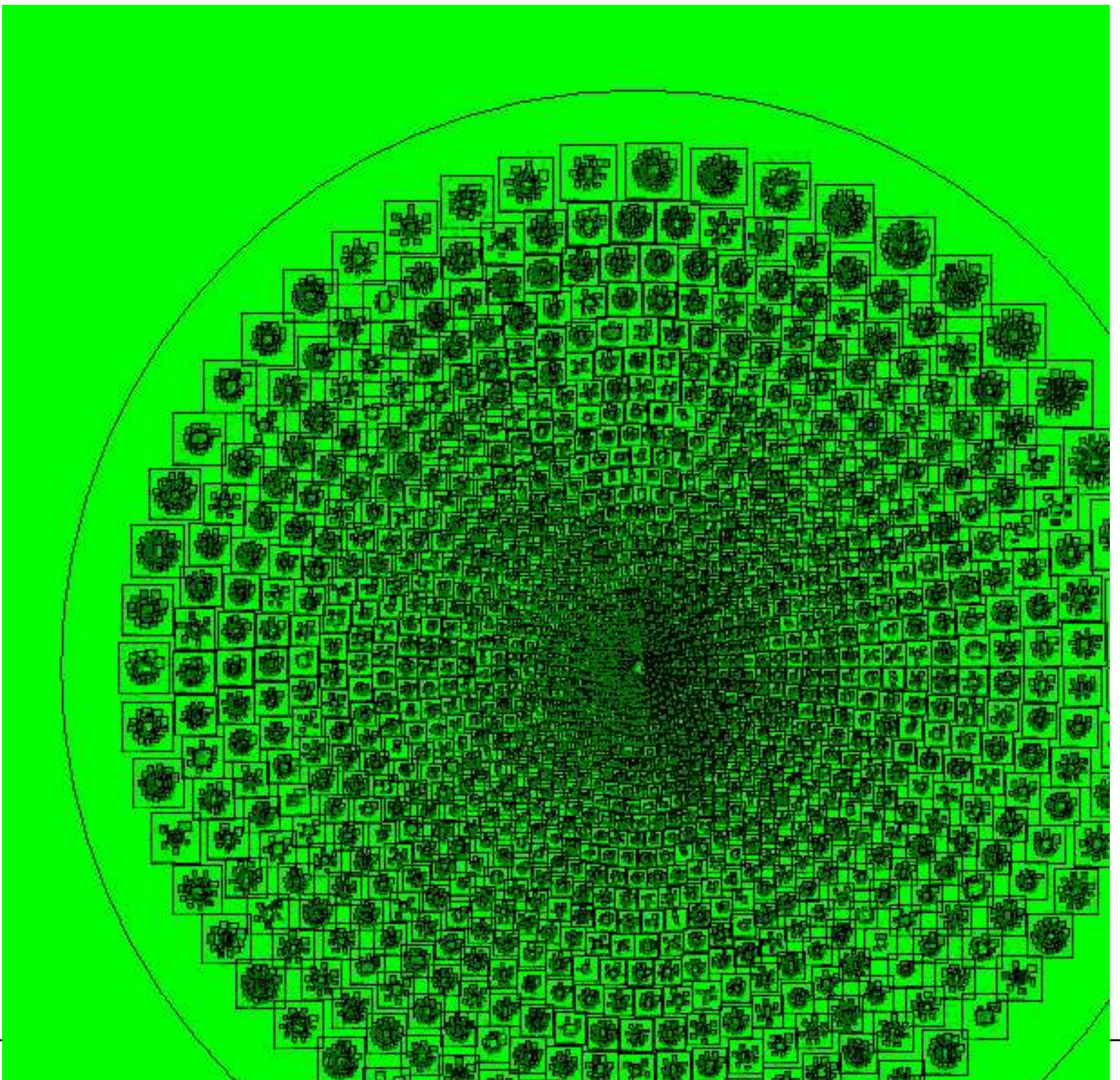


The Whole Linux Internet

UCL Computer Science Department
Jon Crowcroft
jon@cs.ucl.ac.uk

GLUUG Talk, 23rd February, 2001

Hooray for Magicpoint!!



Talk Abstract/Outline

Linux is used for desktop and server systems, and is being explored as an OS for embedded systems for appliances and so on. It is also used for site and access routers. However, there's no real reason not to use Linux for high-end routers. This talk is about the functionality in the 2.4 kernel that has been developed by the community, for high performance, and state-of-the-art traffic management for routing. In the talk, I will also mention for comparison, the Kame system from Japan widely used in BSD Unix routers for similar functions, including

- Fast Forwarding Table lookup
- Nice API from Routing Daemon to Kernel space tables
- Integrated&Differentiated Services
 - (Qdisc/Classifier structures)
- Multicast
- IPv6
- Firewalls/NATs/Masquerading etc

- I'll then talk about what is missing (a replacement for Gated!).

- So linux can take on yet another borg (after microsoft):

Cisco:-)

History and Philosophy

**BBN did 4.1c for Vax,
-> Berkeley**

**Berkeley did 2.8 and 4.2 for PDP 11 and
Vaxen**

Most the world takes their code

MS, Solaris, HPUX all someway derived

Linux doesn't....

Note Spider and SCO and other rewrites
as did Sun/mentat and even MS eventually
(see free v6 winNT stack)

10 years after: That gets us to the end of the 1980s. Entropy!!!

IP evolves - adding

- multicast
- mobile
- ipv6
- integrated services&rsvp
- differentiated services
- NAT
- telepathy

Cisco struggle to keep up

IOS is 10M lines of code on a 300 line RTOS:-)

- Like Nortel telephone exchange s/w,
- like HP printers,
- like Oracle dbase
- like MS - its a mess.....

Modern OS gives us several nice things

- Modules
- Clean User Space/Kernel split
- SMP
- Device driver abstraction
- Good for simple server or router

But, til recently, rather naive network api

- So both free systems of interest languish as somewhat amateur offerings....
- on the other hand, the "professional" offerings accrete more and more junk....

Sugar and Spice:- Router

user space (wont say a lot here)

- routing daemons
- routed
- mrouted
- gated (was NSFnet T3 code - merit/michigan/ibm -
encumbered)
- various projects
 - Linux router project
 - routing toolkit
 - xorp (at&t/berkeley/intel)
- MIBs
- File system for config/logs etc

Model (IETF:-) -

- Forces

Players and Talkers

WIDE Project (wrote magicpoint - good chasps)

- Kenjiro Cho does KAME
 - IPv6
- does altq
 - perf

Linux Project (wrote lotsa stuff)

- Alexey Kuznetsov does FIB and qdisc code
 - IPv6 (?)
- almesberger does nice
- q work

Lets review split between user space and kernel

- /dev/kmem
- /proc
- rtnetlink
- callbacks

RTNETLINK!

`rtnetlink_socket =`

- `socket(PF_NETLINK, int socket_type, NETLINK_ROUTE);`

`int RTA_OK(struct rtattr *rta, int rtabuflen);`

`void *RTA_DATA(struct rtattr *rta);`

`unsigned int RTA_PAYLOAD(struct rtattr *rta);`

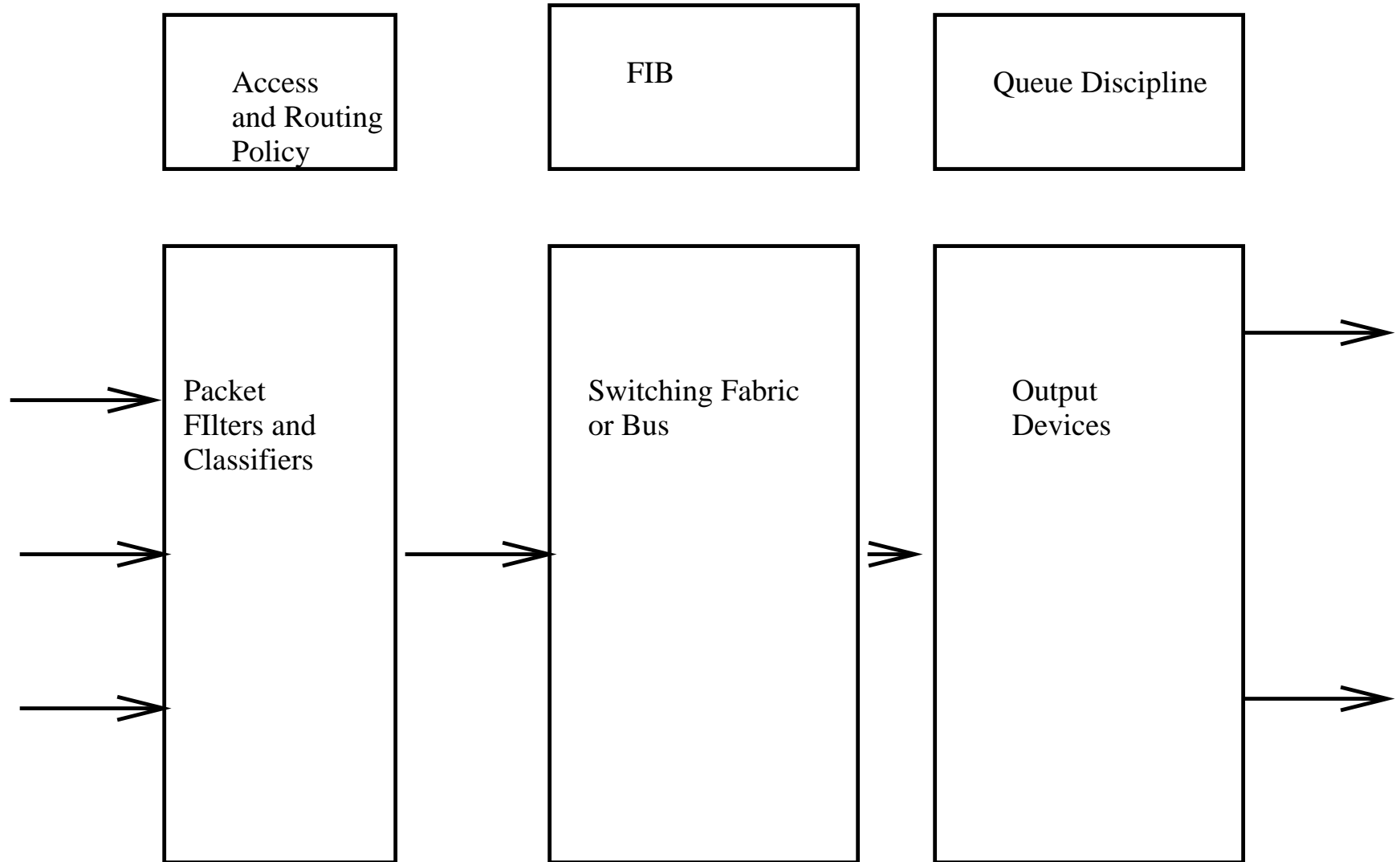
`struct rtattr *RTA_NEXT(struct rtattr *rta, unsigned int rtabuflen);`

`unsigned int RTA_LENGTH(unsigned int length);`

`unsigned int RTA_SPACE(unsigned int length);`

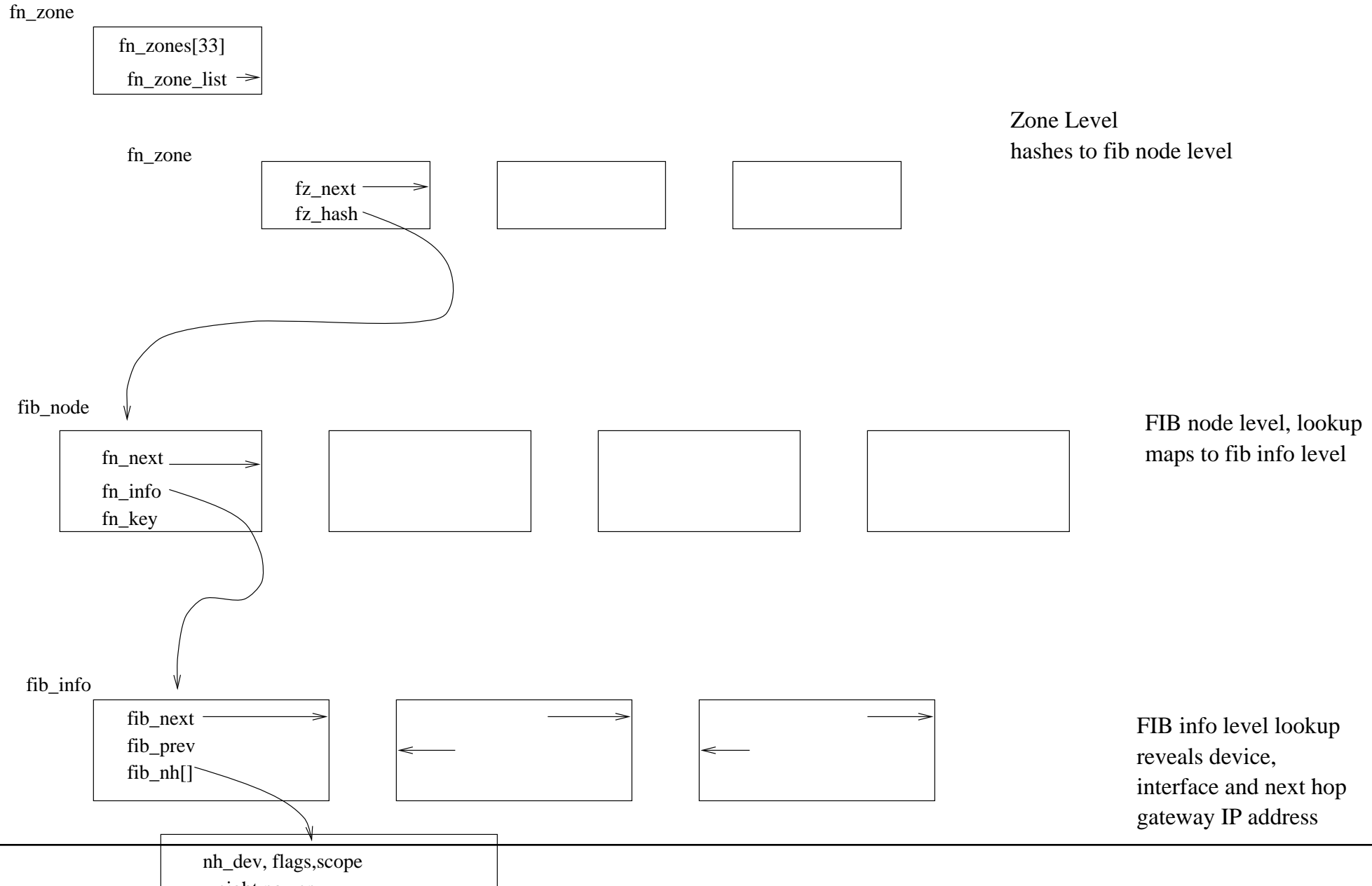
- RTM_NEWLINK, RTM_DELLINK, RTM_GETLINK
Create, remove or get information about a specific network interface.
- RTM_NEWADDR, RTM_DELADDR, RTM_GETADDR
Add, remove or receive information about an IP address associated with an interface.
- RTM_NEWROUTE, RTM_DELROUTE, RTM_GETROUTE
Create, remove or receive information about a network route.
- RTM_NEWNEIGH, RTM_DELNEIGH, RTM_GETNEIGH
Add, remove or receive information about a neighbour table entry

Forwarding and Treatment!

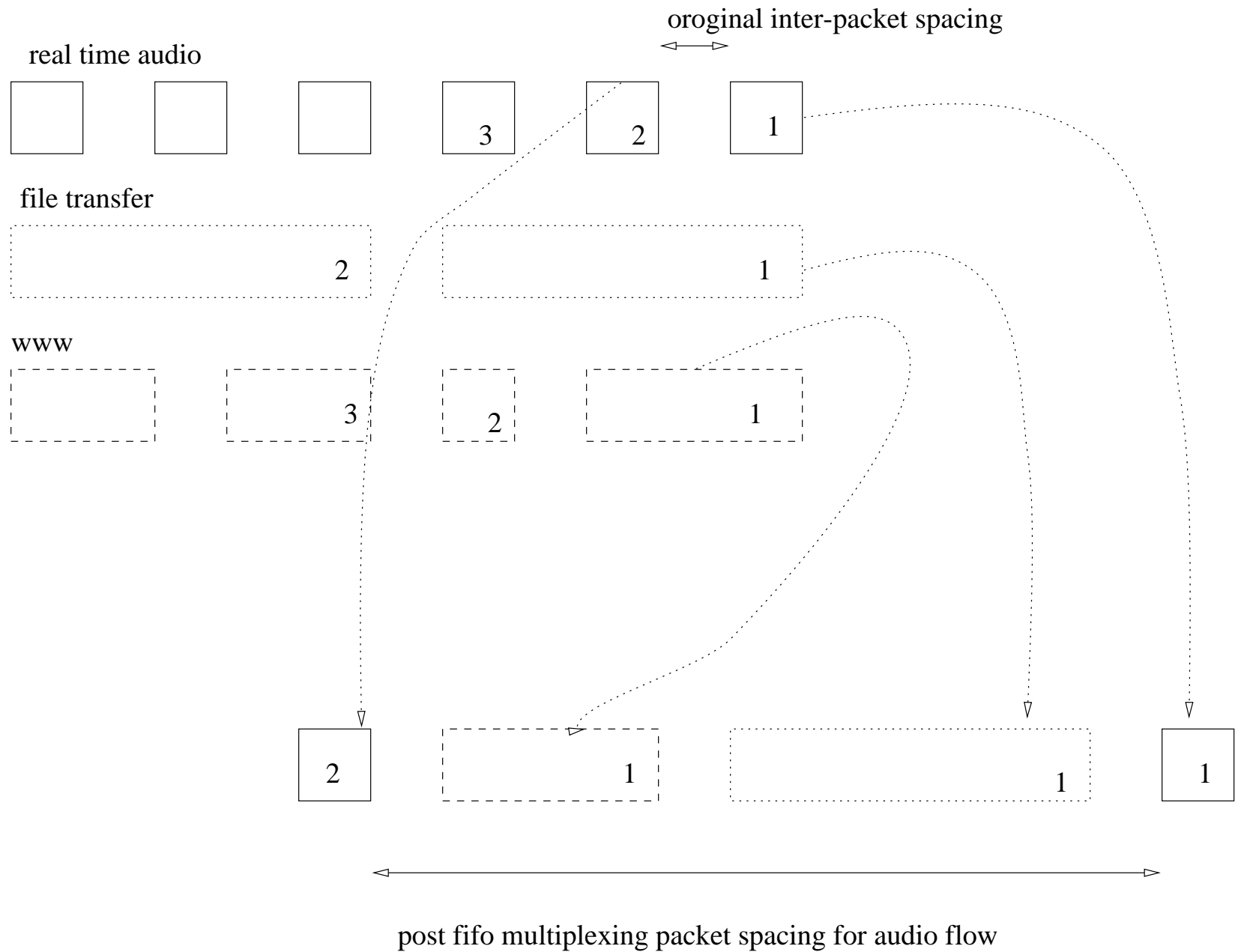


Rats and Snails...

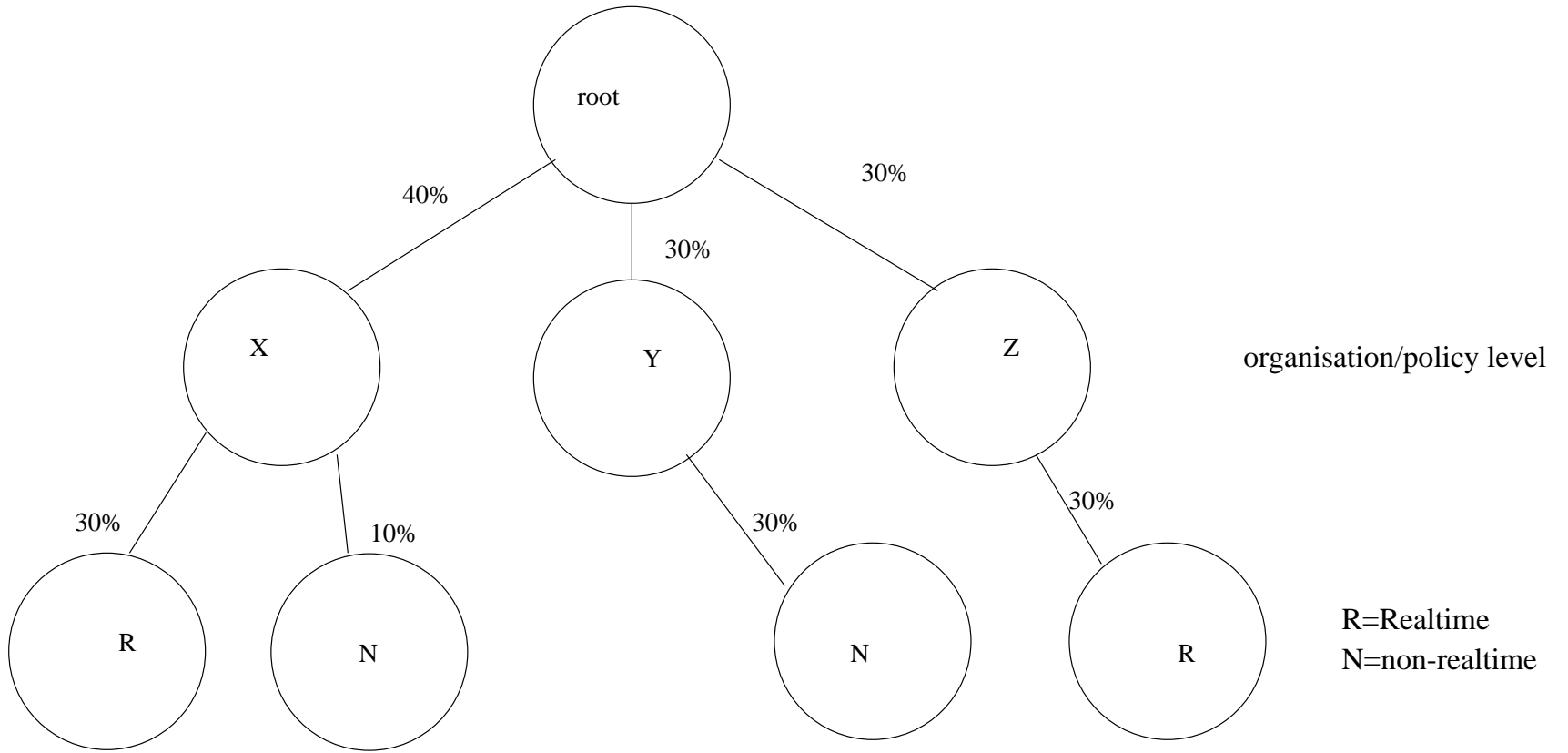
Linux 2.2-2.4 FIB



Packetisation - Jitter Bugs:-)



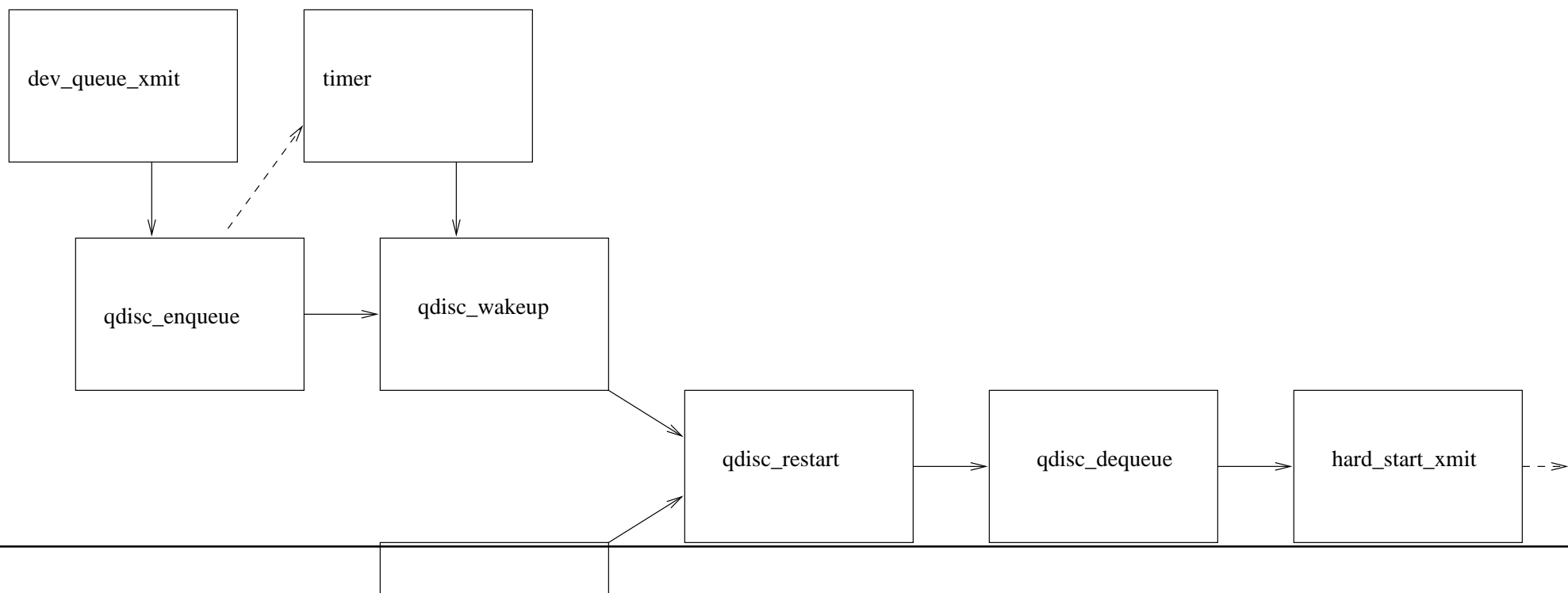
CBQ reminder



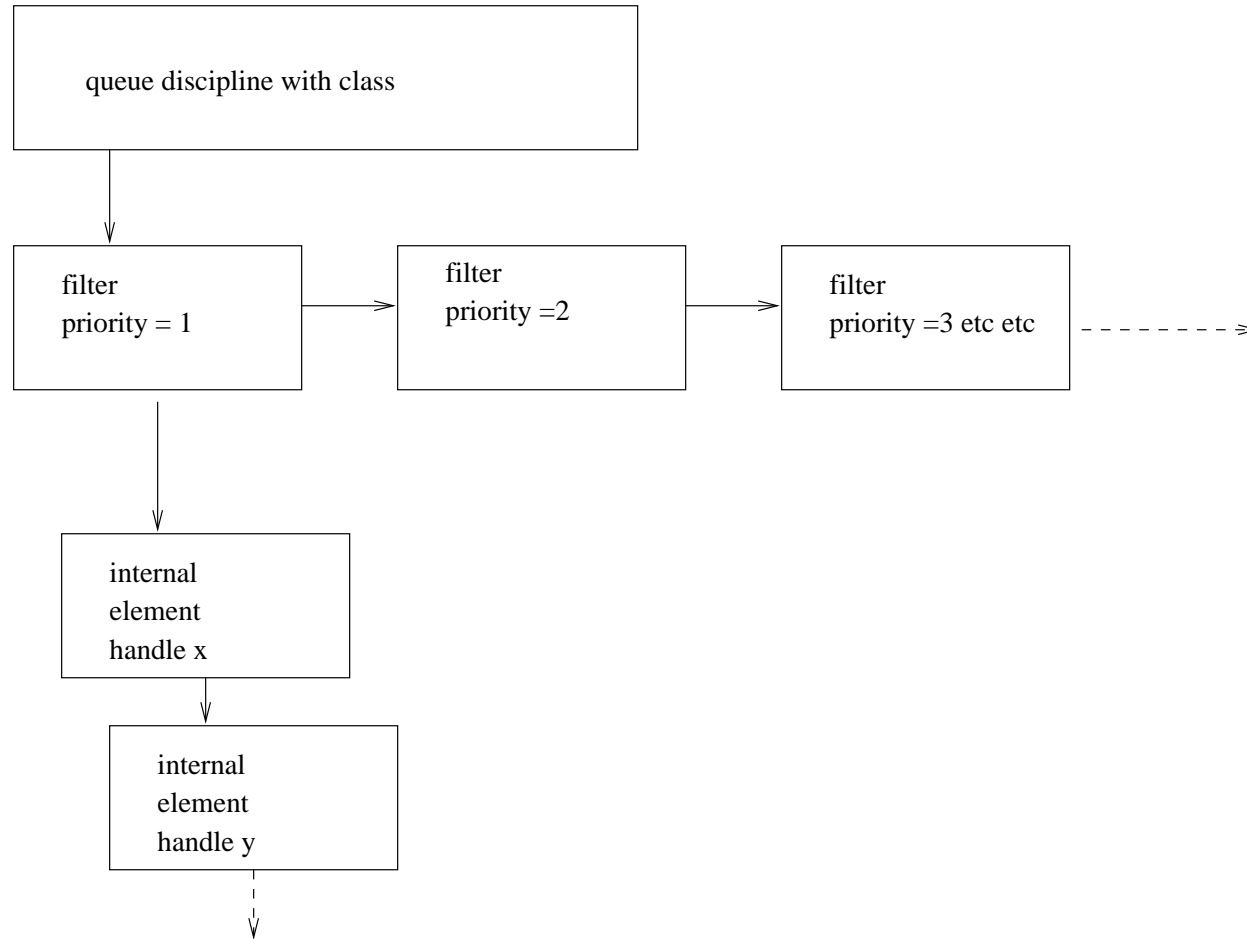
Linux Queueing Discipline - Class Based (OO OO OO :-)

Also Known as Forwarding Treatment

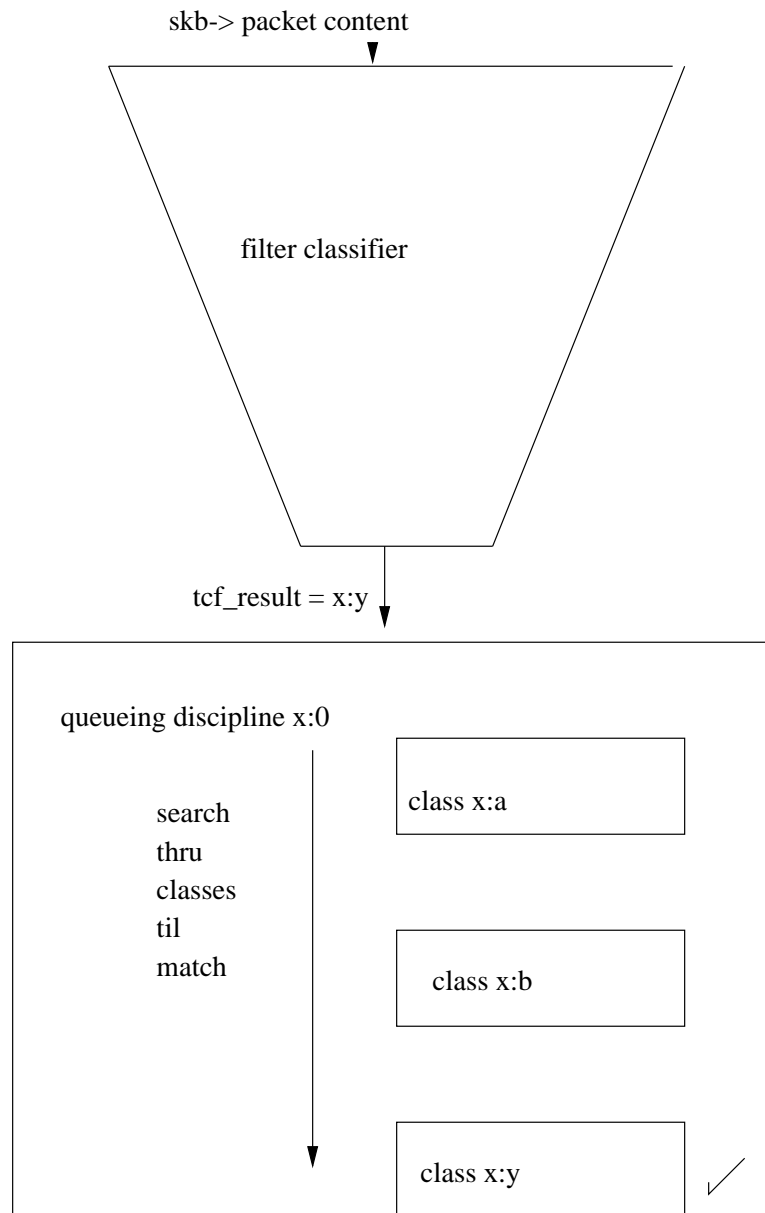
This stuff does int-serv, diff-serv, and makes excuses from coming home late from work



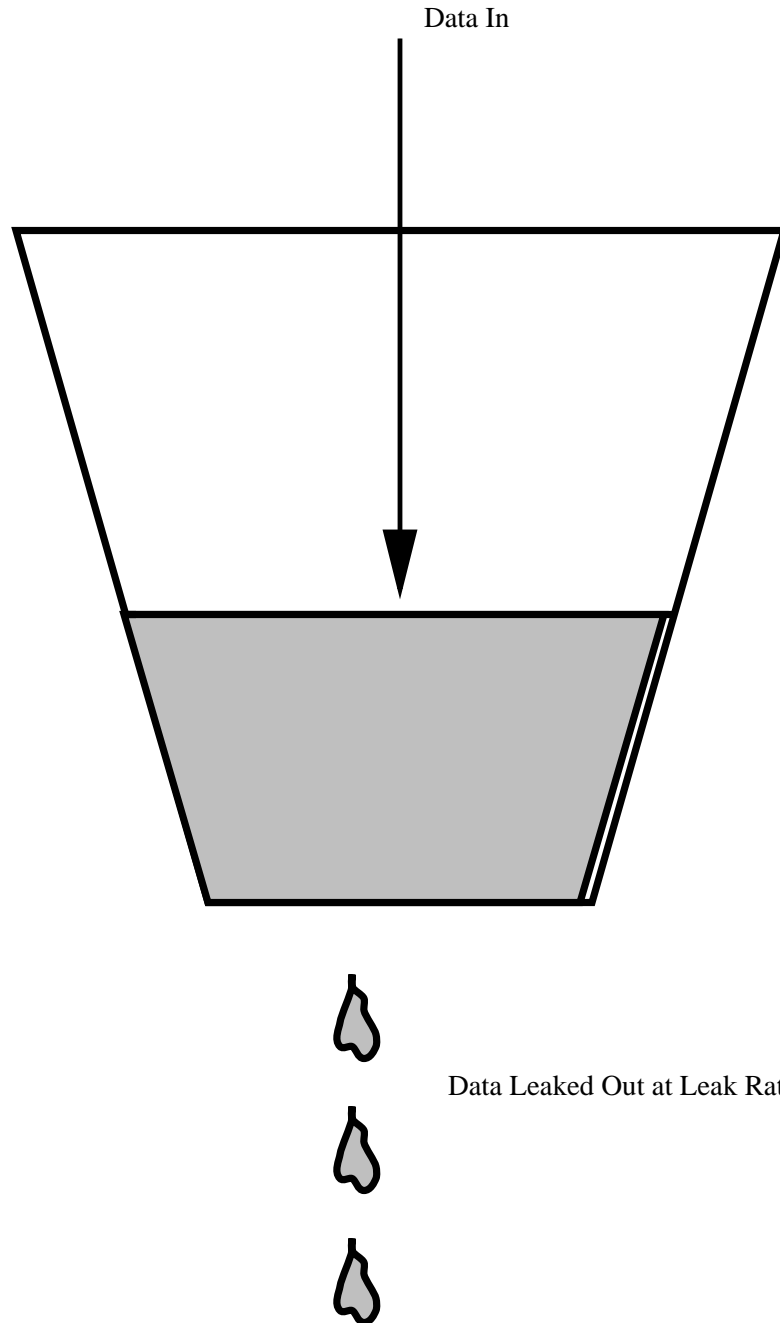
Filter



Apply



For example...leaky bucket...



And puppy dog's tails...

So what does FreeBSD provide for all this?

- **Radix FIB**
 - uses 1 level hashing - like half the FIB
- **ALTQ**
 - heavily CBQ
 - ifconfig/interface oriented
 - hard to see if one can do input q
 - meter/police/drop interface tricky
- **tools**
 - ifconfig!

What are we missing?

- XORP - build opensource platform for modular routed
- Need dijkstras and similar
- Need route mibulator
- Need TE modules
- other?

Have a nice day at UCL!

- Visit

- <http://www.cs.ucl.ac.uk/staff/jon/li/>

- for further information.