National Institute of Information and Communications Technology

Bridging Computing Resource Management and Network Resource Management

Eiji Kawai

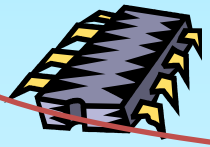
NICT

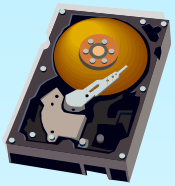
Sharing computing resources

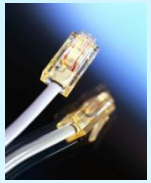
High

Cost in Remote Sharing

Low





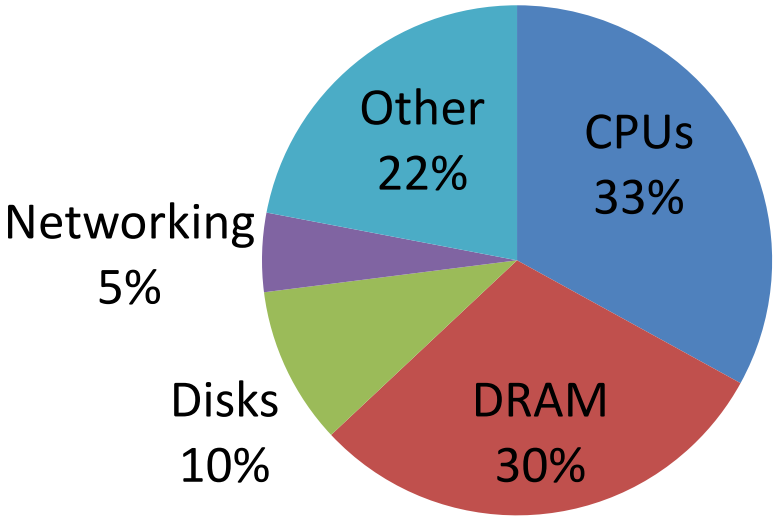


- CPU
  - Local: multi-tasking, multi-OS (VM)
  - Remote: RPC
- Memory
  - Local: virtual memory (paging), multi-OS (VM)
  - Remote: DSM (distributed shared memory)
- Storage
  - Local: file system
  - Remote: NFS, SAN, Distributed FS
- Network
  - Local: socket
  - Remote: VLAN, MPLS, overlay

Our target

Why memory?

- Issues
  - Memory provisioning is often too conservative
  - Huge power consumption in memory devices



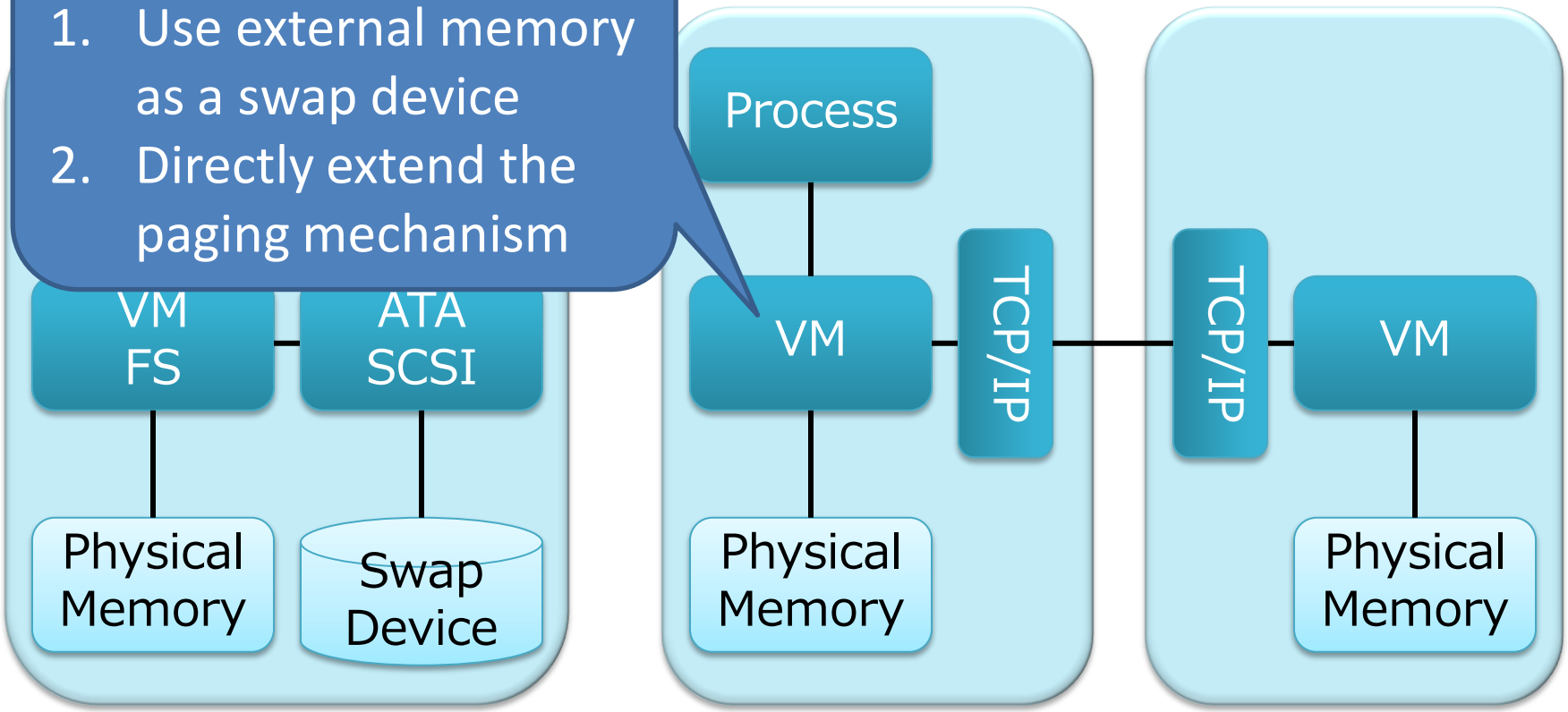
Approximate distribution of peak power usage by hardware subsystem
(Barroso and Hölzle, “The Datacenter as a Computer”)

Approaches to remote memory sharing

- VM Live Migration
  - Migrate a VM to a host with enough free memory
  - Pros: easy to use (many VMM already implement the live migration mechanism)
  - Cons: tight locality (networking and management)
- Distributed shared memory
  - Program shared memory objects through DSM APIs
  - Pros: high flexibility (programmability)
  - Cons: high implementation cost and low manageability
- Remote memory mapping
  - Import a remote memory area into a local virtual memory space
  - Pros: high manageability
  - Cons: high deployment cost (no de fact standard tech.)

Remote memory mapping

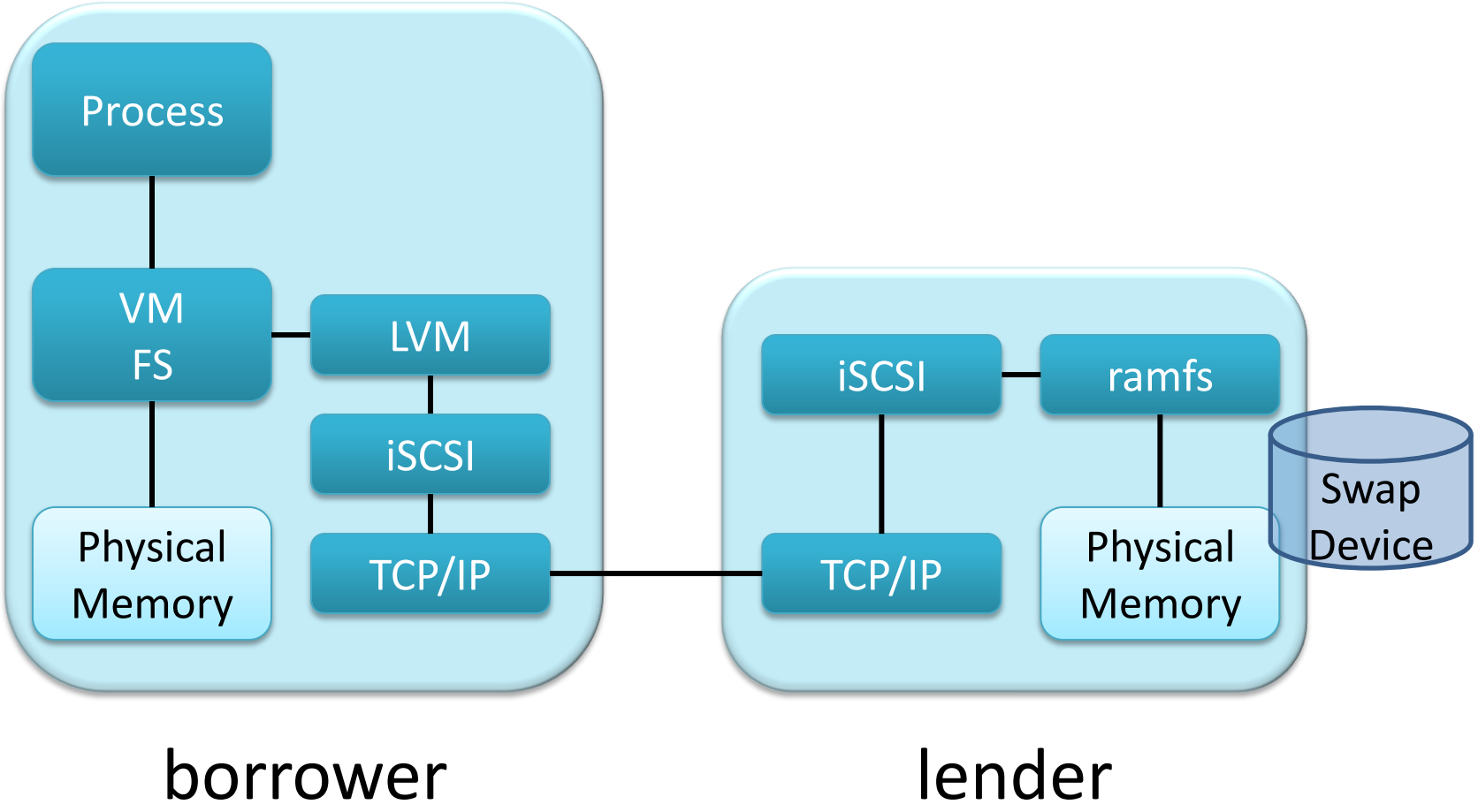
Two typical approaches
1. Use external memory as a swap device
2. Directly extend the paging mechanism



Traditional System

Extended System

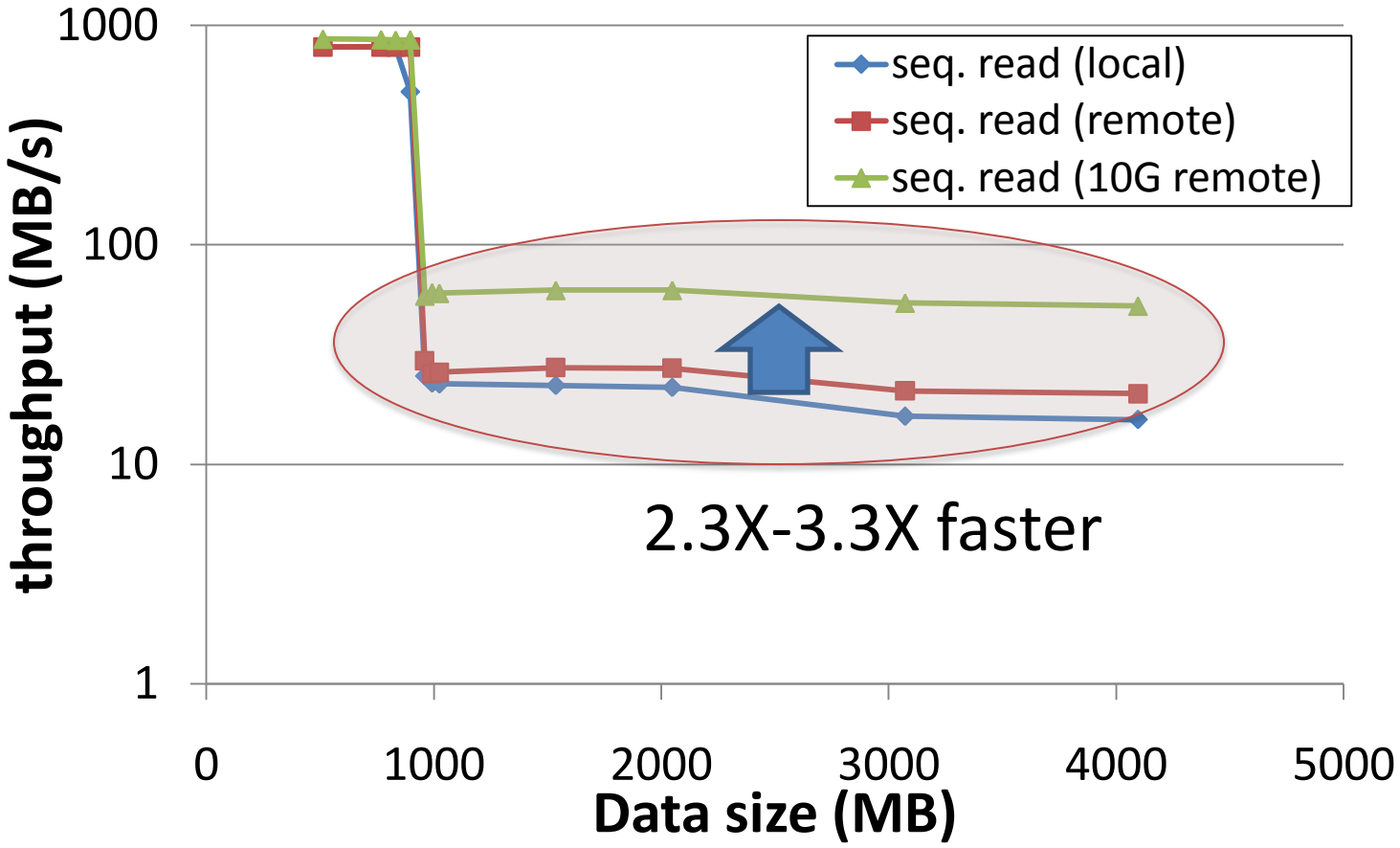
Our approach: Indirect approach with Ramfs, LVM, iSCSI



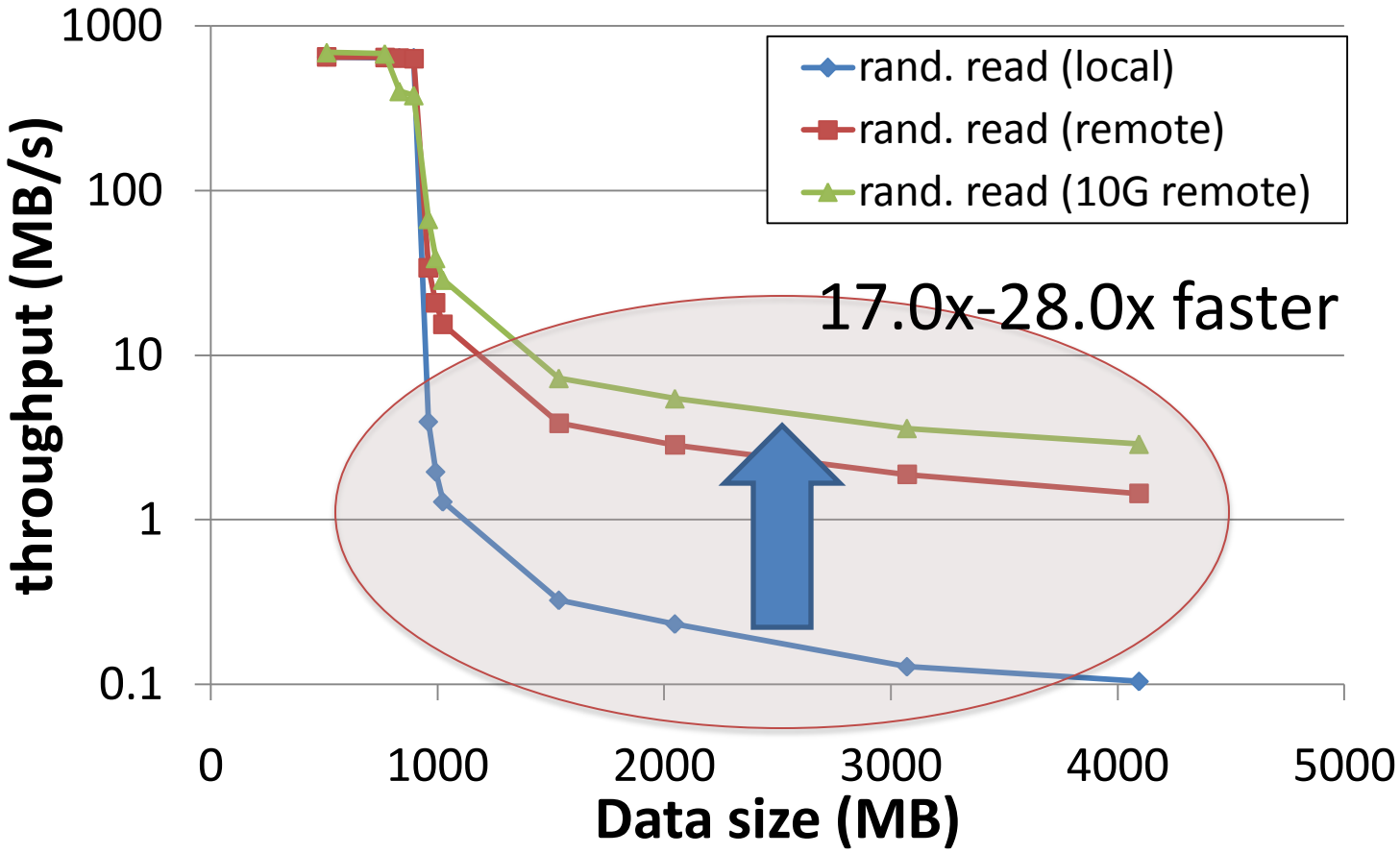
Benefits

- Based on de fact standard technologies
  - Many OSs implements the memory file system mechanism, the logical volume management mechanism, and iSCSI functions
  - Work in a heterogeneous environment (multi-OS environment)
  - Long product life
- Flexible configuration
  - LVM can switch swap devices dynamically
  - RAID is possible
- Integration to VMM
  - Invisible from guest OSs
  - Integrated management framework

Performance: sequential read



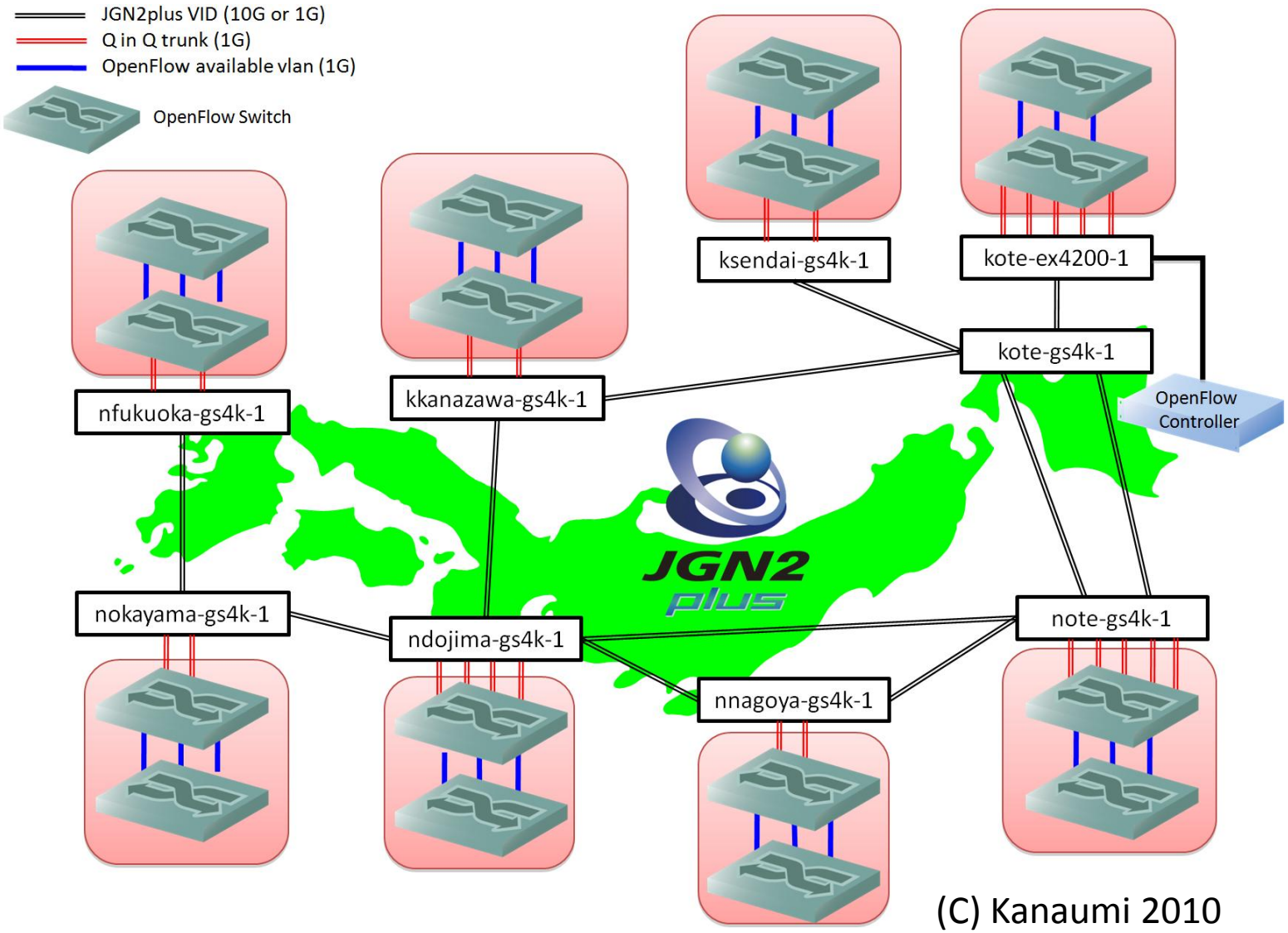
Performance: random read



Next step

- Optimize networking
  - Two control layers
    - Memory sharing
    - Networking
  - The remote host from which a remote swap is mounted is unpredictable
  - Which host has enough free memory?
  - Which host has a path with enough bandwidth?
- OpenFlow can control the networks explicitly

OpenFlow testbed on JGN2plus



Thank you!!