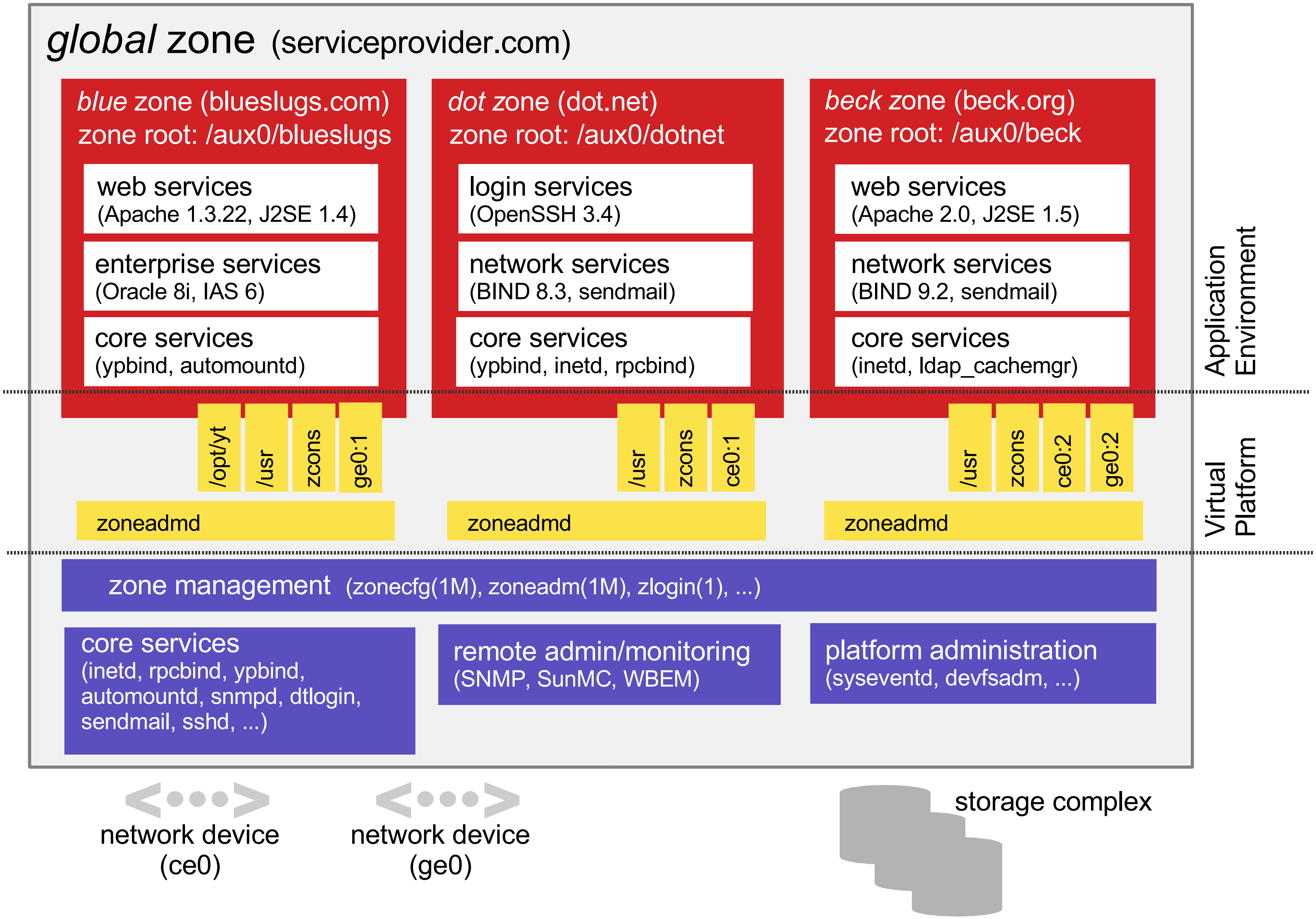
Solaris Zones

- Provides virtualized OS environments, each looking like a Solaris instance
  - Implemented via a lightweight layer in the OS
  - Processes in one zone can only see or affect processes in the same zone
  - Processes can be run while requiring global resources like IP port space or identity
  - Details of physical resources are hidden
  - Each zone can be administered independently
  - No porting as the ABI/API is the same

Zones Security Overview

- Each zone has a security boundary around it
- Runs with subset of `privileges(5)`
- A compromised zone is unable to escalate its privileges
- Important name spaces are isolated
- Processes running in a zone are unable to affect activity in other zones

Zones Block Diagram



global zone (serviceprovider.com)

blue zone (blueslugs.com)
zone root: /aux0/blueslugs

web services
(Apache 1.3.22, J2SE 1.4)

enterprise services
(Oracle 8i, IAS 6)

core services
(ypbind, automountd)

/opt/yt
/usr
zcons
ge0:1

zoneadmd

dot zone (dot.net)
zone root: /aux0/dotnet

login services
(OpenSSH 3.4)

network services
(BIND 8.3, sendmail)

core services
(ypbind, inetd, rpcbind)

/usr
zcons
ce0:1

zoneadmd

beck zone (beck.org)
zone root: /aux0/beck

web services
(Apache 2.0, J2SE 1.5)

network services
(BIND 9.2, sendmail)

core services
(inetd, ldap_cachemgr)

/usr
zcons
ce0:2
ge0:2

zoneadmd

Application Environment

Virtual Platform

zone management (zonecfg(1M), zoneadm(1M), zlogin(1), ...)

core services
(inetd, rpcbind, ypbind, automountd, snmpd, dtlogin, sendmail, sshd, ...)

remote admin/monitoring
(SNMP, SunMC, WBEM)

platform administration
(syseventd, devfsadm, ...)

network device
(ce0)

network device
(ge0)

storage complex

Processes

- Certain system calls are not permitted or have restricted scope inside a zone
- From the global zone, all processes can be seen but control is privileged
- From within a zone, only processes in the same zone can be seen or affected
- `proc(4)` has been virtualized to only show processes in the same zone

File Systems

- Each zone is allocated its own root file system and cannot see that of others
- Processes cannot escape out of a zone, unlike `chroot(2)`
- File systems like `/usr` can be inherited in a read-only manner
- File systems such as `autofs(4)` have been virtualized per zone

Networking

- Single TCP/IP stack for the system as a whole so zones are shielded from most configuration like routing and devices
- Each zone has its own IP port space and is assigned IPv4/IPv6 addresses
- Applications can bind to INADDR_ANY and will only get traffic for that zone
- Zones cannot see the traffic of others

Identity

- Each zone controls its node name, RPC domain name, time zone and locale
- Each zone can use a different naming service such as DNS, LDAP and NIS
- Separate `/etc/passwd` files means that root can be delegated to the zone
- User ids may map to different names when domains differ (as with NFS now)

IPC

- Expected IPC mechanisms such as System V IPC, STREAMS, sockets, `libdoor(3LIB)` and loopback transports are available inside a zone
- Key name spaces virtualized per zone
- Inter-zone communication is available using the network (software loopback)
- Global zone can setup rendezvous too

Devices

- Zones see an subset of “safe” pseudo devices in their `/dev` directory
- Devices like `/dev/random` are safe but others like `/dev/kmem` are not
- Zones can modify the permissions of their devices but cannot `mknod(2)`
- Physical device files like those for raw disks can be put in a zone with caution

N1 Grid Containers: Zones with Resource Management

- Zones do not require dedicated hardware resources
- Resources like CPUs can be partitioned with a arbitrary granularity
- Multiple zones can be multiplexed over a resource pool or a zone can be bound to a pool for service guarantees
- Resource limits can be set on a zone

Conclusion

- Additional feature work is ongoing
- Zones released in Software Express for Solaris 2/04
- BigAdmin site launched simultaneously
- Documentation is also available now from `http://docs.sun.com/`