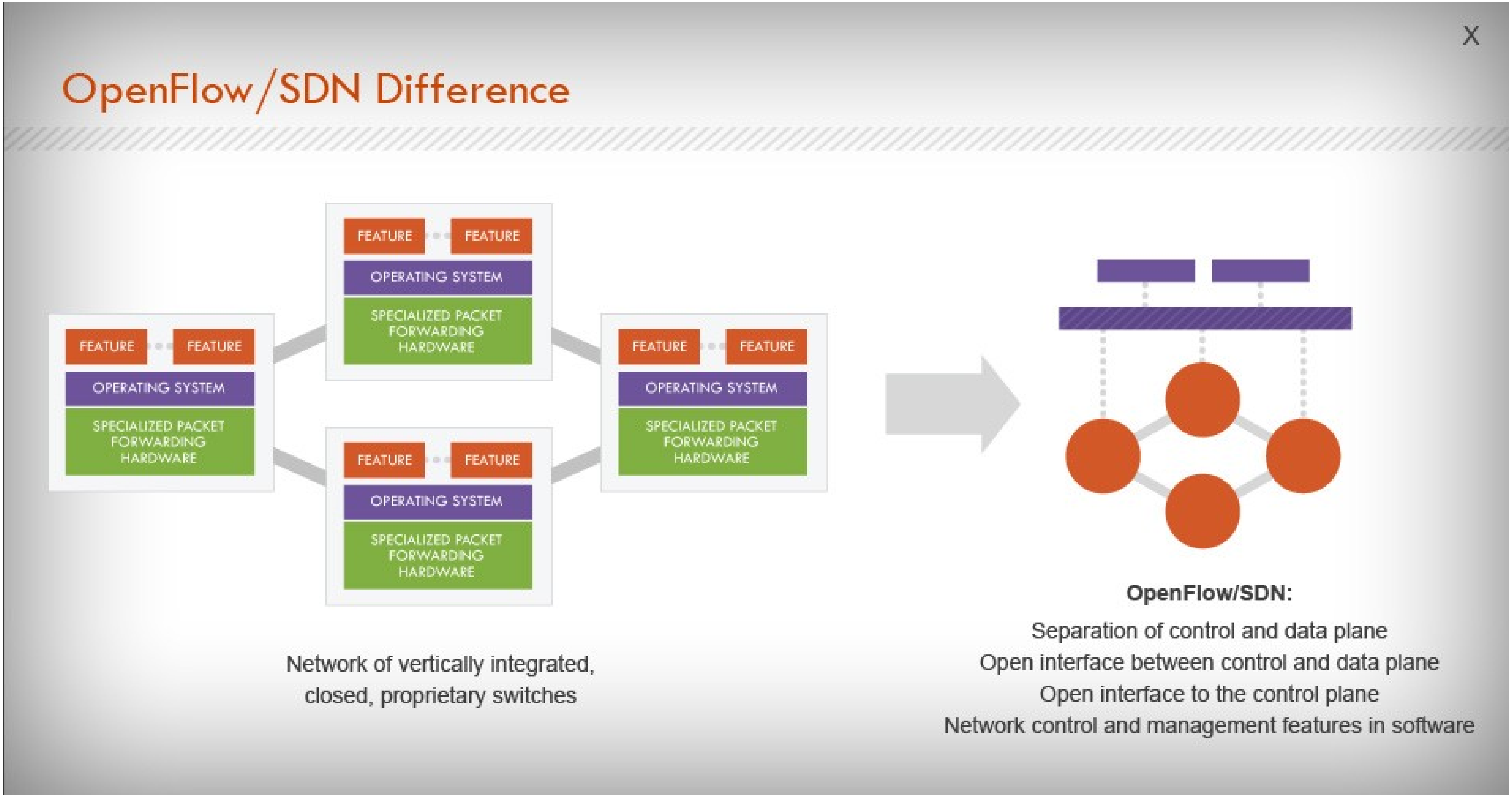
Embedded

- Many arechtectures
  - ARM, PowerPC
- Embedded
  - Power consumption
  - Big.LITTLE architecture
  - Less overhead
- Realtime
  - Hard-realtime
  - Soft-realtime
  - latency

Networking

- Networking is behind other areas in virtualization
- Open vSwitch
- OpenFlow
- SDN
- Optimization
  - Multiqueue
- Tunneling
  - VXLAN
  - NVGRE
  - STT

OpenFlow/SDN


X
FEATURE
FEATURE
OPERATING SYSTEM
SPECIALIZED PACKET FORWARDING HARDWARE
FEATURE
FEATURE
OPERATING SYSTEM
SPECIALIZED PACKET FORWARDING HARDWARE
FEATURE
FEATURE
OPERATING SYSTEM
SPECIALIZED PACKET FORWARDING HARDWARE
FEATURE
FEATURE
OPERATING SYSTEM
SPECIALIZED PACKET FORWARDING HARDWARE

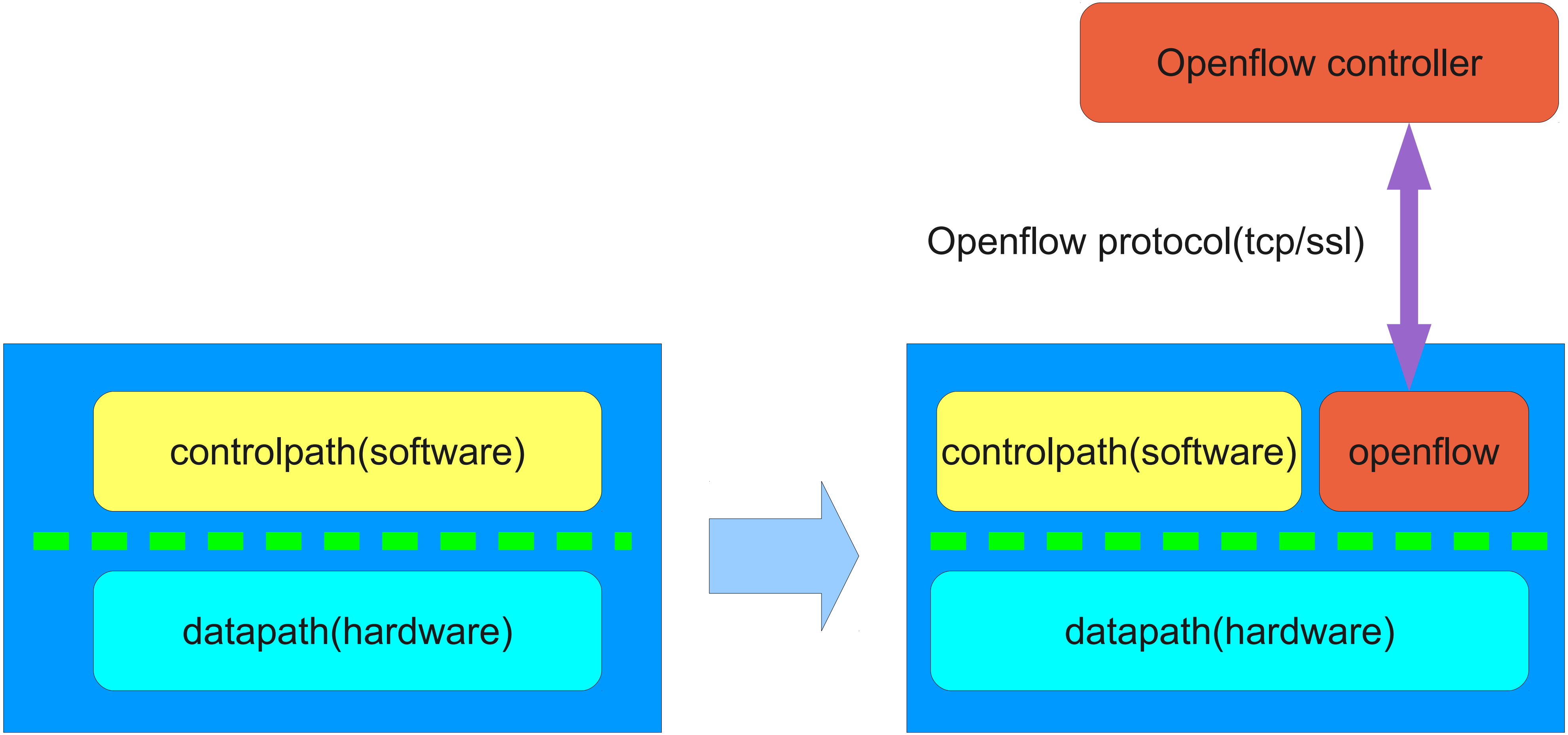
OpenFlow/SDN Difference

Network of vertically integrated,
closed, proprietary switches

OpenFlow/SDN:

Separation of control and data plane
Open interface between control and data plane
Open interface to the control plane
Network control and management features in software

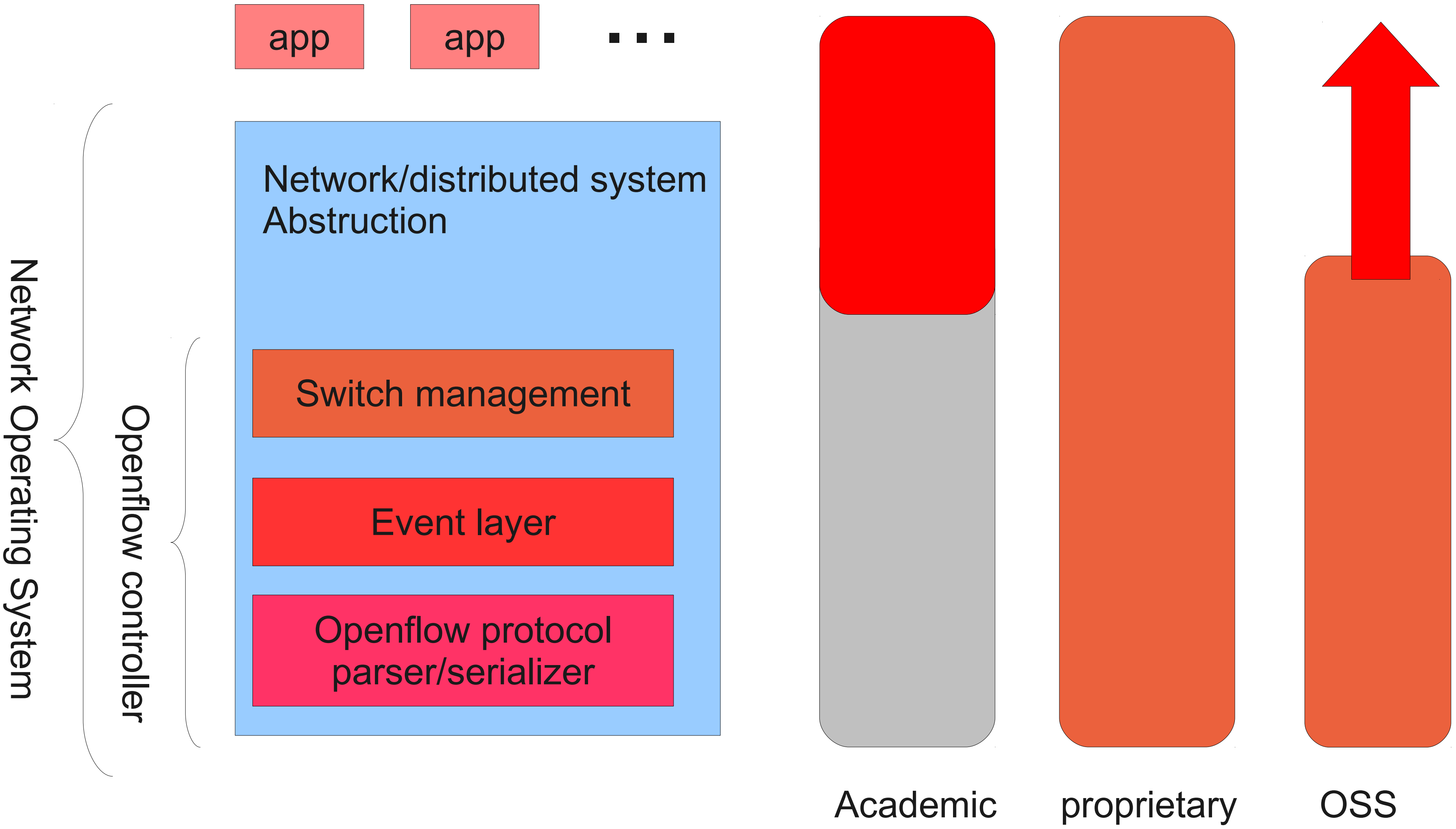
Openflow



Ethernet switch

Openflow ethernet switch

SDN and OSS


app
app
...
Network/distributed system
Abstruction
Switch management
Event layer
Openflow protocol
parser/serializer
Network Operating System
Openflow controller
Academic
proprietary
OSS

Other areas to investigate

- RAS
  - Inject errors into guest
  - Hardware partitioning
- HA, FT
- Security
  - Disaggregating security domain
  - Check pointing
- Nested virtualization
  - IOMMU
- GPU virtualization

Summary

- Virtualization has become common and widely accepted
- The developing area has shifted from core virtualization technology to related area
- There are many hot areas to contribute in virtualization

Thank you

Questions?