Inside Linux Router

Dr. Ying-Dar Lin

High Speed Network Lab.

Department of Computer Information Science

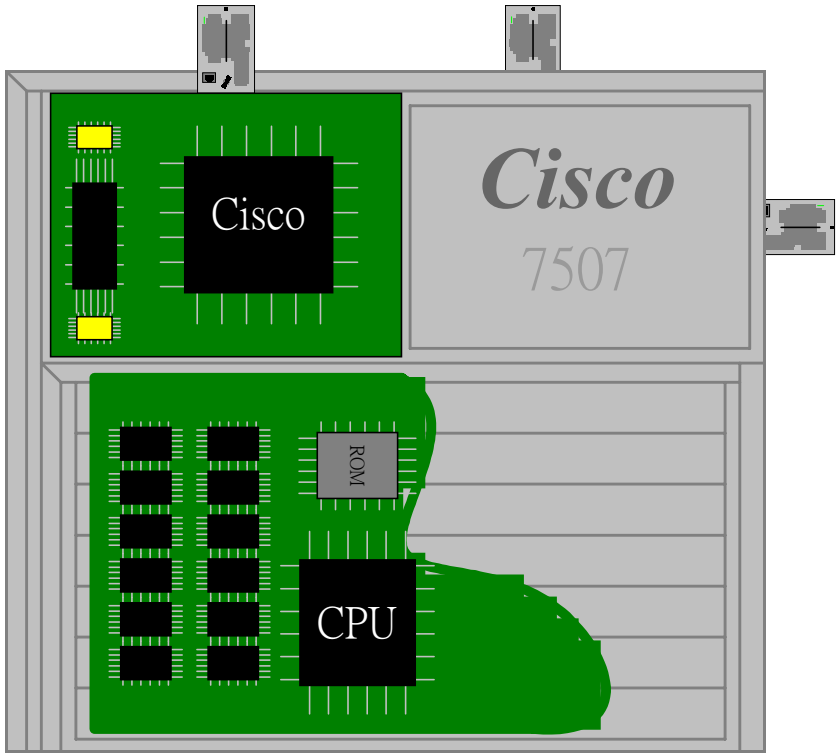
National Chiao Tung University

May 15, 1999

Content

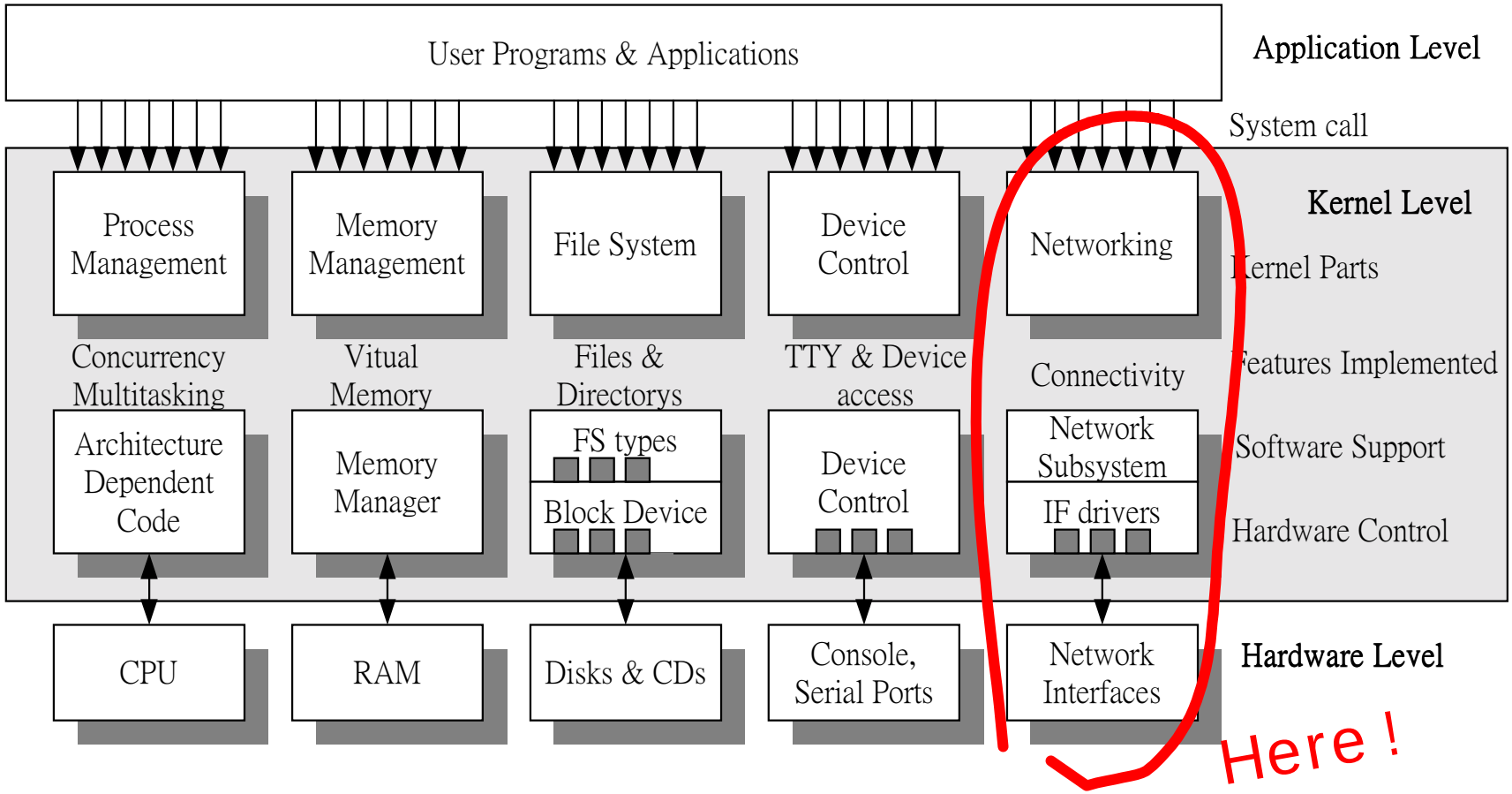
- What's inside the router box ?
- Linux Architecture
- Modules and Daemons
- Protocols and Algorithms
- Packet Flows: User/Control-plane
- New feature: QoS

What's inside the router box ?

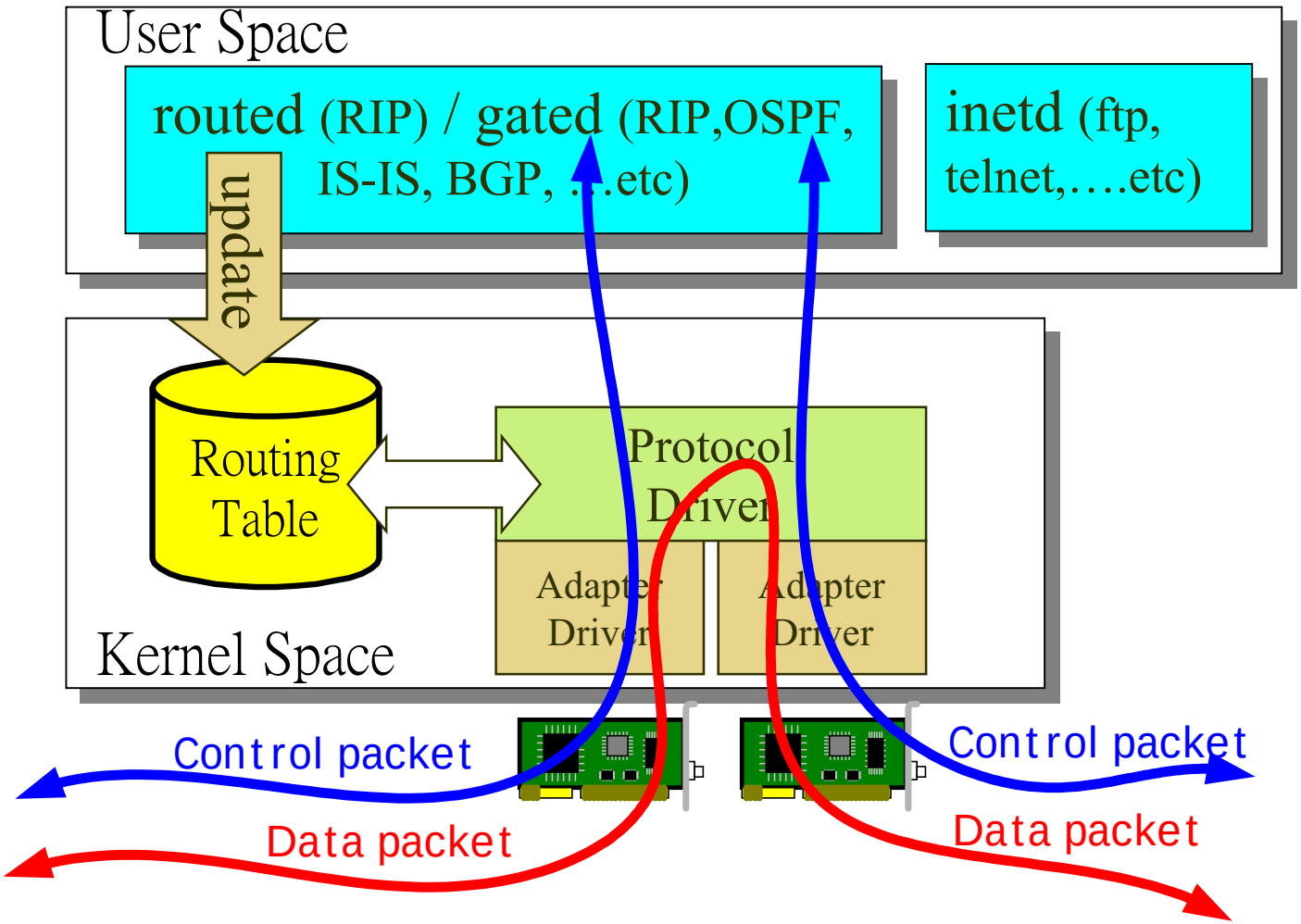


- Much the same as the PC with "LINUX" inside !!

Linux Architecture



Modules & Daemons

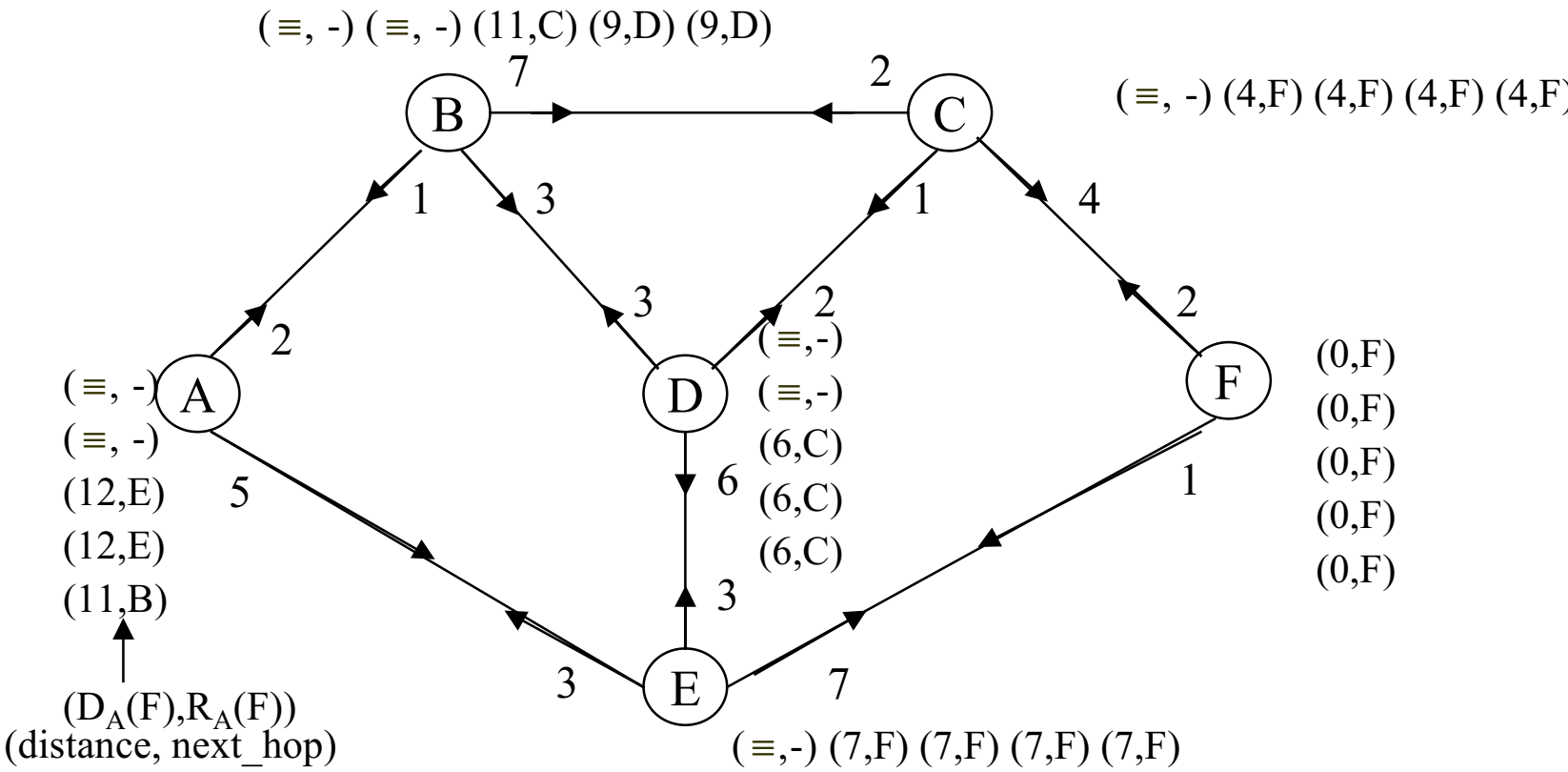


Protocols and Algorithms

- **Standard**
  - Protocol: **ICMP, RIP, OSPF, IS-IS, BGP,….etc**
  - Algorithm:
    - Shortest path computation
      - Distance Vector : Bellman-Ford
      - Link State : Dijkstra
- **Non-standard**
  - Protocol: **IGRP** (cisco)
  - Algorithms:
    - checksum computation
    - Routing Table lookup

Distance Vector Routing (Bellman-Ford)

For destination F

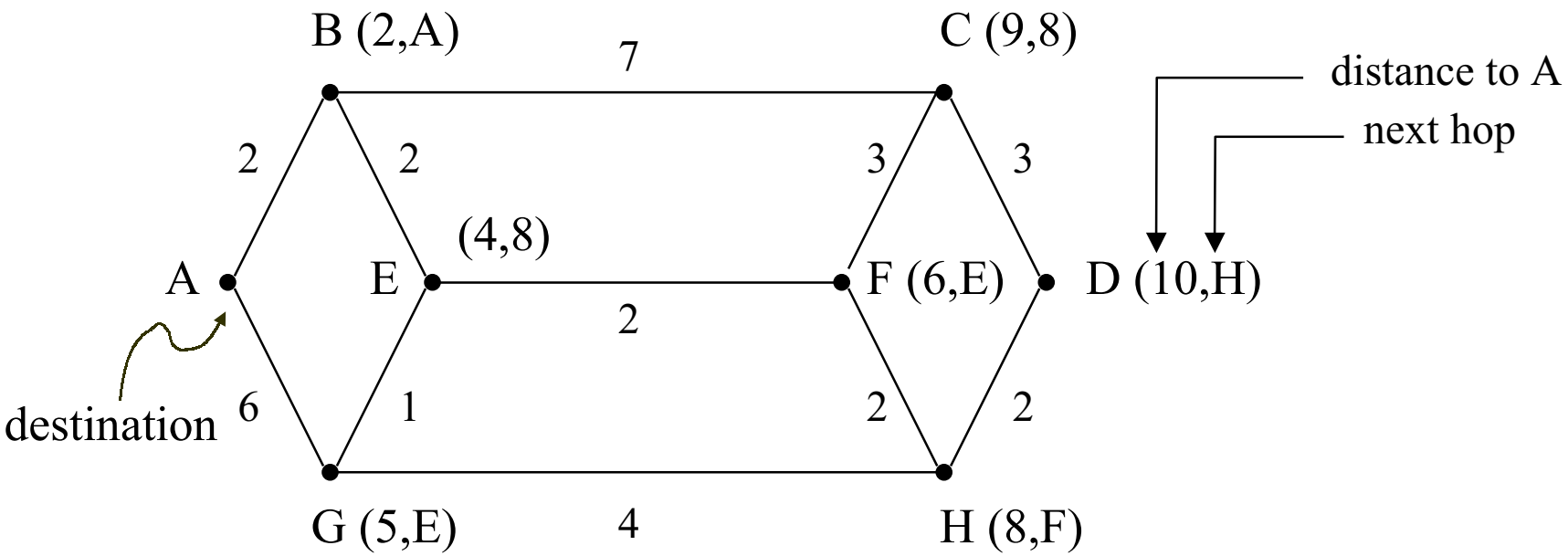


Problems with Distance Vector

- **No link-bandwidth consideration**
  - only cares instantaneous queue length
  - instability & oscillation
- **Only rapidly to Good News**
  - travel at the rate of one hop per exchange
- **But leisurely to Bad News**
  - count to infinite
  - ∵ No router ever has a value a few more higher than the minimum of all its neighbors

Link State Routing (Dijkstra)

For destination A

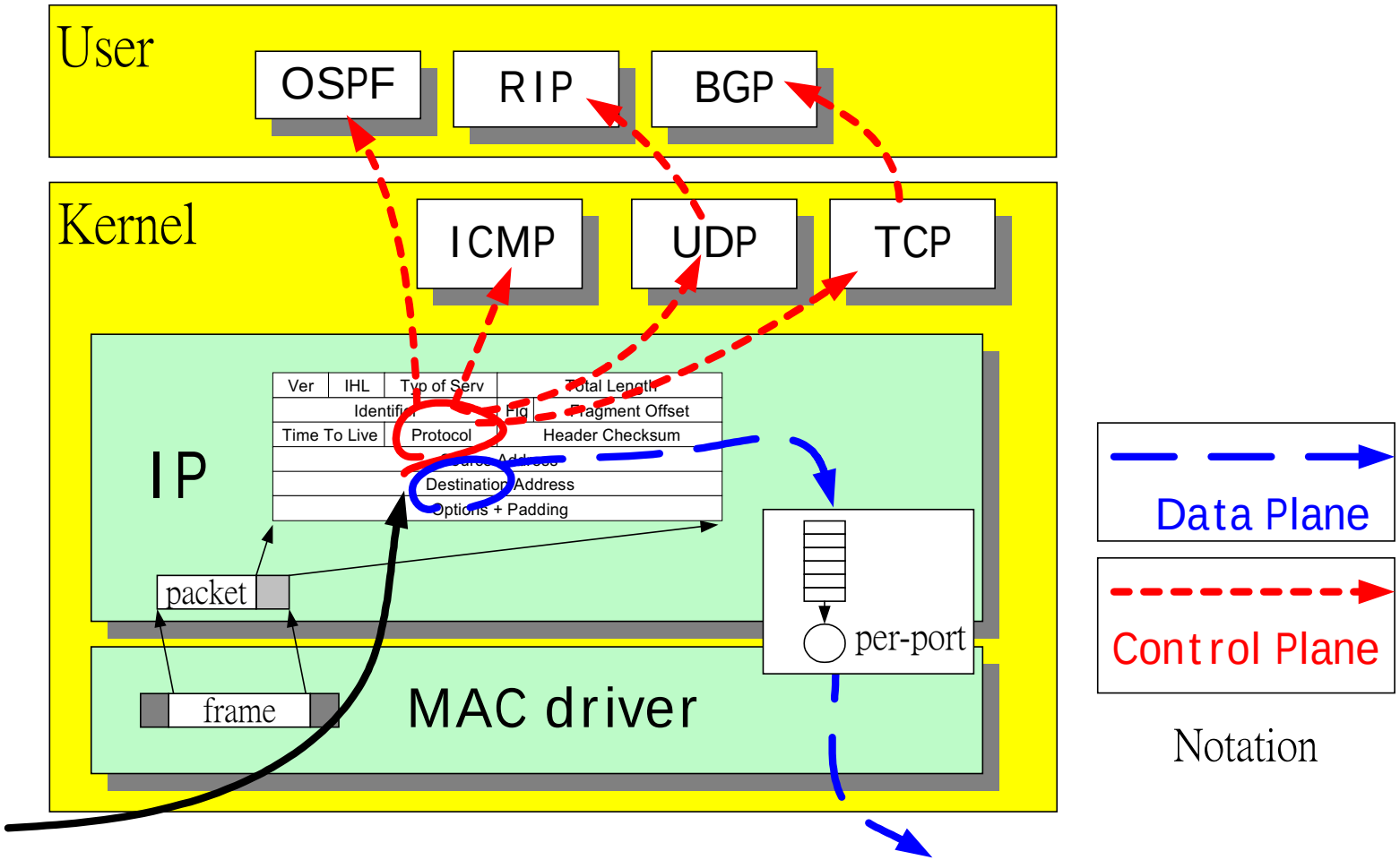


Link State Routing

- **Ex: IS-IS, OSPF**
- **Learn neighbors & their network addresses**
  - **(HELLO packet)**
- **Measure link state**
  - **(ECHO packet)**
- **Building link state packets**
  - **(router id, sequence, age, (neighbors, cost), ….)**
- **Distribute link state packets to all other routers**
  - check and update the table

    (source router, sequence, age, send flags, ACK flags)
- **Compute new routes**
  - run Dijkstra's algorithm locally

Packet Flows - User/Control Plane



Control-Plane : Shortest Path(1/2)

- **RIP in routed - Bellman-Ford**

RIP Routing Table

Destination	Next Hop	Metric	Interface

RIP Header

0 7	8 15	16 31
Command	Version	*Routing Domain
Address Family		*Routing Tag
IP Address		
*Subnet Mask		
*Next Hop IP Address		
Metric		
Repeat of previous 20 bytes		

* Only in RIP-2

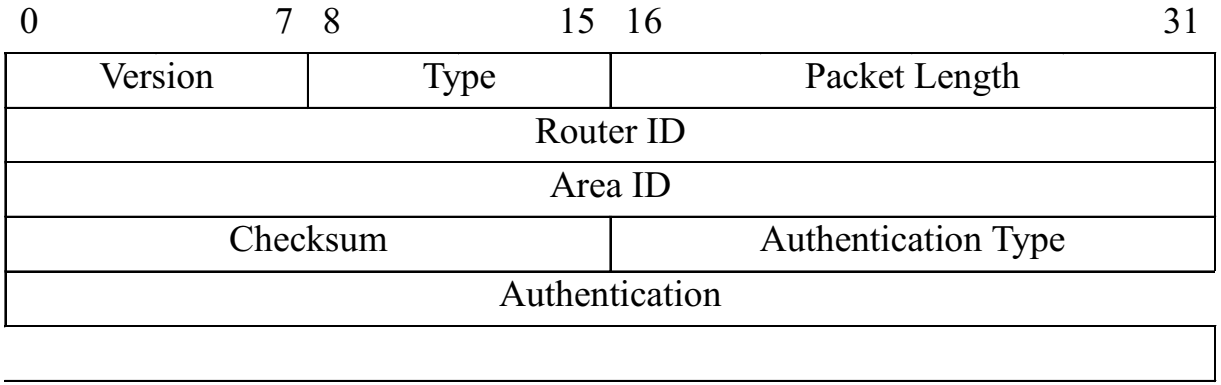
Control-Plane : Shortest Path(2/2)

- **OSPF in gated - Dijkstra** (Dynamic Programming)

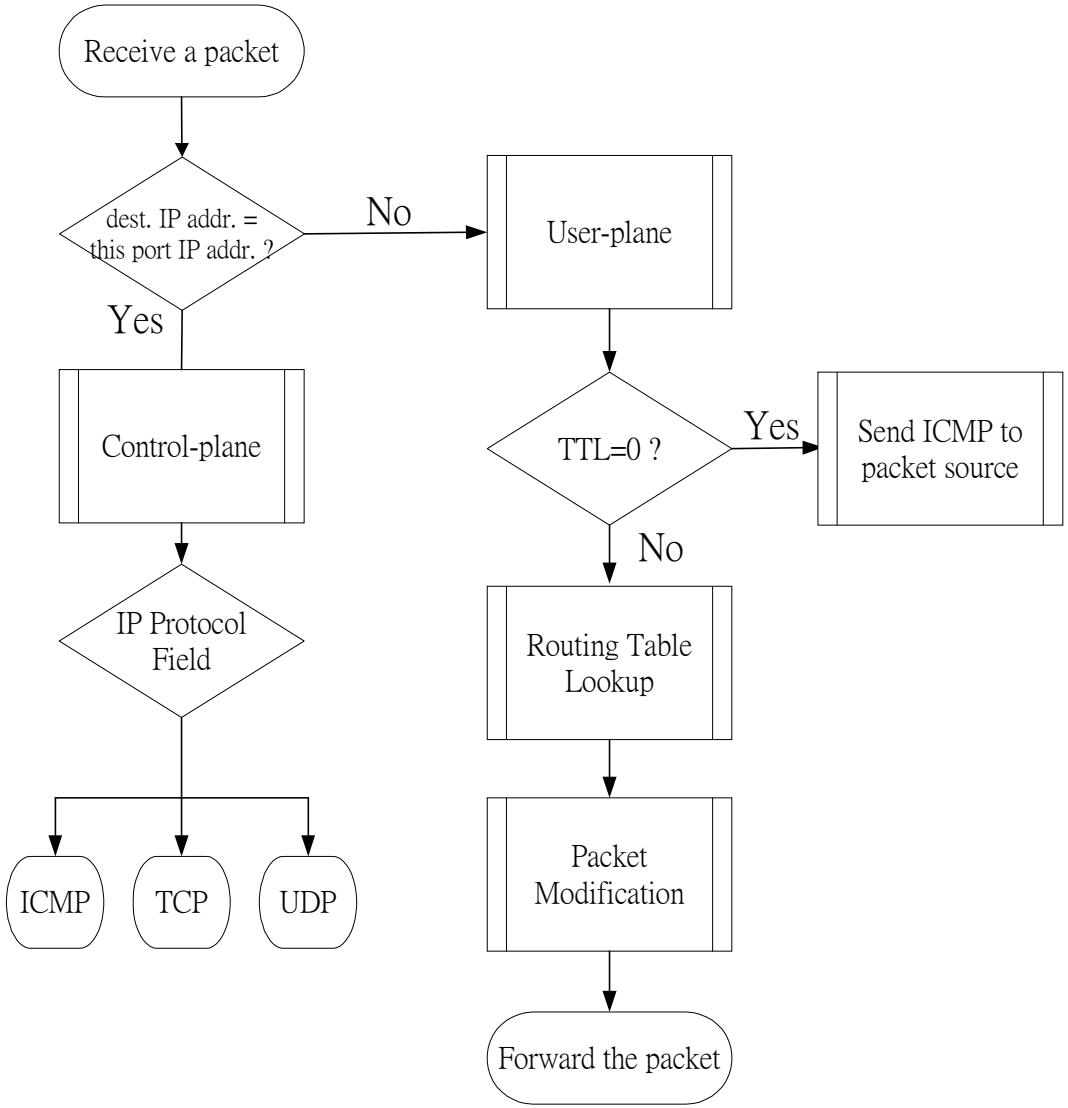
OSPF Routing Table

Destination	Next Hop	Distance metric

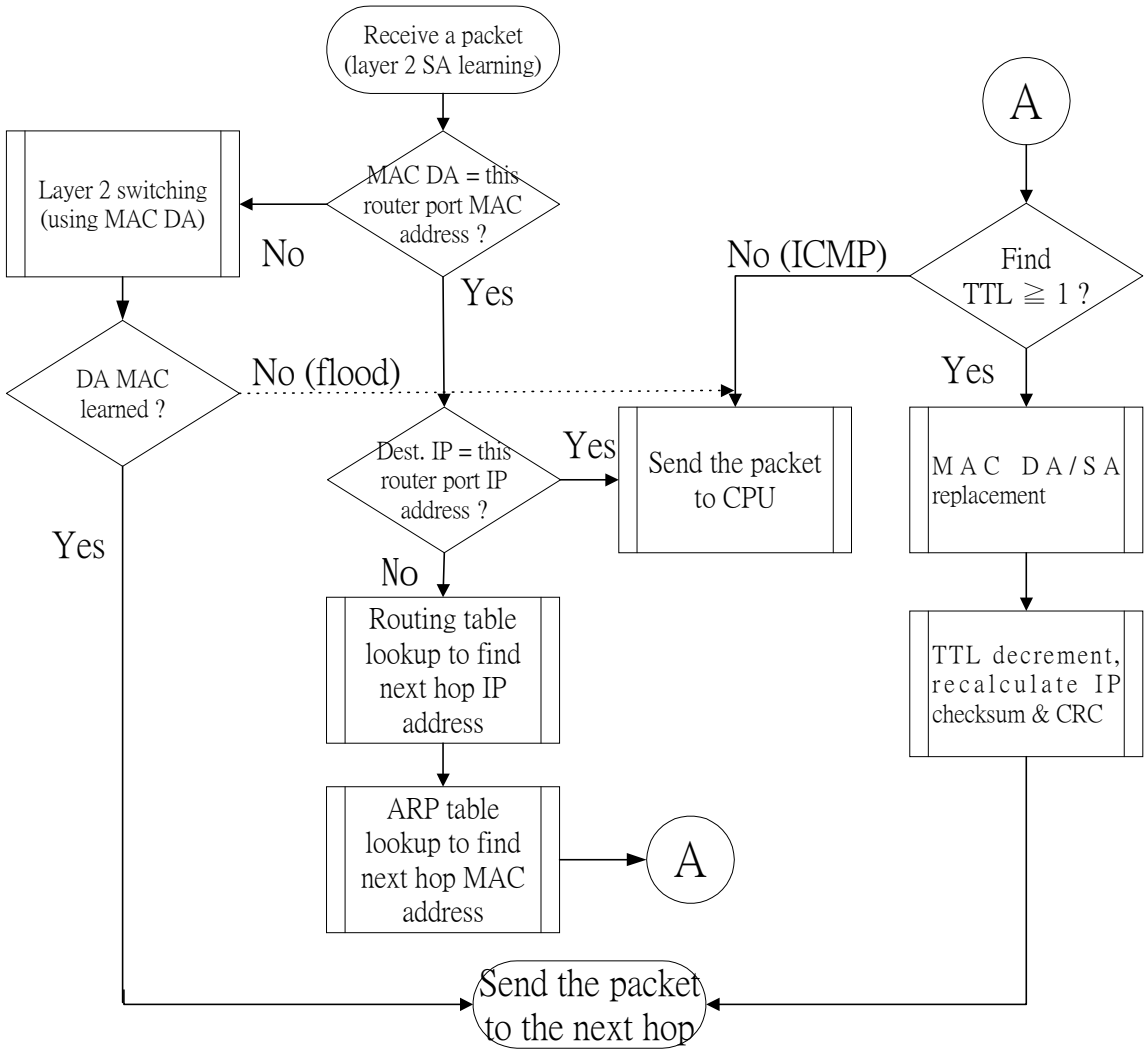
OSPF Header



User-Plane Processing (1/2) - Linux Router

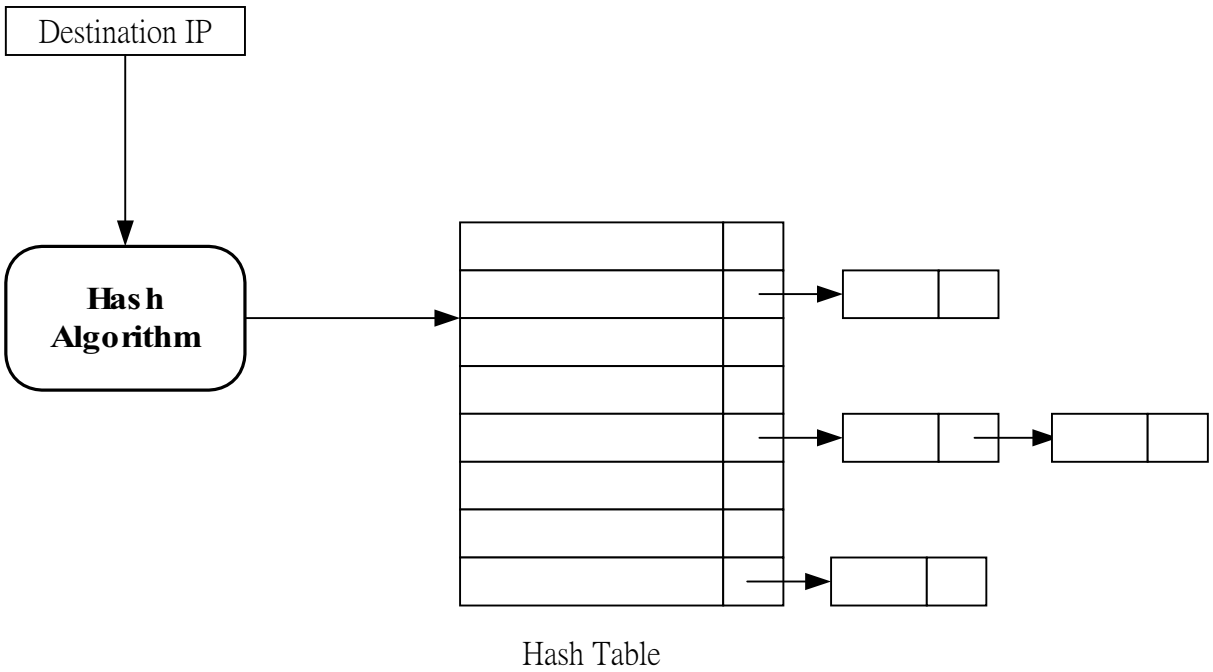


User-Plane Processing (2/2) - Layer 3 switch



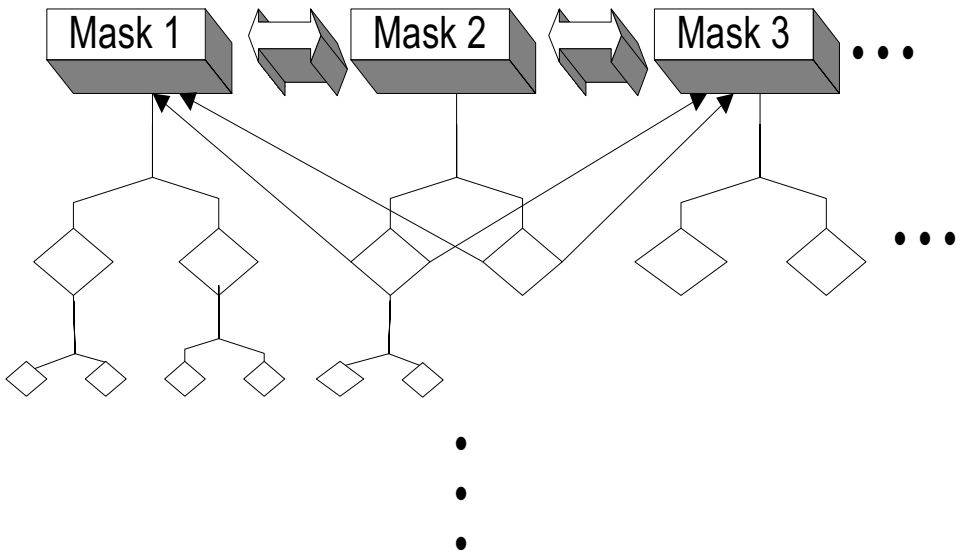
User-Plane : Table Lookup (1/2)

- Routing table in Linux kernel
  - organized as a hash table with linked lists

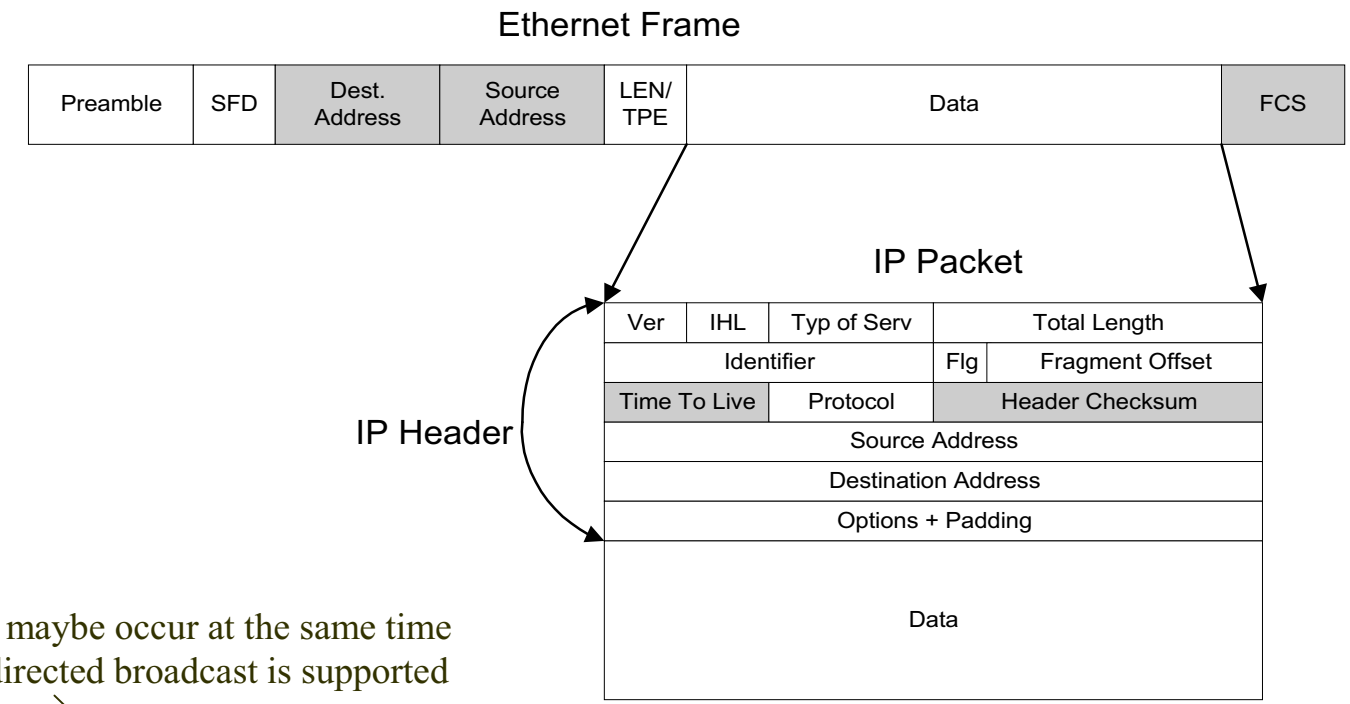


User-Plane : Table Lookup (2/2)

- **Routing table in phase-2 router code**
  - organized as hash table with trees
- **Methodology**
  - **Hash to each tree by IP mask**
  - **Binary search with IP address**
  - **Not Found : Search another tree via forward pointer**



User-Plane : Packet Modification



These two maybe occur at the same time if subnet directed broadcast is supported

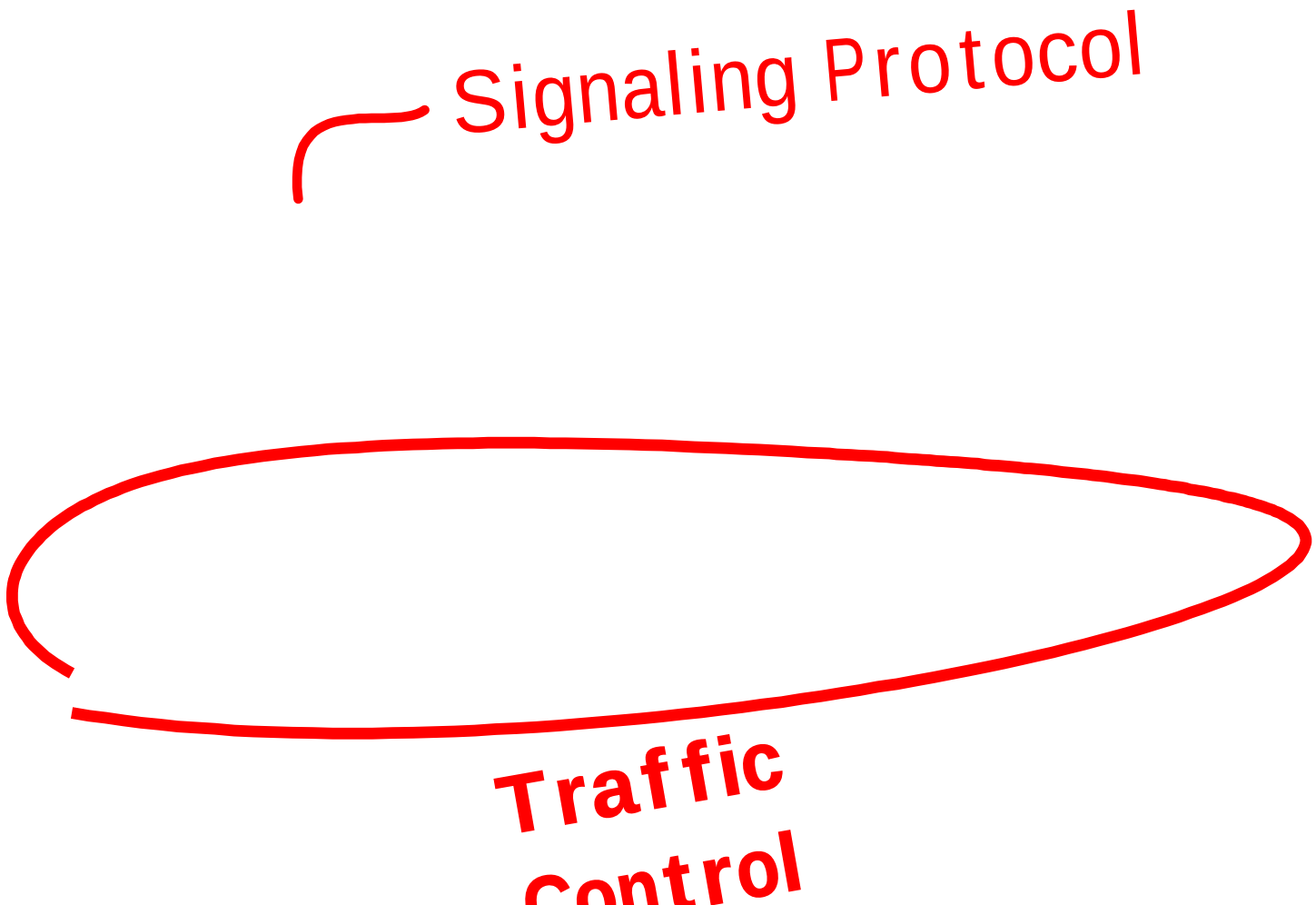
These two maybe occur at the same time in a multi-layer switch

Packet modification summary

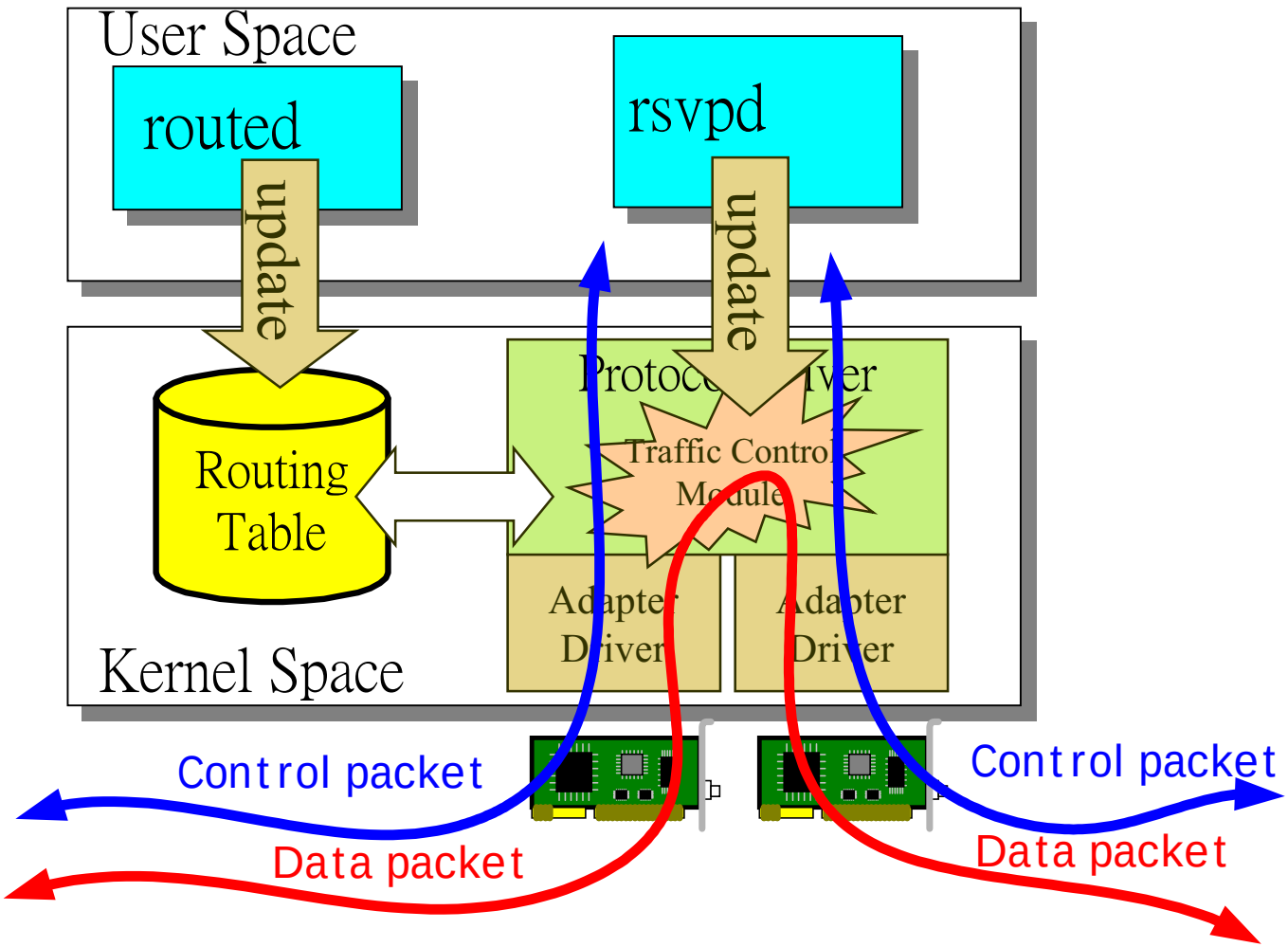
	MAC DA	MAC SA	TTL	Checksum	CRC (Org. Vtag)	CRC(Vtag Changed)
Same subnet	•	•	•	•	•	Recalculate
L3 unicast	Next hop	Router	Decrement	Recalculate	Recalculate	Recalculate
L3 subnet directed BC	•	Router	Decrement	Recalculate	Recalculate	Recalculate
L3 multicast	•	Router	Decrement	Recalculate	Recalculate	Recalculate

New Feature : QoS

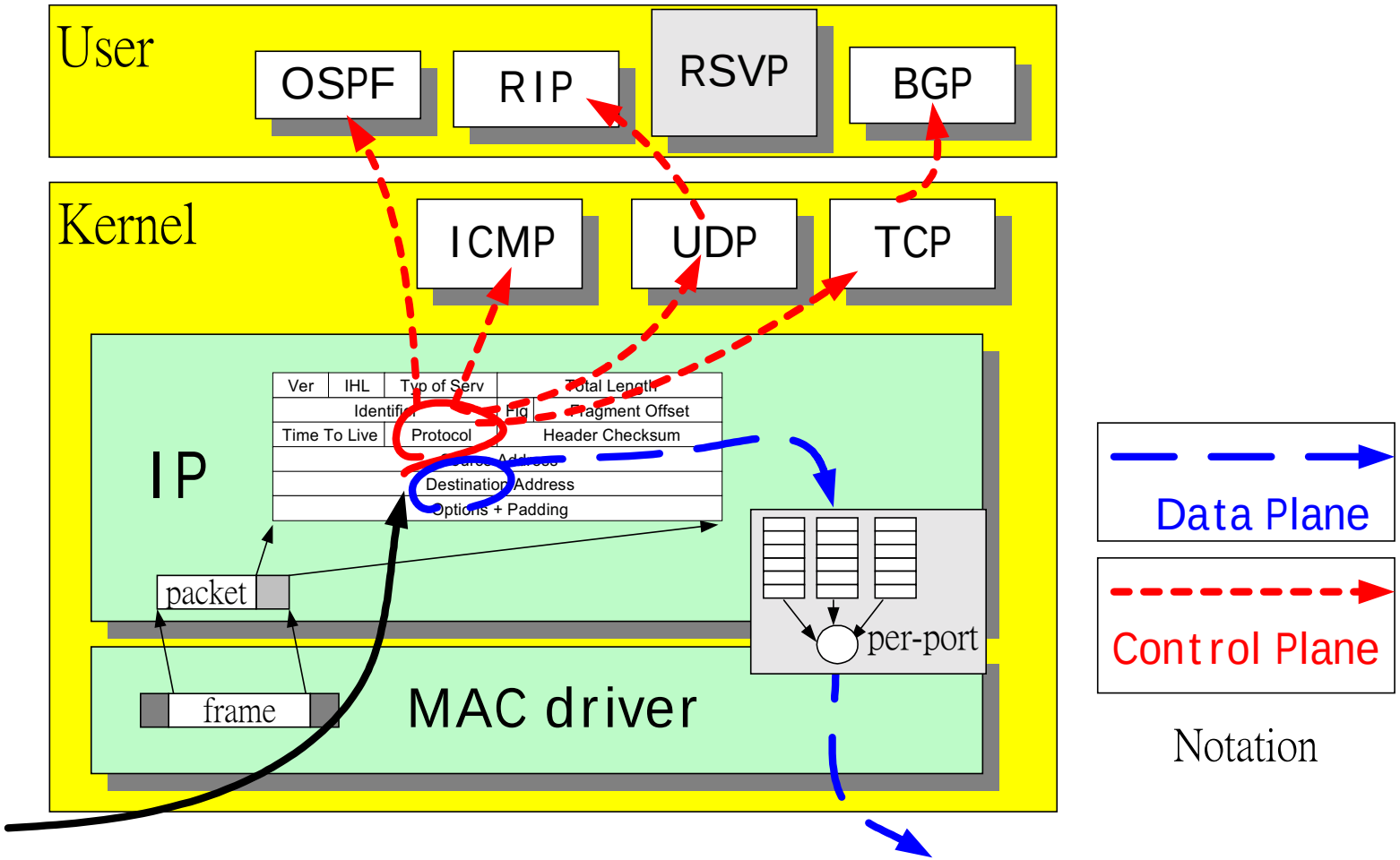
- InterServ: RSVP



QoS Modules & Daemons



Packet Flows - User/Control Plane



New Control-Plane Module

- rsvpd
  - by Information Sciences Institute (ISI)
    - Use CBQ as traffic scheduler
    - link aggregation for CL service
    - no traffic control modules
  - patched by Alexey Kuznetsov
    - traffic control function:

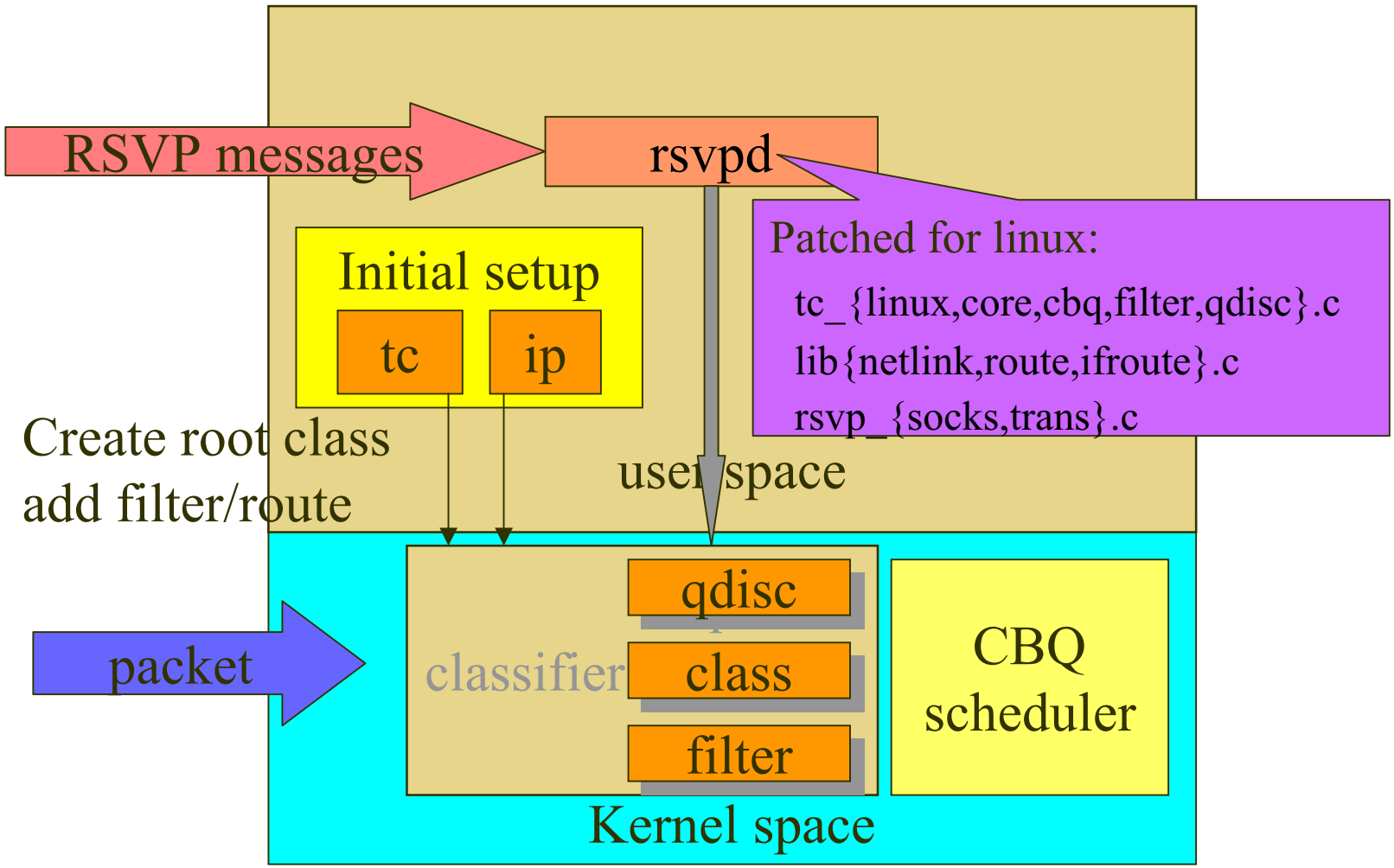
TC_AddFlowspec()	TC_AddFilter()
TC_ModFlowspec()	TC_DelFilter()
TC_DelFlowspec()	TC_Advertise()

Needs admission control to admit

New User-Plane Modules

- **Scheduler:**
  - CBQ **(Class-based Queuing)**
  - CSZ **(Clark-Shanker-Zhang)**
  - PRIO **(n-band priority queue)**
- **Rate estimator**
  - a base for statistical multiplexing for CL service
- **Classifier:**
  - Routing table based
  - Firewall based
  - U32

Rsvpd 4.1 & Linux Kernel 2.2



Known Bugs

- rsvpd compilation
  - don't use IPv6
- kernel modules
  - don't support auto-load yet
    - sch_cbq.o
    - cls_u32.o