User-Mode-Linux

Trent “Lathiat” Lloyd
Lathiat@sixlabs.org

What is UML?

What is UML?

- Linux Kernel
- User-Mode Process
- No root privileges required
- No hardware emulation (CPU etc)

Alternatives

- VMWare
  - Slower (CPU/Hardware Emulation/Virtualization)
  - Needs root privileges for networking
- Bochs/Plex86
  - Emulate Hardware
  - Can run any OS and runs on many OS'
- FreeBSD Jail/Chroot Jail
  - Restrictive Networking
  - Creates files on host file systems
  - Requires root setup

Features / Uses

- Virtual Disks
- Virtual Networking
- Virtual PCI
- Hostfs
- Virtual Serial Lines
- Management Console
- Kernel Debugging SMP
- Multiple Instances
- No root needed
- Low resource requirements
- Virtual Hosting
- Security/Isolation
- Secure access for un-trusted users
- Debugging Kernels
- Trying dangerous code
- Disaster Simulations
  - I.e. rm –rf /

Getting UML

- Website: http://user-mode-linux.sf.net/
  - Compile
  - Binaries
  - RPMs
- Do not use the Debian Package
  - Evil and possessed

Compiling UML

- Get uml-patch-2.4.20-4.bz2, linux-2.4.20.tar.bz2

```
lathiat@seven:~/uml$ tar jxf linux-2.4.20.tar.bz2
lathiat@seven:~/uml$ cd linux-2.4.20
lathiat@seven:~/uml/linux-2.4.20$ bzcat ../uml-patch-2.4.20-4.bz2 | patch –p1
                    [patch output]
lathiat@seven:~/uml/linux-2.4.20$ make menuconfig ARCH=um
                    [configure]
lathiat@seven:~/uml/linux-2.4.20$ make dep ARCH=um
lathiat@seven:~/uml/linux-2.4.20$ make modules ARCH=um
lathiat@seven:~/uml/linux-2.4.20$ make linux ARCH=um
```

The Root Filesystem

- Debian
  - Download pre-made image
  - Deboostrap
  - Install on a partition then boot in UML
- Redhat
- Mandrake
- Others

Creating a Debian RootFS

```
uml@seven:~/isha2$ dd if=/dev/zero of=root bs=1M seek=1024 count=1
1+0 records in
1+0 records out
1048576 bytes transferred in 0.031510 seconds (33277573 bytes/sec)
uml@seven:~/isha2$ /sbin/mke2fs root
mke2fs 1.33 (21-Apr-2003)
root is not a block special device.
Proceed anyway? (y,n) y
uml@seven:~/isha2$ sudo su
seven:/home/uml/isha2# mount -o loop root /uml
seven:/home/uml/isha2# deboostrap sid /uml ftp://ftp.uwa.edu.au/mirrors/linux/debian
I: Base system installed successfully.
seven:/home/uml/isha2# cd /uml/etc
```

fstab

```
# /etc/fstab: static file system information.
#
# <file system> <mount point>   <type>  <options>               <dump>  <pass>
/dev/ubd/0      /               ext3    errors=remount-ro       0       1
proc            /proc           proc    defaults                0       0
#dev            /dev/           devfs   defaults                0       0
```

Other Files

- /etc/hostname

isha

- /etc/hosts

127.0.0.1 isha localhost

The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
ff02::3 ip6-allhosts

- /etc/apt/sources.list

deb ftp://ftp.uwa.edu.au/mirrors/linux/debian sid main contrib non-free

Consoles

- /etc/inittab

```
1:2345:respawn:/sbin/getty 38400 ttys/0
#2:23:respawn:/sbin/getty 38400 tty2
#3:23:respawn:/sbin/getty 38400 tty3
#4:23:respawn:/sbin/getty 38400 tty4
#5:23:respawn:/sbin/getty 38400 tty5
#6:23:respawn:/sbin/getty 38400 tty6
```

- /etc/securetty

```
ttys/0
```

Starting it UP

```
./linux
        tracing thread pid = 15380
Linux version 2.4.20 (jathiat@seven) (gcc version 3.2.3)#1 Sun May 11 14:56:13 WST 2003
On node 0 totalpages: 8192
zone(0): 8192 pages.
zone(1): 0 pages.
zone(2): 0 pages.
Kernel command line: ubd0=ishe/root_mode=ft con0= fd:0,fd:1 con=null root=/dev/ubd0
Calibrating delay loop... 204.22 BogoMIPS
Memory: 29340k available
Dentry cache hash table entries: 4096 (order: 3, 32768 bytes)
Inode cache hash table entries: 2048 (order: 2, 16384 bytes)
Mount cache hash table entries: 512 (order: 0, 4096 bytes)
Buffer-cache hash table entries: 1024 (order: 0, 4096 bytes)
Page-cache hash table entries: 8192 (order: 3, 32768 bytes)
Checking for host processor cmov support...No
Checking for host processor xmm support...No
Checking that ptrace can change system call numbers...OK
Checking that host ptys support output SIGIO...Yes
Checking that host ptys support SIGIO on close...No, enabling workaround
POSIX conformance testing by UNIFIX
Linux NET 4.0 for Linux 2.4
Based upon Swansea University Computer Society NET3.039
Initializing RT netlink socket
Starting kswapd
VFS: Disk quotas version dquot_6.4.0 initialized
devfs: v1.12c (20020818) Richard Gooch (rgooch@atnf.csiro.au)
devfs: boot_options: 0x1
JFFS version 1.0, (C) 1999, 2000 Axis Communications AB
JFFS2 version 2.1. (C) 2001 Red Hat, Inc., designed by Axis Communications AB
pty: 256 Unix98 ptys configured
SLIP: version 0.8.4-NET3.019-NEWTTY (dynamic channels, max=256).
RAMDISK driver initialized: 16 RAM disks of 4096K size 1024 blocksize
loop: loaded (max 8 devices)
PPP generic driver version 2.4.2
Universal TUN/TAP device driver 1.5 (C)1999-2002 Maxim Krasnyansky
SCSI subsystem driver Revision: 1.00
scsi0 : scsi_debug, Version: 0.61 (20020815), num_devs=1, dev_size_mb=8, opts=0x0
  Vendor: Linux    Model: scsi_debug       Rev: 0004
  Type:   Direct-Access                    ANSI SCSI revision: 03
blkmtd: error, missing `device' name
Initializing software serial port version 1
mconsole (version 2) initialized on /home/jathiat/.uml/Ccgkbj/mconsole
Partition check:
 ubd: unknown partition table
UML Audio Relay (host dsp= /dev/sound/dsp, host mixer = /dev/sound/mixer)
Initializing stdio console driver
NET4: Linux TCP/IP 1.0 for NET4.0
IP Protocols: ICMP, UDP, TCP
IP: routing cache hash table of 512 buckets, 4Kbytes
TCP: Hash tables configured (established 2048 bind 4096)
NET4: Unix domain sockets 1.0/SMP for Linux NET4.0
VFS: Mounted root (ext2 filesystem) readonly.
Mounted devfs on /dev
INIT: version 2.84 booting
Activating swap.
Checking root file system...
fsck 1.27 (8-Mar-2002)
/dev/ubd/0: clean, 11992/131072 files, 276477/262144 blocks
System time was Sun May 11 10:20:37 UTC 2003.
Setting the System Clock using the Hardware Clock as reference...
hwclock is unable to get I/O port access:  the iopl(3) call failed.
System Clock set. System local time is now Sun May 11 10:20:38 UTC 2003.
Calculating module dependencies... depmod: Can't open /lib/modules/2.4.20/modules.dep for writing
done.
Loading modules:
modprobe: Can't open dependencies file /lib/modules/2.4.20/modules.dep (No such file or directory)
Checking all file systems...
fsck 1.27 (8-Mar-2002)
Setting kernel variables.
Loading the saved-state of the serial devices...
Mounting local filesystems...
mount: /dev already mounted or /dev busy
Running 0dns-down to make sure resolv.conf is ok...done.
```

Login

INIT: Entering runlevel: 2
Starting system log daemon: syslogd.
Starting kernel log daemon: klogd.
Starting internet superserver: inetd.
Starting PCMCIA services: module directory
/lib/modules/2.4.20/pcmcia not found.
Starting deferred execution scheduler: atd.
Starting periodic command scheduler: cron.

Debian GNU/Linux 3.0 isha ttys/0

isha login:

SKAS

- Separate Kernel Address Space
- Runs all UML processes as three processes
  - Kernel Space
  - User Space
- UML is in address space of its processes
- Processes can read it, UML jail's it so at least it can't be written
- Noticeable Speedup
- Requires a patch to the host kernel (can be module)

Networking Methods

- uml_switch
- TUN/TAP
  - Bridging
- TAP
- Slirp
- SLIP
- Multicast

Uml_switch

- Provides a totally virtual network
- No connection to a host network
- Needs a UML to route for outside access or can be attached to a TAP device
- Eth0=daemon,ip,socktype,controlsocket,datasocket
  - Can leave all them off except 'daemon'

TUN/TAP

- Creates a separate device on the host
- Allows direct network access for any protocol
- Can be bridged to another network
- Can use uml_net helper to setup IP forwarding, routing, proxy ARP etc
- Preconfigured TAP devices
- eth0=tuntap,,,192.168.0.12

Configuring a TAP Device

- The ‘tunctl’ program is used
- tunctl –u <user/id>
  - Outputs device name
  - Can be deleted with tunctl -d

```
ifconfig tap0 192.168.0.1 up
route add -host 192.168.0.253 dev tap0
host# bash -c 'echo 1 > /proc/sys/net/ipv4/ip_forward‘
host# route add -host 192.168.0.253 dev tap0
host# bash -c 'echo 1 > /proc/sys/net/ipv4/conf/tap0/proxy_arp‘
host# arp -Ds 192.168.0.253 eth0 pub
```

Bridging with TUN/TAP

- Configuring bridges allows direct connection to the network
- No ARP Proxy
- Can speak protocols host doesn't route.

Setting up a Bridge

host# brctl addbr br0
host# ifconfig eth0 0.0.0.0 promisc up
host# ifconfig tap0 0.0.0.0 promisc up
host# ifconfig br0 192.168.0.1 netmask 255.255.255.0 up
host# brctl stp br0 off
host# brctl setfd br0 1
host# brctl sethello br0 1
host# brctl addif br0 eth0
host# brctl addif br0 tap0
Then setup the UML internal interface as if it was on your lan

Debian Magic

```
iface br0 inet static
        address 130.95.13.25
        netmask 255.255.255.192
        network 130.95.13.0
        broadcast 130.95.13.63
        gateway 130.95.13.3
        bridge_ports eth0
        bridge_fd 1
        stp eth0 off
iface br0 inet6 static
        address 2001:388:7094:4080::7
        netmask 64
        gateway 2001:388:7094:4080::1
```

Startup/Shutdown Magic

```
#!/bin/sh
TAPDEV=`sudo tunctl -u $2 |cut -d\' -f2`
echo $TAPDEV > /var/run/uml/tap-$2
sudo brctl addif br0 $TAPDEV
sudo su $2 "screen -c /dev/null ~uml/bin/linux ubd0=$1/root
    mode=tt con0=fd:0,fd:1 con=null &&
    /home/uml/cleanup_tap"
#!/bin/sh
USR=`/usr/bin/id|/usr/bin/cut -d\( -f2 |/usr/bin/cut -d\) -f1`
TAPDEV=`cat /var/run/uml/tap-$USR`
/usr/bin/sudo brctl delif br0 $TAPDEV
/usr/bin/sudo ifconfig $TAPDEV down
/usr/bin/sudo tunctl -d $TAPDEV
```

IPv6

- Bridging(+TAP) is good for IPv6
  - Allows direct connection to other routers
- TAP
  - Second best, because SLIP/daemon/multicast etc don't allow/support it as easily or at all

Slirp

- Allows a network with no root privs
- Virtual SLIP server, allows NAT-like functions, port forwarding

SLIP

- eth0=slip,1.2.3.4
- Interface is ‘umn’ inside UML
- Can only have one SLIP device

Multicast

- Let's UML's talk to each other without root
- No outside access
- Like a hub and all UMLs are in promiscuous
- Needs multicast in kernel
- Generates lots of traffic

Management Console

- Controls UML
- Sysrq
  - Reboot
  - Sync
  - Etc
- Ctrl+Alt+del
- Add network/disk/IO lines

Linux From Scratch

- Saves reboots

Questions?

Thanks to:

David Coulson (Hosted UML)
Abdul Basit/NextGenCollective.net
William Stearns (Hosted UML)
Paul Day/Bur.st Networks

THE END

1ca::6 - IPv6 Mini-Conference
Linux.conf.au 2004
Adelaide
http://conf.sixlabs.org/