

OS Circular: Internet Client for Reference

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

Kuniyasu Suzaki, Toshiki Yagi, Kengo Iijima, Nguyen Anh Quynh
National Institute of Advanced Industrial Science and Technology

Contents

- Motivation
- Implementation
 - VM-Loader “VMKnoppix”
 - Stackable Virtual Disk “Trusted HTTP-FUSE CLOOP”
- Optimization
 - For fragmentation and download methods
- Current Status & Future Work

Motivation

- Please assume the time of OS update.
- Users want to know the feasibility of the new OS before updating the current OS.
 - Example: When I update “glibc”, I want to check the availability of “glibc” for the current applications.
- Users want to get back to old OS image, because they want to use old applications.
 - Example: Sometimes I want to use Apache1.3, because the security check is looser than Apache 2.* .

OS Circular

- OS Circular is a framework of *Internet Disk Image Distributor* for Virtual Machine.
 - It enables to boot OS on VM **without installation**.
 - It isn't live migration.
 - Deal with network disconnection for mobile computing
 - Parts of disk Image can be cached on a local storage.
 - *The OS is periodically updated* for Security.
 - OS Circular allows to rollback to previous image.
- OS Circular is **Client Centric System** which utilizes virtualization technology.
 - “Virtual Machine” + “Stackable Virtual Disk”
 - We developed VMKnoppix as VM Loader of OS Circular 4

Requirement for VM

- Unified Device Model
 - The guest OS only have to support these devices.
 - The unified device model enables to *share the disk images on the other virtual machines.*
 - QEMU-DM (Device model)
 - QEMU(KQEMU), KVM and Xen-HVM assumes the same devices.
 - They are RealTek RTL8029 for NIC, Cirrus Logic GD5446 for Video Card and etc.
- Full Virtualization
 - It enables to use normal installer and security management for Guest OS.
 - We can update the kernel with package manager on the virtual machine.

“VMKnoppix” as VM Loader

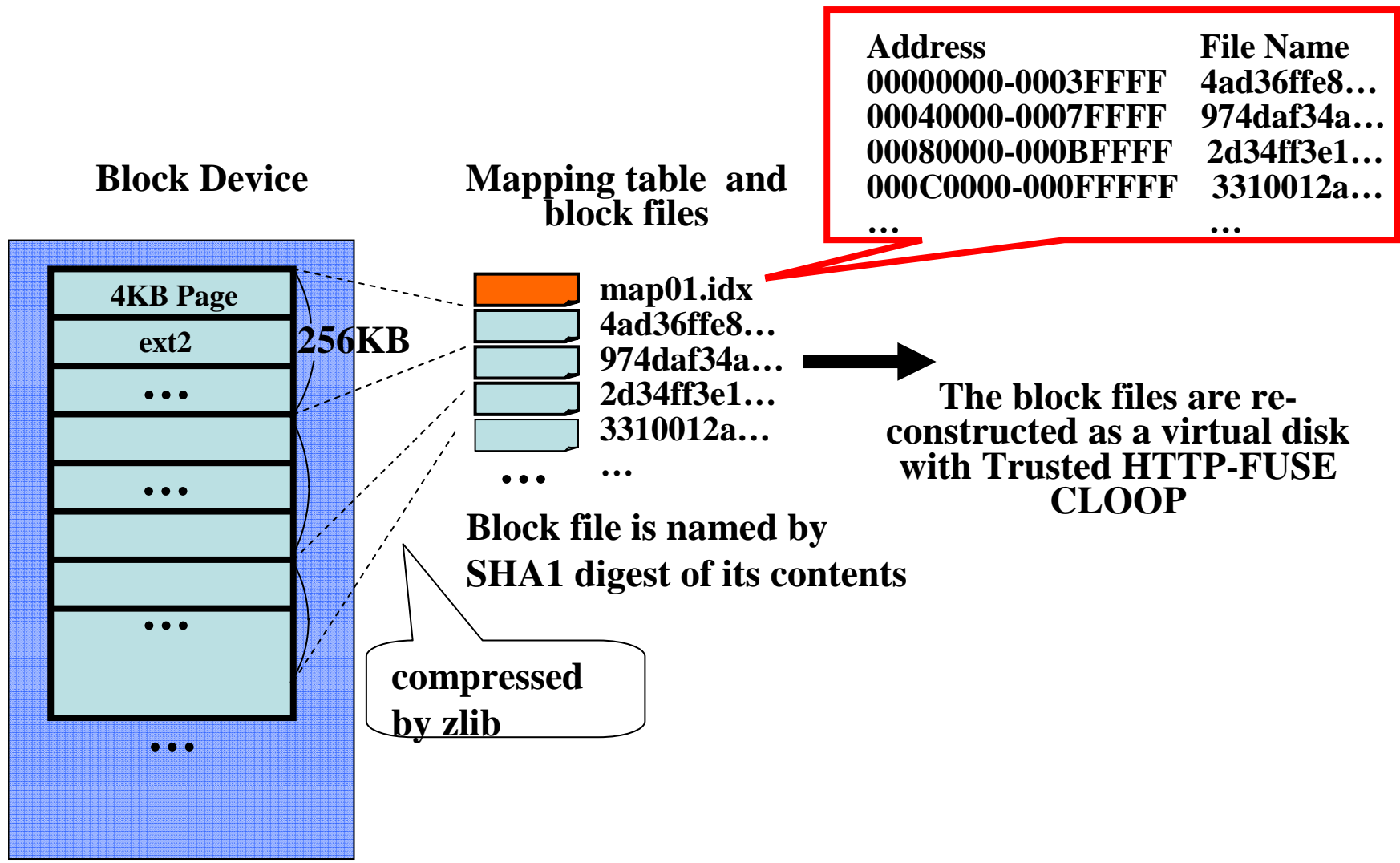
- VMKnoppix = KNOPPIX(1CD Linux) with VM Software (QEMU, KQEMU, KVM, and Xen-HVM) and driver of stackable virtual disk.
 - KNOPPIX boots as host OS (Domain0).
 - KNOPPIX (AutoConfig) prepares device drivers on anonymous PC.
 - VM software (QEMU, KQEMU, KVM, or Xen-HVM) launches a virtual machine with the stackable block device.

Requirements of Virtual Disk

- Virtual Disk is Block Level Abstraction.
- The requirement for OS Migration. (Pfaff[NSDI'06])
 - Versioning
 - Partial update & Rollback
 - Globalization
 - World Wide Deployment (Internet)
 - Network/Storage Transparent
 - Handle network (dis/re)-connection for mobile computing
 - Security
 - Virtual disks have to keep validness of contents
- We developed “Trusted HTTP-FUSE CLOOP”.

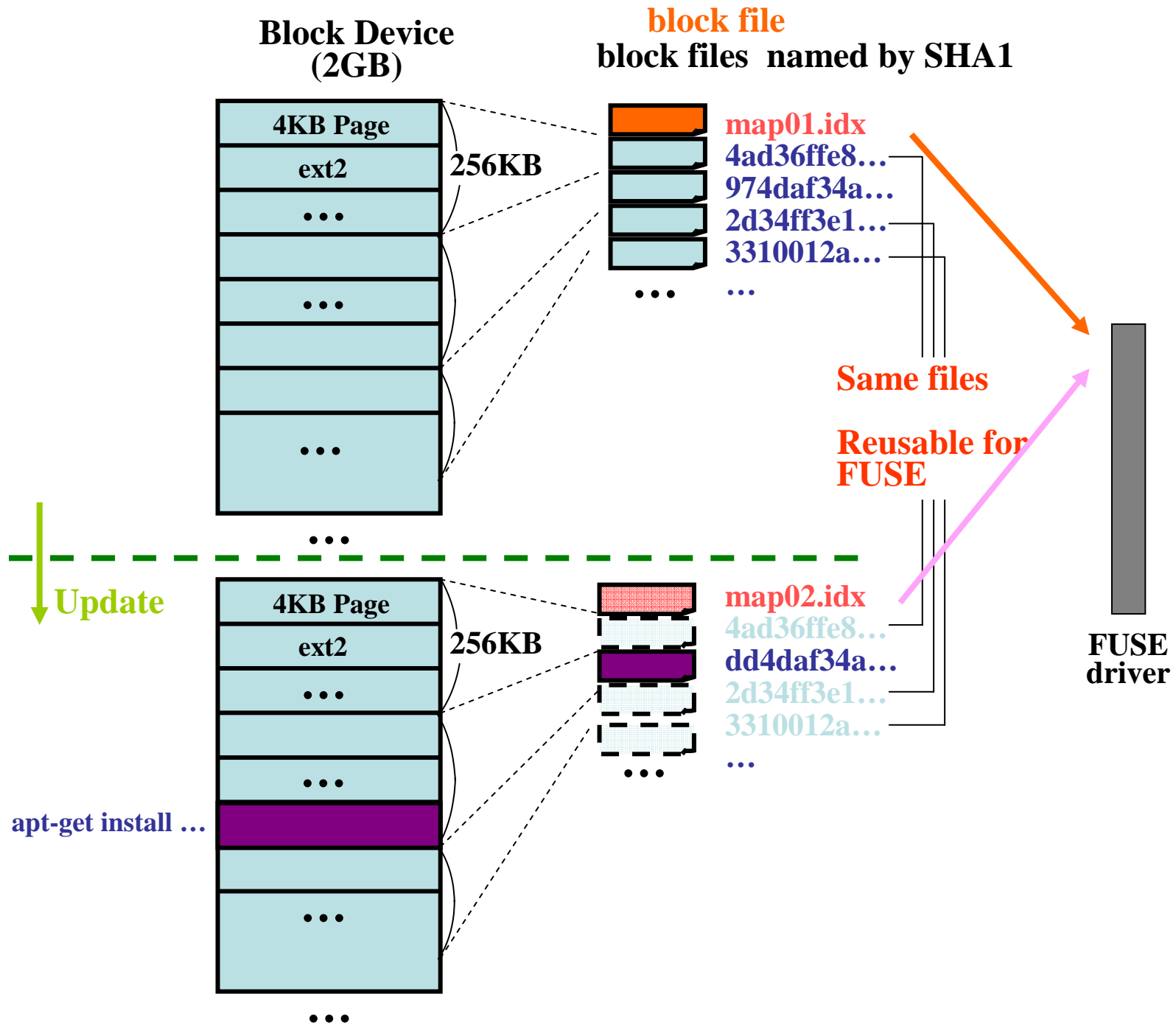
Trusted HTTP-FUSE CLOOP (1/2)

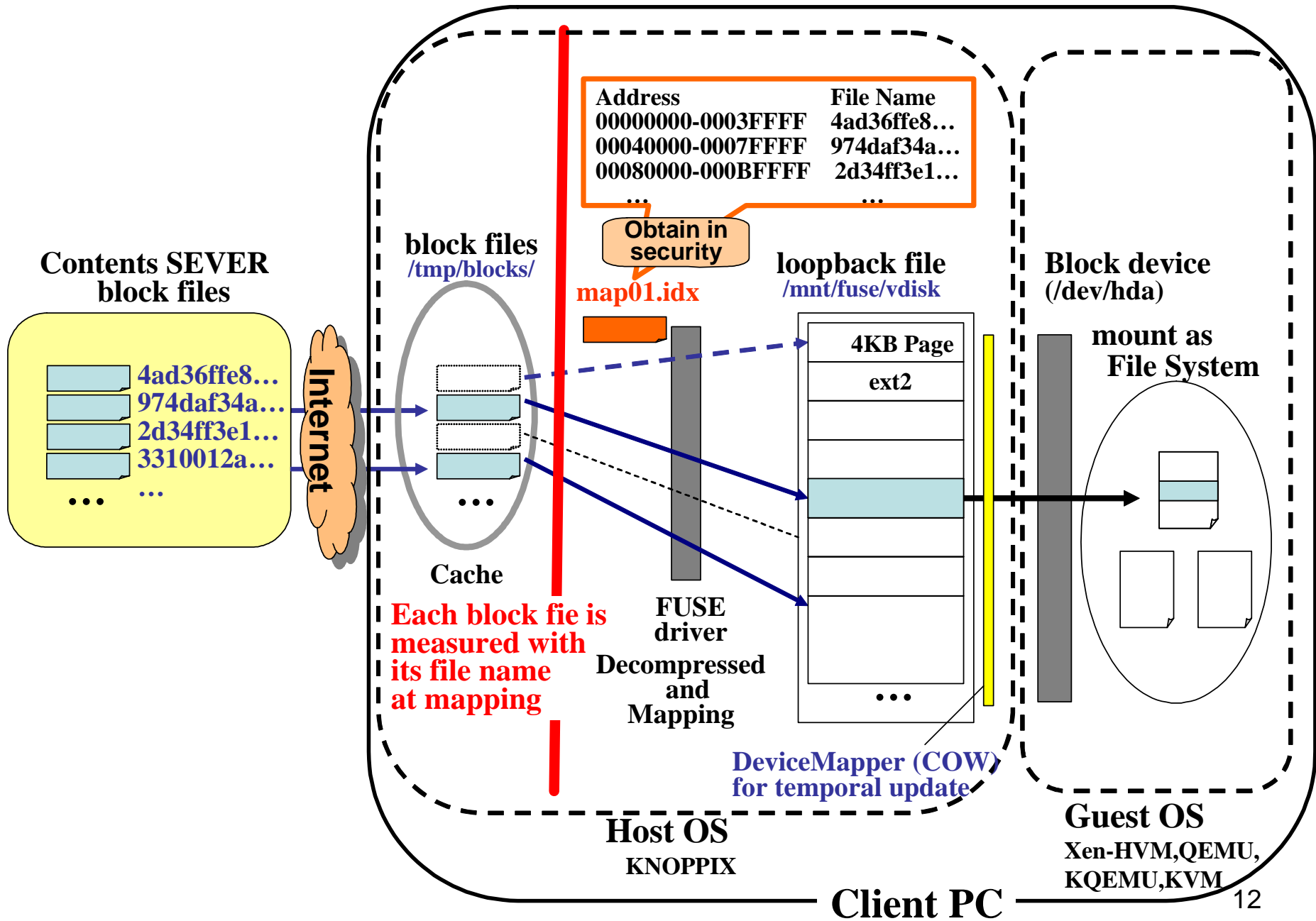
- The image of Trusted HTTP-FUSE CLOOP is made from existing normal block device.
- Original block device is split by 256KB and compressed by zlib.
Each data is saved to each “block file”.
- Block file name is a SHA1 value of its contents.
 - If there are same contents in blocks, they are expressed by 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.

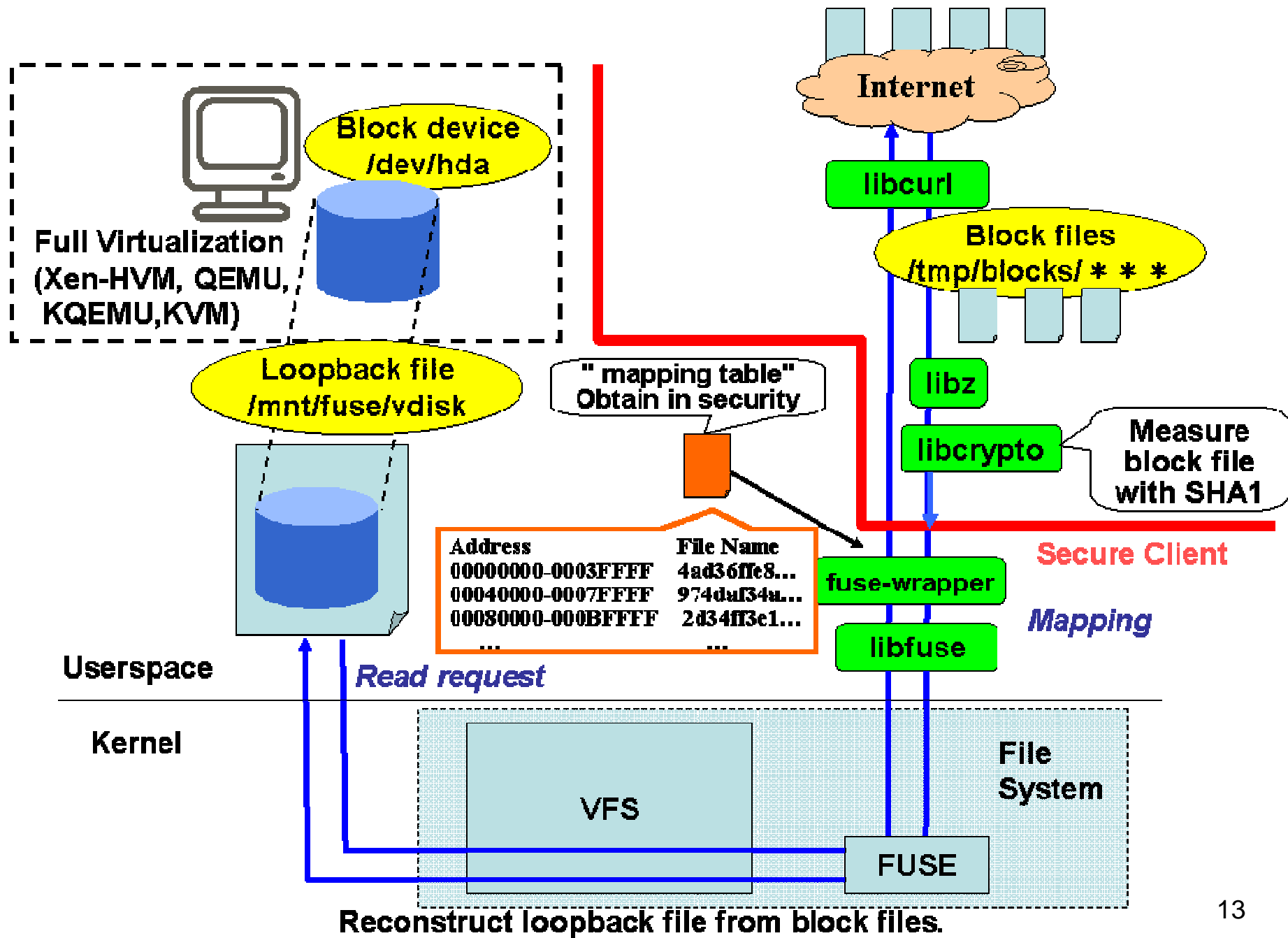


Trusted HTTP-FUSE CLOOP (2/2)

- 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.







Log of Trusted HTTP-FUSE CLOOP (/var/log/fs_wrapper_PID.log)

```
1150452051.109: #00000000(845b31ded38e15c1fa8febf97fe0781f23af98c3) :missed.  
1150452051.112: #00000000(845b31ded38e15c1fa8febf97fe0781f23af98c3) :hits.  
1150452051.112: #00000001(166cbaedbb1cc836e7c95d7d9943efde5a53829e) :missed.  
1150452051.113: #00000002(29c4e363dbad648072751ca1f856e5780dd2981d) :missed.  
1150452051.114: #00000003(fa8ad05b713a9cf8a701636ca6c353dc58fd6bfd) :missed.  
1150452051.114: #00000004(1f82a543fa9310c44eff6a13618beca3cacffc12) :missed.  
1150452051.128: #00000004(1f82a543fa9310c44eff6a13618beca3cacffc12) :hits.  
1150452051.128: #00000005(916f62a6e2caedc1279a0a74975a406ddb60ec25) :missed.  
1150452051.129: #00000006(19111dfc877a4fe241e125d10176d85a99b4bb86) :missed.  
1150452051.130: #00000007(950c1d7623b374f8e03309a93041f5adfa3af80f) :missed.  
1150452051.130: #00000008(486472b0ee27157d755bd59d623179cfc0034747) :missed.
```

falsified

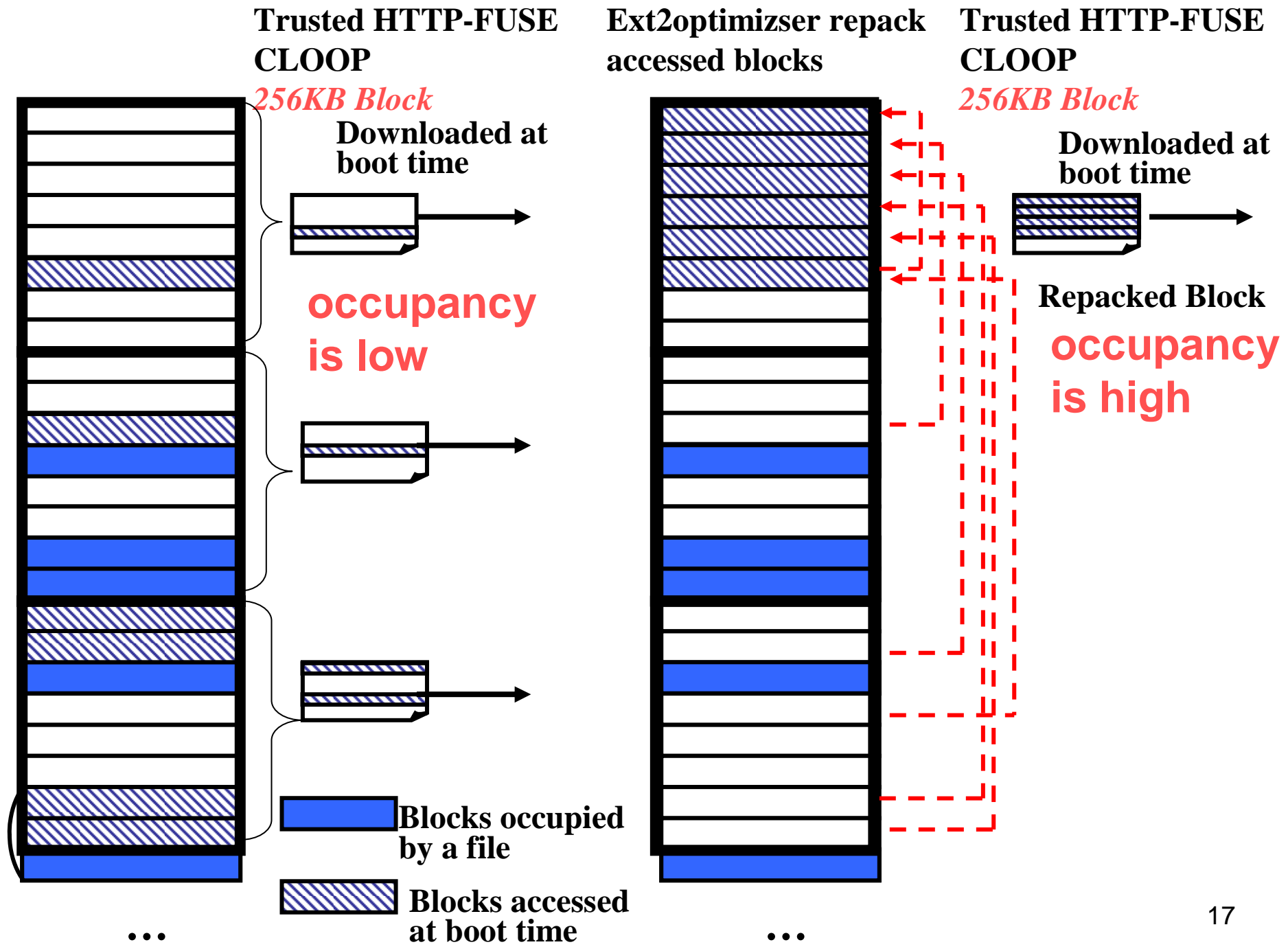
```
1150452375.989: #00000000(845b31ded38e15c1fa8febf97fe0781f23af98c3) :missed.  
1150452375.993: #00000000(845b31ded38e15c1fa8febf97fe0781f23af98c3) :hits.  
1150452375.993: #00000001(166cbaedbb1cc836e7c95d7d9943efde5a53829e) :missed.  
1150452375.994: #00000002(29c4e363dbad648072751ca1f856e5780dd2981d) :missed.  
1150452375.995: #00000003(fa8ad05b713a9cf8a701636ca6c353dc58fd6bfd) :missed.  
1150452375.996: #00000004(1f82a543fa9310c44eff6a13618beca3cacffc12) :missed.  
1150452375.997: #00000004(1f82a543fa9310c44eff6a13618beca3cacffc12) :hits.  
1150452375.997: #00000005(916f62a6e2caedc1279a0a74975a406ddb60ec25) :missed.  
1150452375.998: #00000006(19111dfc877a4fe241e125d10176d85a99b4bb86) :missed.  
E: can't validate block.
```

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.

World Wide Deployment of Server

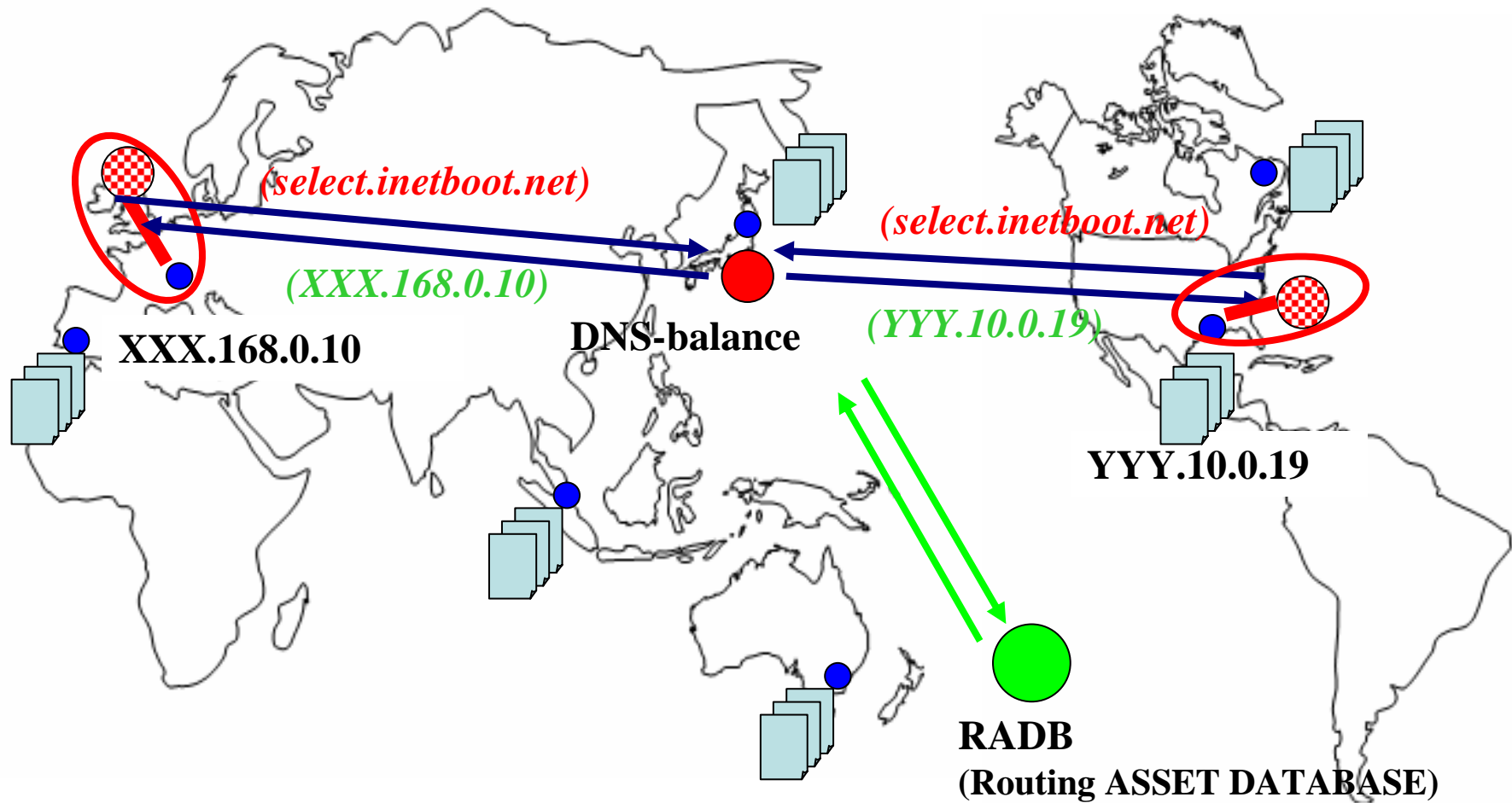
- We utilize inexpensive Web Hosting Service.
 - 5GB/ month from \$10



 Client

 Web server for Block Files

Resolve *select.inetboot.net*
by DNS-Balance(*ns.inetboot.net*).



Current Implementation of OS Circular

- VM Loader
 - VMKnoppix
 - KNOPPIX 5.1.1 + VM (QEMU090,KQEMU, KVM16, and Xen3.1.0)
 - Driver of Trusted HTTP-FUSE CLOOP
 - Setup script for OS Circular
- OS Images is obtained by Trusted HTTP-FUSE CLOOP
 - Contents
 - Debian GNU/Linux Etch (07/Dec/06, 11/Dec/06)
 - Periodically updated with “apt-get” command
 - Ubuntu Linux (6.06LTS, 6.10, 7.04)
 - CentOS5 Linux
 - FreeBSD
 - Download Sites
 - 10 US, 2 EU, and 7 Japan.

Performance

- ThinkPAD T60(Core2 Duo T7200 2Ghz, 2GB Memory)
 - Measure the boot time of Debian GNU/Linux till GDM
 - Compare 2 network delays (0 msec, 30 msec)
 - Compare 3 types of optimization (no, ext2opt, ext2opt+DLAHED)
 - Compare 2 virtual machines (Xen-HVM, KQEMU)

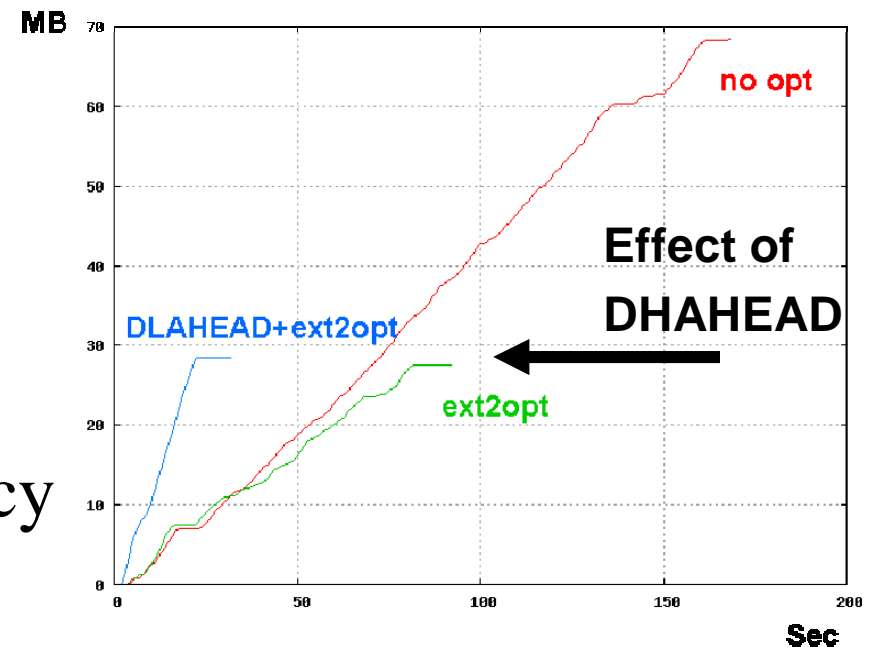
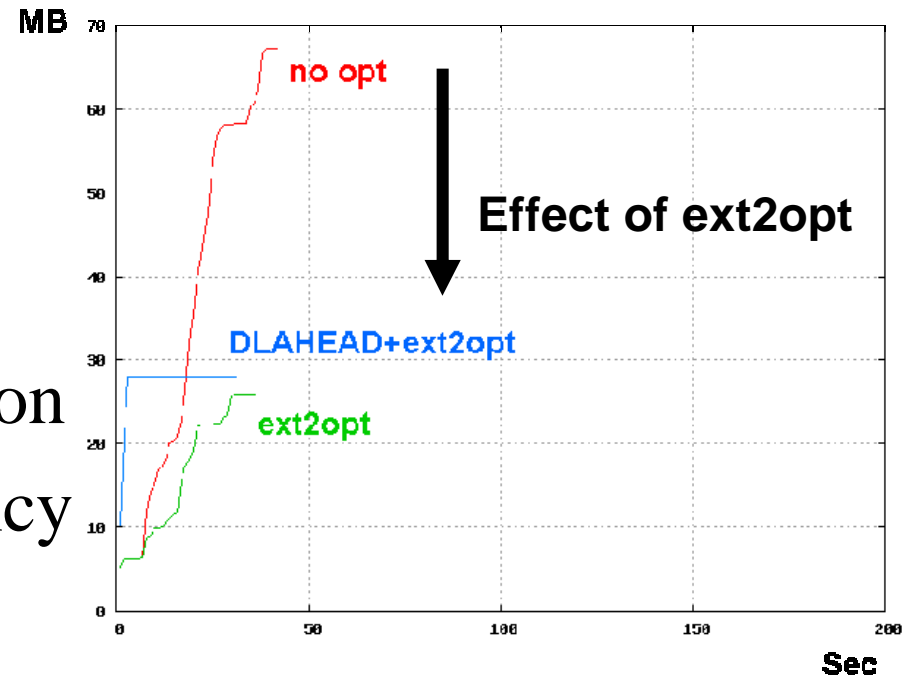
Compare network latency

Xen-HVM on 0 msec latency

•Result

- Ext2opt reduced the total download to half.
- DLAHAED was effective on long latency.
- The combination of Ext2opt and DLAHEAD increased slight download because of double download of Trusted HTTP-FUSE CLOOP and DLAHEAD.

Xen-HVM on 30 msec latency

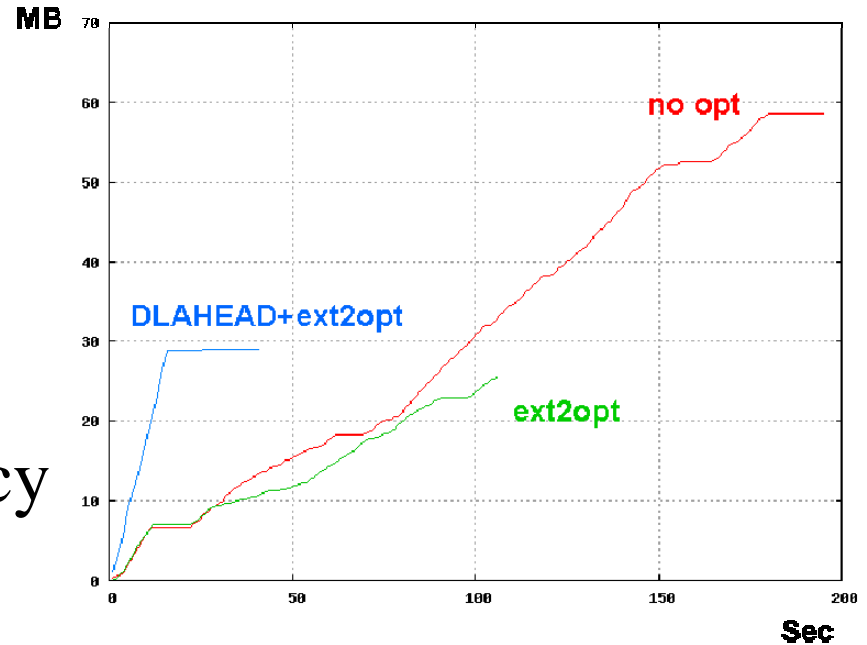


Compare virtual machine

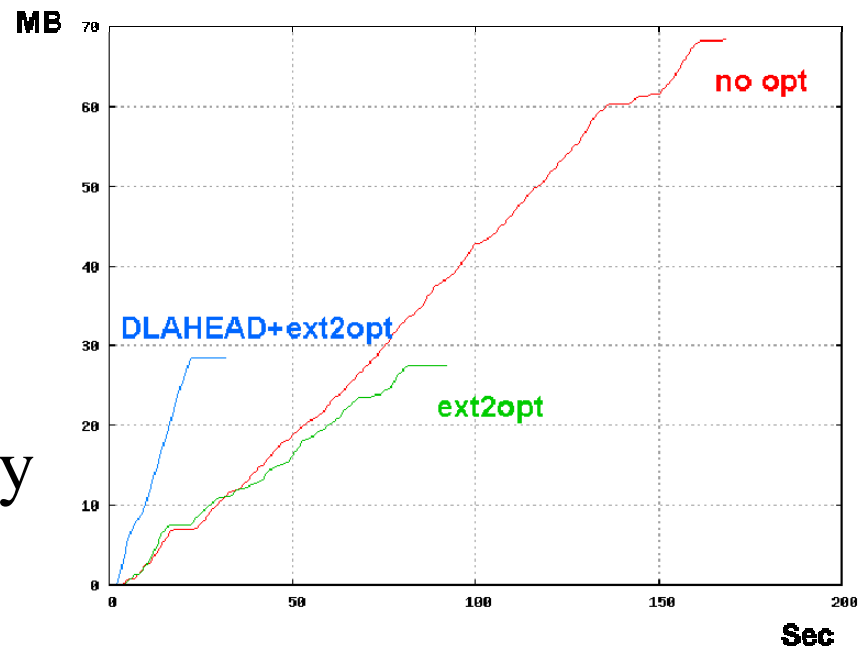
KQEMU on
30 msec latency

•Result

- KQEMU is slower than Xen-HVM in general but the performance is dominated by latency.
- Ext2opt and DLAHEAD makes close performance.



Xen-HVM on
30 msec latency



Related Work (OS Migration)

- OS Zoo <http://www.oszoo.org/>
 - Distribute Virtual Disk files of QEMU for Linux, Minix, Plan9, OpenSolaris, etc.
 - No ongoing maintenance
- Collective [HostOS'03, NSDI'05]
 - Cache based System Management
 - Based on COW file of VMware
 - COW files are shared by NFS over SSH
- LivePC of Moka5 <http://www.moka5.com>
 - Streaming download of Virtual Disk

Discussion (To be Trust)

- Current implementation has some problems
 - A) Current implementation has to trust VMKnoppix
 - Can't prevent Virtual Machine Based Rootkit (Subvirt[SSP'06])
 - We plan to integrate Trusted Computing to OS Circular.
 - Integrate Intel TXT/AMD-SVM(skinit)
 - B) Mapping table file has to distribute in secure way.
 - Distribute with TNC (Trusted Network Connect).
 - C) Current implementation has no way to authenticate correct update of the guest OS
 - The Guest OS should be checked by Vulnerability Database.
 - CVE (Common Vulnerabilities and Exposures) offers the database.
 - <http://cve.mitre.org/>

Conclusions

- OS Circular is a framework of Internet Disk Image Distributor for virtual machines.
 - OS Circular is consisted of VM Loader “VMKnoppix” and Stackable Virtual Disk “Trusted HTTP-FUSE CLOOP”.
- The current targets are some distributions of Linux and FreeBSD.
 - They are updated semi-automatically.
- The service is available.
 - <http://openlab.jp/oscircular/>