



OpenSolaris

Erik Fischer

Principal Engineer

erik.fischer@sun.com

OpenSolaris (1)

- A Solaris operációs rendszer forráskódja
- Fázisos release
- CDDL licenc – a Mozilla Public License módosítása
- www.opensolaris.org
- 2005. június óta létezik
- 40-47 community projekt
 - > Pl. Solaris PowerPC port

Solaris

- Sun Microsystems fejlesztése, mintegy 24 éve
- Valamikor BSD alapú volt
- Ma leginkább a System V-hoz hasonlít
- Kellemes egyvelege a világ UNIX-ainak...
- Moduláris kernel

OpenSolaris (2)

- Jelenleg a Solaris 11 (Nevada) forráskódja
- Ami teljes
 - > Kernel
 - > Unix parancsok, daemon-ok
 - > lib-ek
 - > FC stack
 - > SFW – Solaris FreeWare
- Ami nem
 - > XSun
 - > JDS
 - > Lokalizáció
 - > man

OpenSolaris (3)

- ~60MB tar.bz2
- ~100MB bináris archive és extra tool per platform
- ~600MB compiler per platform
 - > Sun Studio 11
 - > Ingyenes (!)
- Platformok
 - > SPARC – 64-bit
 - > X86 – 32- és 64-bit

Disztribúciók

- www.genunix.org
- SPARC és x86
 - > Solaris 10 – kis csalással
 - > Solaris Express
- x86
 - > SchilliX
 - ISO és forrás
 - > BeleniX
 - Sun India indította
 - LiveCD, install és USB bootable image
 - > Nexenta
 - LiveCD, install és VMWare image

Alapvető ismeretek a Solaris-ról (1)

- Alapvetően szerver operációs rendszer
- Nagyságrendileg 600-700 projekt integrálódik két évente
- Tulajdonképpen a Solaris 9 óta ingyenes
- A Solaris 9 óta sokat javult az interaktív felület
 - > JDS – GNOME alapú GUI
- Az x86-os device támogatás határozottan és szinte napról-napra javul
- Minimális hardver
 - > x86: P3, 256MB memória, 4GB HDD, lehetőleg ne a legújabb SLI vagy CrossFire grafikus kártyával...
 - > SPARC: US-II, III, IIIi, IV, IV+, T1, T2, 512MB memória, 4GB HDD

Alapvető ismeretek a Solaris-ról (2)

- Installáció
 - > SPARC – szinte triviális
 - > x86 – ha szerencsénk van, akkor szintén szinte triviális
 - > Szükséges információk:
 - Gépnév
 - IP cím, netmask, default router
 - Diszk layout
 - Címtár szolgáltatás függő adatok
 - > Többé-kevésbé képes meglévő OS verziót is upgradeelni

Alapvető ismeretek a Solaris-ról (3)

- Lokalizáció
 - > Az OS természetesen HU locale ismerő
 - > Ennek ellenére nem beszél magyarul
 - > A JDS-nek van magyar lokalizációja
- Kezelői felület
 - > A hard core hackereknek command line
 - > A megrögzött öregfiúknak CDE
 - > A kényelmet szeretőknak JDS (GNOME)

Alapvető ismeretek a Solaris-ról (4)

- Támogatás
 - > Solaris
 - Ingyen csak biztonsági hibajavítások
 - A támogatás alapvetően kedvező
 - > Solaris Express
 - Nincs, mert ez technology preview
 - Újabb build letöltése
 - > Egyéb OpenSolaris disztribúciók
 - Alapvetően nincs, hiszen még nem végleges az operációs rendszer verzió
 - Újabb build letöltése

Alapvető ismeretek a Solaris-ról (5)

- Talán az egyik legbiztonságosabb Unix
- Talán az egyik legmodernebb is
- Biztosan a legjobban skálázódó (144 CPU, lineáris skálázódással)
- Biztosan a legdinamikusabb rendszer
 - > Ha a hardver támogatja, akkor menet közben MINDEN komponenst ki tudunk venni a rendszerből
- Talán az egyik leginkább NUMA-aware
- Mintegy 15000 szoftver (nem Open Source)

Alapvető ismeretek a Solaris-ról (6)

- Dokumentáció
 - > man
 - > docs.sun.com
 - html és pdf formátumban
 - Több 10,000 oldal
 - Solaris 2.4-ig visszamenőleg
 - > opensolaris.org
 - > Google

Adminisztráció

- Szoftver telepítés
 - > Package-ek, patch-ek
 - > Command line
- Rendszer konfiguráció
 - > Klasszikus UNIX filozófia, apró csavarral
 - > /etc könyvtár
 - > vi
- Gyengébbek kedvéért webes GUI

OpenSolaris különlegességek

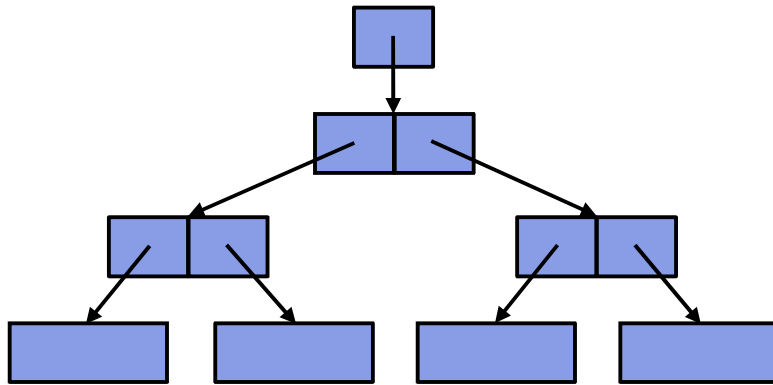
- ZFS fájlrendszer
- DTrace
- Zónák
- Privilege rendszer
- Az SMF és a 10 contract alrendszer
- Branded zónák (az év vége, jövő év eleje)
 - > RH és derivatív Linux-ok

ZFS

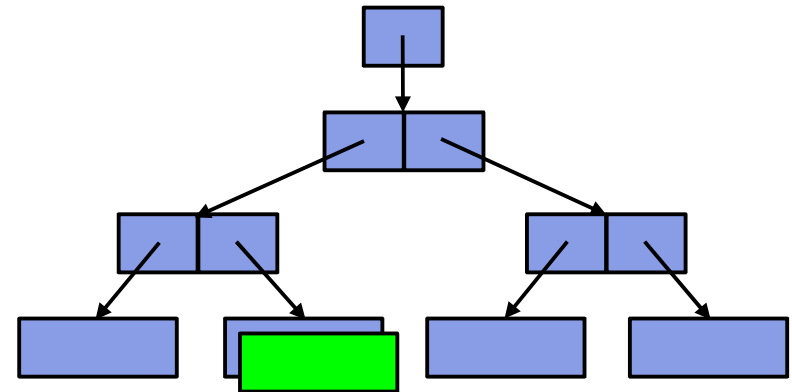
- A diszk VM-je, nincs szükség kötet kezelőre
- Tranzakcionális
- Teljesen integritás védett
- 128-bites (256 quadrillió zettaB)
- Dinamikus metaadatok
- Beépített tömörítés, titkosítás
- Diszk scrubbing
- Replikáció
- Minden művelet copy-on-write
- Gyors snapshot, verziózás
- Big-endian/little-endian konverzió

COW

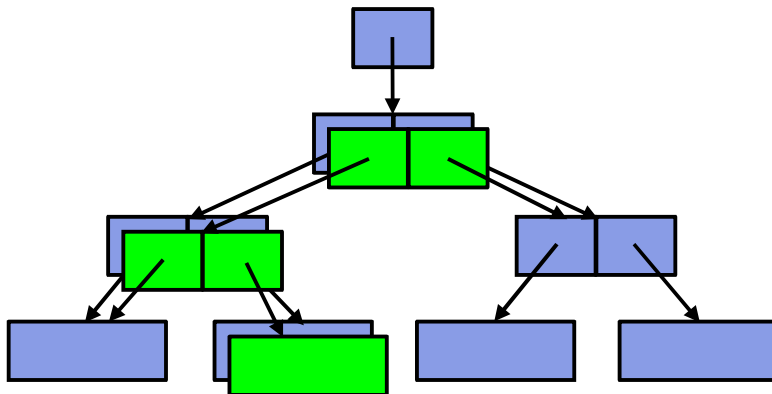
1. Az eredeti



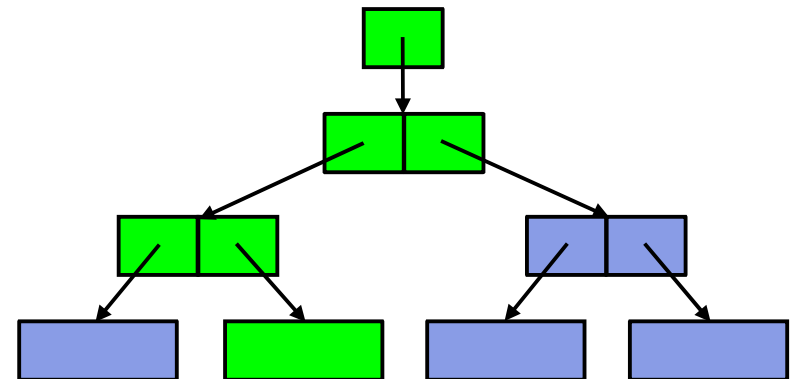
2. Adat blokk COW



3. Indirekt blokk COW

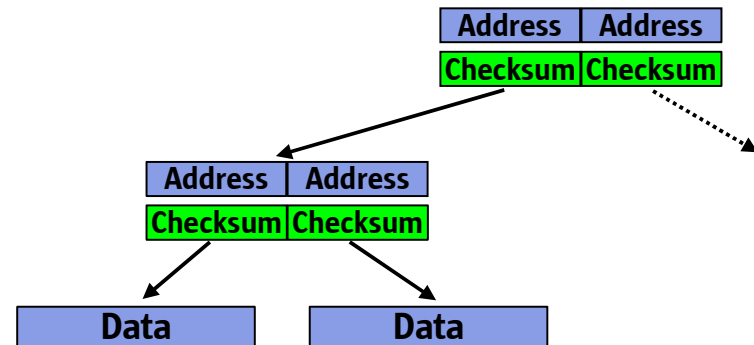


4. Überblock módosítás (atomi)

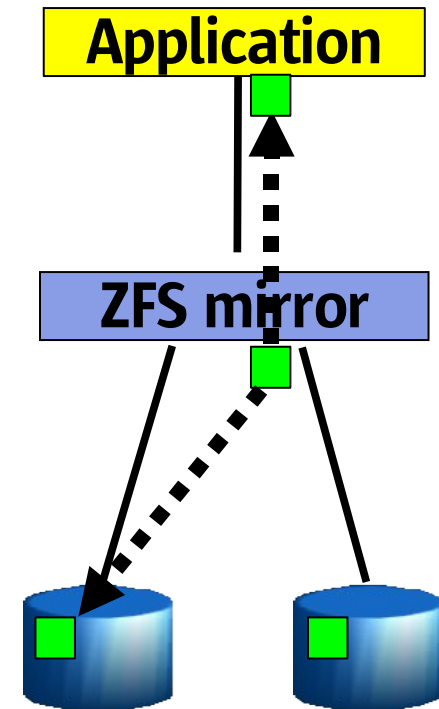
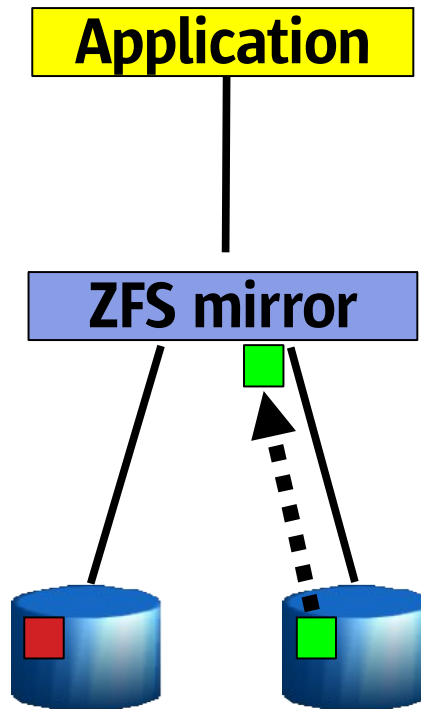
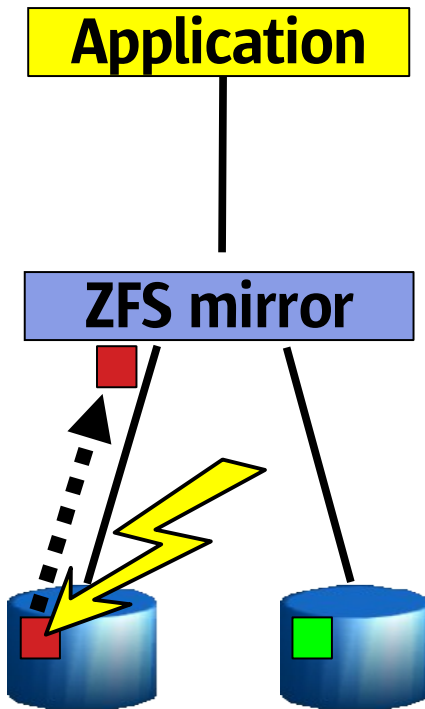


Checksum

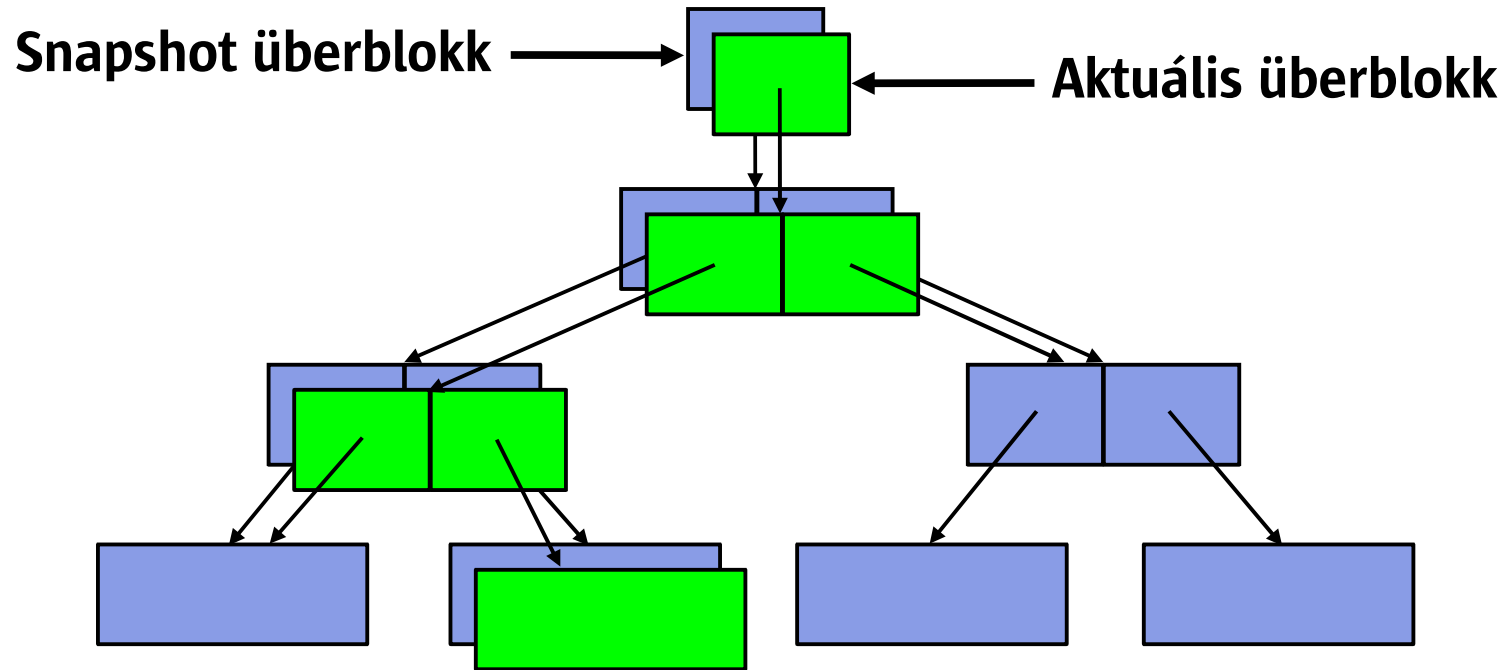
- Bit rot
- Phantom writes
- Misdirected reads and writes
- DMA parity errors
- Driver bugs
- Accidental overwrite



Önjavítás



Snapshot



Hibakeresés – in vitro

- Végzetes, nem reprodukálható hibák
 - > Core dump
 - > Postmortem mdb(1), dbx(1)
- Tranziens hibák
 - > Programozási hibák vagy percepcionális teljesítmény problémák
 - Ad hoc technikák vagy szinte semmi
 - truss(1), mdb(1), *stat(1M), sar(1), prstat(1)
 - Sun Studio Performance Analyzer

Hibakeresés – in vivo

- Invazív technikák
 - > Bináris instrumentálás
 - > Forrás szintű instrumentálás
 - > Interposer library-k
 - > Debug library-k
 - > Debug kernel
 - > Általában durva beavatkozások
 - > Általában lassú
 - > Általában nagy az additív hibák injekciójának esélye

Elvárások

- Az ideális dinamikus hibakereső rendszer
 - > Képes információ gyűjtésére
 - > Az adatszerkezetek módosítása
 - > Mindezt éles rendszereken is
 - > Teljesen biztonságosan
 - > Teljesítmény veszteség nélkül

DTrace

- Interpretált probe alapú nyelv, prédikátumokkal és akciókkal
- Dinamikus függvény be- és kilépési pont instrumentációs rendszer
- Jellemzők:
 - > User és Kernel rétegben is működik
 - > 40000+ probe egy átlagos Solaris 10 rendszeren
 - > Alapállapotban csak root-ként működik
 - > 3 privilégiumot igényel: `dtrace_kernel`, `dtrace_proc` és `dtrace_user`
 - > 410 oldalas dokumentáció

Probe

- Az instrumentáció pontos helye egy hierarchiában (leírója egy 'n'-es)
- Egy provider bocsájtja a rendelkezésünkre
- Minden provider modulokra és egy függvényekre tagolódik
- A probe-nak van neve (általában entry és return)
- `dtrace -l`

`<provider, module, function, probe_name>`

Provider-ek

- Szép számban akadnak
 - > fbt – szintem minden entry és return a kernelben (~39000 db)
 - > syscall – syscall tábla (~450 db)
 - > profile
 - > lockstat
 - > proc
 - > sysinfo
 - > vminfo
 - > sched
 - > io
 - > fpuinfo
 - > sdt (~190 db)
 - > mib – TCP/IP-hez kapcsolódó függvények (~430 db)

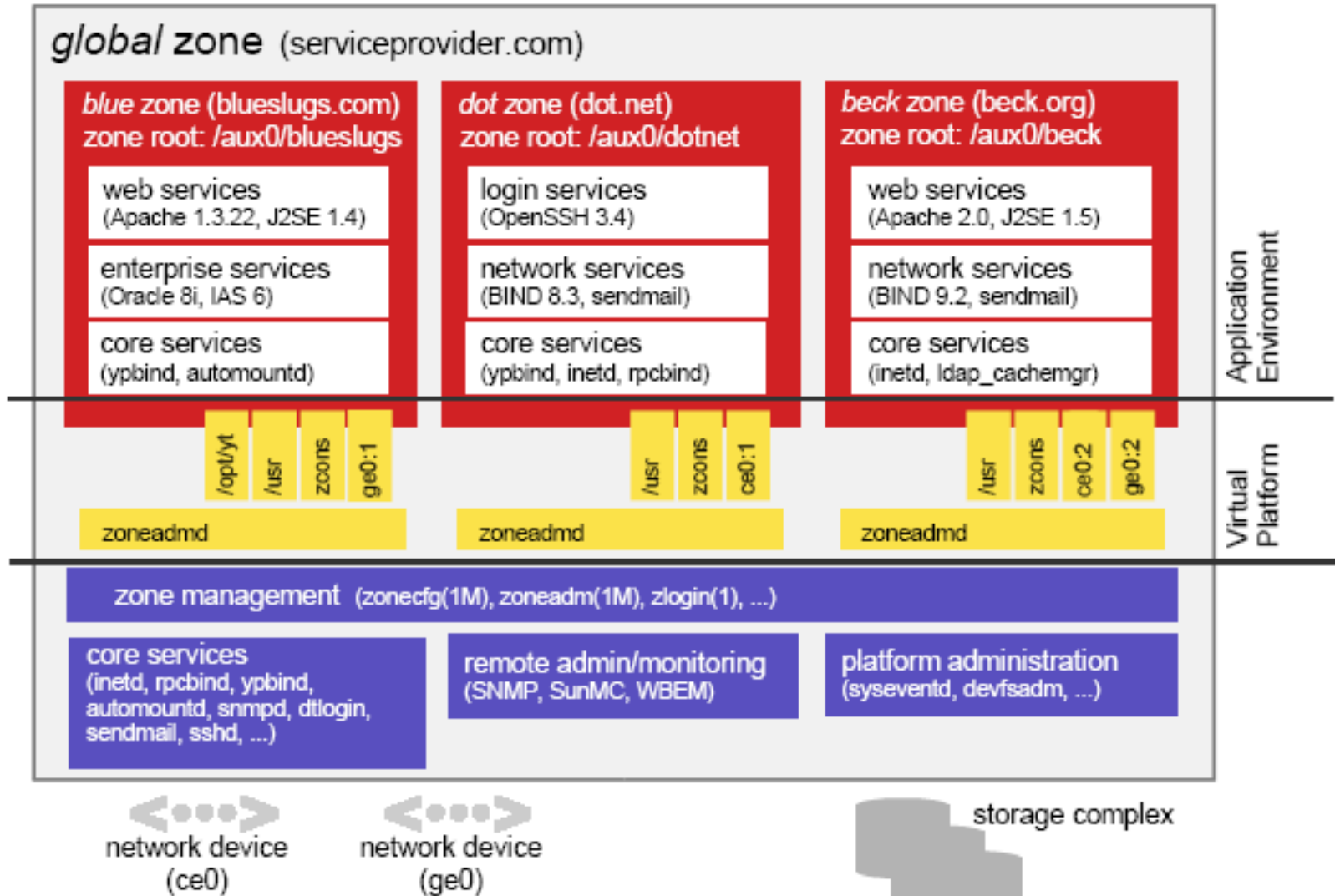
A D nyelv

- C és awk keveréke
- Teljes hozzáférés a kernel C típusaihoz
- Teljes hozzáférés a statisztikákhoz és globális információkhoz
- String támogatás

Zónák

- Egyetlen kernel megosztása több virtualizált applikációs konténer között
- Processz “bedobozolás”
 - > Erőforrás és biztonsági izoláció
- Virtualizált hardver, de nem virtuális gép
- Kintről önálló operációs rendszernek tűnik

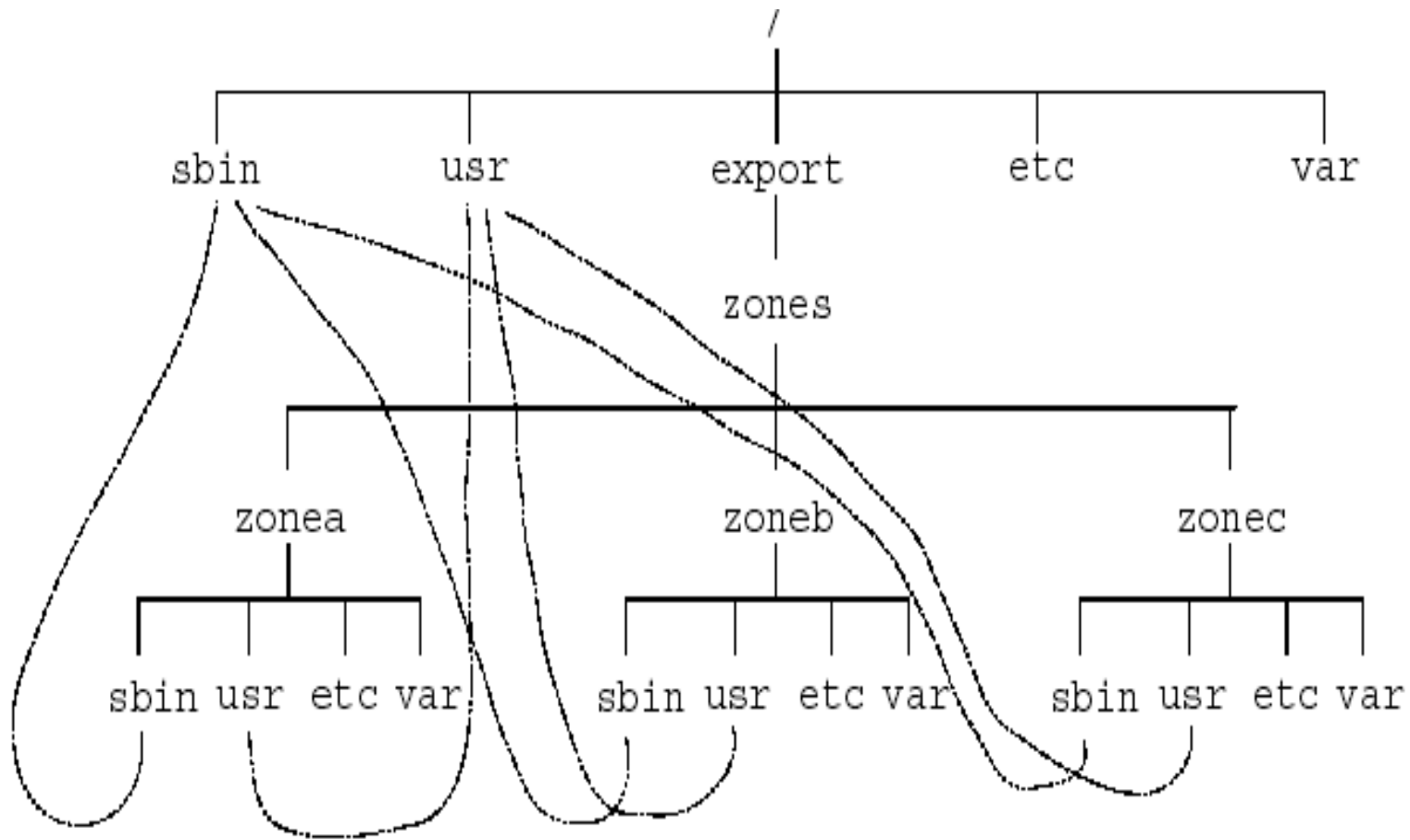
Zónák – blokk diagram



Zónák – virtualizáció

- Önálló IP cím(ek), multiplexelt kommunikáció
- Önálló címtér és címtár
- Önálló IPC tér
- Többszörös syslog
- Többszörös pkg, patch adminisztráció
- Virtualizált /proc, /etc/mnttab
- Önálló fájlrendszer
- Privát fájlrendszer pontok (/ , /var, /etc)
- Bizonyos fájlrendszerek read-only módon beszármazhatnak egy zónába (pl. /usr)
- Korlátozott mount/umount lehetőségek
- Korlátozott privilégiumok (globális, nem globális)

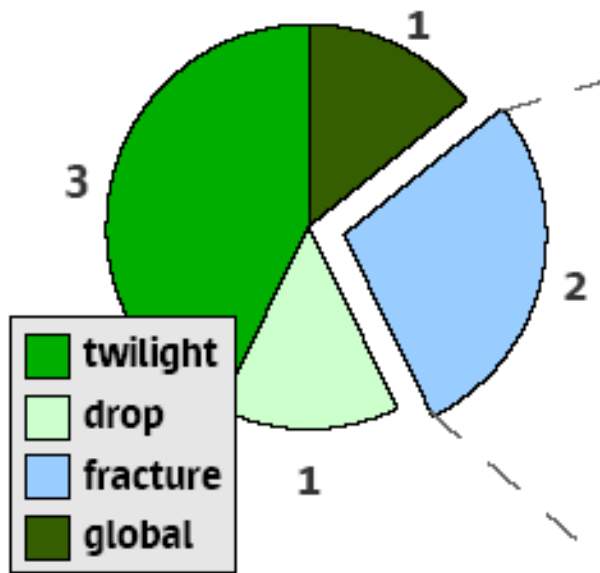
Zónák – fájlrendszer virtualizáció



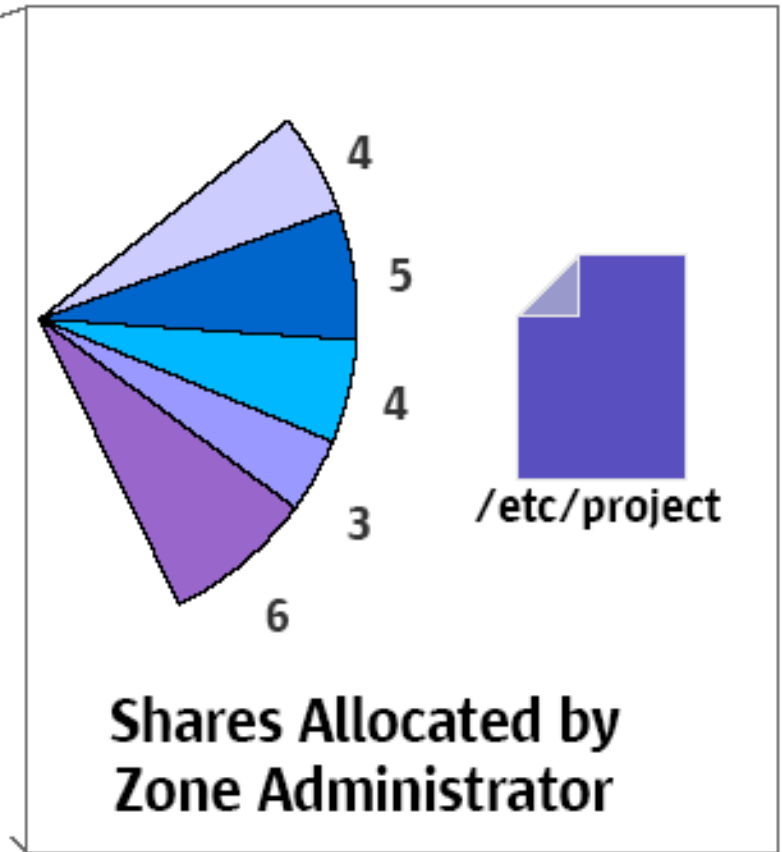
Zónák – erőforrás menedzsment

- A Solaris 9 továbbfejlesztett erőforrás menedzsmentje
- Projekt alapon működik
- Dinamikus erőforrás limitek definiálhatók a zónákhoz

Zónák – erőforrás menedzsment



Shares Allocated to Zones



Shares Allocated by Zone Administrator

RBAC

- Solaris 8 óta az operációs rendszer része
- Átlagos felhasználó, de csak su-val közelíthető meg
- A szerepek 4 fájlban definiálódnak és parancs szinten konfigurálhatóak
- A Solaris 10-től privilégium kiegészítést is kaptak

Privileges - alapgondolat

- Least privilege
- A mindenható root jogainak szétszedése apró darabokra
 - > Egyenként elvenni vagy odaadni
- Sok minden fut root-ként, aminek nem kellene
 - > Pl. alacsony portokra bind-oló web szerver

Megvalósítás

- A kernel nem $UID==0$ -ra vizsgál, hanem az éppen szükséges privilege-re
- 47 egyedileg ki-bekapcsolható privilege
- Bővíthető
- Alapvetően privilege halmazokkal operál
- Kompatibilitás: root minden privilege-vel rendelkezik
- Privilege-aware programozás

Kriptográfiai keretrendszer

- User és kernel egyaránt
- digest, mac, encrypt, decrypt funkciók
- Algoritmusok
 - > User: DES, 3DES, AES, RC4, RSA, DSA, D-H, SHA-1, MD5
 - > Kernel: DES, 3DES, AES, Blowfish, SHA-1, MD5
- Signózott elf (!)
- Hardver RNG támogatás
- Hardver accelerator támogatás

Egyéb biztonsági funkciók

- Kerberos v5
- PAM
- Integrált packet filter
- IPsec
- Integrált baselining
- C2-es audit
- Trusted Extension (B2) az év végétől, jövő év elejétől

SMF

- A teljes OS infrastruktúra egységesítése
- /etc/rc*.d, /etc/inetd.conf és millió más szolgáltatás indító/konfiguráló hely halála
- Automatikus, függőségeket is figyelembe vevő szolgáltatás indítás
- Új szolgáltatás azonosítás

FMRI: Fault Management Resource Identifier

> svc://gép/alrendszer/szolgáltatás:

SMF

- XML repository
- Egyszerű és könnyen bővíthető
- Szolgáltatás ki és bekapcsolás
 - > `svcadm disable system/cron:default`
 - > `svcadm enable network/ssh:default`
- Profilok

SMF – diagnosztika

```
# svcs -x
```

```
svc:/network/ntp:default (Network Time Protocol  
(NTP).)
```

```
State: maintenance since Mon Oct 18 13:58:42 2004
```

```
Reason: Start method exited with  
$SMF_EXIT_ERR_CONFIG.
```

```
See: http://sun.com/msg/SMF-8000-KS
```

```
See: ntpq(1M)
```

```
See: ntpdate(1M)
```

```
See: xntpd(1M)
```

```
Impact: 0 services are not running.
```


SMF – diagnosztika

```
# svcs -x -v
```

```
svc:/application/print/server:default (LP Print Service)
```

```
State: disabled since Mon Oct 18 16:17:27 2004
```

```
Reason: Disabled by an administrator.
```

```
See: http://sun.com/msg/SMF-8000-05
```

```
See: man -M /usr/share/man -s 1M lpsched
```

```
Impact: 1 service is not running:
```

```
    svc:/application/print/rfc1179:default
```

SMF – függőségek

- Függőségi információk

```
% svcs -d network/smtp:sendmail
```

STATE	STIME	FMRI
online	18:20:14	svc:/system/identity:domain
online	18:20:26	svc:/network/service:default
online	18:20:27	svc:/system/filesystem/local:default
online	18:20:27	svc:/milestone/name-services:default
online	18:20:27	svc:/system/system-log:default
online	18:20:30	svc:/system/filesystem/autofs:default

```
% svcs -D network/smtp:sendmail
```

STATE	STIME	FMRI
online	18:20:32	svc:/milestone/multi-user:default

Contract alrendszer

- A SMF egyik háttere
- Garantálja a “szerződött” alkalmazások “szerződésének” betartását
- A szerződés az elérhetőségre és a működés alapvető paramétereire vonatkozik

Contract alrendszer

- Tetszőleges felhasználói program szerződhető
 - > `ctrun(1)`
 - > Milyen eseményekre kell újraindítás: `core`, `exit`, `hwerr`
 - > Mi legyen a gyermek processzekkel
 - > Hány újraindítási próbálkozás történjen
 - > Pl.

```
ctrun -r 0 -t -f core,exit,hwerr httpd
```



Erik Fischer
Principal Engineer
erik.fischer@sun.com