HTTP-FUSE PS3 Linux: an internet boot framework with kboot

http://openlab.jp/oscirclar/

Kuniyasu Suzaki and Toshiki Yagi

National Institute of Advanced Industrial Science and Technology

Background

- PS3Linux makes us cell programming.
- Installing Linux to PS3 is painful
  - Installation time is too long!
- Then, how about Live-CD?
  - Yes, but they can never upgrade itself.

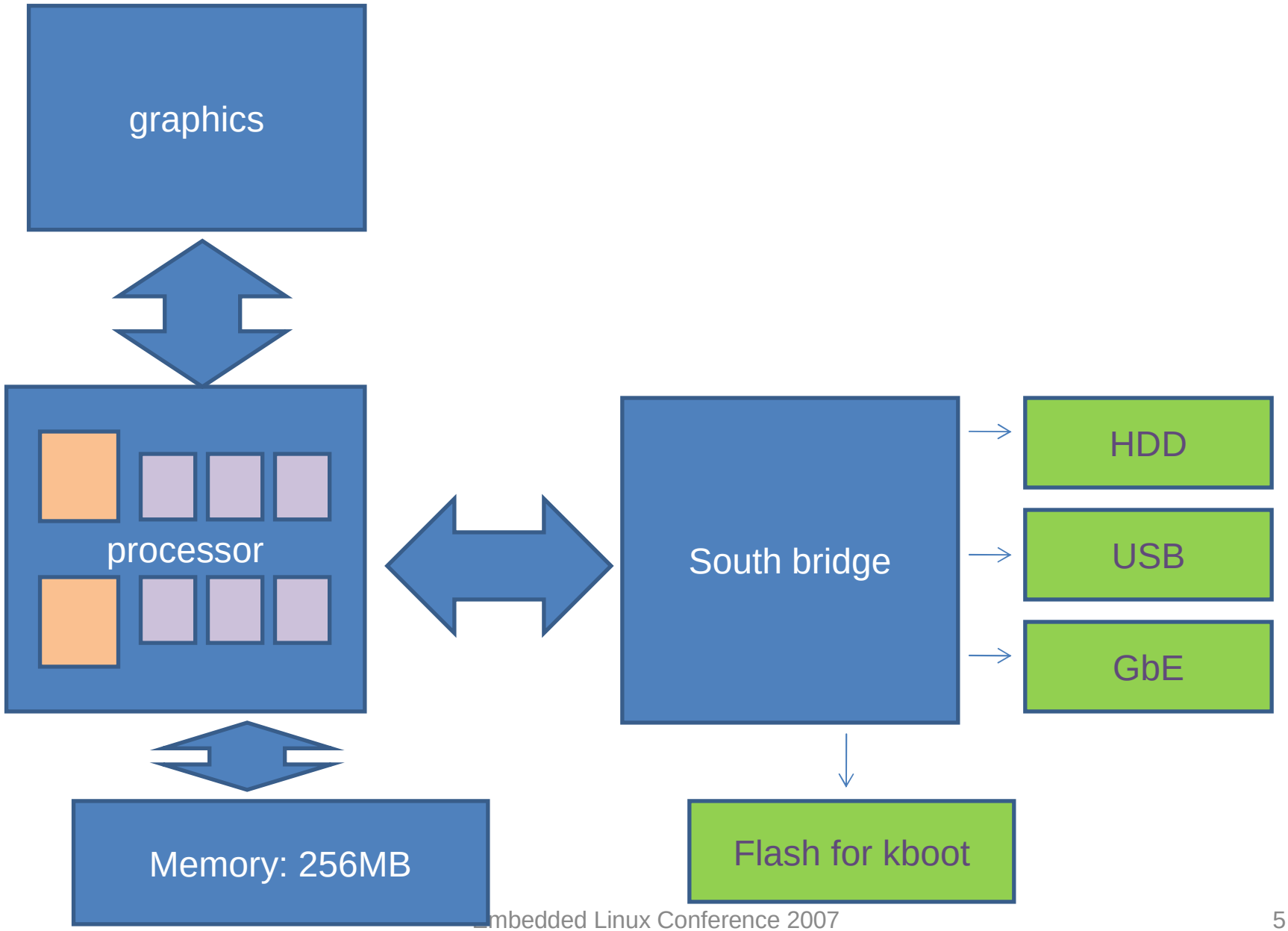
Network booting

- With network booting, it’s no need to install Linux to HDD.
- NFS boot is the most popular network booting method.

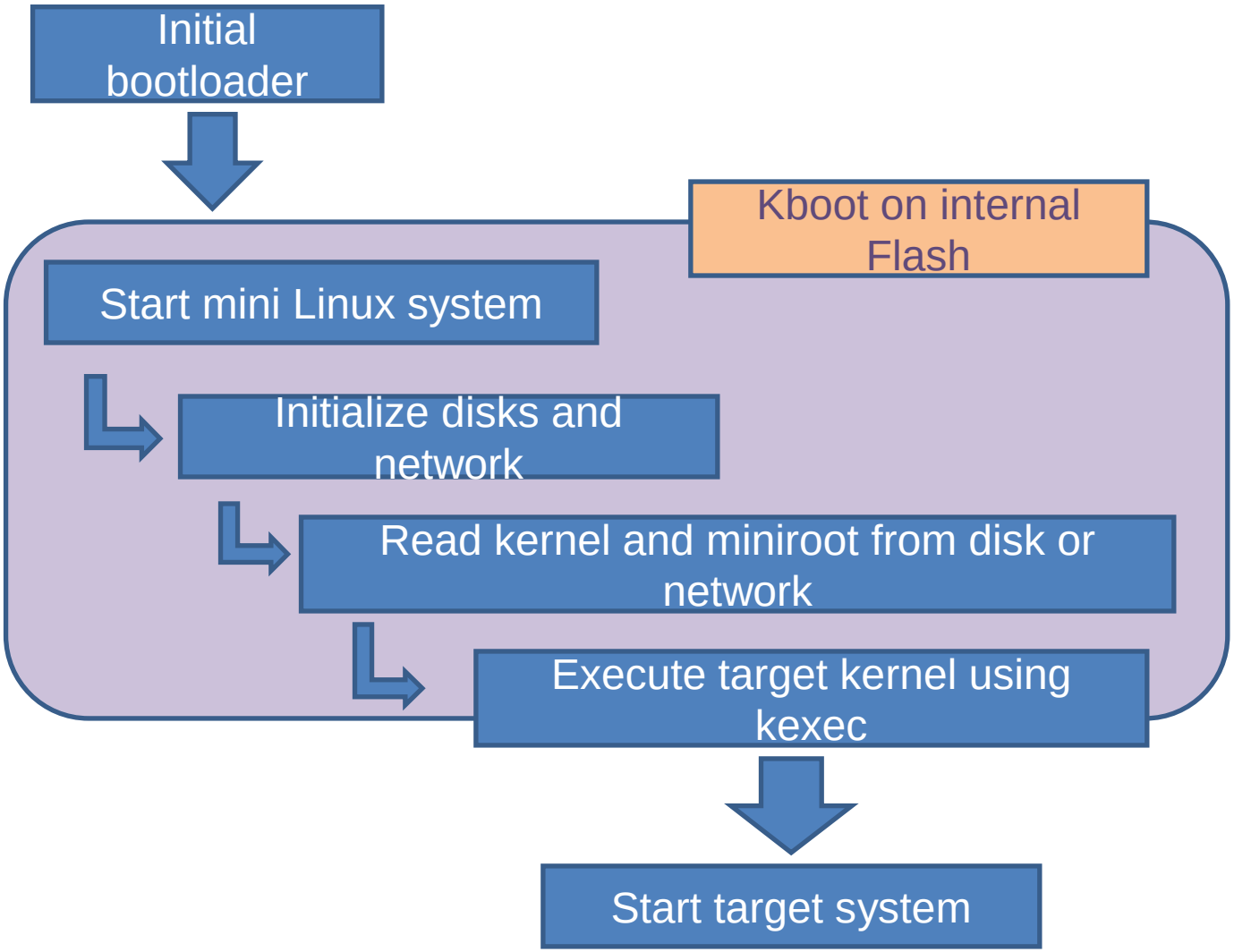
- But NFS service does not match to distribute worldwide.

About PS3 as Linux machine

- Processor: 1 PPE and 6 SPEs
  - PPE is 64-bit Power Architecture processor core
  - SPE is simple processor which optimized for compute-intensive workloads.
- Memory: 256MB
- 4MB Flash memory to boot loader
- Internal SATA 2.5inch HDD


graphics
processor
South bridge
HDD
USB
GbE
Memory: 256MB
Flash for kboot

Boot sequence of PS3 Linux



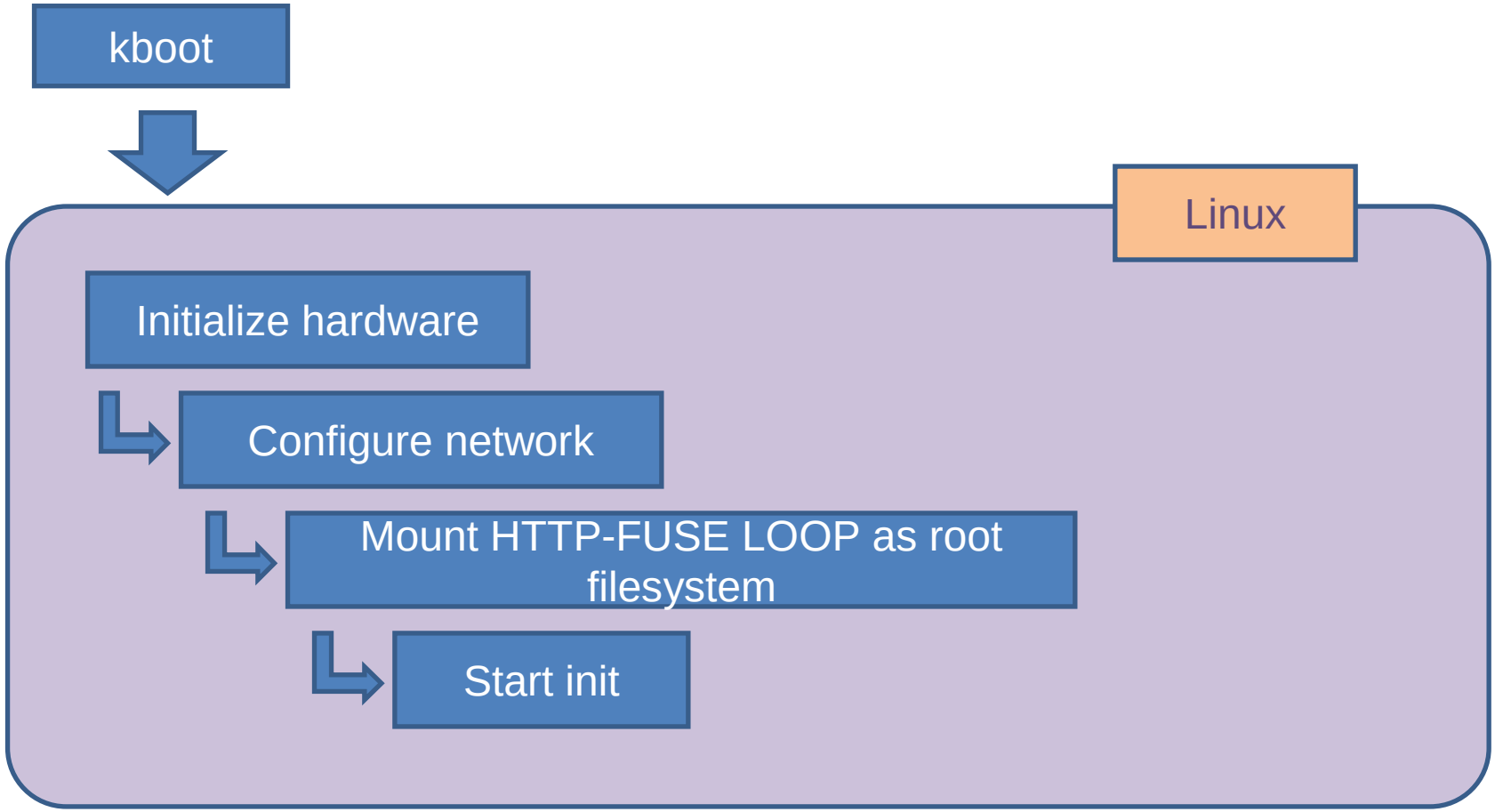
Internet booting using HTTP

- We developed Internet virtual block device “HTTP-FUSE LOOP”.
- Boot loader “kboot” can boot from kernel and miniroot on remote HTTP server.
- It requires DHCP to get kernel and miniroot from HTTP server.

Booting kernel via HTTP

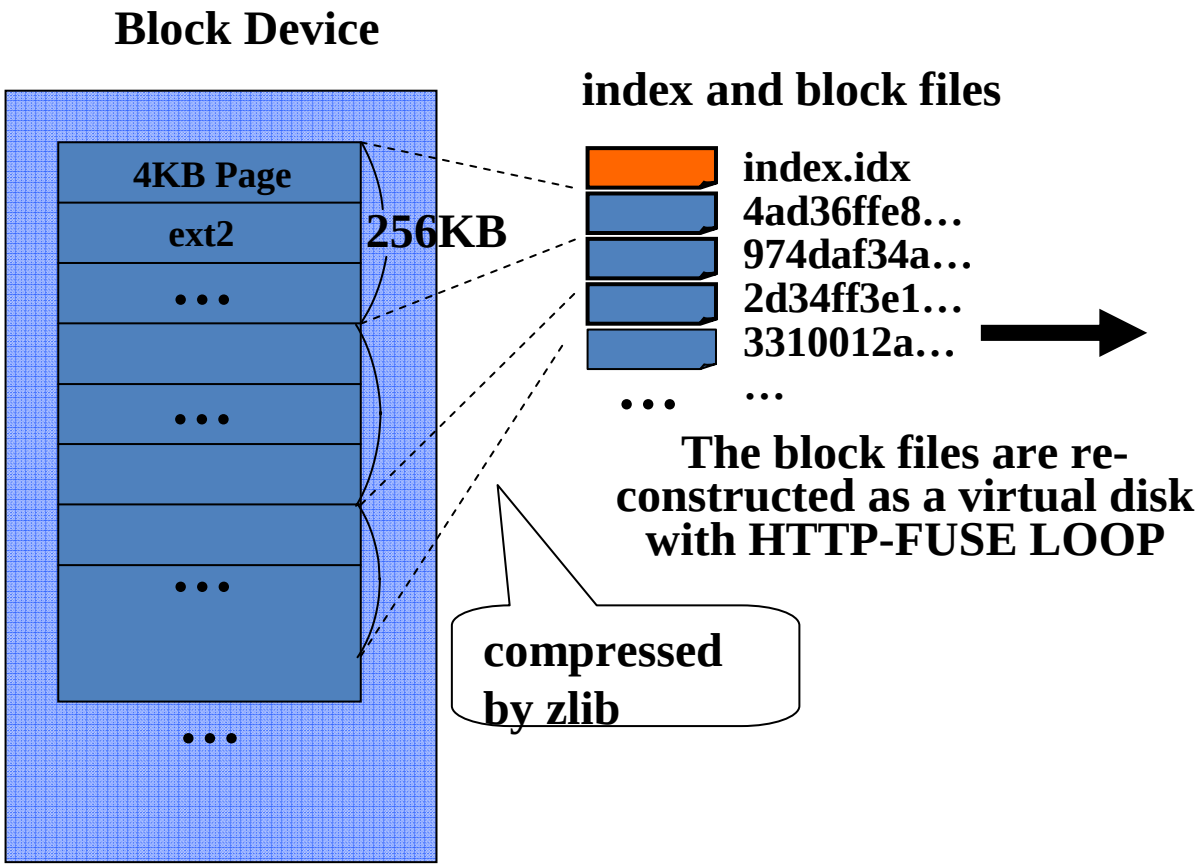
- On kboot command prompt, you can download file on HTTP server, and boot from it.
- Example:
  kboot: http://www.example.org/ps3/vmlinux
  initrd=http://www.example.org/ps3/initrd.img
  video=ps3fb:mode:5

After booting kernel



Data structure of HTTP-FUSE LOOP

- Original block device is split by a constant size (currently 256kB) and compressed by zlib.
- Each data is saved to a block file, its name is a SHA1 value of its contents.
  - If there are same contents, they are held together one block file and reduce total file space.
- Block files are managed by “index” file.

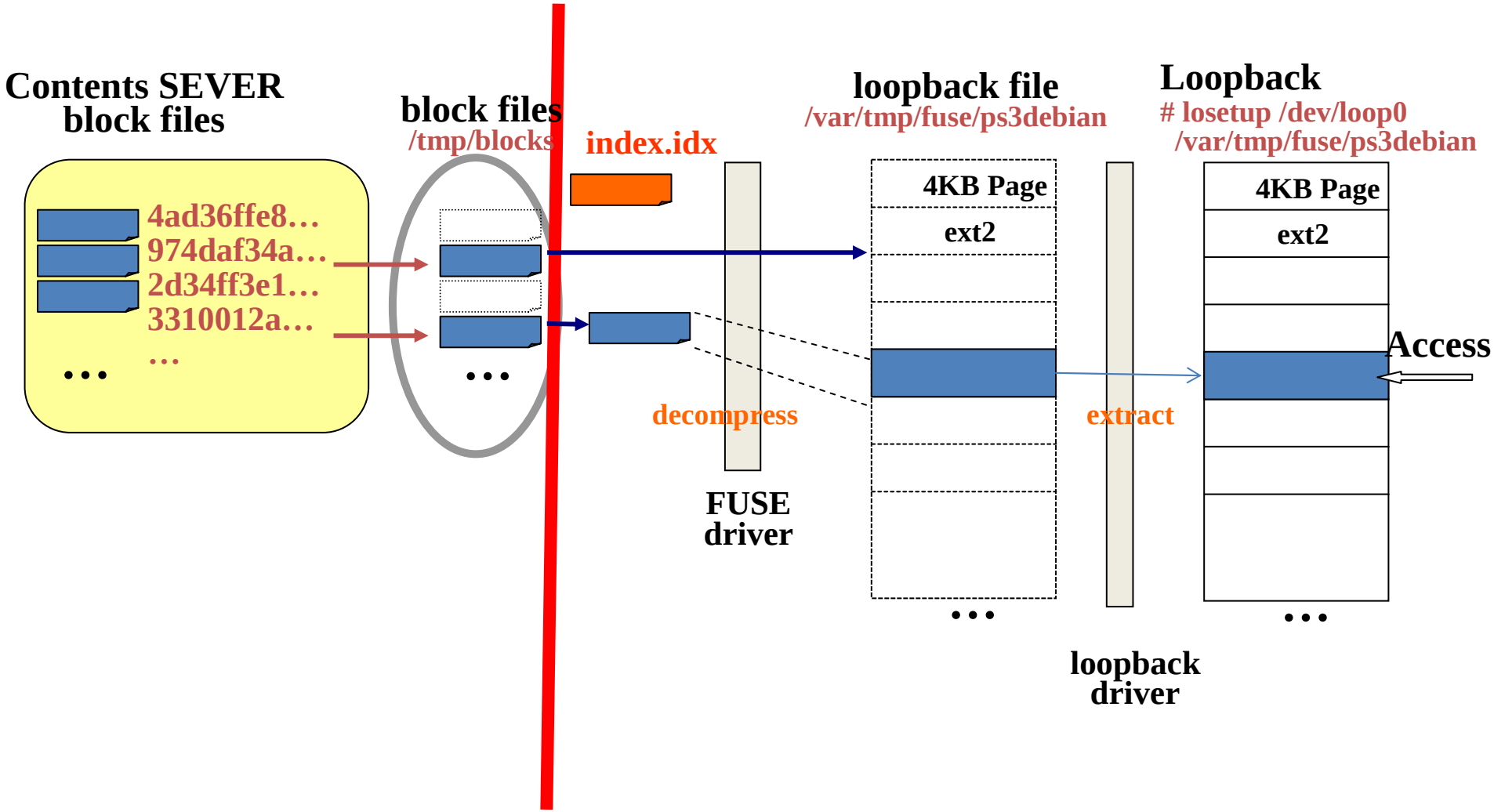

Block Device
4KB Page
ext2
256KB
index and block files
index.idx
4ad36ffe8…
974daf34a…
2d34ff3e1…
3310012a…
…
The block files are re-constructed as a virtual disk with HTTP-FUSE LOOP
compressed by zlib

Mounting HTTP-FUSE LOOP

- When mounting HTTP-FUSE LOOP, it downloads an “index” file and makes virtual loopback file from the qe“index” information.
- HTTP-FUSE LOOP is virtual loopback file
  - Use “mount –o loop” or “losetup” command to mount HTTP-FUSE LOOP

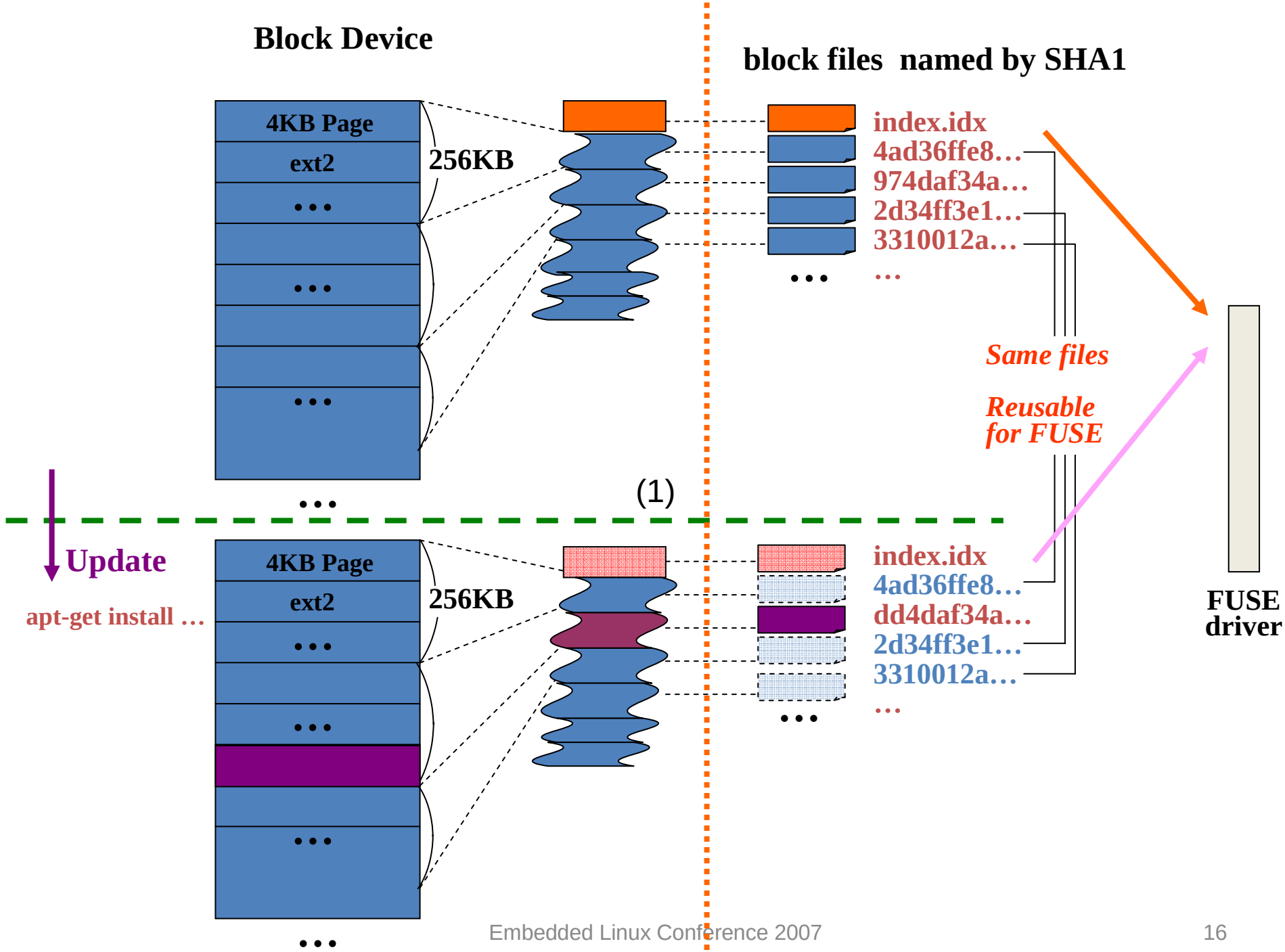
Reading from HTTP-FUSE LOOP

- When reading request occurred to HTTP-FUSE LOOP, it gets relevant block file.
  - If block file is available on cache, HTTP-FUSE LOOP use it, otherwise, downloads it from HTTP server.
- HTTP-FUSE LOOP driver extracts block file, and maps it to satisfy the reading request.

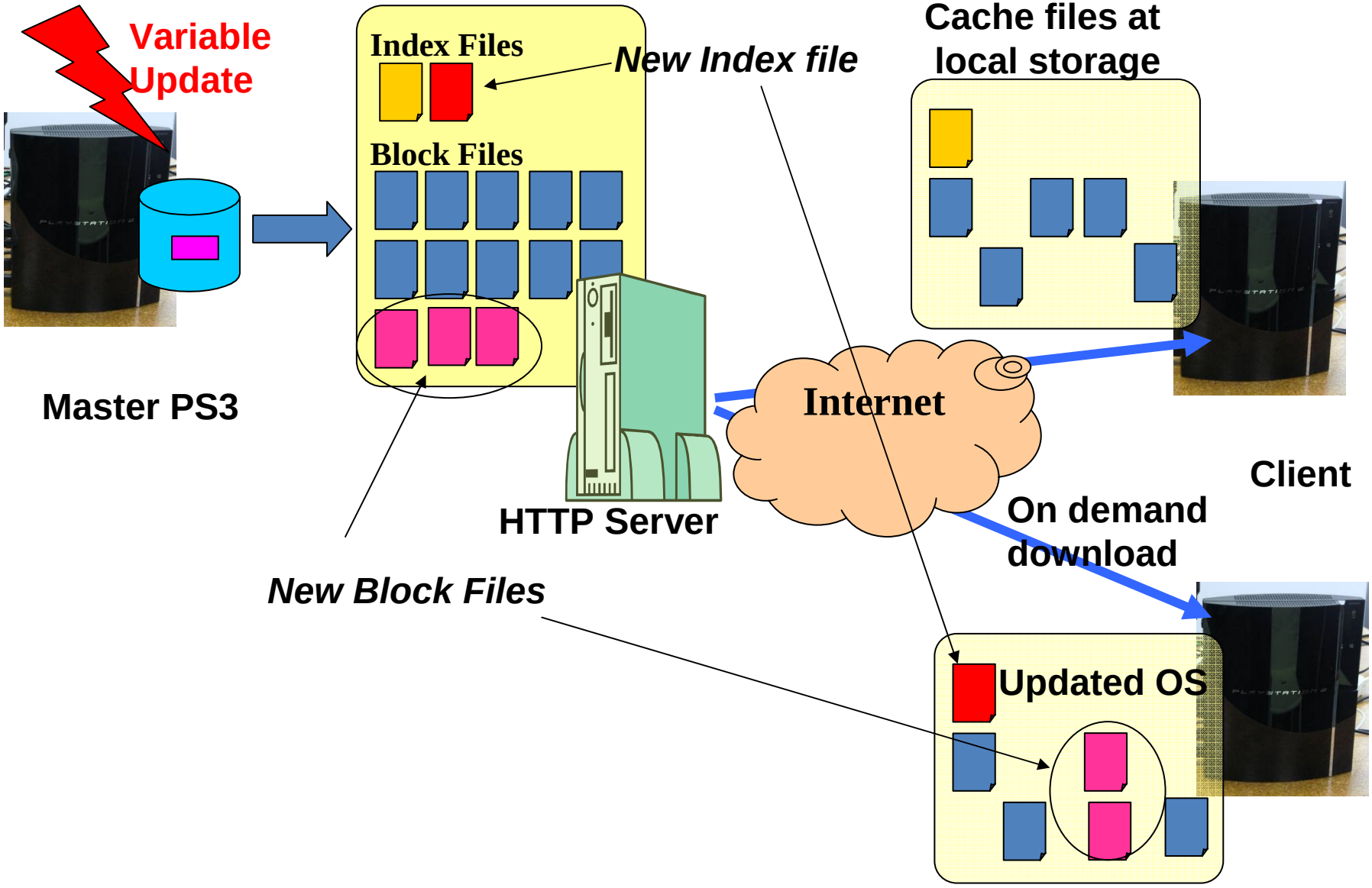

Contents SEVER
block files
4ad36ffe8…
974daf34a…
2d34ff3e1…
3310012a…
…
block files
/tmp/blocks
index.idx
decompress
FUSE
driver
loopback file
/var/tmp/fuse/ps3debian
4KB Page
ext2
extract
loopback
driver
Loopback
losetup /dev/loop0
/var/tmp/fuse/ps3debian
4KB Page
ext2
Access

Updating virtual device(1)

- When upgrade occurs to original storage,
  - Block file newly created at modified location.
  - Old block file are used at unmodified location.
  - New “index” file created.
- Adding new block files and “index” file, updating virtual device is almost done.
  - When HTTP-FUSE LOOP driver restarted with new “index” file, updating is completely done.


Block Device
block files named by SHA1
4KB Page
ext2
256KB
index.idx
4ad36ffe8…
974daf34a…
2d34ff3e1…
3310012a…
…
Same files
Reusable
for FUSE
(1)
Update
apt-get install …
4KB Page
ext2
256KB
index.idx
4ad36ffe8…
dd4daf34a…
2d34ff3e1…
3310012a…
…
FUSE
driver

Partial Update



HTTP-FUSE LOOP issue

- It is (currently) read-only.
- It uses temporary directory for cache.
- Its speed depends on network bandwidth and latency.

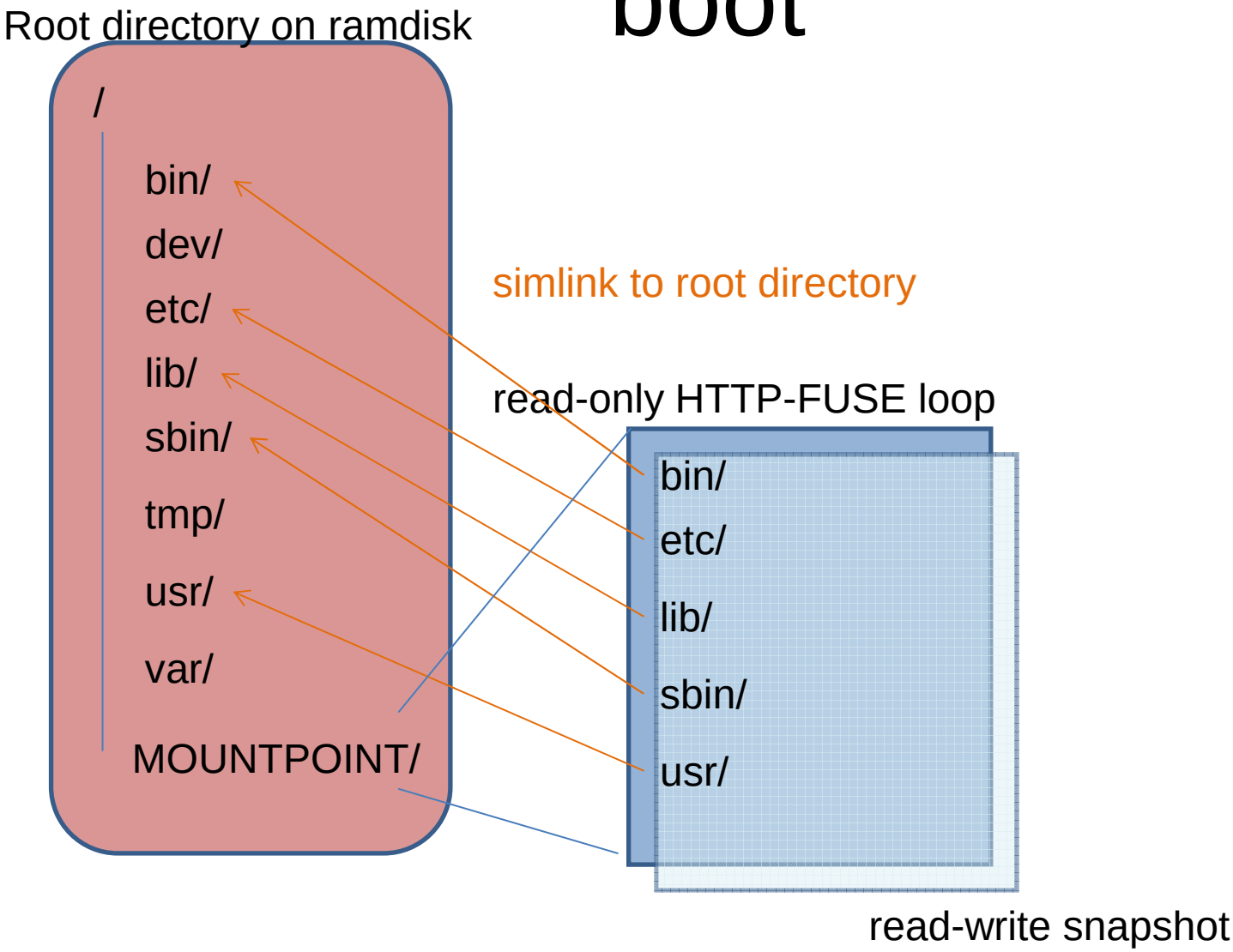
Read-only issue

- Normal distribution needs Read-write disk access.
  - If you have only read-only access, normal distribution falls to single-user mode.
- To solve read-only issue, we introduced device-mapper snapshot.

“Pivot-root” issue

- Normal distribution remounts HDD partition to root directory at init starts.
  - Root directory before init moved another directory or just deleted.
  - HTTP-FUSE LOOP uses temporary directory for cache, remounting causes trouble.

After HTTP-FUSE PS3 Linux boot



Network bandwidth and latency

- Performance of HTTP-FUSE LOOP affected by bandwidth and latency.
  - We solved this problem with worldwide mirror sites.

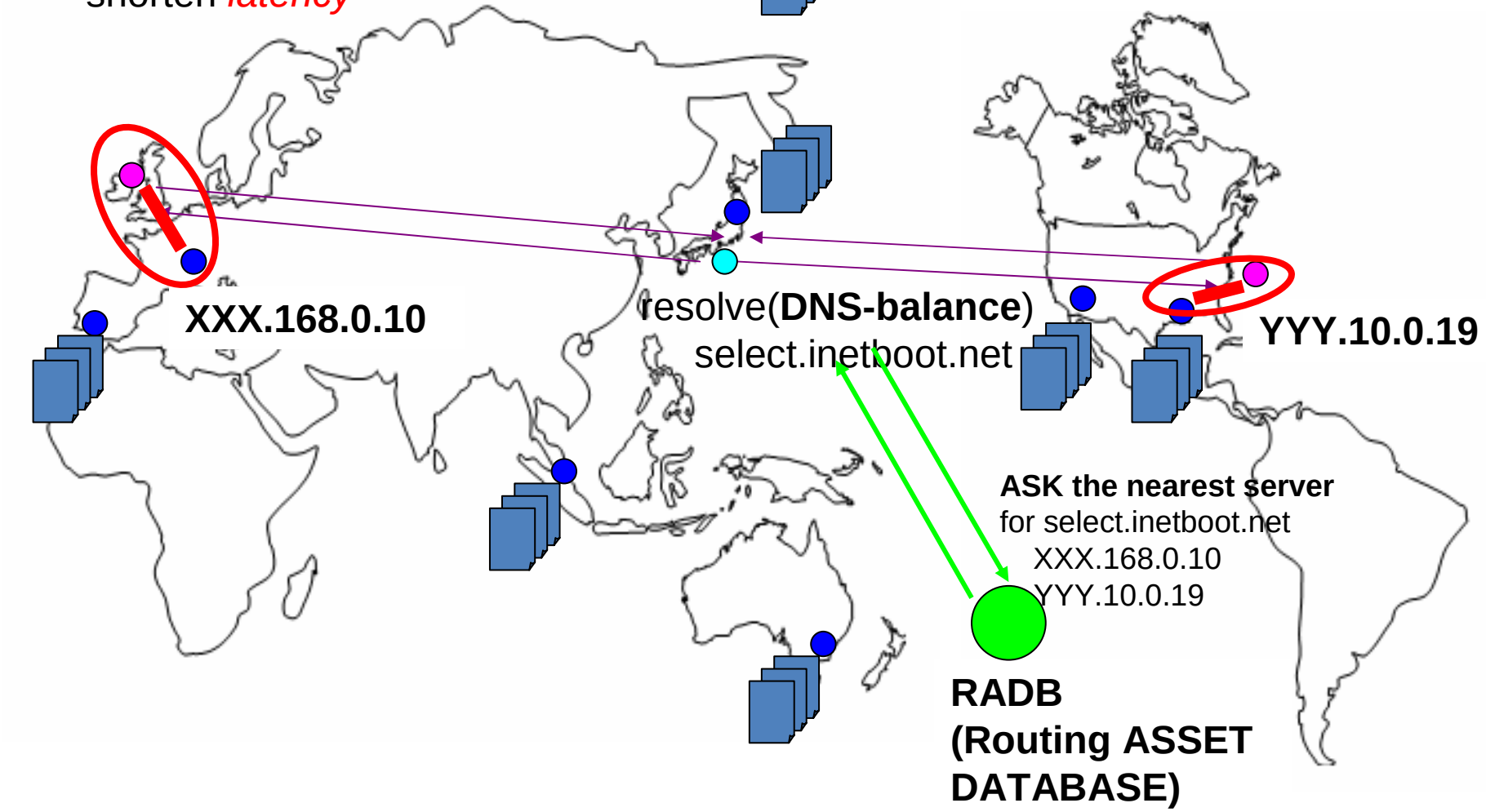
World Wide Deployment of Server

- We utilize inexpensive Web Hosting Service.
  - 5GB/ month from $10

DNS-Balance

- Client
- Web server for HTTP-FUSE Xenoppix
- DNS server: **ns.inetboot.net**
- Block files

DNS request
Resolve **select.inetboot.net** to shorten *latency*



Current status

- It can boot PS3 Debian.
- It uses 32MB, 1/8 of total memory of PS3 as space for device-mapper snapshot.
- Amount of download is about 170MB
  - From boot to finish starting gnome

Future plan

- More distributions
  - Fedora core, gentoo and more
- Add Cell/SDK and some demos
  - For SPE Programming.
- Auto setup external swap device

Conclusion

- We made HTTP-FUSE PS3 Linux, which completely boot from remote HTTP server.
- The current target is Debian GNU/Linux.

- This work is a apart of OS-Circular.
  - http://openlab.jp/oscircular/