Know your tools
KVM

Dariusz Puchalak
Dariusz_Puchalak < at > ProbosIT.pl

O mnie.

Linux/Unix – 14+ lat
KVM ~ 2.5lat

IT „Consulting” ~ 10 lat

I'm NOT a programmer!

Historia

Pierwsza stabilna wersja:

- Linux kernel 2.6.20 2007.02.04 11:10
- Intel VT-x

egrep '^flags.*(vmx|svm)' /proc/cpuinfo

Wirtualizacja

- QEMU
- VirtualBox
- Xen
- KVM
- Vserver/OpenVZ
- LXC
- Bochs
- UML
- VMWare
- Lguest
- Wine
- Dosemu/DosBox

Po co wirtualizacja?

?

Typy wirtualizacji

- Para wirtualizacja
- Pełna (sprzętowa) wirtualizacja

Dlaczego KVM?

?

Dlaczego KVM?

- Wsparcie 32/64 bity
- SMP host i guest
- Migration + Live migration
- Upstream
- Bezpieczeństwo
- Guest swapping
- All host features :)
- ...

Bezpieczeństwo

–Kernel zabezpiecza przed „nie root'ami”
–Procesy innych użytkowników – izolacja
–Selinux/AppArmor – dodatkowa izolacja
–Minimalny kod w nadzorcy (Hypervisor)
• ~2k linii - x86 emulator
• ~3k linii – obsługa MMU
–Wykorzystuje istniejące mechanizmy i kod linuksowy

Format wirtualnych dysków

kvm-img (qemu-img)

Supported formats: cow qcow vmdk cloop dmg bochs vpc vvfat qcow2 parallels nbd host_cdrom host_floppy host_device raw tftp ftps ftp https http

kvm-img

- check [-f fmt] filename
- create [-F base_fmt] [-b base_image] [-f fmt] [-o options] filename [size]
- commit [-f fmt] filename
- convert [-c] [-f fmt] [-O output_fmt] [-o options] [-B output_base_image] filename [filename2 [...]] output_filename
- info [-f fmt] filename
- snapshot [-l | -a snapshot | -c snapshot | -d snapshot] filename

Trochę praktyki

```
kvm -m 512 \
-net nic,model=rtl8139 \
-net user \
-vga std \
-no-kvm-pit \
-usbdevice tablet -usb \
-boot c -localtime \
-hda XP.img \
-name XP_Test1
```

Trochę praktyki -net

-net nic[,vlan=n][,macaddr=mac]
[,model=type][,name=name][,addr=addr]
[,vectors=v]

models:
"virtio", "i82551", "i82557b", "i82559er",
"ne2k_pci", "ne2k_isa", "pcnet", "rtl8139",
"e1000", "smc91c111", "lance", "mcf_fec"

vlan:
np. VLAN=1234567

Trochę praktyki -net

-net user [,option][,option][,...]
np.
hostfwd=[tcp|udp]:[hostaddr]:hostport-
[guestaddr]:guestport"

-net tap[,vlan=n][,name=name][,fd=h]
[,ifname=name][,script=file]
[,downscript=dfile]

Trochę praktyki -net

-net socket[,vlan=n][,name=name][,fd=h]
[,listen=[host]:port][,connect=host:port]

-net socket[,vlan=n][,name=name][,fd=h]
[,mcast=maddr:port]

-net vde[,vlan=n],[,name=name],
[sock=socketpath][,port=n]
[,group=groupname][,mode=octalmode]

Trochę praktyki

-net dump[,vlan=n][,file=file][,len=len]

-net none

-vga type
type = cirrus, std, vmware, none

-vnc display[,option[,option[,...]]]
-sdl
-snapshot

Trochę praktyki -drive

-drive option[,option[,option[,...]]]
"file=file"
"if=interface" ide,scsi, sd, mtd, floppy,
pflash, virtio
"media=media"
"cyls=c,heads=h,secs=s[,trans=t]"
"snapshot=snapshot"
cache=cache none,writeback,writethrough
"format=format" raw
"serial=serial"

Trochę praktyki

```
kvm -m 256 \
-net nic,model=virtio... -net .... -vga std\
-drive file=...system.img,if=virtio,boot=on \
-drive file=...data.img,if=virtio \
-name "Debs mirror system" \
-daemonize -ctrl-grab \
-monitor unix:...sys.mon,server,nowait \
.....
```

KVM a VDE

```
auto sw1-dmz
iface sw1-dmz inet manual
 vde2-switch -t sw1-dmz -n 256 -group dmz
 post-up /usr/sbin/brctl addif br0 sw1-dmz
 post-up /sbin/ifconfig sw1-dmz up
 post-down /usr/sbin/brctl delif br0 sw1-dmz
```

VDE w większej skali

dpipe - bi-directional pipe command
vde_plug - Virtual Distributed Ethernet plug
(two plugs creates a vde cable)

dpipe vde_plug = vde_plug /tmp/vde2.ctl

dpipe vde_plug = ssh remote.org vde_plug

KVM a Linux

ps auxwf | grep kvm

1000 9422 100 0.2 484576 10312 ?
Sl 15:09 0:05 kvm -m 256 -net nic,.....

KVM a CPU

cpulimit -e qemu-system-x86_64 -l 50

cpulimit -p 1234 -l 50

cpulimit -P /usr/bin/qemu-system-x86_64 -l 50

KVM a control groups

cgroups:
- CPU
- Memory

Disk

Wydajność

VirtIO

Parawirtualizacja

Paravirtualization is dead!

Parawirtualizacja

Oczywiście poza:
- I/O
- Zegarem
- generator liczb losowych

KSM

Kernel Shared Memory

Można upakować do 300% więcej maszyn (dla windowsów).
Gorzej dla Linuksów.

Zerowanie zwolnionej pamieci.

Zarządzanie

- CLI
- Libvirt
- Ovirt
- Cobbler
- Libguestfs
- ConVirt
- .....

Problemy KVM

Wersje:
- kvm 88
- qemu
- kvm-qemu

- „running target” (np. -vga std vs -std-vga)
- Brak współdzielenia clipboard
- PCI pass through (wymaga VT-d)
- Real mode dla intela

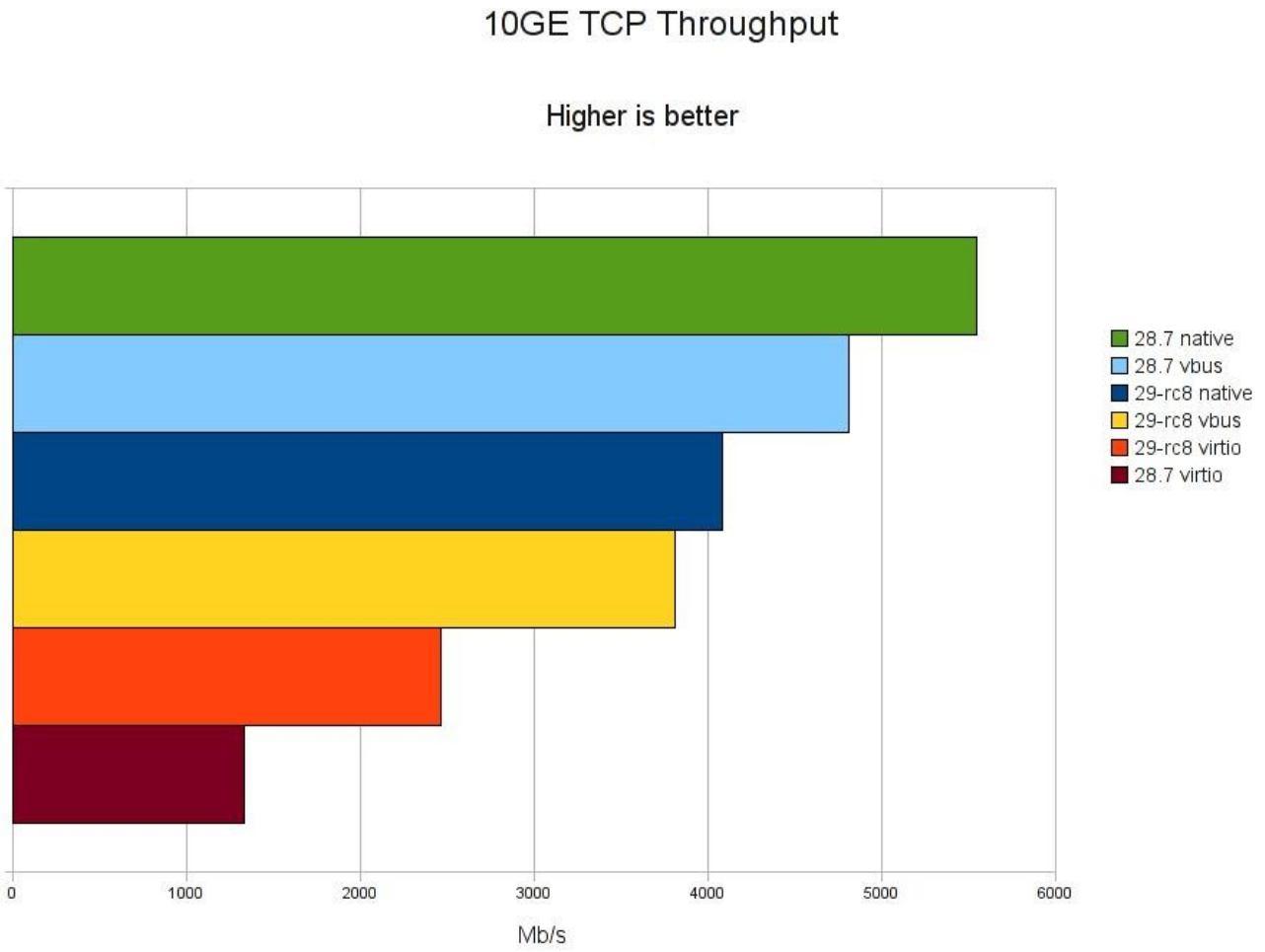
Przyszłość

Vmchannel

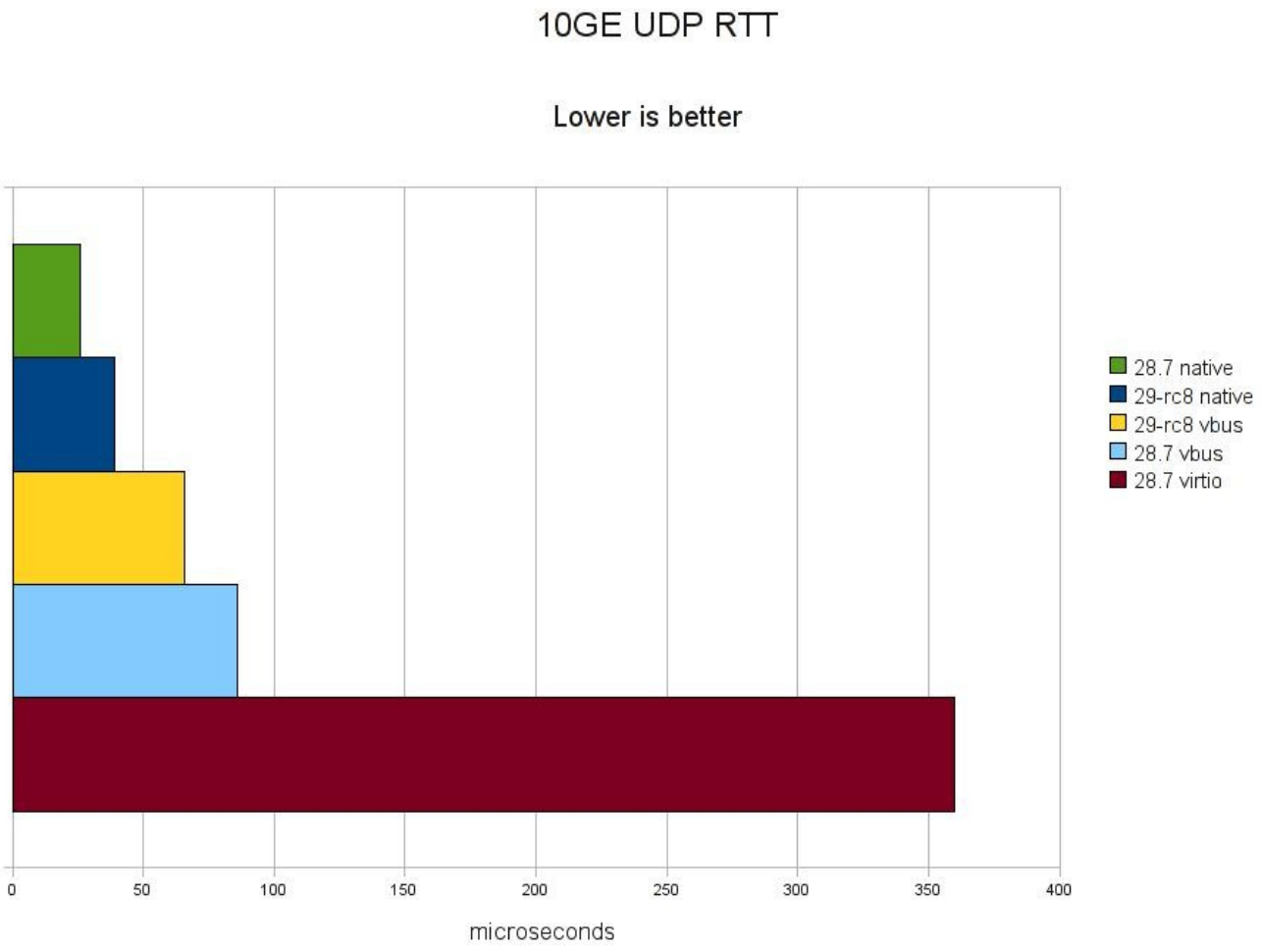
Spice http://www.spice-space.org/

virtual-bus / AlacrityVM

Wydajność


10GE TCP Throughput
Higher is better
28.7 native
28.7 vbus
29-rc8 native
29-rc8 vbus
29-rc8 virtio
28.7 virtio
0
1000
2000
3000
4000
5000
6000
Mb/s

Opóźnienia



*29-rc8 virtio results of 4016us excluded due to likely bug

Pytania?

?

Dziękuje

Dariusz_Puchalak@proboslT.pl