Solaris Zones revisited

Arbeitsgemeinschaft Rechnerbetrieb

Oliver Schonefeld
oschonef@techfak.uni-bielefeld.de

6. Juni 2006

Übersicht

Was sind Solaris Zones?

- Virtualisierung von System Resourcen
  - “operating system-level virtualization”
- ähnlich wie FreeBSD Jails oder Virtuozzo (Linux vServer)
- mehrere Software Partitionen mit einer Betriebssystem Instanz
- Resource Managment Facilities
- eine *global* und mehrere *non-global* Zones

Ziel: Konsolidierung von Systemen und/oder “physikalische” Trennung von Diensten.

“Zone”, “Container” oder beides?

- *Zone* $\Rightarrow$ eine virtuelle Abstraktion des Betriebssystem; stellt eine geschützte Umgebung bereit, in der Anwenungen ausgeführt werden
- *Container* $\Rightarrow$ eine Zone, die auch die Resource Managment Facility verwendet

$\rightarrow$ im Folgenden der Einfacheherheit Synonym verwendet.

Was geht mit Zones ...

- ein oder mehrere Netzwerk-Interfaces
  - werden als Aliase auf Interfaces der global Zone gelegt
- direkter Zugriff auf Devices durch non-global Zones möglich
  - z. B. Festplatten, Partitionen, optische Laufwerke, Audio-Geräte, Framebuffer, ...
- global und non-global Zones können sich "package directories" teilen (*sparse zones*)
- Patches werden auch automatisch in Zonen installiert
- Directories als Filesysteme via `lofs(7FS)` von global in non-global Zones mounten (ro/rw)
- Verwaltung und Konfiguration mit `zonecfg(1M)`, `zoneadm(1M)`
- `zlogin(1)` für den Zugang zu einer Zone
- viele Programme "zone-aware" durch -z und/oder -Z Option/en (z. B. `ps(1)`, `prstat(1)`, ...)

... und was sind die Limitierungen[1]

- theoretisch maximal 8192 Zones (1 global, 8191 non-global)
- keine hierarchischen Zonen
  - non-global Zones können keine weiteren Zonen "enthalten"
- kein shared memory zwischen den Zonen
  - Kommunikation nur über IP
  - IPFilter für Packet Filtering zwischen non-global Zones und anderen Computern
  - IPFilter kann nicht den Traffic zwischen Zonen filtern
    - `-reject` oder `-blackhole` Routen verwenden
- keine unterschiedlichen Zeiten ("clock sources")
  - aber unterschiedliche Zeitzonen (`TZ`) möglich
- Dienste, die nicht von Zonen angeboten werden können:
  - NFS Server, DHCP, NTP
- Einschränkungen bei X-Windows
  - kein X-Server auf der Konsole

[1] Stand August 2005

Arbeiten mit Zones

Typische Arbeitschritte

- Zone konfigurieren
- Anpassen der Konfiguration
- Zone installieren (und irgendwann auch deinstallieren)
- Zone starten und stoppen
- Einloggen in eine Zone

Zone konfigurieren (1)

Kommado: `zonecfg -z <zonename> <subcommand>`

Grundüberlegungen:

- IP-Adresse oder nicht? Welche?
- *sparse* oder *non-sparse* Zone?
  (d. h. "package directories" mit global Zone geteilt oder nicht)
- Wo soll die Zone im Dateissystem der global Zone liegen?
- Sollen andere Dateisysteme (via lofs) gemountet werden?
- Soll Zugriff auf andere Devices ermöglicht werden?
- Soll ein Zone Template verwendet werden oder eine "blanke" Konfiguration verwendet werden?

Zone konfigurieren (2)

```
# zonecfg -z test
zonecfg:test> create
zonecfg:test> export
create -b
set autoboot=false
add inherit-pkg-dir
set dir=/lib
end
add inherit-pkg-dir
set dir=/platform
end
add inherit-pkg-dir
set dir=/sbin
end
add inherit-pkg-dir
set dir=/usr
end
```

Zone konfigurieren (3)

```
zonecfg:test> set zonepath=/zones/test
zonecfg:test> set autoboot=true
zonecfg:test> add net
zonecfg:test:net> set address=192.168.1.7/24
zonecfg:test:net> set physical=bge0
zonecfg:test:net> end
zonecfg:test> add fs
zonecfg:test:fs> set dir=/zones/test/homes
zonecfg:test:fs> set special=/export/homes
zonecfg:test:fs> set type=lofs
zonecfg:test:fs> add options [nodevices]
zonecfg:test:fs> end
zonecfg:test> verify
zonecfg:test> commit
zonecfg:test> quit
#
```

Zone installieren (1)

Kommando: `zoneadm -z <zonename> <subcommand>`

Arbeitsschritte:

- Erstellen des Zonepath Verzeichnisses
- Ggf. andere Verzeichnisse erstellen (für `lofs`-Mounts)
- Zone installieren
- Zone booten und System initial konfigurieren (`sysidtool(1M)`)
  - kann durch Erstellen einer `sysidcfg(4)` Datei automatisiert werden
- Ggf. weitere Konfiguration des Systems

Zone installieren (2)

```
# zoneadm -z test verify
# zoneadm -z test install
Preparing to install zone <test>.
Creating list of files to copy from the global zone.
Initializing zone product registry.
Determining zone package initialization order.
Preparing to initialize <779> packages on the zone.
Initialized <779> packages on zone.
Successfully initialized zone <test>.
# zoneadm list -vci
ID NAME             STATUS         PATH
 0 global           running        /
 2 test             installed      /zones/test
```

Zone booten und mit der Konsole verbinden

Kommando: `zlogin [-C] [-l <username>] <zonename>`

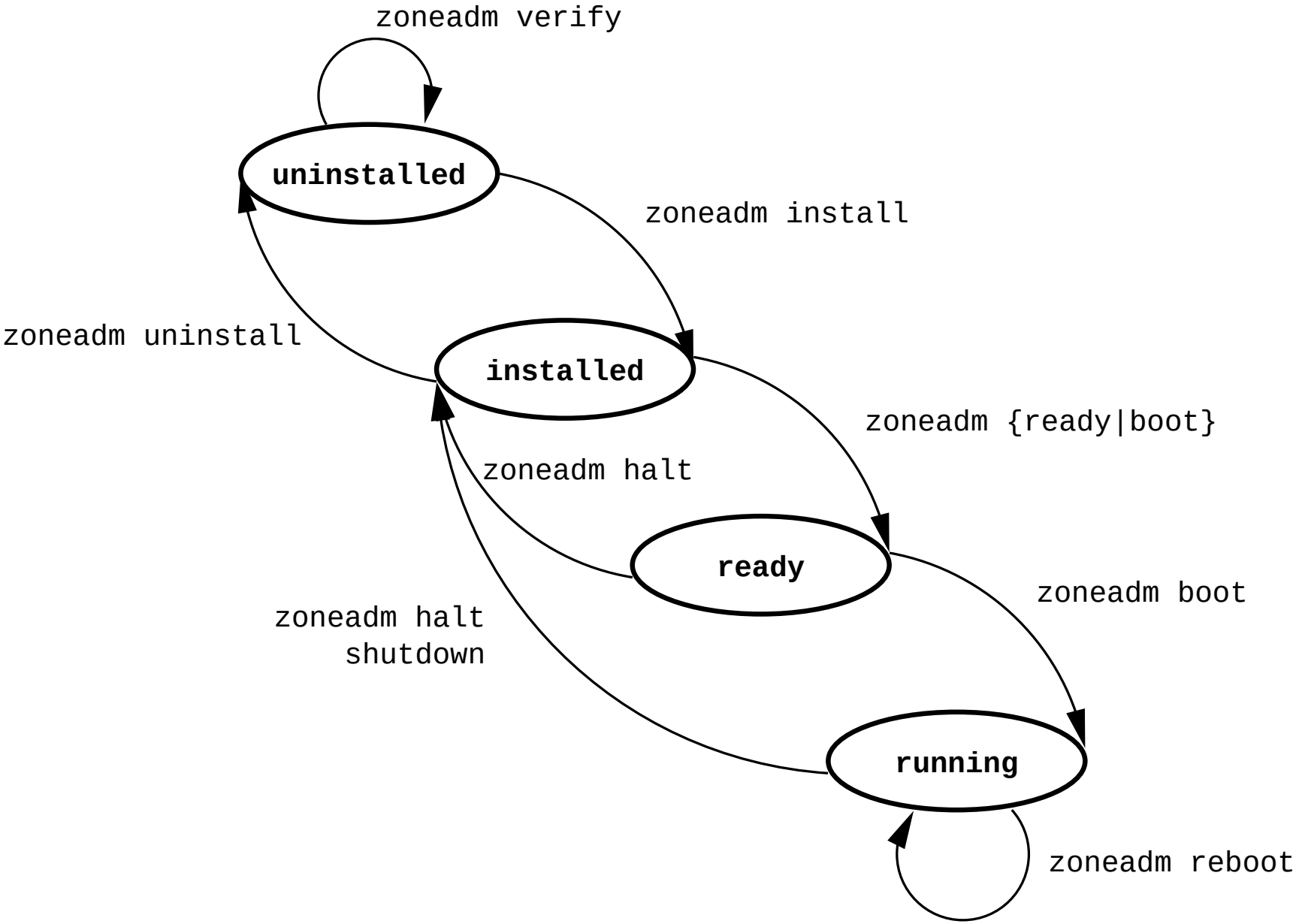
- ähnlich wie `rsh(1)`
- `zlogin -C` attached an die Konsole einer (auch nicht bebooteten) Zone
  - Escape Sequenz ist standardmäßig `~.`

```
# zoneadm -z test boot
# zlogin -C test
[ sysidtool ]
~.
# zoneadm list -vci
ID NAME             STATUS         PATH
 0 global           running        /
 2 test             running        /zones/test
```

Zone starten und stoppen

- Zonen mit `autoboot=true` werden automatisch mit dem System gestartet
- `zoneadm -z <zonename> boot` startet eine Zone manuell
- Zone stoppen und neustarten
  - "durch gutes Zureden"
    `zlogin <zonename> shutdown -i<state>`
  - "mit Gewalt"
    `zoneadm -z <zonename> halt` und `zoneadm -z <zonename> reboot`
    Shutdown Scripte werden nicht ausgeführt!

Verschiedene Zustände einer Zone



Resource Managment Facility

- Resource Pools
- bislang nur CPU als Resource
  - eine oder mehrere CPU(s) kann/können exklusiv einer Zone zugewiesen werden
  - eine oder mehrere CPU(s) kann/können mehreren Zonen zugewiesen werden
- Fair Share Scheduling
  - garantiert faire CPU-Zeit-Verteilung innerhalb eines Pools
- `pooladm(1M)` und `poolcfg(1M)`
- Zone einem Pool zuweisen
  - `set pool=<poolname>` in der Zonen-Konfigurtion

Ausblick

- BrandZ ("branded Zone")
  - Zonen mit "non-native operating enviorments"
  - "lx brand" $\Rightarrow$ Linux; distributionsunabhänige Bereitstellung des Kernel Syscall-Interfaces
- Zone Migration
  - *halted* Zone kann auf ein anderes System migriert werden
  - integriert in Solaris Express 04/06 (Nevada Build 36)
- Resource Managment
  - RAM und Swap
  - Anzahl der Prozesse
  - Anteil an CPU Zeit auf einer CPU (z. B. 50% einer CPU)
  - System V IPC (shared memory, semaphores, message queues)

Links

- http://www.opensolaris.org
- http://www.opensolaris.org/os/community/zones/
- http://www.opensolaris.org/os/community/zones/faq/
- http://www.sun.com/bigadmin/features/articles/solaris_zones.html
- http://learningsolaris.com/archives/2006/01/16/zones_howto/
- http://www.opensolaris.org/os/community/brandz/