"OS Circular": A Framework of Internet boot with virtual machine

http://openlab.jp/oscircular/

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

• Purpose of OS Circular
• What is OS Circular?
  – 2 components
    • VM-Loader "VMKnoppix"
    • Internet Virtual Disk "HTTP-FUSE CLOOP"
  – "Guest OS" is periodically updated and keep security.
• Current Implementation
• Future Work
  – Trusted Boot with TPM (Trusted Platform Module) chip
  – OS Circular for PlayStation3.
• Conclusions
Purpose of OS Circular

- **Internet Disk Image Distributor** for Virtual Machine
  - Boot any OSes on anonymous PC without installation.
  - Deal with Disconnect network for mobile computing
    - Parts of disk Image are cached on local storage.
    - *The OS is periodically updated and keep security.*
      - But it allows to rollback to previous image.

- **OS Circular is used as a reference installation.**
  - New OS image enables to check the feasibility of new functions before installing it on the hard disk.
  - Old OS image keeps old application and enables to open old format files.
Strategy for deployment

• Policy
  – Utilize popular/inexpensive Internet service.
    • Most researches assume special services on the server. They prevent world wide deployment.

• Client Centric System
  – Current CPU is powerful and runs VM easily.
  – Disk Image is obtained via HTTP servers.
    • Reduce requirements for server and makes world wide scalability.
Related Work (OS Migration)

• OS Zoo
  – Distribute Virtual Disk files of QEMU.
    • Linux (Debian, Ubuntu etc), Minix, Plan9, OpenSolaris, etc.
    • No ongoing maintenance on the downloaded disk image
• FLOZ (Free Live OS Zoo)
  – The GUI of QEMU is transferred by VNC to Web Browser on Client PC.
  – **Server Centric System**
    • Bad response because the server exits in Italy
    • Require server power
    • Guest OS has no network service because of security
• Collective [HostOS'03][NSDI'05]
  – Cache based System Management
    • It update the disk image by the COW(CopyOnWrite) of VMware
    • COW files are shared by NFS over SSH
## Comparison

<table>
<thead>
<tr>
<th>OS</th>
<th>VM</th>
<th>Type of Management</th>
<th>Disk Image</th>
<th>Partial update</th>
<th>Transfer</th>
<th>Security on Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS Circular</td>
<td>Xen-HVM</td>
<td>Client Centric</td>
<td>Trusted HTTP-FUSE CLOOP</td>
<td>Log Structured blocks on HTTP-FUSE CLOOP</td>
<td>HTTP</td>
<td>Blocks are checked on Client</td>
</tr>
<tr>
<td>Collective</td>
<td>VMware</td>
<td>Client Centric</td>
<td>VMware disk file</td>
<td>COW of VMware</td>
<td>NFS over SSH</td>
<td>SSH</td>
</tr>
<tr>
<td>OS Zoo</td>
<td>Xen-HVM</td>
<td>Client Centric</td>
<td>QEMU disk file</td>
<td>HTTP &amp; FTP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLOZ</td>
<td>QEMU</td>
<td>Server Centric</td>
<td>Server Cenric</td>
<td>VNC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
OS Circular

- OS Circular is Client Centric System which utilizes virtualization technology.
  - "VM Loader" + "Internet Virtual Disks"
  - Client PC boots with the VM Loader and get Guest-OS image via Internet Virtual Disk
    - Guest-OS images are maintained on the server and keep security.
VM Loader

- VM Loader is consisted of "Virtual Machine software" and host OS which runs on real the real machine.

- The host OS supports real device drivers
  - KNOPPIX is used by SoulPAD [Mobisys'05] and Collective [NSDI'05], because KNOPPIX automatically detects devices and loads the appropriate Linux drivers.

- Virtual Machine offers an **Abstraction Layer** which is a common PC environment on any PCs.
  - The abstraction is popular for OS migration.
    - SoulPAD [Mobisys'05], Collective [NSDI'05], Internet Suspend/Resume [WMCSA'02] use VMware
Abstraction of VM&OS

- **Common Device Model (VM Abstraction)**
  - Guest OS only have to prepare fixed drivers.
  - QEMU-DM is popular
    - Xen-HVM, KVM, KQEMU are based on QEMU-DM.

- **Full Virtualization (VM Abstraction)**
  - It enables to use normal installer and security management for Guest OS.

- **Unify x86 Architecture (OS Abstraction)**
  - Recent OSes offer universal i386 packages for x86.
  - The difference of CPU type doesn't cause problem.
    - The CPU architecture of Xen-HVM depends on the real CPU but it doesn't cause problem.
"VMKnoppix" as VM Loader

- VMKnoppix is a collection of Virtual Machine softwares on KNOPPIX.
  - Xen, QEMU, KQEMU, KVM, VirtualBox, UserModeLinux
  - KNOPPIX boots as the host OS (Domain0 of Xen).
    - KNOPPIX prepares the device drivers for virtual machine software.
    - Virtual Machine boots with an Internet Disk Image.
Internet Virtual Disk

- Virtual Disk is Block Level Abstraction.
- Virtual Disk has requirements for OS Migration. (Pfaff[NSDI'06])
  - Versioning
    - Partial update & Rollback
  - Globalization
    - World Wilde Deployment
    - Network/Storage Transparent
      - Handle network (dis/re)-connection for mobile computing
  - Security
    - Virtual disks have to keep validness of contents
      - Prevent the falsification of block.
    - (Option) OS itself should be maintained by Security Software
    - (Option) Virtual machine has to be trustable
- We developed "Trusted HTTP FUSE CLOOP".
Trusted HTTP-FUSE CLOOP (1/3)

- It is made from normal block device which has root file system.
- Block device is split by 256KB and compressed by zlib. Each data is saved to each block file.
- Block files are managed by "index" file which is a mapping table of block files.
  - "index" file works as a header of CLOOP.
- Block file name is a SHA1 value of its contents.
  - If there is a same contests blocks, they are held together a same name file and reduce total storage space.
  - The basic idea is resemble to "Venti of Plan9"[USENIX'02]
The block files are reconstructed as a virtual disk with HTTP-FUSE CLOOP.

Block Device

<table>
<thead>
<tr>
<th>Address</th>
<th>File Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000000-0003FFFF</td>
<td>4ad36ffe8…</td>
</tr>
<tr>
<td>00040000-0007FFFF</td>
<td>974daf34a…</td>
</tr>
<tr>
<td>00080000-000BFFFF</td>
<td>2d34ff3e1…</td>
</tr>
<tr>
<td>000C0000-000FFFFF</td>
<td>3310012a…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

4KB Page
- ext2
- …
- …
- …
- …

index and block files

index.idx
- 4ad36ffe8…
- 974daf34a…
- 2d34ff3e1…
- 3310012a…
- …
Trusted HTTP-FUSE CLOOP (2/3)

• Block files are reconstructed to a CLOOP 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 of index file when it mapped to CLOOP. It can keep security.
Contents SEVER block files

Each block file is measured with its file name at reading

block files
/tmp/blocks

index.idx

CLOOP file(700MB)
/cdrom/KNOPPIX

CLOOP driver

Cache
64KB

FUSE driver

Obtain in security

header (location data)

Address
00000000-0003FFFF    4ad36ffe8…
00040000-0007FFFF    974daf34a…
00080000-000BFFFF    2d34ff3e1…
000C0000-000FFFFF    3310012a…
...

File Name
4ad36ffe8…
974daf34a…
2d34ff3e1…
3310012a…
...

Obtain in security

Loopback(2GB)
# losetup /dev/CLOOP /cdrom/KNOPPIX

Access

4KB Page
000000
040000
080000

ext2

CLOOP driver
Cache
64KB

decompress

extract

...
Trusted HTTP-FUSE CLOOP (3/3)

• When original block device is updated, the relevant block files are newly created with new SHA1 name. The "index" file is also renewed.
  – Old block files are reusable, because most parts of block is unchanged.

• HTTP for file deliver
  – Most popular and well designed.
    • 80 port is usually opened.
      – Other network block devices use special port which is usually closed.

• Block files are network/storage transparent.
  – Block files are cached and reused on local storage.
  – If necessary block files are stored in a local storage, network connection is not necessary.
Block Device (2GB)

4KB Page
ext2
...
...
...

256KB

CLOOP file

index.idx
4ad36ffe8...
974daf34a...
2d34ff3e1...
3310012a...
...

Same files
Reusable for FUSE

FUSE driver

apt-get install ...

Update
Xen HVM
disk=["phy:/dev/cloop0,hda1,ro"]

Internet
/libcurl
/tmp/blocks/**

//Obtain in security

"Index file"

/libcrypto
/fuse-wrapper
/libfuse
/glibc

Measure each block file

Reconstruct CLOOP file from block files.

ext3
VFS
FUSE

File System

Block Device

User-space
Security on OS Circular

• The Guest OS image has to keep security
  – The OS is updated periodically by package manager.
    • "yum" of Fedora or "apt-get" of Debian
      – It should be Semi-automatic.
  – The target includes the kernel.
    • We have to use Full Virtualization.

• The way of distribution
  – Index file is obtained in secure way because it is used to check block file by SHA-1 Digest on a Client.
  – Block files are downloadable from un-trusted servers.
Partial Update

Security Update

Block Files

Cache files at local storage

Index Files

HTTP Server

On demand download

Internet

New Block Files

Updated OS

New Index file

Master PC

Client PC

Xen HVM

Client PC
World Wide Deployment of Server

- We utilize inexpensive Web Hosting Service.
  - 5GB/ month from $10
Care for network latency

- **DNS-Balance**
  - DNS-Balance is a kind of name resolver which suggests *near mirror servers* with routing information offered by RADB.net
- **DLAHEAD (DownLoad AHEAD)**
  - Take a profile of downloaded block files at boot time.
  - *The necessary block files are downloaded in advance* with extra download connections
DNS-Balance

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

**Client**

**Web server for HTTP-FUSE Xenoppix**

**DNS server:** **ns.inetboot.net**

**Block files**

ASK the nearest server for **select.inetboot.net**:

- XXX.168.0.10
- YYY.168.0.19

**RADB** (Routing ASSET DATABASE)
Current Implementation of OS Circular

- VMKNOPPIX (VM Loader)
  - Based on KNOPPIX 5.1.1
  - Xen 3.0.4, QEMU0.9.0, KQEMU, KVM16
  - The driver for Trusted HTTP-FUSE CLOOP
    - Setup script for OS Circular

- OS Images is obtained by Trusted HTTP-FUSE CLOOP
  - Debian GNU/Linux
    - Periodically updated with "apt-get" command
  - FreeBSD
Performance

- Client: ThinkPAD T60 (Core Solo T1300 1.67Ghz)
  - Xen 3.0.4 HVM Domain
  - Debian GNU/Linux boot till GDM
- Server: Web hosting (network latency: 10msec)
To be Trust (1/3)

- There are some problems
  A) Current Version has to trust Xenoppix.
     - Can't prevent Virtual Machine Based Rootkit (Subvirt[SSP'06])
  B) Index file have to distribute in secure.
  C) There is no way to authenticate that vulnerable applications are replaced correctly on the updated OS.
To be Trust (2/3)

- **Trusted Boot with TPM (Trusted Platform Module)**
  - The boot procedure is measured and the measurements are registered on TPM.
  - The measurements are reported to Remote Attestation
- **Trusted Network Connect**
  - Allow connection for the Client which is authenticated by Remote Attestation

- The measurement by TPM are used for authentication of Xenoppix. **(A)**
- The "index" file is obtained using Trusted Network Connect **(B)**
  - It enables to boot commercial OS on OS Circular.
To be Trust (3/3)

• Vulnerability Database
  – The update of Guest OS should be checked by Vulnerability Databases.
  – CVE (Common Vulnerabilities and Exposures)
    – http://cve.mitre.org/

• We can check that the OS image doesn't include vulnerable software. (C)
OS Circular for PlayStation3

- OS Circular is not only applied to virtual machines but also applied to the game machine "PlayStation 3". It is called HTTP-FUSE PS3 Linux.
  - PlayStaion3 offers us the same device model.
- PlayStation3 uses "kboot" as a boot-loader, which is stored at build-in 4MB Flash memory.
  - Kboot sets up IP address and enable to download kernel and initrd via HTTP form Internet.
Architecture of PlayStation 3

- Graphics
- SPE
  - PPE
  - CELL processor
- Memory: 256MB
- South bridge
- Block device aren't required at boot time
- 4MB Flash (kboot)
- Boot starts from 4MB Flash
- HDD
- USB
- GbE
Boot sequence of kboot

Initial bootloader

Kboot on internal Flash

Start mini Linux system

Initialize network

Download kernel and initrd from network

Execute the kernel using kexec

Start Linux

kboot: http://***/vmlinux initrd=http://***/initrd

DHCP (LAN)

HTTP (Internet)

vmlinux (kernel for cell)

Initrd (HTTP-FUSE CLOOP)
Boot sequence of PS3 Linux

- kboot
- Initialize hardware
- Configure network
- Mount HTTP-FUSE LOOP as root filesystem
- Start init

Mapping on demand

HTTP (Internet)
- Index Files
- Block Files
Status of HTTP-FUSE PS3 Linux

• It boots PS3 Debian.
• Traffic at boot time is about 170MB
  – From boot to finish starting gnome
• Release soon!

• Future Plan
  – We will includes the SDK for SPE (Synergistic Processing Element). It is used to develop the de-compress engine for cloop.
    • Machida, "SPE-assisted User Level Device Driver on Cell Processor ", Embedded Linux Conference April/2007
    • Tsukamoto, "BOF: Distributed cluster computing on Cell/PS3", Ottawa Linux Symposium June/2007
Conclusions

- OS Circular is Framework of Internet Client.
- Internet Virtual Disk "Trusted HTTP-FUSE CLOOP" offers us disk images for Virtual Machine.
- The current Guest OSes are Debian GNU/Linux and FreeBSD.
  - Debian is updated by "apt-get" semi-automatically.
- OS Circular is applied to PlayStation3.

- Future work
  - OS Circular will integrate Trusted Boot and guest OS will link to Vulnerability Database.
Related Presentation

- **Linux Symposium 2006**
  - "HTTP-FUSE Xenoppix"

- **Linux Kongress2006**
  - "Trusted Boot of HTTP-FUSE KNOPPIX"
  - [http://www.linux-kongress.org/2006/abstracts.html#4_2_2](http://www.linux-kongress.org/2006/abstracts.html#4_2_2)

- **Virtualization Miniconf at Linux.Conf.Au 2007**
  - OS Circulation environment "Trusted HTTPFUSE Xenoppix"

- **Embedded Linux Conference 2007**
  - HTTP-FUSE PS3 Linux which is internet boot framework with kboot