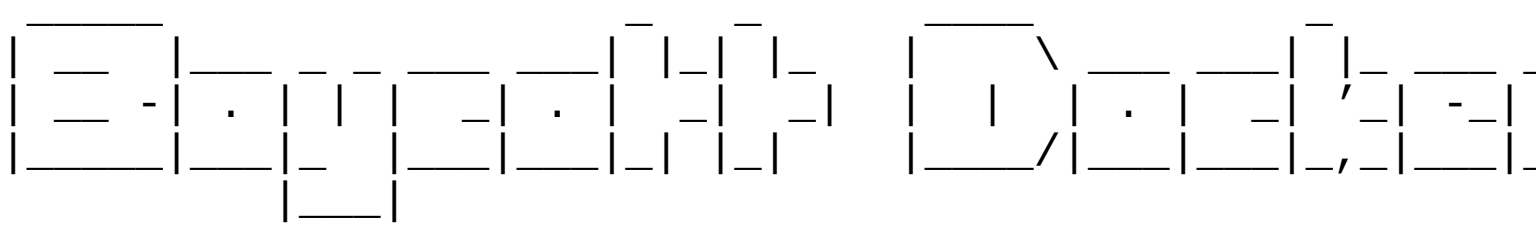

Eivind Uggedal

Hype

"Best features of Docker:
- Easy to get a project to the top of HackerNews
- New Github stars for old ideas
- Has lots of Twitter followers"

twitter.com/hipsterhacker/status/577587054642536449

Containere

- Namespace i en kernel for at ulike samlinger av prosesser kan kjøre og kontrolleres uavhengig av hverandre:
  - prosess (PID)
  - nettverk
  - mount
  - brukere (UID/GID)
  - ...

Tidslinje

- 1967: VM - IBM CP-40
- 1979: chroot(2) - Version 7 AT&T UNIX
- 2000: jail(2) - FreeBSD 4.0
- 2001: VServer - Linux patchset
- 2005: zones - Solaris 10
- 2005: OpenVZ - Linux patchset
- 2007: WPARs - IBM AIX 6.1
- 2007: SRP - HP-UX 11i
- 2008: LXC - Linux
- 2009: Heroku - PaaS
- 2010: dotCloud - PaaS
- 2013: Docker - Linux

Linux

- Utvidelser i clone(2) flag
  - 2002: CLONE_NEWNS - mount
  - 2006: CLONE_NEWUTS - hostname
  - 2006: CLONE_NEWIPC - SysV IPC / POSIX MQ
  - 2008: CLONE_NEWPID - prosess
  - 2009: CLONE_NEWNET - nettverk
  - 2012: CLONE_NEWUSER - bruker
- Control Groups (cgroups)
  - 2008: cpuset - pinning av CPU/minne-enheter
  - 2008: memory - begrense minne
  - 2008: devices - hviteliste enheter
  - 2010: blkio - disk I/O begrensning
  - 2012: net_prio - prioritering av nettverstrafikk
  - 2015: pids - antall prosesser

Docker arkitektur

- Master prosess, kjørt som root
- Kontrolleres via RPC HTTP grensesnitt
  - Unix domain socket
  - TCP
- Klient (fri flyt ved tilgang til socket/TCP)

PID 1

- Upstream anbefaling
  - Din app som PID 1
  - Ikke supervisord/runit/s6/etc
- Hva med zombie prosesser?
  - Forby child prosessser?
- Trenger du et OS
  - Bruk LXC og la init(8) bli PID 1
  - Få med cron(8) og syslogd(8) på kjøpet

Distroer

- Mainstream distroer ikke designet for containere:
  - Node: 642.2 MB
  - Java: 641.9 MB
  - Ubuntu: 187.9 MB
  - CentOS: 172.3 MB
  - Debian: 125.1 MB

- Alternativ:
  - Alpine: 5.2 MB
  - Bygg din egen
    - buildroot.net
    - custom

Sikkerhet

- Ressursoptimalisering - ikke sikkerhetsisolasjon!
  - Linux kernel
    - 4.3: 20 621 558 linjer
  - Docker root daemon

- Sikkerhetsoppdateringer

User namespace

- Fjerner nødvendighet av root
- Sikrere
- Enklere
- LXC: støttet i over 2år
- Docker: eksperimentell (ubrukelig) versjon tilgjengelig forrige måned

Tillit
------

o hub.docker.com...

WTF
---

o --privileged=true

Ytelse

- Kjappere enn VM?
  - clearlinux.org - VM booter på under 150ms
  - Hardware virtualisering + virtio er generelt raskere på IO/net mens litt tregere på ren CPU ytelse
- Kjappere enn LXC?
  - Docker trenger ikke starte init med oppstartsscript
  - I bredre distroer er disse optimalisert for container bruk

Effektivitet

- Lagdeling og underliggende FS tenknologi (COW)
  tregere og bruker ofte mer plass enn naiv en/to-lags
  implementasjon (LXC)

- Manglende garbage collection

Community

- Bidro selv i starten
- 1000+ åpne issues
- $$$
  - RedHat
  - Microsoft
- CoreOS drama
  - rkt
  - opencontainers.org

"All problems in computer science can be solved by
 another level of indirection...
 Except for the problem of too many layers of indirection."

   David Wheeler

Misbruk

"Containers will not fix your broken architecture.
You are welcome."

twitter.com/littleidea/status/659445920954642432

"Not even two dockers could contain the shittiness
of this application."

twitter.com/sadserver/status/625663479370883073

Egnethet

- Ikke putt alt i en docker container

- Hvorfor må din app containes?
  - Er den for kompleks til å håndteres uten?

- Dockerizing
  - Antipattern

- Kveg vs kjæledyr
  - Er din app klar til å bli skutt i hodet
    på tilfeldige tidspunkt?

- Stateless vs stateful

- Distribuert system vs failover

- Docker gjør ikke tar(1) obselete

- Container specifications
- Container runtimes
- Container management
- Container definition
- Registries
- Container OS
- VM management
- Scheduling
- Cluster definition
- Security
- Log management
- Service discovery
- Dynamic config management
- Container orchestration plattforms
- Hosted container platforms
- Container platform management
- Container-based PaaS
- Networking
- Monitoring
- Data layer
- Continous integration/deployment
- Getting started aides

Open Container Spec, Open Container Initiaitve, Docker, CoreOS, AppC, runc, libcontainer, rkt, Docker Engine, Docker daemon, Docker client, rkt CLI, Docker Image, Dockerfile, ACI (App Container Image), Docker Registry, Amazon EC2 Container Registry, Docker Hub, Google Container Registry, Quay.io, Docker Trusted Registry, CoreOs Enterprise Registry, boot2docker, RancherOS, Project Atomic, Ubuntu Core "Snappy", SmartOS, Photon OS, Docker Machine, Hashicorp Vagrant, Hashicorp Otto, Docker Swarm, fleet, Chronos, Docker Compose, fleet unit file, etcd, Marathon, Hashicorp Consul, Apache ZooKeeper, confd, Consul Template, Apache Mesos, Kubernetes, Hashicorp Nomad, Amazon EC2 Container Service (ECS), Google Container Engine, Tutum, RedHat Openshift, Joyent Triton, Giant Swarm, ProfitBricks, Modulus, Project Orca, Rancer, ContainerShip, panamax, Shipyard, Joyent SmartDataCenter, Mesosphere DCOS, CoreOS Tectonic, Nirmata, ContainerShip Enterprise, StackEngine, AppFormix, Deis, Flynn, RedHat Openshift Origin, Cisco Mantl, Deis Dokku, Docker port expose, Docker linking, libnetwork, flannel, Weave, Calico, Docker stats API, sysdig, cAdvisor, Weave Scope, Sysdig Cloud, ClusterHQ Flocker, Docker logs, logspout, Shippable, Wercker, Twistlock, Scarlock, Conjur, Docker Kinematic

Kompleksitet

- Docker engine: 580 511 linjer

Alternativer

- LXC
- Lettvektsløsninger:
  - unshare(1)
  - github.com/arachsys/containers
- Direkte i app server (uWSGI)

```
#include ...

int main(int argc, char argv)
{
        uid_t uid = getuid();
        uid_t gid = getgid();

        unshare(CLONE_NEWUSER|CLONE_NEWNS|CLONE_NEWPID|CLONE_NEWNET);

        int fd = open("/proc/self/uid_map", O_RDWR);
        dprintf(fd, "%u %u 1\n", 0, uid);
        close(fd);
        fd = open("/proc/self/gid_map", O_RDWR);
        dprintf(fd, "%u %u 1\n", 0, gid);
        close(fd);

        setgroups(0, 0);

        chdir(argv[1]);
        mount("/dev", "./dev", 0, MS_BIND|MS_REC, 0);
        mount("/proc", "./proc", 0, MS_BIND|MS_REC, 0);
        chroot(".");

        execv(argv[2], argv+2);
}
```

Unikernels

- Fremtiden?
- Fjerner deler av OS i motsetning til å legge til et lag over
  - Mindre
  - Sikrere
  - Kjappere
  - Mer effektivt
- Rumprun
  - NetBSD
- MirageOS
  - OCaml

"The most revered feature of the modern operating system is support for running existing applications. Minimally implemented application support is a few thousand lines of code plus the drivers, as we demonstrated with the Rumprun unikernel. Therefore, there is no reason to port and cram an operating system into every problem space."

Antti Kantee. "The Rise and Fall of the Operating System".
USENIX ;login: October 2015, Vol. 40, No. 5

Slides
------

o http://git.uggedal.com/presentations/tree/boycott-docker