Containers

LWN articles

The failure of operating systems

- OSes manage physical memory!
  - But LRU means one process' use causes another to slow down
- OSes manage the CPU
  - But applications have many processes
  - More processes == more CPU
- OSes provide performance isolation
  - Global denial of service attacks are easy
    - Directory bomb, fork bomb
- Difficult to get accurate application statistics
- Difficult to have to applications use the same port

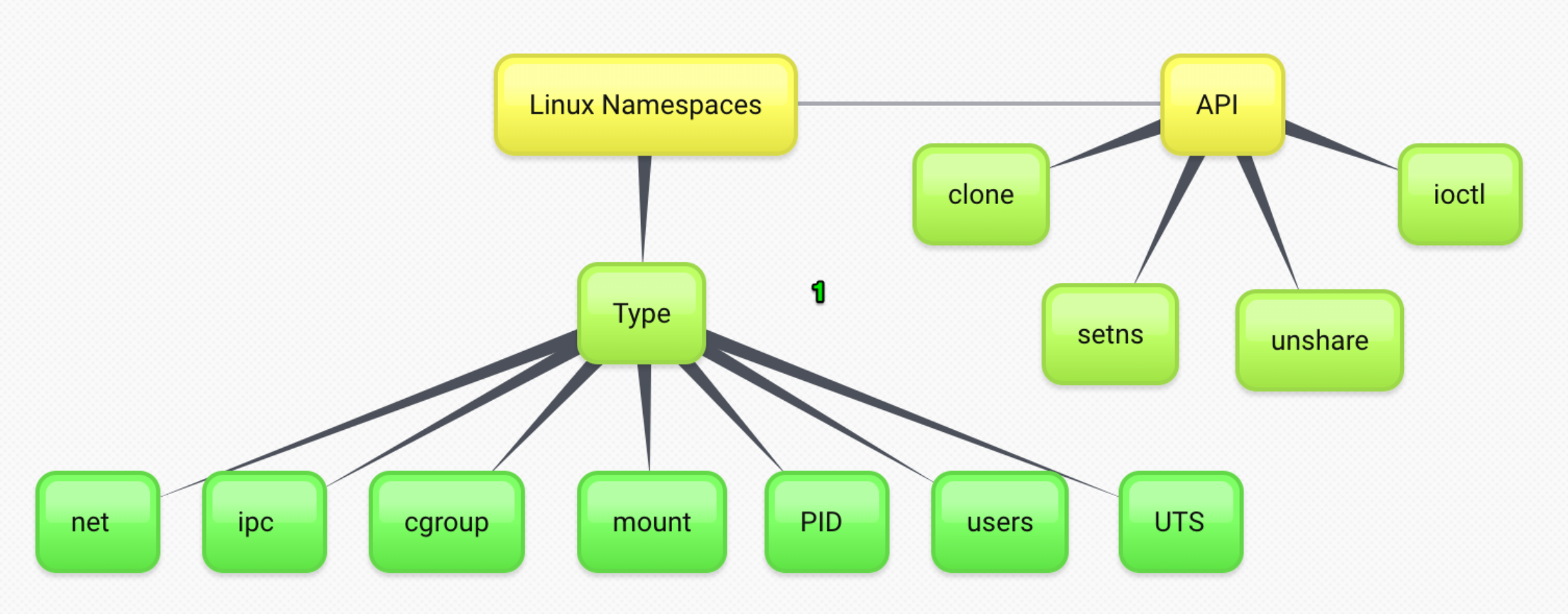
Containers to the rescue

- cgroups – control group
  - Hard limit on CPU
    - Schedule hierarchically – cgroups first, then subgroups, then processes
  - Hard limit on physical memory
- Namespaces allow security & multiplexing
  - If you can’t name a resource, you can’t control it
  - Network namespace
    - Cgroups have a network device, independent port numbers
  - mount namespace
    - Different containers see different file system namespace
  - User IDs, group ID namespace
    - Includes init process

Namespaces

- The purpose of each namespace is to wrap a particular global system resource in an abstraction that makes it appear to the processes within the namespace that they have their own isolated instance of the global resource
- UTS – nodename and domainname
- IPC – semaphores, pipes, POSIX message queues
- Network – IP address, routing

Namespace figure



- Why is container start time so much faster than VM boot?