Kontejnerová virtualizace na Linuxu

Pavel Šimerda
pavel.simerda@netinstall.cz

40. konference EurOpen.CZ

http://data.pavlix.net/europen/40/

Specifika kontejnerů

- Sdílený kernel, iptables a další
- Jmenné prostory / skupiny prostředků
- Absence exaktního přidělování prostředků
- Pouze limity na využívání prostředků
- Minimální režie, možnost spustit více služeb

Implementace

- Linux Containers (LXC)
- OpenVZ
- Linux-VServer (?)

OpenVZ

- OpenVZ patches
- Kernel 2.6.32 pro RHEL
- OpenVZ kernel v Debianu

Instalace OS

- Vytvoření virtuálu dle tarballu (`vzctl create`)
- Šablony z wiki.openvz.org
- Konverze existujících systémů
- Debootstrap a podobné, chroot
- Nutné úpravy: `udev`, `MAKEDEV`
- Jaderné moduly (iptables)

Virtuální filesystém

- Jednoduchý `chroot`
- Úložiště: `/var/lib/vz/*/private`
- Přípojný bod: `/var/lib/vz/*/root`
- Možnost mountování jincých filesystémů
- Data na LVM
- Disková kvóta (`--diskspace`)

Obrazy a migrace

- Filesystém a konfigurace
- Vytvoření a zrušení snímku (`vzctl checkpoint`)
- Živá migrace (`vzmigrate --online`)
- Kombinace s dynamickým routingem (OSPF)

Rozdělení systémových prostředků

- Rozdělení procesorového času (`--cpuunits`)
- Absolutní omezení procesorového času (`--cpulimit`)
- Paměťové limity (`/proc/bc/resources`)

Primární limity

- Procesy (`--numproc`)
- TCP sockety (`--numtcpsock`)
- Ostatní sockety (`--numothersock`)
- Garantovaná alokovatelná paměť (`--vmguarpages`)

Sekundární limity

- Jaderná paměť (`--kmemsize`)
- Suma odchozích bufferů TCP (`--tcpsndbuf`)
- Suma příchozích bufferů TCP (`--tcprcvbuf`)
- Suma lokálních bufferů (`--othersockbuf`)
- Suma příchozích UDP bufferů (`--dgramrcvbuf`)
- Garantovaná paměť při OOM (`--oomguarpages`)
- Limit na alokaci RAM (`--privvmpages`)

Pomocné parametry

- Sdílená paměť (`--shmpages`)
- Skutečně využité stránky (`--physpages`)
- Souborové deskriptory (`--numfile`)
- Souborové zámky (`--numflock`)
- Pseudoterminály (`--numpty`)
- Signály (`--siginfo`)
- Pravidla v iptables (`--numiptent`)
- Swapovaná paměť (`--swappages`)

Virtuální konzole

- Přístup nezávislý na síti (`vzctl enter`)
- Možnost pouštět samostatné příkazy (`vzctl exec`)

Virtuální IPv4/IPv6

- IP adresa se konfiguruje „zvenčí“
- Minimum pro fungování IP vrstvy
- Fungují globální IPv4 i IPv6 adresy
- Dostatečné pro provoz serverových aplikací
- `--ipadd`, `--ipdel`
- Vyžaduje zapnutý IPv4/IPv6 forwarding
- Problémy s dynamickým routingem

Virtuální Ethernet

- Virtuální ethernetové rozhraní
- Neposkytuje pokročilé konfigurace
- Nefunguje IPv6 SLAAC
- Nefunguje IPsec tunnel mode
- Nefungují některé kombinace vlan a bridge
- Objevuje se spousta dalších neočekávaných chyb
- `--netif_add`, `--netif_del`

Konvergence OpenVZ a LXC

- Kontejnery v Linuxu
- OpenVZ patche
- OpenVZ nástroje
- LXC nástroje

Bezpečnost kontejnerů

- Kontejnery versus fyzické stroje
- Kontejnery versus plná virtualizace či paravirtualizace
- Firewall
- SELinux
- Možné zranitelnosti