Are containers that we have now secure enough?

Pavel Emelyanov, Vladimir Davydov
Parallels, Inc.

kmem accounting

- What is kmem?
  - kernel stack pages, page tables, dcache/icache, other slab allocations
- Why limit it?
  - fork bomb, sparse address space, dcache pressure can slow down or even kill the system
- How to do that?
  - kernel memory accounting & limit – **already done** as part of the memory cgroup – see memory.kmem.limit_in_bytes
  - slab shrinker support (dcache/icache) – **in progress**

veth

- What is veth?
  - network device pair, works as a pipe
  - added to a bridge to create network of containers
- Problems with that?
  - works at the link layer, so a container can do stuff it may not be supposed to do (traffic stiffing, IP spoofing)
  - to avoid this, use ebtables, but they come at extra costs
- What about vnet?
  - assign multiple IPs to the same physical device
  - pass through an **IP** to a net ns
  - route packets on the **network** layer

Thank you

Questions?

Pavel Emelyanov, xemul@parallels.com
Vladimir Davydov, vdavydov@parallels.com