

Go And Linux Namespaces Don't Mix Well

Martynas Pumputis
@martyns



\$ whoami

- Vilnius Native
- <3 Systems
- Experimental Go port to Barrelfish OS

SEARCH

PROCESSES

BY NAME

CONTAINERS

BY DNS NAME

BY IMAGE

PODS

REPLICA SETS

DEPLOYMENTS

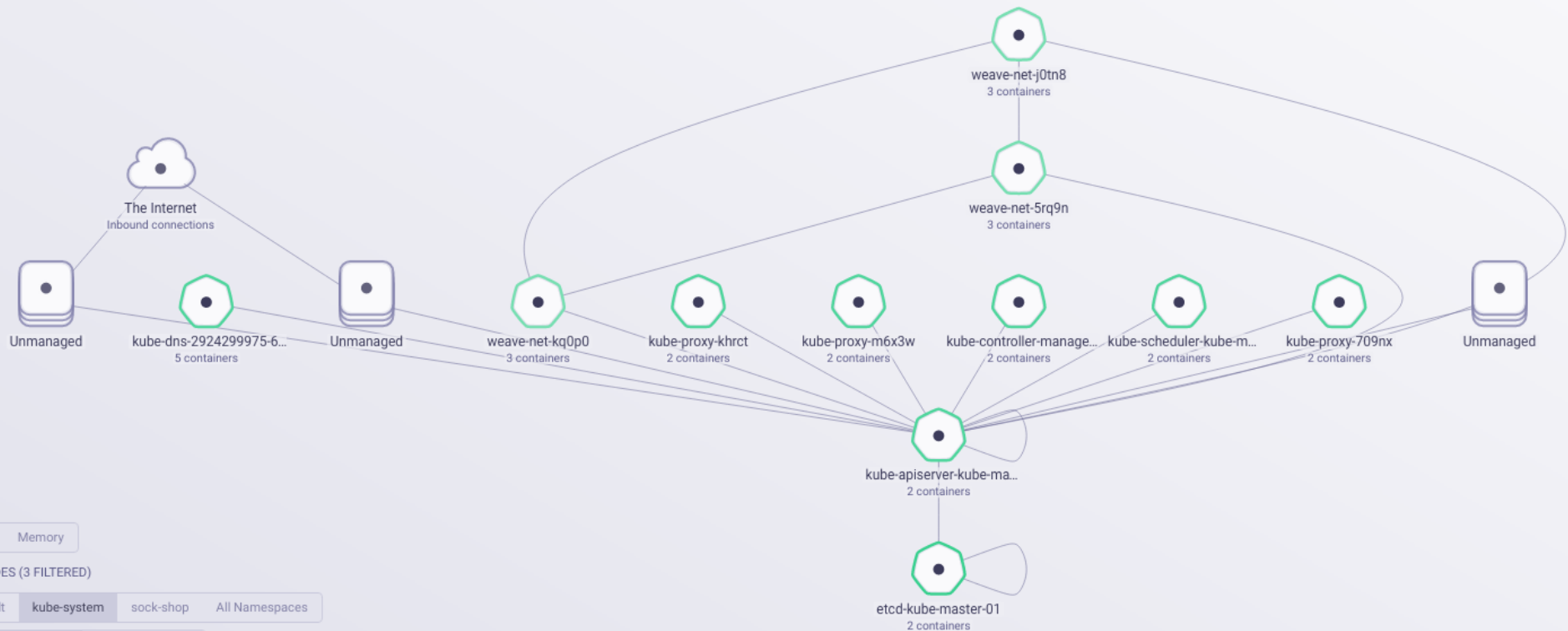
SERVICES

HOSTS

WEAVE NET

GRAPH

TABLE



<https://weave.works>

Background

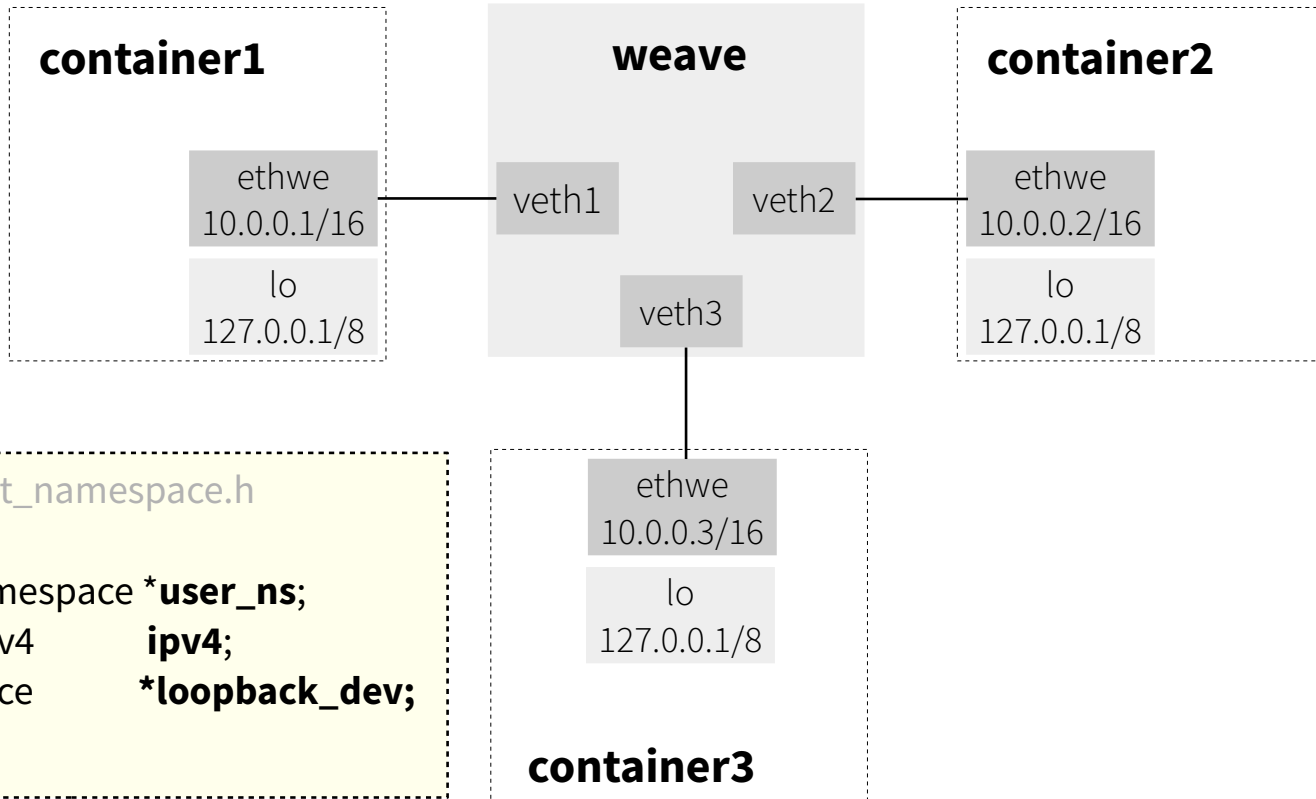
Weave Net

github.com/weaveworks/weave



```
host1 $ weave launch  
host2 $ weave launch host1  
c1 $ curl http://c4
```

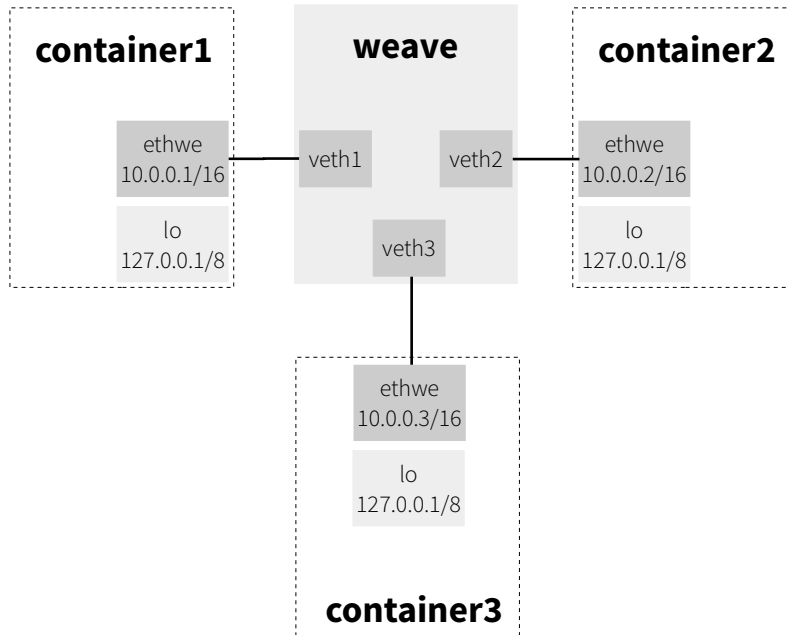
host1



```
// include/net/net_namespace.h
struct net {
    struct user_namespace *user_ns;
    struct netns_ipv4      ipv4;
    struct net_device      *loopback_dev;
    ...
}
```

```
// include/linux/netdevice.h
struct net_device {
    struct net *nd_net;
    ...
}
```

Problem

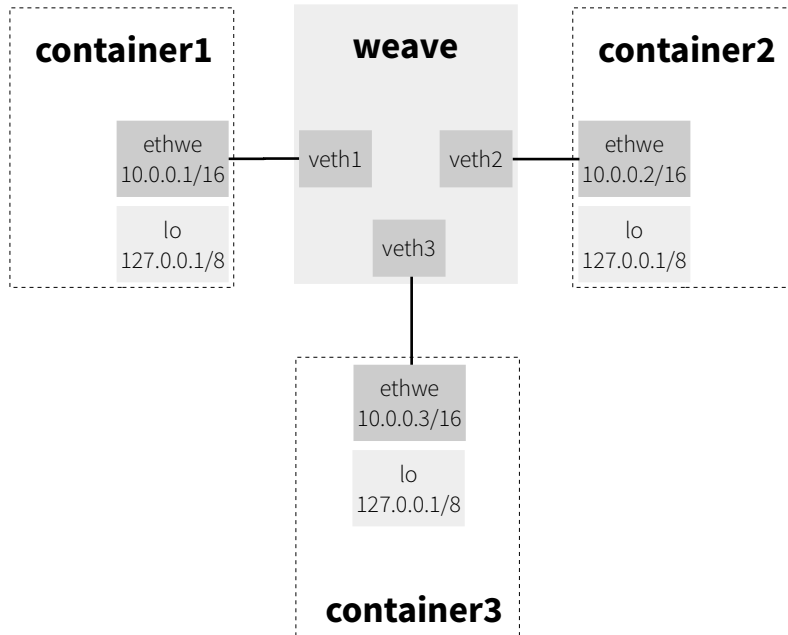


\$ weave ps

container1 10.0.0.1/16

container2 10.0.0.2/16

container3 10.0.0.3/16



\$ weave ps

Cannot find weave bridge: Link not found

Debugging

```
func printContainerIPs(containerID string) {  
    bridge, err := netlink.LinkByName("weave")  
    if err != nil {  
        fmt.Fatalf("Cannot find weave bridge: %s", err)  
    }  
  
    runtime.LockOSThread()  
    defer runtime.UnlockOSThread()  
  
    ns, _ := netns.GetFromPid(docker.GetPid(containerID))  
    hostNetNs, _ := netns.Get()  
    netns.Set(ns)  
    defer netns.Set(hostNetNs)  
  
    links, _ := netlink.LinkList()  
    fmt.Println(getContainerIPs(links, bridge))  
}
```

```
$ for i in $(seq 1 100); do docker $(weave config) run -td alpine /bin/sh; done  
<..>  
$ for i in $(seq 1 10); do weave ps >/dev/null; done  
Cannot find weave bridge: Link not found
```

\$ dmesg

<..>

```
[ 33.124702] device vxlan-6784 entered promiscuous mode
[ 48.159653] weave: port 1(vethwe-bridge) entered forwarding state
[ 74.775754] device veth0875d7f entered promiscuous mode
[ 74.872823] eth0: renamed from veth0a2892f
[ 74.873760] docker0: port 1(veth0875d7f) entered forwarding state
[ 74.873770] docker0: port 1(veth0875d7f) entered forwarding state
[ 75.245843] device vethwepl3971 entered promiscuous mode
[ 75.259164] ethwe: renamed from vethwepg3971
[ 75.262244] weave: port 2(vethwepl3971) entered forwarding state
[ 75.262253] weave: port 2(vethwepl3971) entered forwarding state
```

```
$ github.com/socketpair/nltrace weave ps
```

```
<..>
```

```
----- BEGIN NETLINK MESSAGE -----
```

```
[NETLINK HEADER] 16 octets
```

```
.nlmsg_len = 44
```

```
.type = 18 <route/link::get>
```

```
.flags = 5 <REQUEST,ACK>
```

```
.seq = 1
```

```
.port = 0
```

```
[PAYLOAD] 16 octets
```

```
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
```

```
[ATTR 03] 6 octets
```

```
77 65 61 76 65 00          weave.
```

```
[PADDING] 2 octets
```

```
00 00          ..
```

```
<..>
```

strace

\$ strace -f weave ps

<..>

[pid 3361] openat(AT_FDCWD, "/proc/17526/ns/net", ...) = 61

[pid 3361] openat(AT_FDCWD, "/proc/3357/task/3361/ns/net", ...) = 62

[pid 3361] setns(61, CLONE_NEWNET) = 0

<..>

[pid 3361] socket(AF_NETLINK, SOCK_RAW, NETLINK_ROUTE) = 63

[pid 3361] bind(63, {sa_family=AF_NETLINK, ...}, 12) = 0

[pid 3361] sendto(63, "\x20\x00\...", ...) = 32

<..>

[pid 3361] recvfrom(63, "\x4c\x00\...", ...) = 236

<..>

[pid 3361] clone(...) = 3365

<..>

[pid 3361] setns(62, CLONE_NEWNET) = 0

<..>

[pid 3365] sendto(65, "\x2c\x00\...", ...) = 44

[pid 3365] recvfrom(65, "\x40\x00\...", ...) = 64

[pid 3365] write(2, "Cannot find weave bridge: Link not found\n", 41

\$ strace -f weave ps

<..>

[pid 3361] openat(AT_FDCWD, "/proc/17526/ns/net", ...) = 61

[pid 3361] openat(AT_FDCWD, "/proc/3357/task/3361/ns/net", ...) = 62

[pid 3361] setns(61, CLONE_NEWNET) = 0

<..>

[pid 3361] socket(AF_NETLINK, SOCK_RAW, NETLINK_ROUTE) = 63

[pid 3361] bind(63, {sa_family=AF_NETLINK, ...}, 12) = 0

[pid 3361] sendto(63, "\x20\x00\...", ...) = 32

<..>

[pid 3361] recvfrom(63, "\x4c\x00\...", ...) = 236

<..>

[pid 3361] clone(...) = 3365

<..>

[pid 3361] setns(62, CLONE_NEWNET) = 0

<..>

[pid 3365] sendto(65, "\x2c\x00\...", ...) = 44

[pid 3365] recvfrom(65, "\x40\x00\...", ...) = 64

[pid 3365] write(2, "Cannot find weave bridge: Link not found\n", 41

Solution


- *runtime.LockOSThread()*?

Noup

- Self-exec with *os/exec*?

Noup

syscall: Setuid/Setgid doesn't apply to all threads on Linux #1435

 **Open** gopherbot opened this issue on Jan 21, 2011 · 68 comments



Ian Lance Taylor

6/30/16



Other recipients: mart...@gmail.com

- show quoted text -

You aren't missing anything. Doing this correctly requires runtime support, and that support does not exist. It's not even obvious how to write it, at least not to me.

lan

```
func WithNetNS(nsPath string, cmd string, args ...string) ([]byte, error) {  
    var stdout, stderr bytes.Buffer  
  
    args = append([]string{"--net=" + nsPath, "weaveutil", cmd}, args...)  
    c := exec.Command("nsenter", args...)  
    c.Stdout = &stdout  
    c.Stderr = &stderr  
    if err := c.Run(); err != nil {  
        return nil, fmt.Errorf("%s: %s", string(stderr.Bytes()), err)  
    }  
  
    return stdout.Bytes(), nil  
}
```

Conclusions

- Go for containers software?



- [github.com/vishvananda/{netlink, netns}](https://github.com/vishvananda/netlink)



- strace **FTW!**

Questions ?



@martyns