

BOF

“OS Circular”

Bootable disk image archive on the Internet

<http://openlab.jp/oscircular/>

Kuniyasu Suzuki

National Institute of Advanced Industrial Science and Technology

- “OS Circular” is a project for **bootable disk image archive on the Internet**.
 - Old and New disk image is saved on the *stackable virtual disk*
 - User boots OSEs from the Internet without installation **on real and virtual machine**.
 - User prepares small bootable image only.
 - User can roll-back and roll-forward the OS image.
 - It is used for *Reference Installation*.
 - **LiveCD is a kind of bootable disk image archive.**
 - Current distro has LiveCD (Fedora 7,8,9 Ubuntu 7.04, 7.10, 8.04)
 - We must burn a CD-ROM. No security update.
 - Customization is not so easy.
 - We must make a LiveCD from scratch.
 - We can keep the old image using **Stackable File System (NILFS, brtfs)** . It enables to rollback to the old OS image.
 - But it is not so popular and the file system depends on the kernel (OS).
 - It is not Internet File System for anonymous users.

Related Work

- OS Zoo
 - Distribute the disk file of QEMU
 - Linux Distributions, *BSD, Plan9, OpenSolairs, MINIX
 - Big Disk File

– <http://www.oszoo.org>



- LivePC of Moka5
 - Moka5 is a venture company (Stanford “Collective” group)
 - Streaming service of Disk image to the customized VMWare
 - <http://www.moka5.org>

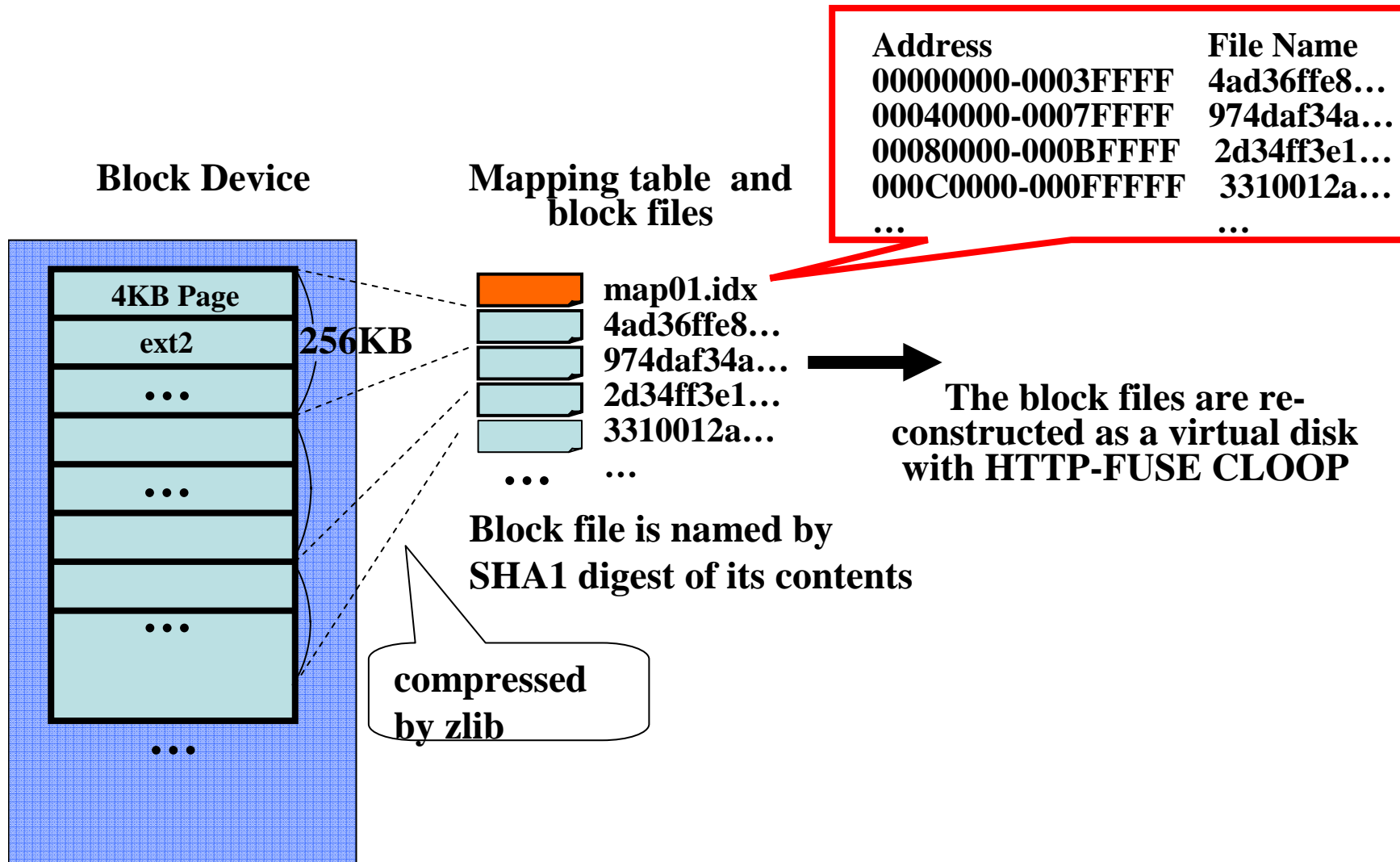
Lineup of Development

- OS Circular (previous HTTP-FUSE KNOPPIX)
- InetBoot: Internet BootLoader
 - HTTPFS version (for LiveCD: KNOPPIX, Fedora, Ubuntu)
 - HTTP-FUSE version
- VMSeed
- HTTP-FUSE PS3 Linux

OS Circular

- OS Circular is a framework of Internet Disk Image Distributor via Internet.
- The disk image is managed by CAS (Content-Addressed Storage) “HTTP-FUSE CLOOP”
 - Venti of Plan9 depends on same idea.
- User boots an OS from Internet.
 - Hard disk works as cache. The cached image is reusable for next boot and applied to Mobile Computing.

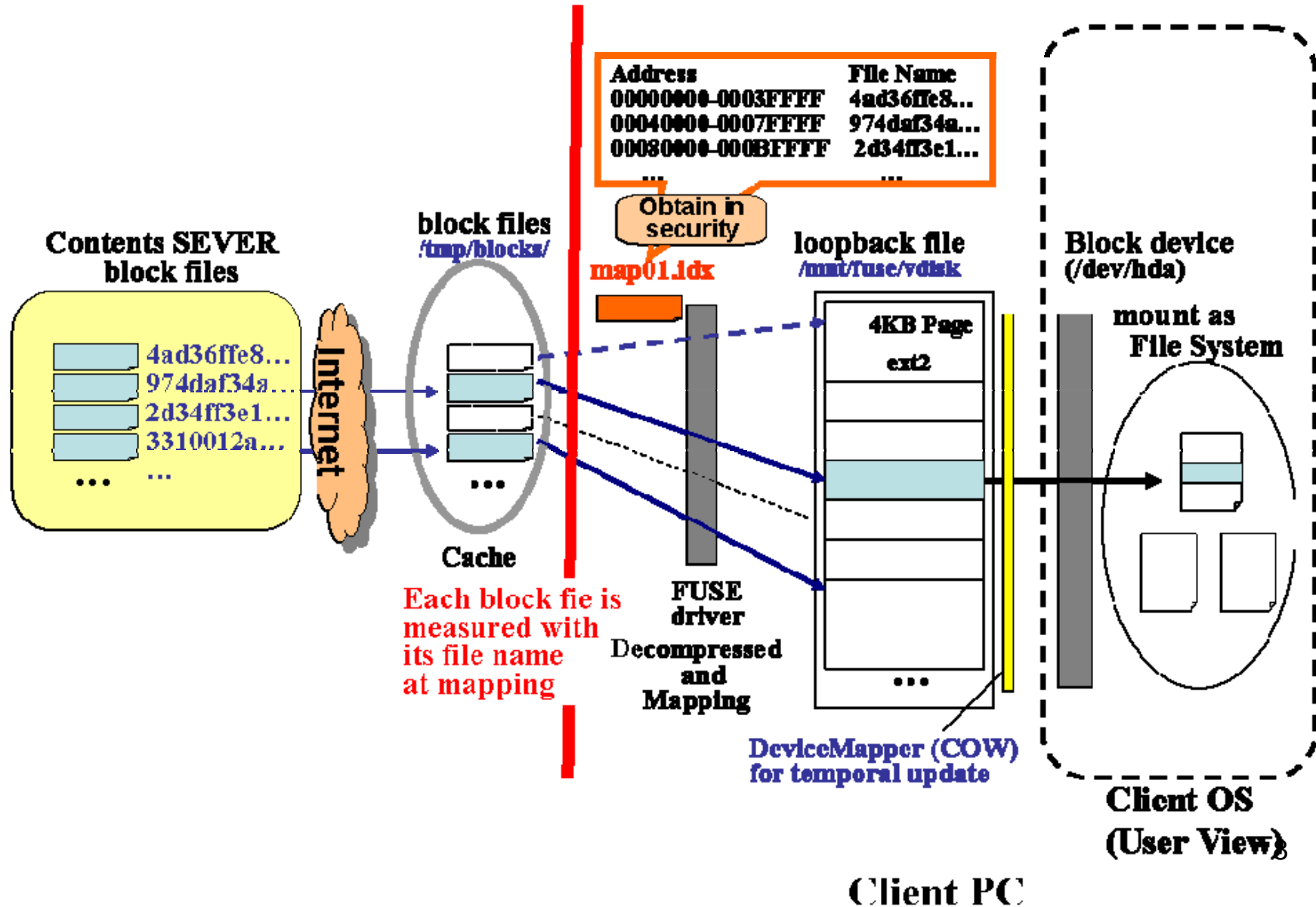
Block files for HTTP-FUSE CLOOP



HTTP-FUSE CLOOP (1/2)

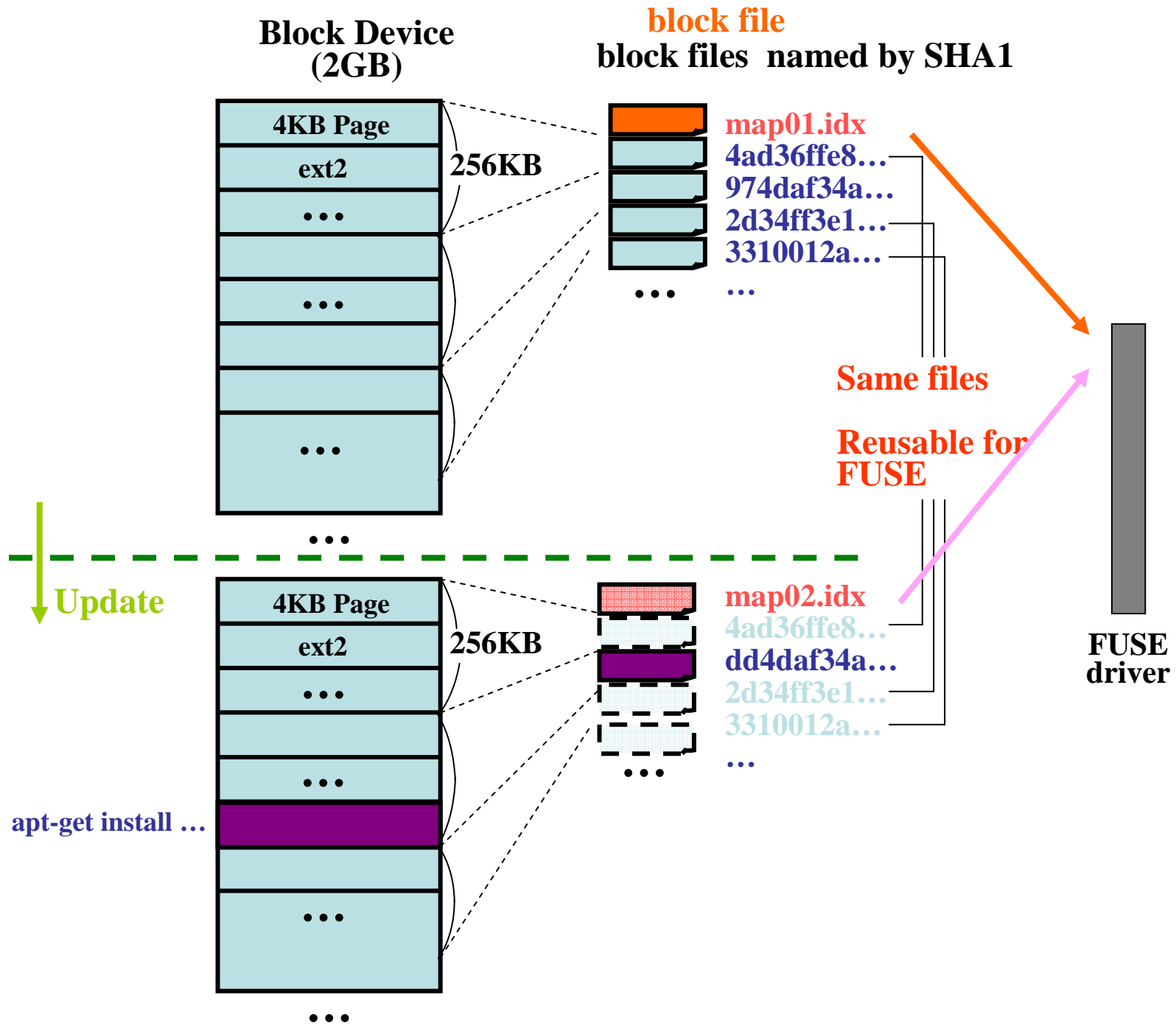
- The image of HTTP-FUSE CLOOP is made from existing normal block device.
- Original block device is split by 256KB and compressed by zlib.
*Each data is saved to each “**block file**”.*
- Block file name is a SHA1 value of its contents.
 - If there are same contents in blocks, they are expressed by one block file and reduce total storage space.
 - *The basic idea is resemble to “Venti of Plan9”[USENIX’02]*
- Block files are managed by *“**mapping table**”* file.
- Block files are reconstructed to a loopback file by FUSE wrapper.
 - FUSE is a User-land File System.
 - <http://fuse.sf.net>
- Each block file is measured with the SHA1 file name when it mapped to loopback file.

Mount FS using HTTP-FUSE CLOOP



HTTP-FUSE CLOOP (2/2)

- When a file is updated or created on an original block device, the relevant block files are newly created with new SHA1 file name. The mapping table file are also renewed.
 - Old block files are reusable.
- HTTP for file deliver
 - Most popular and well designed for Internet.
 - Utilize inexpensive Web hosting services and Proxies and Mirror Servers for world wide deployment.
- Block files are network/storage transparent.
 - **Block files are cached on a local storage and reused.**
 - If necessary block files are stored in a local storage, network connection is not necessary.



Available OSes on OS Circular

- On real machine
 - **KNOPPIX 4.0.2, 5.0.1, 5.1.1**
 - KNOPPIX is advanced at AutoConfig and applied to any PC.
 - **“gPXE” , “Kboot” and “kexec”** can boot them.
 - They download kernel and miniroot and reboot PC them.
- On virtual machine
 - **Plan9 and NetBSD**
 - on Xen 2.0.3 DomU (para-virtualization)
 - presented HTTP-FUSE Xenoppix (Ottawa Linux Symposium 2006)
 - **Debian Etch, Ubuntu6.06/6.10/7.04, CentOS5**
 - on Xen-HVM/KVM/QEMU (full-virtualization)

Lineup of Development

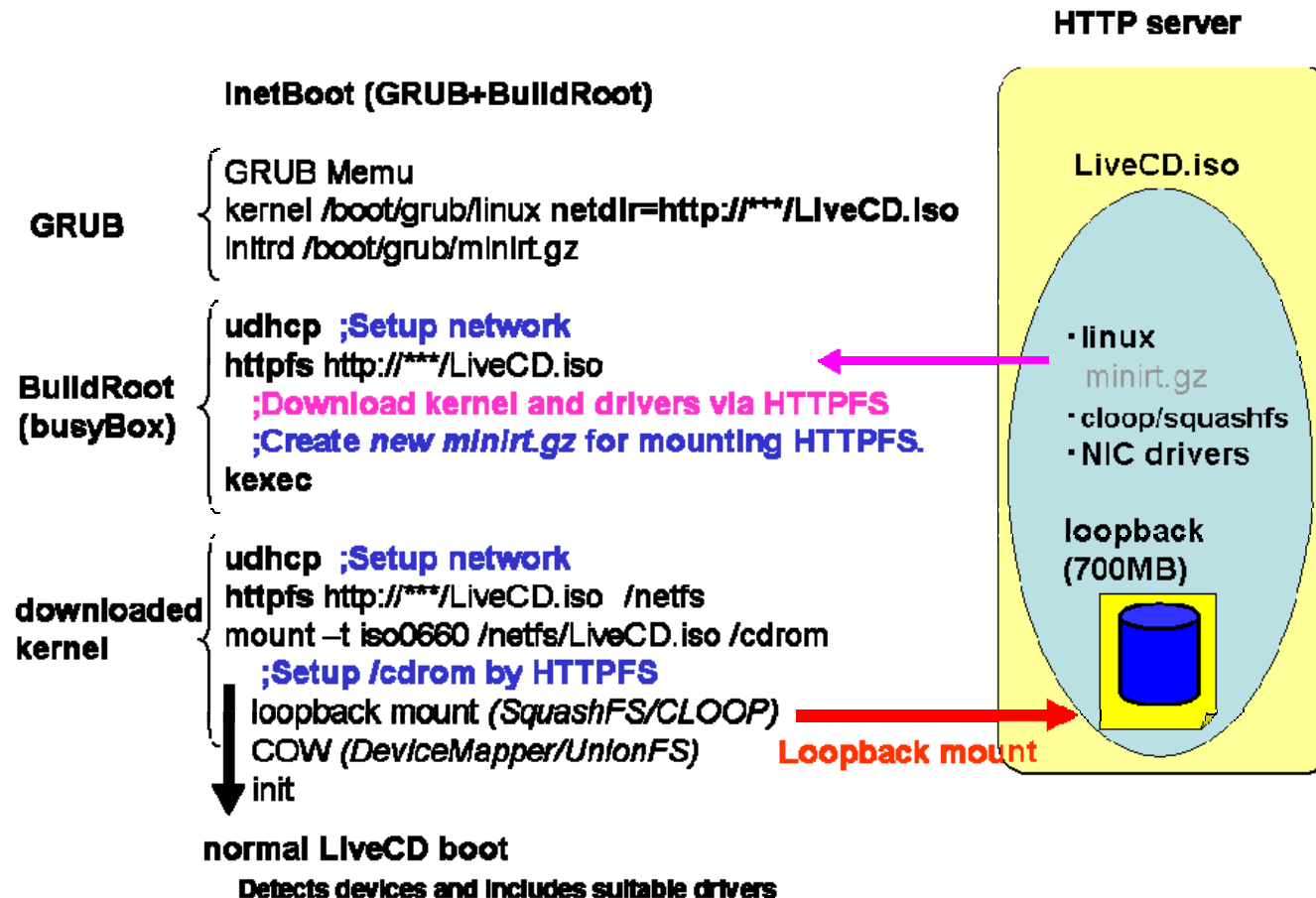
- OS Circular (previous HTTP-FUSE KNOPPIX)
- **InetBoot: Internet BootLoader**
 - HTTPFS version (for LiveCD: KNOPPIX, Fedora, Ubuntu)
 - HTTP-FUSE version
- VMSeed
- HTTP-FUSE PS3 Linux

InetBoot

- “Internet Bootloader” and “Bootable Disk Image on the Internet”
- Internet Bootloader
 - GRUB+BuildRoot(Busybox) 5MB bootable CD
 - The kernel is downloaded and reboot PC with them with “**kexec**”.
 - Double-Deck Boot (Busybox, “kexec” -> downloaded kernel)
- 2 ways to get a Disk Image
 - **HTTPFS**
 - Used to mount a ISO file of LiveCD.
 - *Fedora, Ubuntu, KNOPPIX, VMKnoppix*
 - » X86_64 version will be released soon.
 - Random Read request is transferred to “range” request of HTTP server.
 - **HTTP-FUSE CLOOP**
 - **Developed for OS Circular**

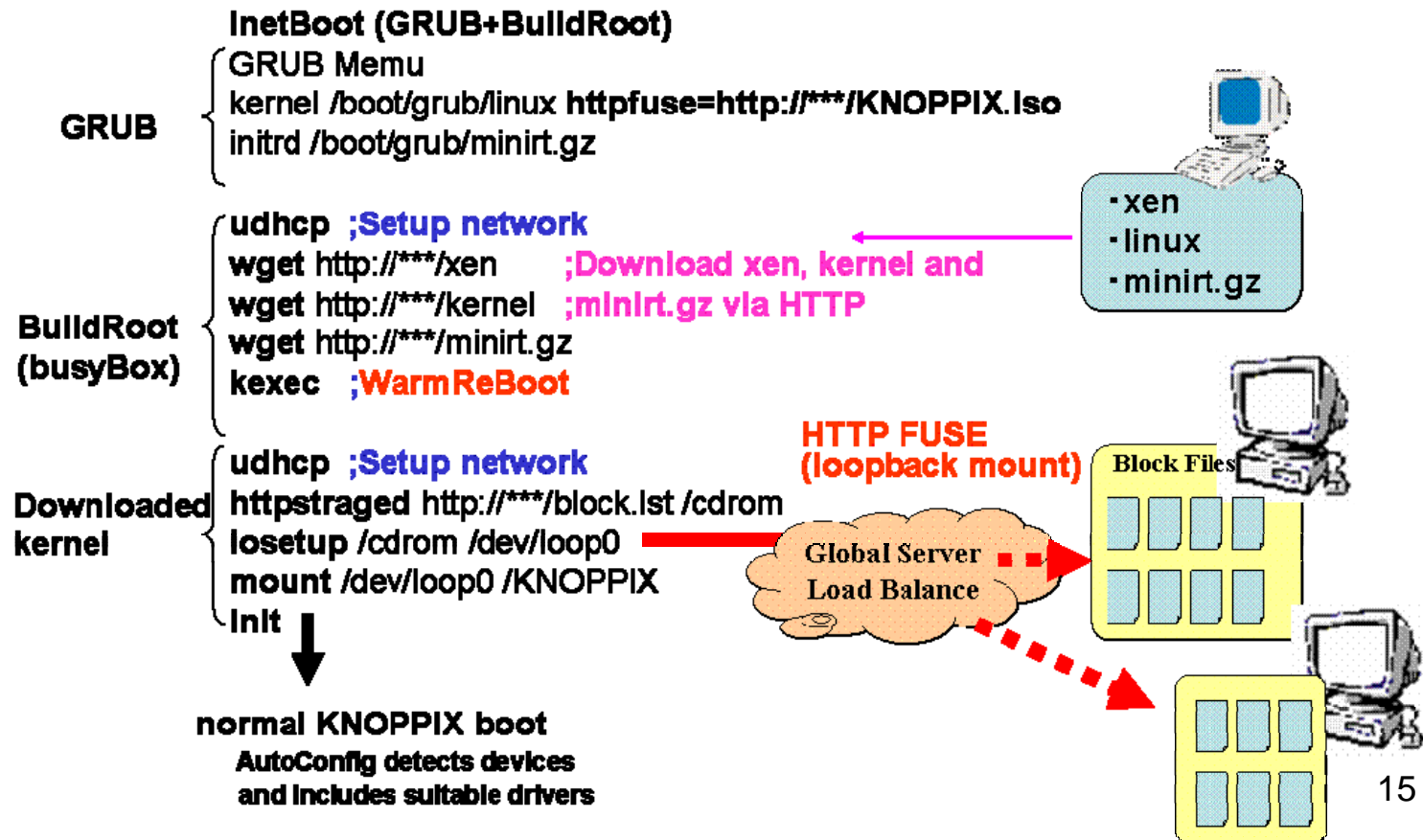
InetBoot with HTTPFS

- Utilize existing ISO file of LiveCD.
 - HTTPFS mount the ISO file on the Internet
 - New Miniroot is customized to mount ISO file as the Root-FS by HTTPFS.
 - It can download hypervisor “Xen” if needed.



InetBoot with HTTP-FUSE CLOOP

- HTTP-FUSE CLOOP re-constructs a virtual disk.
 - Prepare the bootable disk image (reuse the OS Circular image), kernel and miniroot.
 - It can download hypervisor “Xen” if needed.



Available OSes on InetBoot

- HTTPFS version
 - **Fedora 8,9**
 - **Ubuntu 7.04, 7.10, 8.04**
 - **KNOPPIX(531, 511, 501, 402)**
 - **VMKnoppix (Xen: 3.2.0, 3.1.1, 3.1.0, 3.0.4.1, 3.0.4)**
- HTTP-FUSE CLOOP version
 - On Real Machine
 - **KNOPPIX (511,501, 402)**
 - On Virtual Machine
 - **Plan9** and **NetBSD** on DomU of Xen 2.0.3

Lineup of Development

- OS Circular (previous HTTP-FUSE KNOPPIX)
- InetBoot: Internet BootLoader
 - HTTPFS version (for LiveCD: KNOPPIX, Fedora, Ubuntu)
 - HTTP-FUSE version
- **VMSeed**
- HTTP-FUSE PS3 Linux

VMSeed

- VMSeed is an *extension of Inetboot* to virtual machine.
- Utilize **sparse virtual disk format** of each virtual machine.
 - The initial virtual disk includes bootloader, kernel and miniroot only.
 - The disk image is downloaded from Internet and saved to the virtual disk. So **the virtual disk grows by use of the guest OS.**
- Related Project
 - LivePC of Moka5 (www.moka5.com)
 - Streaming service of Disk image to the customized VMWare

VMSeed

- VMSeed is “growing virtual disk image(Guest OS)”
 - Same disk image is applied to many VM; VMware, VMware, VirtualBox, VirtualPC, Parallels, Xen, QEMU, KQEMU, and KVM.

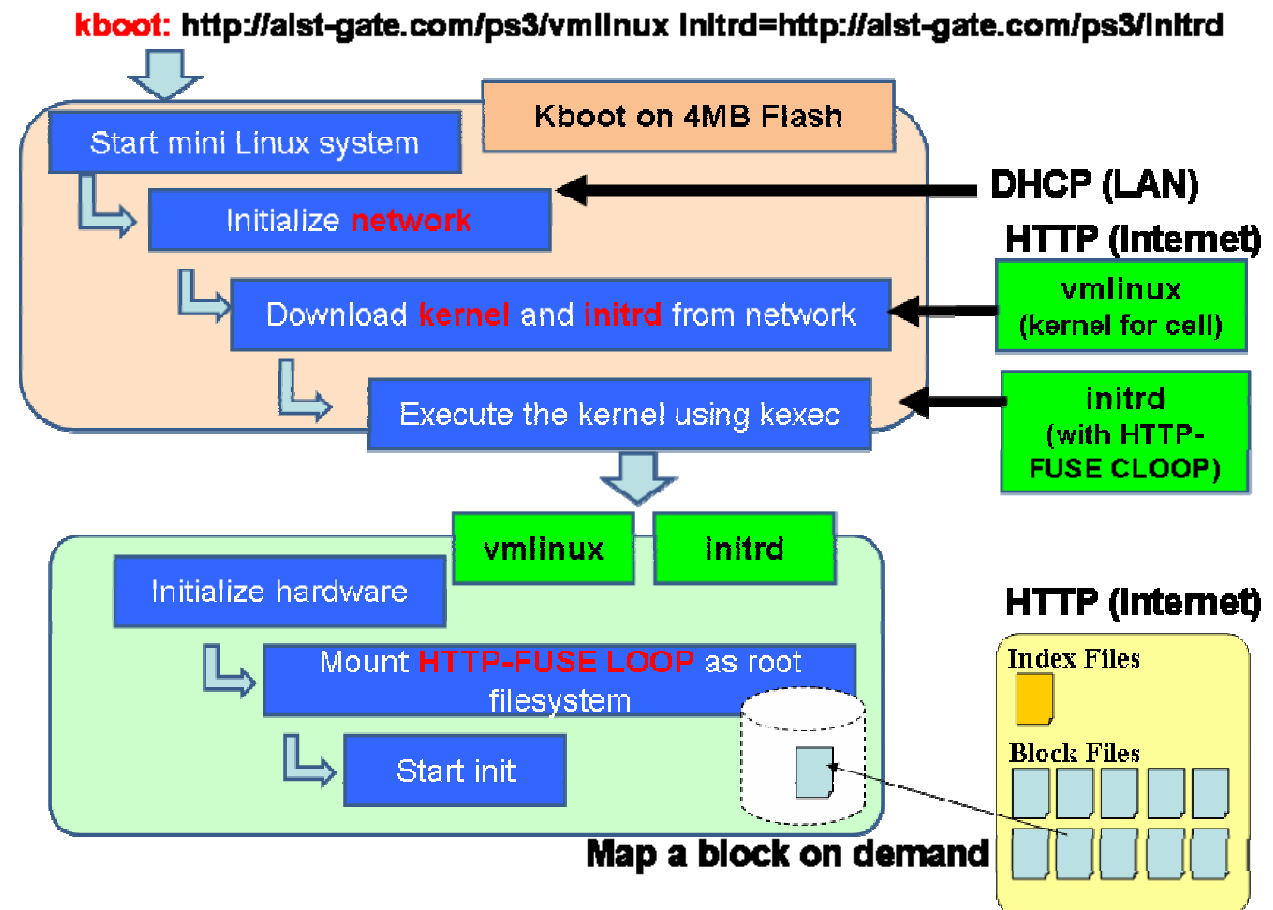
	KNOPPIX 511	KNOPPIX 501	KNOPPIX 402	Xenoppix402 (Xen 3.0.2)	Plan9 (DomU)	NetBSD (DomU)	Initial Disk Size (Virtual 2GB)	Comment
VMware	OK	OK	OK	OK	OK	OK	33MB	
VirtualBox	OK	OK	OK	NG	NG	NG	68MB	
VirtualPC	OK	OK	OK	NG	NG	NG	102MB	
Parallels	OK	OK	OK	OK	OK	OK	32MB	
Xen	OK	OK	OK	OK	OK	OK	31MB	on Sparse FS
QEMU KQEMU	OK	OK	OK	OK	OK	OK	31MB	on Sparse FS
KVM	OK	OK	OK	NG	NG	NG	31MB	on Sparse FS

Lineup of Development

- OS Circular (previous HTTP-FUSE KNOPPIX)
- InetBoot: Internet BootLoader
 - HTTPFS version (for LiveCD: KNOPPIX, Fedora, Ubuntu)
 - HTTP-FUSE version
- VMSeed
- **HTTP-FUSE PS3 Linux**

HTTP-FUSE PS3 Linux

- Utilize "kboot" on 4MB Flash of PlayStation 3
 - kboot can get "kernel" and "initrd" via HTTP.
- The disk image is obtained by Internet Virtual Disk "HTTP-FUSE CLOOP".



Problem for prevailing

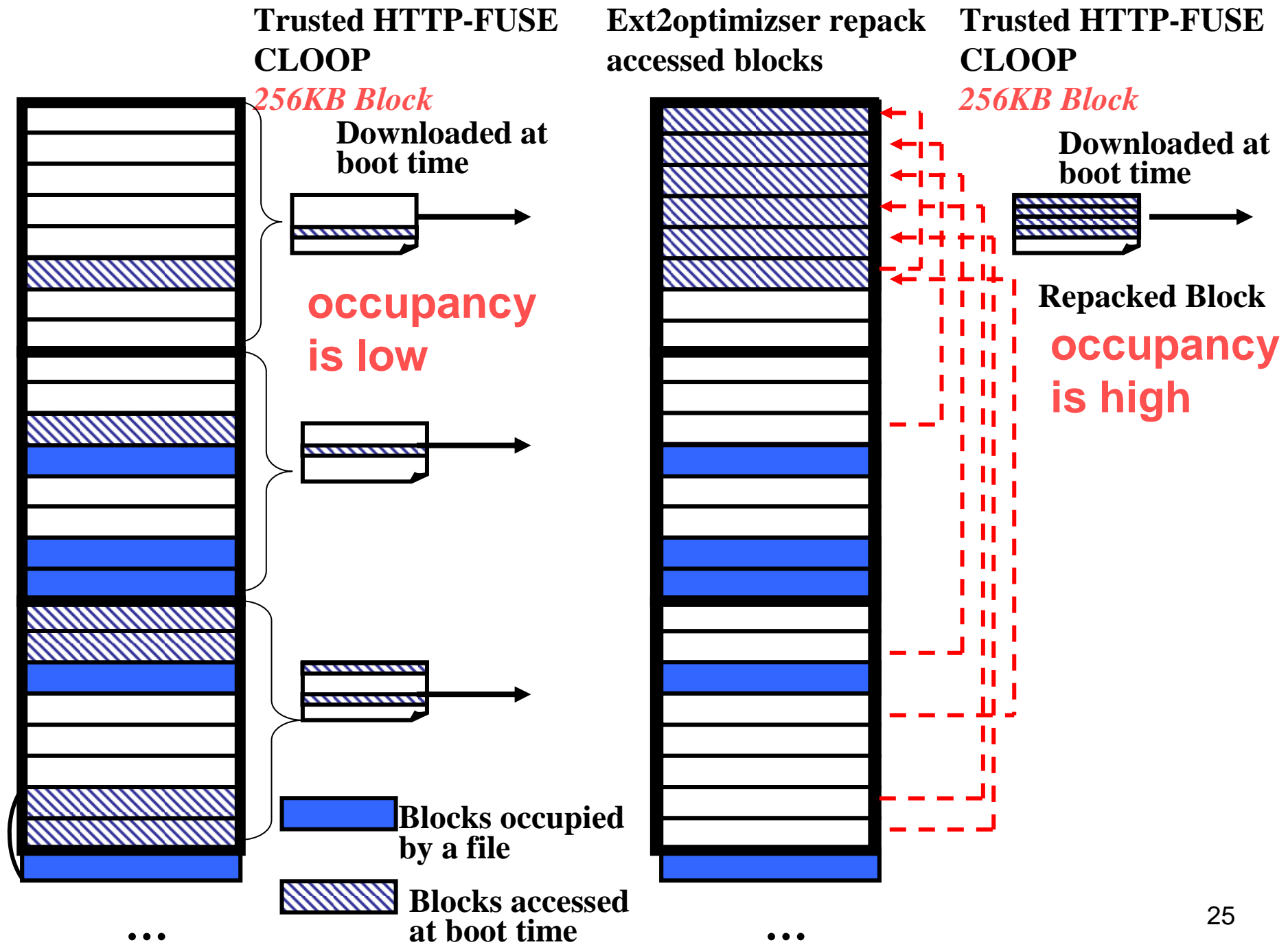
- Network Latency
 - HTTPFS and HTTP-FUSE is sensitive for the network latency
 - Prepare some servers (3 sites in US, 3 sites in EU) but it is not enough.
 - Please contribute your HP server!

Optimization

- Trusted HTTP-FUSE CLOOP is very sensitive for network latency, because small block files are downloaded as the occasion demands.
 - Optimization for fragmentation
 - Optimization for download methods

Optimization for Fragmentation

- Block size mismatch between file system and virtual block device causes *fragmentation*.
 - Trusted HTTP-FUSE CLOOP **256KB**
 - File System (ext2) **4KB**
 - Kitagawa* reported the occupancy of requested blocks at boot time (on HTTP-FUSE KNOPPIX 3.8.2) was 30%.
 - * [Linux Kongress 2006]
- “**ext2optimizer**” repacks the data blocks of ext2 file system to be in line.
 - It is based on the profile of accessed data blocks at boot time.
 - As the results, ext2optimizer reduces the number of block files.



Optimization for download methods

- 2 optimizations
 - DLAHEAD (DownLoad AHEAD)
 - **The necessary block files are downloaded in advance** with extra download connections (default 4).
 - [Preparation] Take a profile of downloaded block files at boot time.
 - DNS-Balance
 - DNS-Balance is a kind of name resolver which suggests **the nearest server** with routing information offered by RADB.net
 - http://openlab.jp/dns_balance/dns_balance.html
 - Users find the nearest download site.
 - It prevents inter-continental download because we offers servers in EU, US, and Japan.

Future Plan

- Link to Trusted Computing.
 - The incidents of boot procedure are stored to TPM (Trusted Platform Module).
 - The incidents are validated by the Trusted Third Party “Remote Attestation”.
 - User can confirm the boot has no **rootkit and malware**.
Furthermore User can **confirm the applications are vulnerable or not**.

Summary

- As you like
 - <http://openlab.jp/oscircular/>

Related paper and presentation

- USENIX Annual Technical Conference 2008, (Poster), “InetBoot and VMSeed: Trusted Internet Bootloader for Hypervisor and Guest OS”
- Thirteenth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '08) (Poster), “TPM + Internet Virtual Disk + Platform Trust Services = Internet Client”
- USENIX LISA 2007 (21st Large Installation System Administration conference), “OS Circular: Internet Client for Reference”
- LinuxConf Europe 2007, “OS Circular on QEMU/KQEMU/KVM/Xen-HVM”
- Linux Konogress 2006, “Trusted Boot of HTTP-FUSE KNOPPIX”
- Linux Symposium2006 “HTTP-FUSE Xenoppix”