# Refeshing basics in Networking, IP and routing

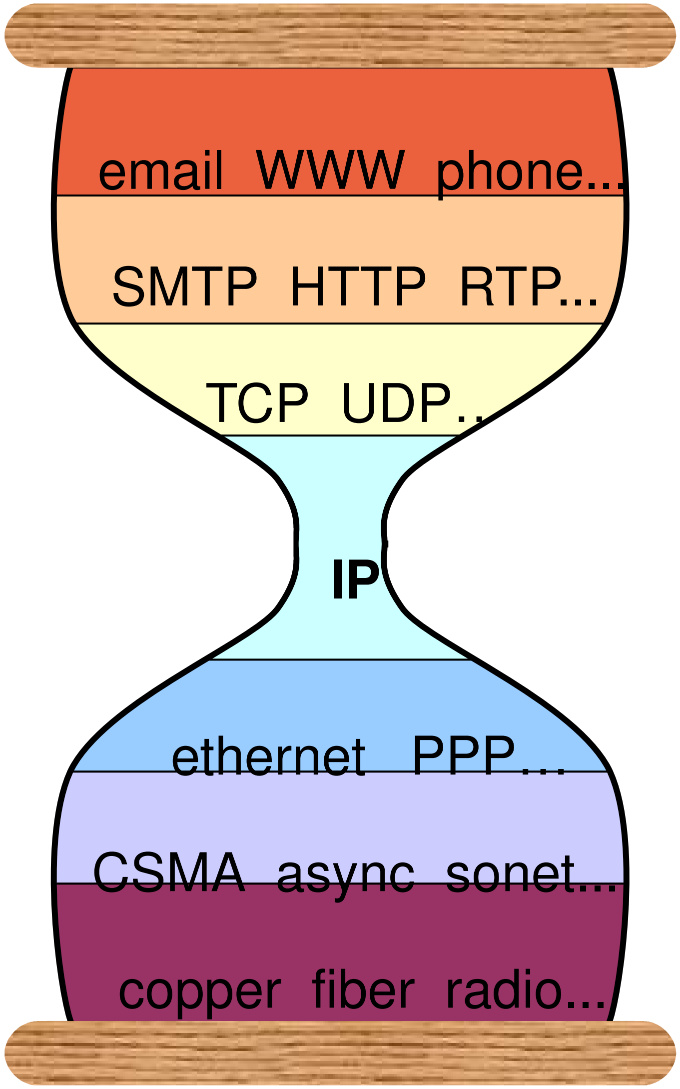
Robert Olsson KTH/CSD

Most slides from:
KTH Computer Science Course

Thanks to: Olof Hagsand KTH/CSC

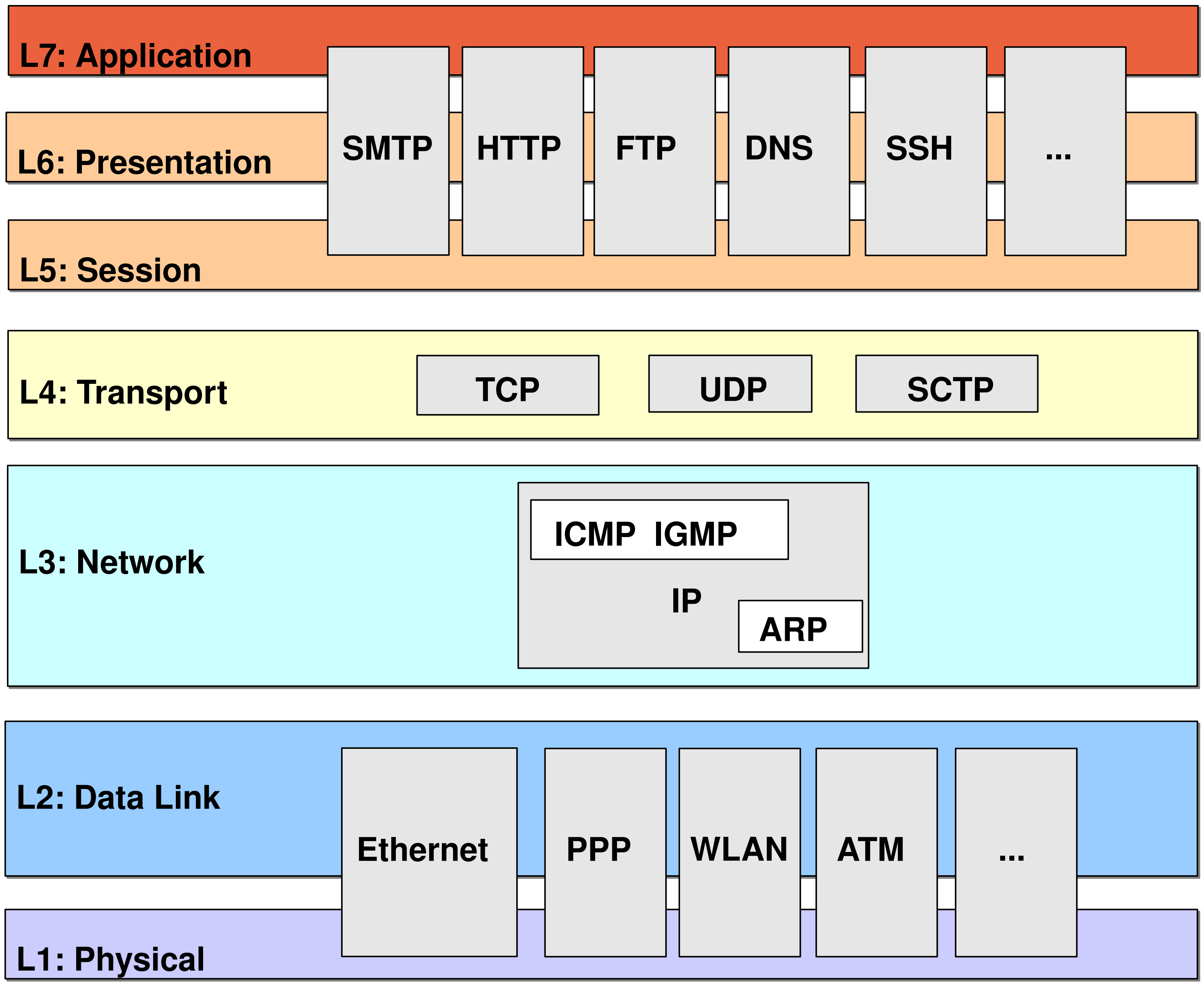
# The Hourglass Model

- Anything over IP – IP over anything
- All applications depend on IP
- IP runs over all networks
- IP is at the heart of all communication

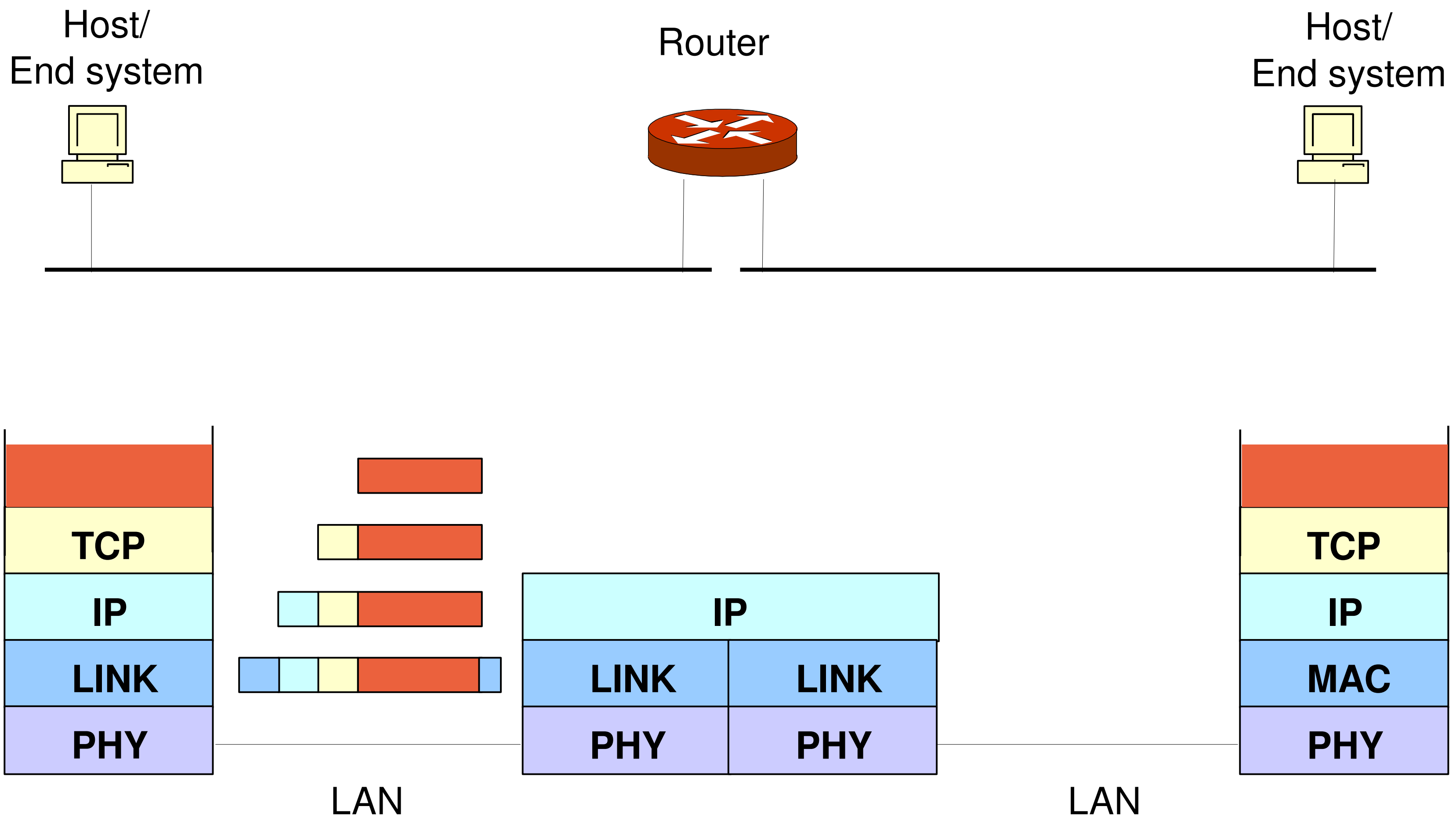



From Steve Deering, 2000

# The TCP/IP stack and OSI ref model



# Layering in TCP/IP

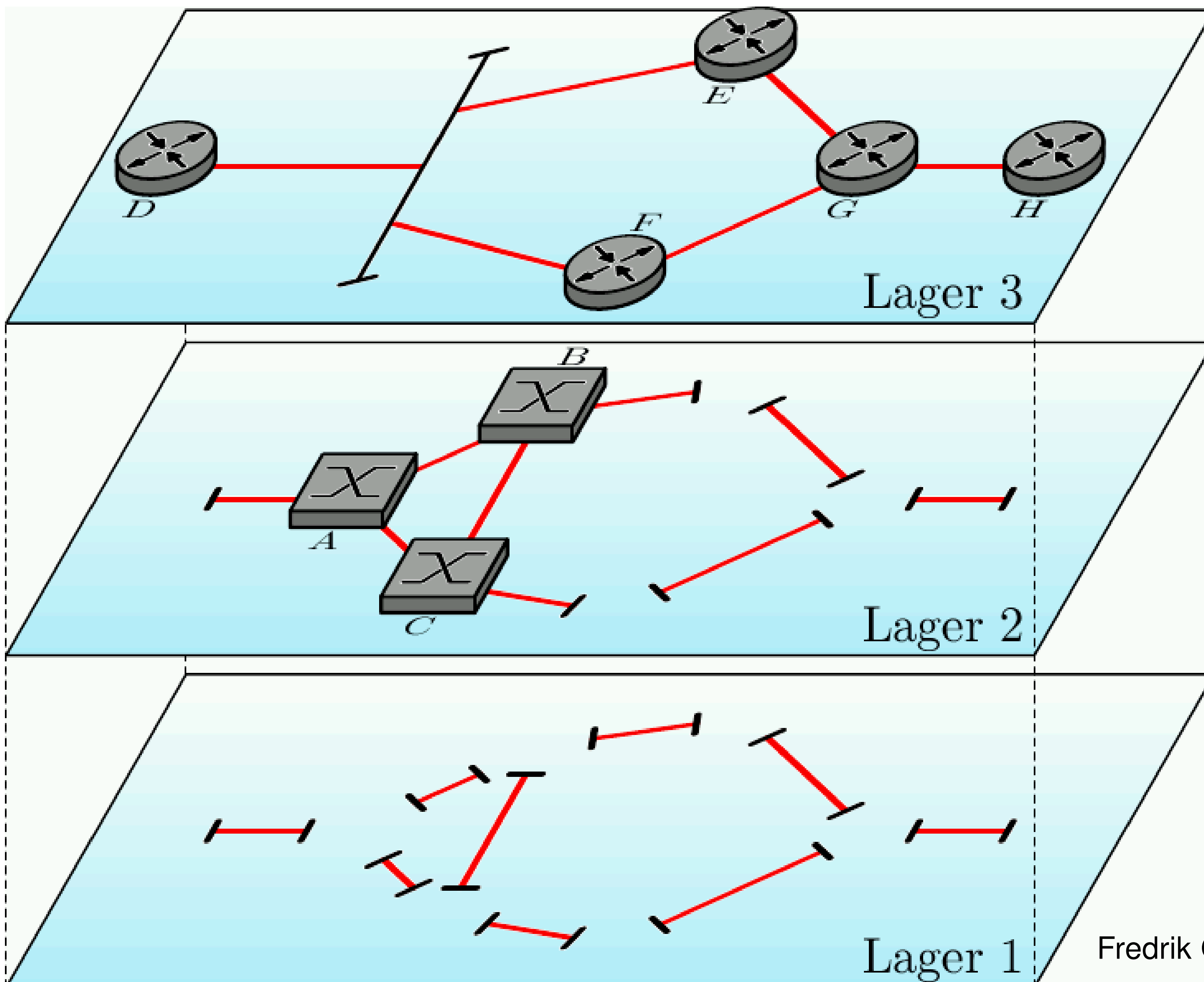


# Layering example

Network

Data Link

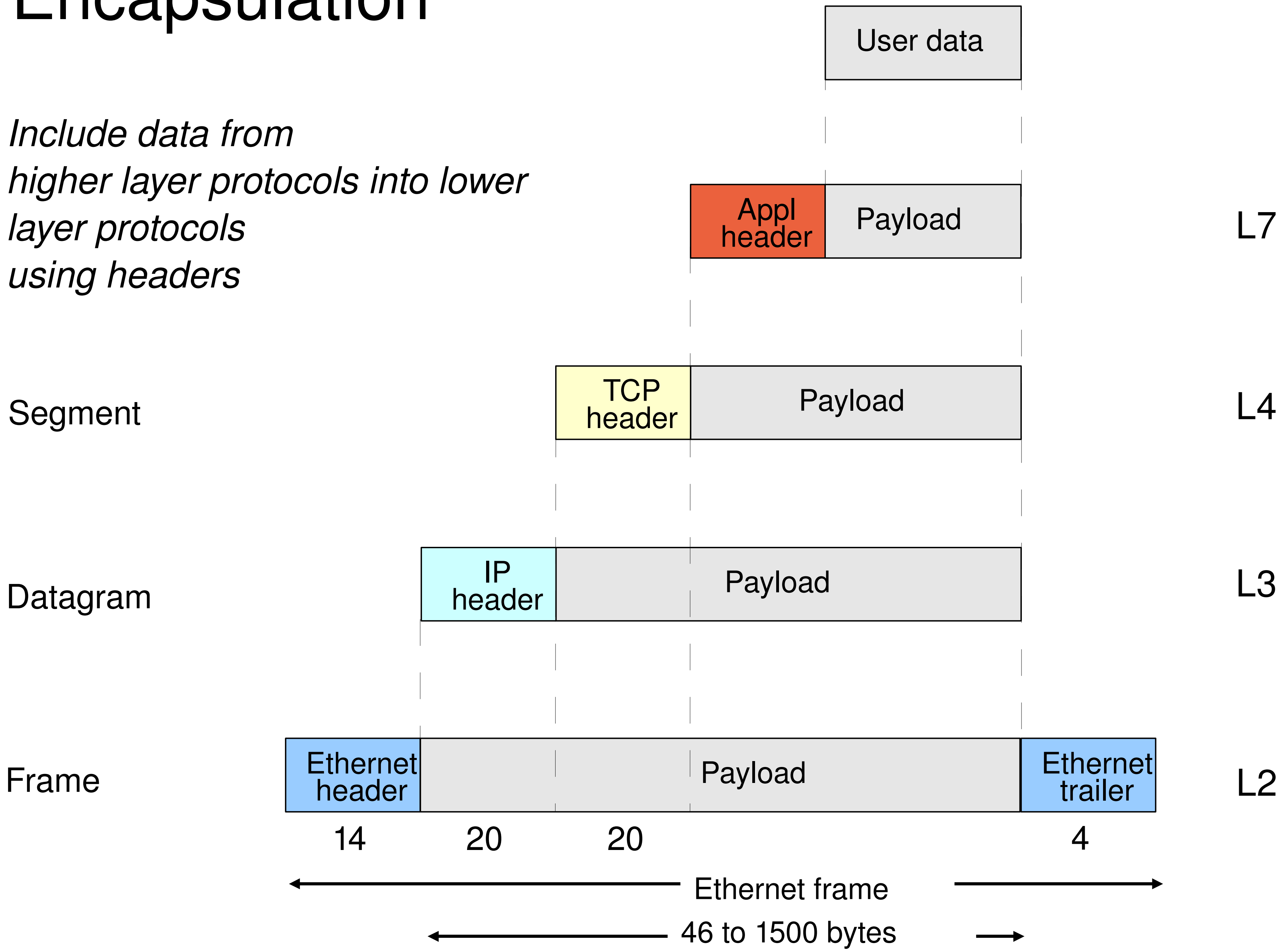
Physical



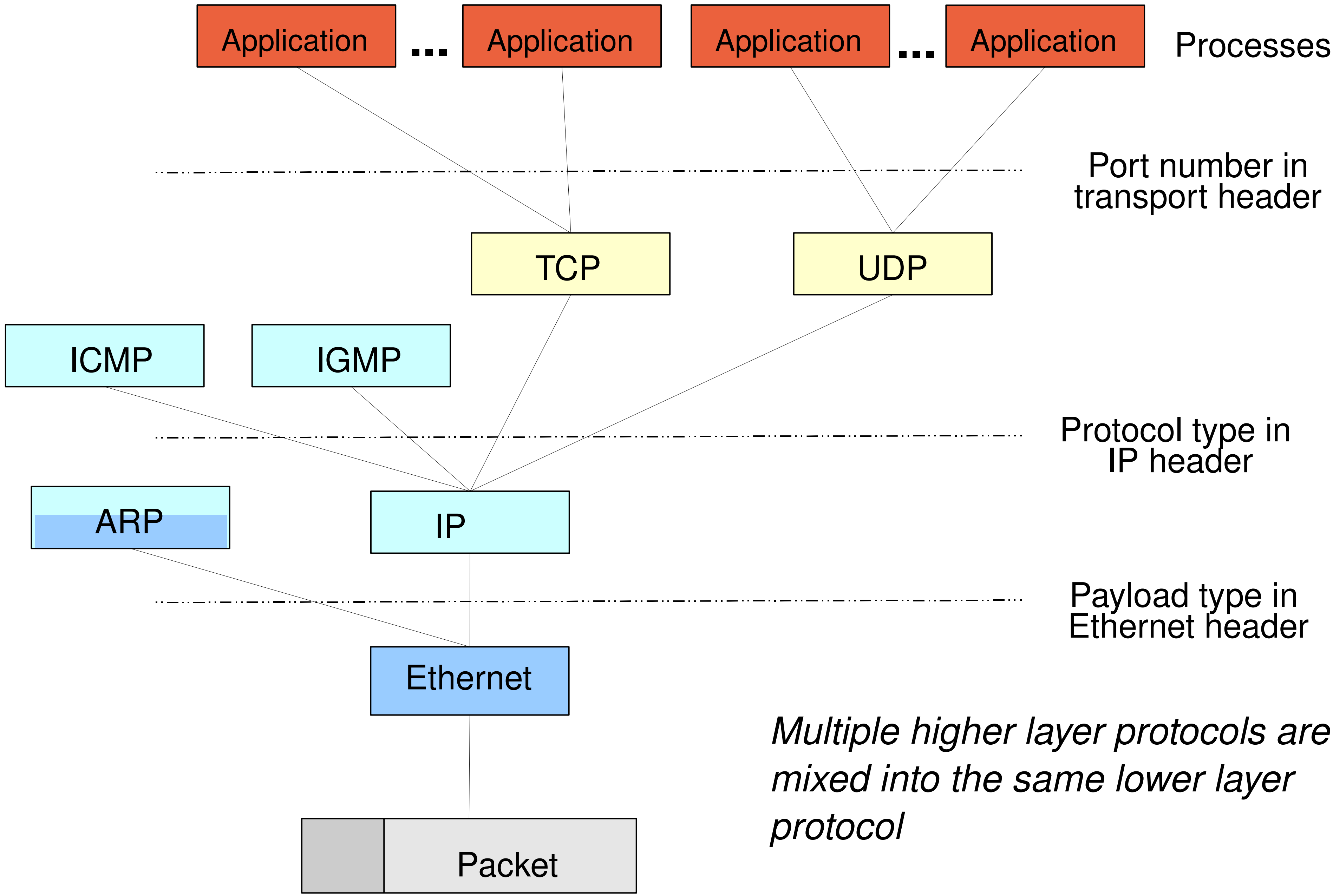

Fredrik Gunger

# Encapsulation

*Include data from higher layer protocols into lower layer protocols using headers*



# Multiplexing and Demultiplexing




*Multiple higher layer protocols are mixed into the same lower layer protocol*

# The End2End Argument

*A specific application-level function should not be built into the lower levels of the system.*

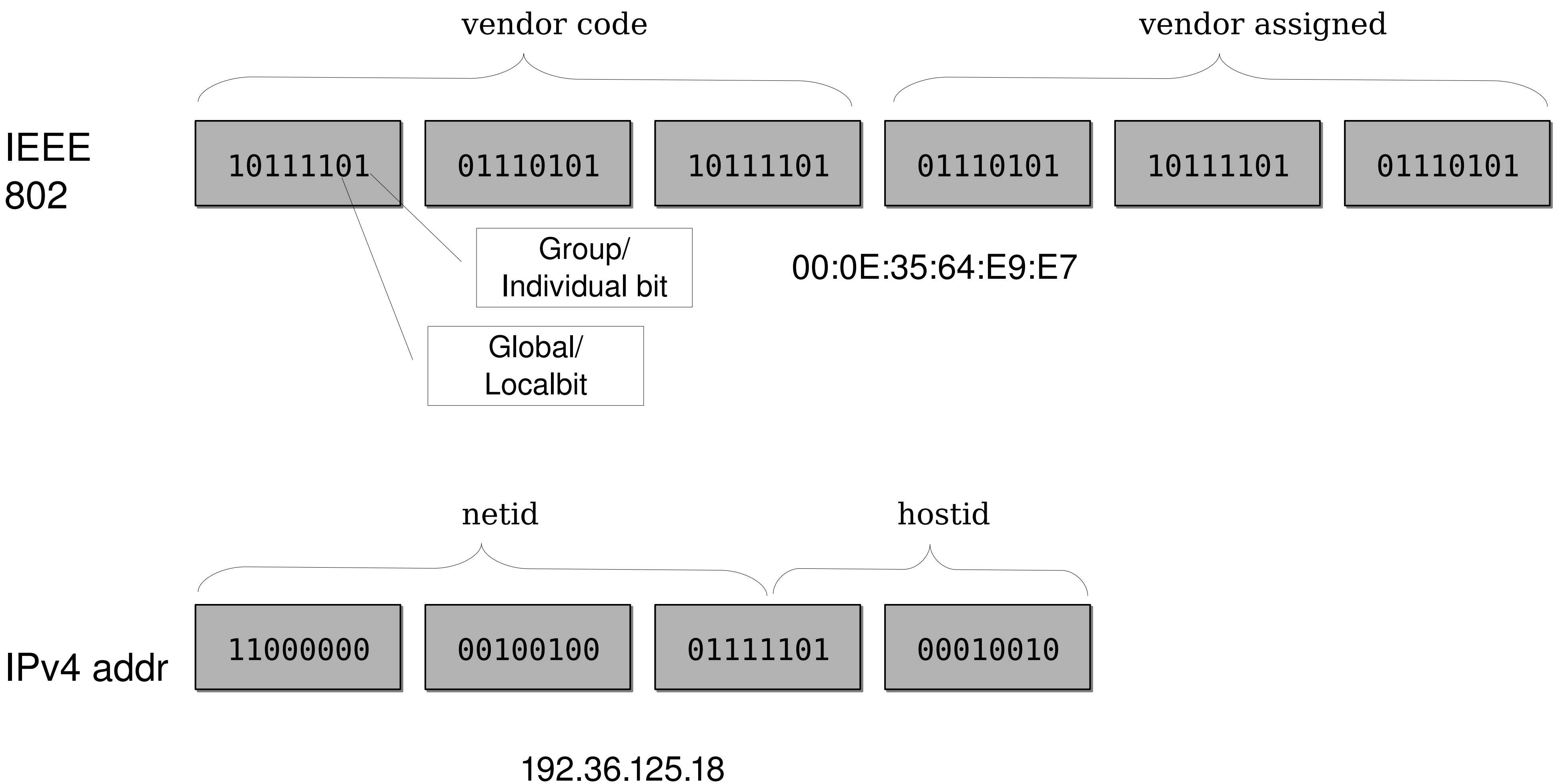
- The functions “in” the Internet are simple and general.
- The bulk of functions are in software at the “edge”.
- The complexity of the core network is reduced.
- Generality in the network increases the chances that new applications can be added.

*Saltzer, Reed, Clark, 1984, Blomenthal & Clark 2001*

# Standards: Organizations

- Internet standard groups
  - ISOC – Internet Society
    - IAB – Internet Architecture Board
    - IETF – Internet Engineering Task Force
  - IANA – Internet Assigned Numbers Authority
    - ICANN – Internet Corporation for Assigned Names and Numbers
- Related
  - ISO – International Standards Organisation
  - IEEE – Institute of Electrical and Electronics Engineers
  - ITU-T – International Telecommunications Union – Telekom sector
  - W3C – World-Wide-Web Consortium
  - ...

# IEEE 802 vs IPv4 addresses



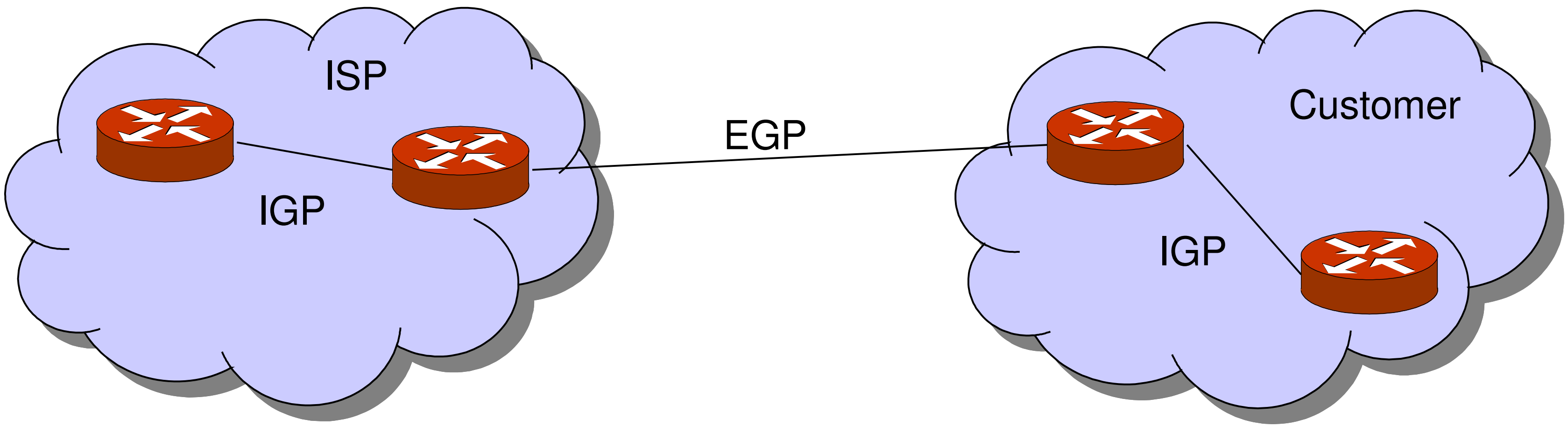
# IP-numbers (private)

```
3. Private Address Space. (RFC1918)

The Internet Assigned Numbers Authority (IANA) has
reserved:

     10.0.0.0        -   10.255.255.255  (10/8 prefix)
     172.16.0.0      -   172.31.255.255  (172.16/12 prefix)
     192.168.0.0     -   192.168.255.255 (192.168/16 prefix)



Used for NAT and experiments
```

# Internal vs external routing

- The Internet is huge
  - Necessary to divide the routing problem into sub-problems.
  - The “Internet” is divided into Autonomous systems (ASs)
  - Each AS is independently managed
- Inter-domain routing / External routing
  - Routing between AS:s
  - Based on commercial agreements – Policies, Service-level-agreements
- Intra-domain routing / Internal routing
  - Routing inside an AS
  - An AS may be further divided into *areas*
  - Best path based on hop metric
- Static vs Dynamic routing

# IGP/EGP



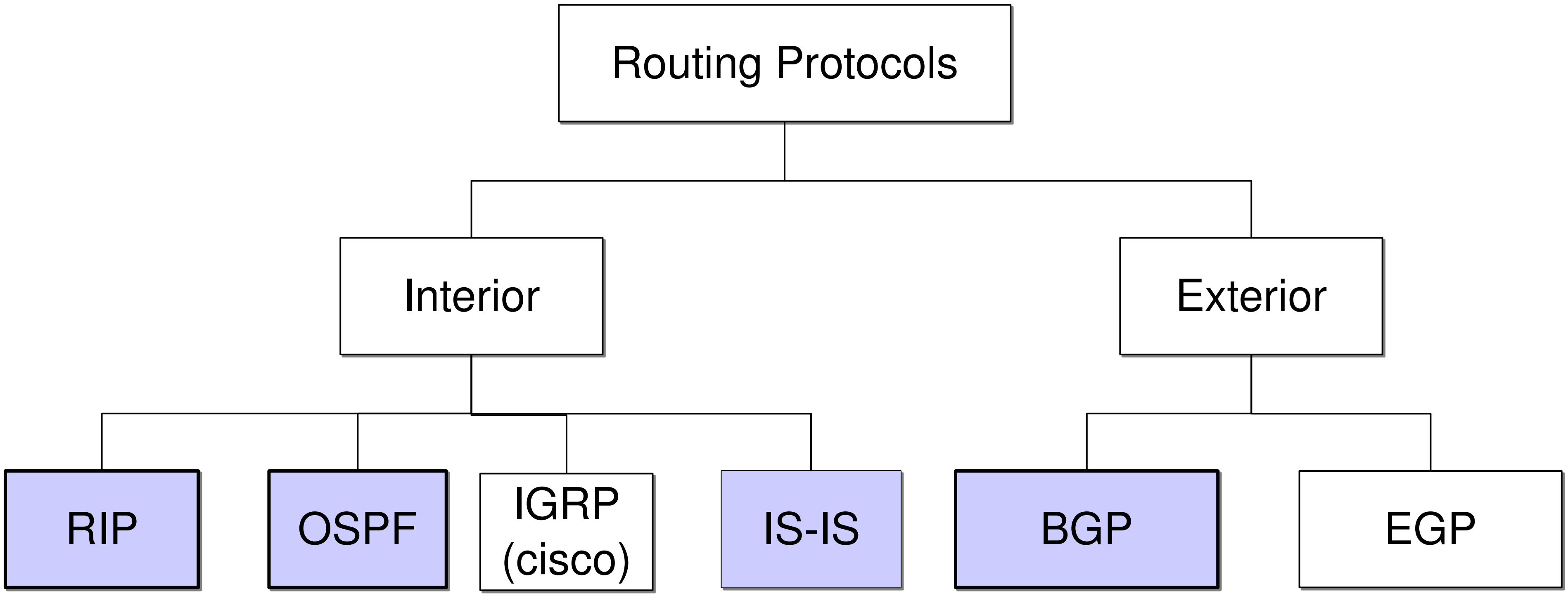

EGP

- Exterior Gateway Protocol.
- Runs between networks/domains (inter-domain)
- Examples: BGP, static routing

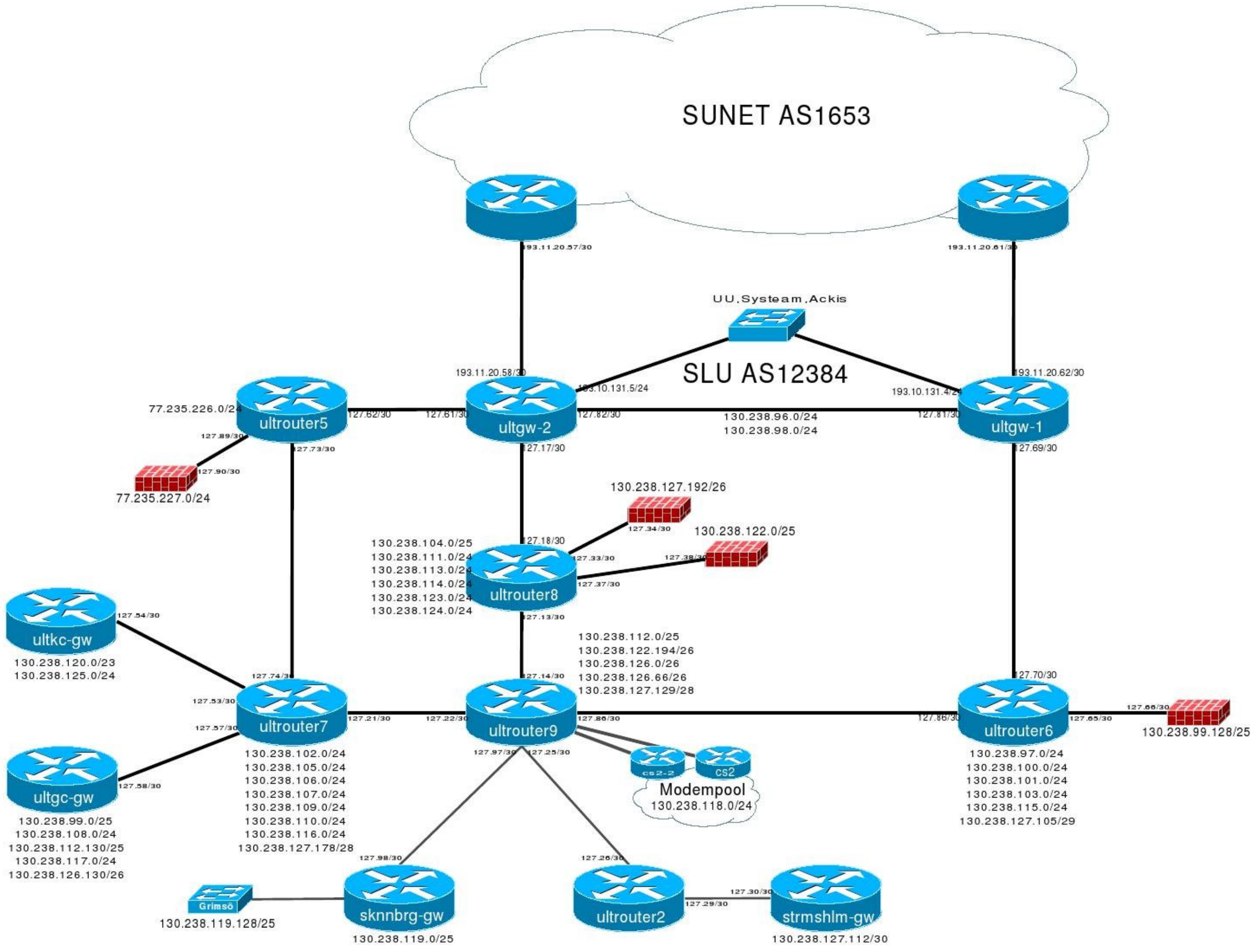
IGP

- Interior Gateway Protocol.
- Runs within a network/domain (intra-domain)
- Examples: RIP, OSPF, IS-IS.

# Popular Unicast Routing Protocols



# IP topology example



# Routers: Cisco CRS-1

CISCO's current flagship:
  Carrier- Routing System

3-stage multi-stage switching plane

  >50% of cost

Trie prefix lookup

7.5kW

Each slot has 40Gbps

32Tbps raw bandwidth

Distributed RP

Several Logical Routers

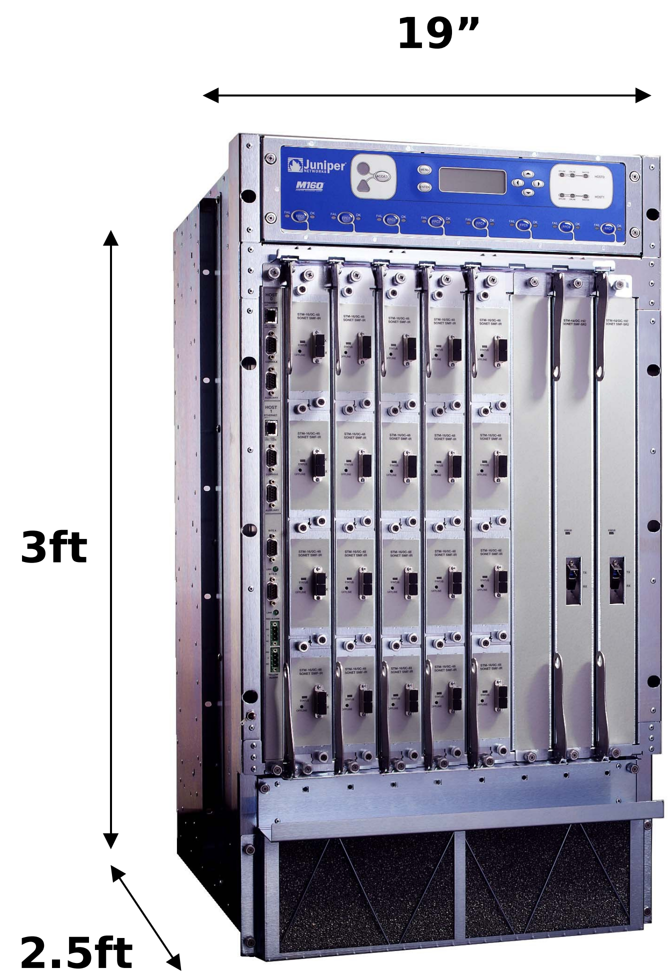
Optical_Electric transitions:
  O-E-O-E-O-E-O

# Routers: Juniper

- M-series
  - Shipping started 1998
  - M5, M10, M20, M40e, M160, M320
  - 8xOC-192 or 32xOC-48 ports in a M160

- T-series
  - Shipping started 2002
  - T320, T640
  - 32xOC-192 or 128xOC-48 ports in a T640

**Capacity: 80Gb/s**
**Power: 2.6kW**




**Juniper M160**

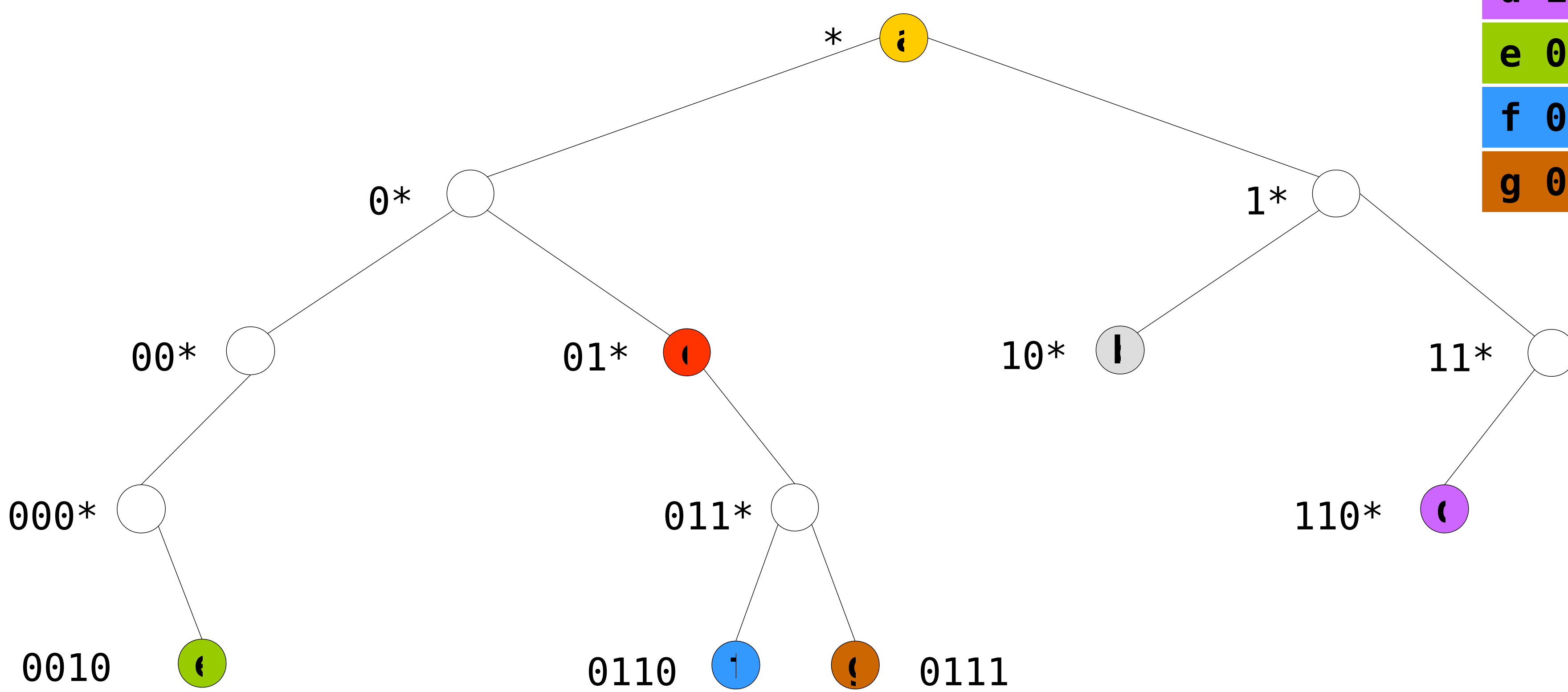
# Routers: Open source

- Linux, BSD platforms
- Most routing protocols exist as open source projects
- But PC HW has traditionally been a limiting factor
- Now quad core CPUs, new buses (PCI express), 10Gbps NICs enables gigabit forwarding speeds.
- Example: the Bifrost open source router (UU/KTH)

# Linear Search on Length Using a Trie

- Binary tree
  - Nodes are prefixes
  - Left branch represents ´0´in the string
  - Right branch represents ´1´

| | |
|---|---|
| a | * |
| b | 10* |
| c | 01* |
| d | 110* |
| e | 0010 |
| f | 0110 |
| g | 0111 |



# LC-trie routing lookup

- Routings algorithms a high researched area

- LC trie contribution by Stefan Nilsson, Gunnar Karlsson @ KTH

*

- Further refined and first implemented and for Linux kernel

  Robert Olsson, Hans Liss, Jens låås, Uppsala University and Swedish

  university of agricultural sciences. About 3 man-years and lots of support

  for Linux Networking Team

- Got Intel Academic Award 2005

# What is BGP?

- Border Gateway Protocol version 4
- Defined in RFC 4271
- An inter-domain routing protocol
- Uses the *destination-based* forwarding paradigm
  - No other relations can be expressed: sources, tos, link load
- Uses *path-vector* routing
- Views the Internet as a collection autonomous systems
- Exchanges information between *peers* using TCP as underlying protocol
- Maintains a database (RIBs) of *network layer reachability information* (NLRI:s)
- Supports a toolkit of mechanisms to express and enforce policies decisions at the AS level

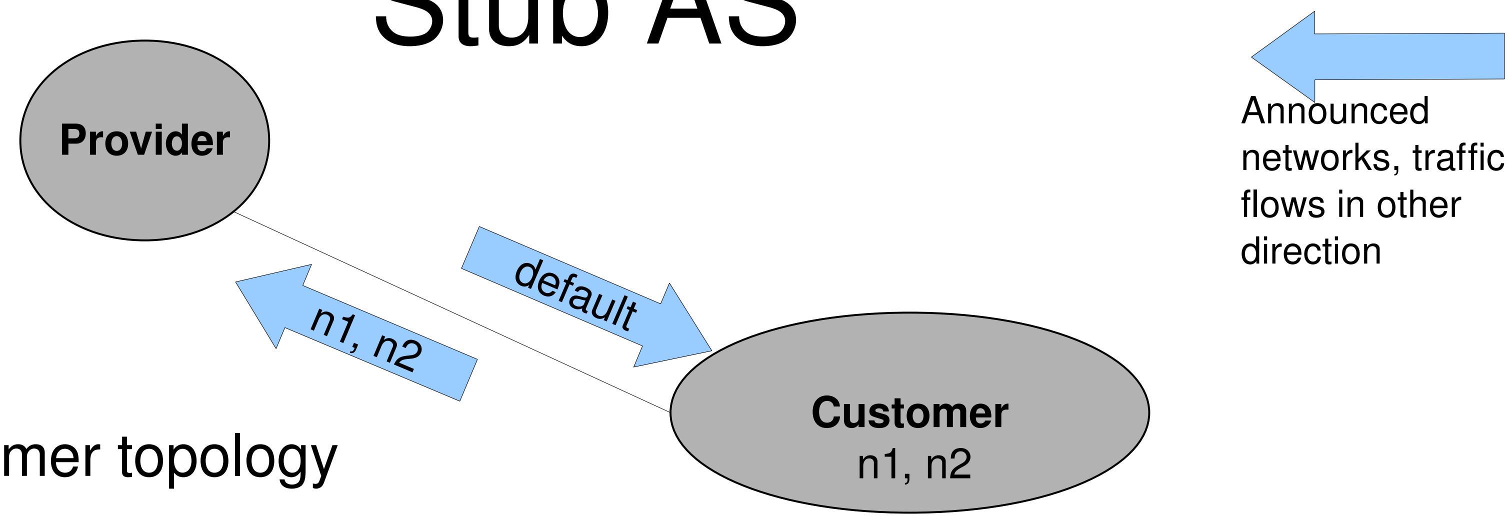
# The routing table

- Currently, backbone IP tables are around 300000 entries.
  - The RIB may be much larger
- Virtual private networks (many customer routing tables) the tables are even larger
- Also, a “routing table” is actually many data-structures:
  - Many different protocols
  - Forwarding information base (FIBs)
  - Routing information base (RIBs)

# Autonomous Systems (AS)

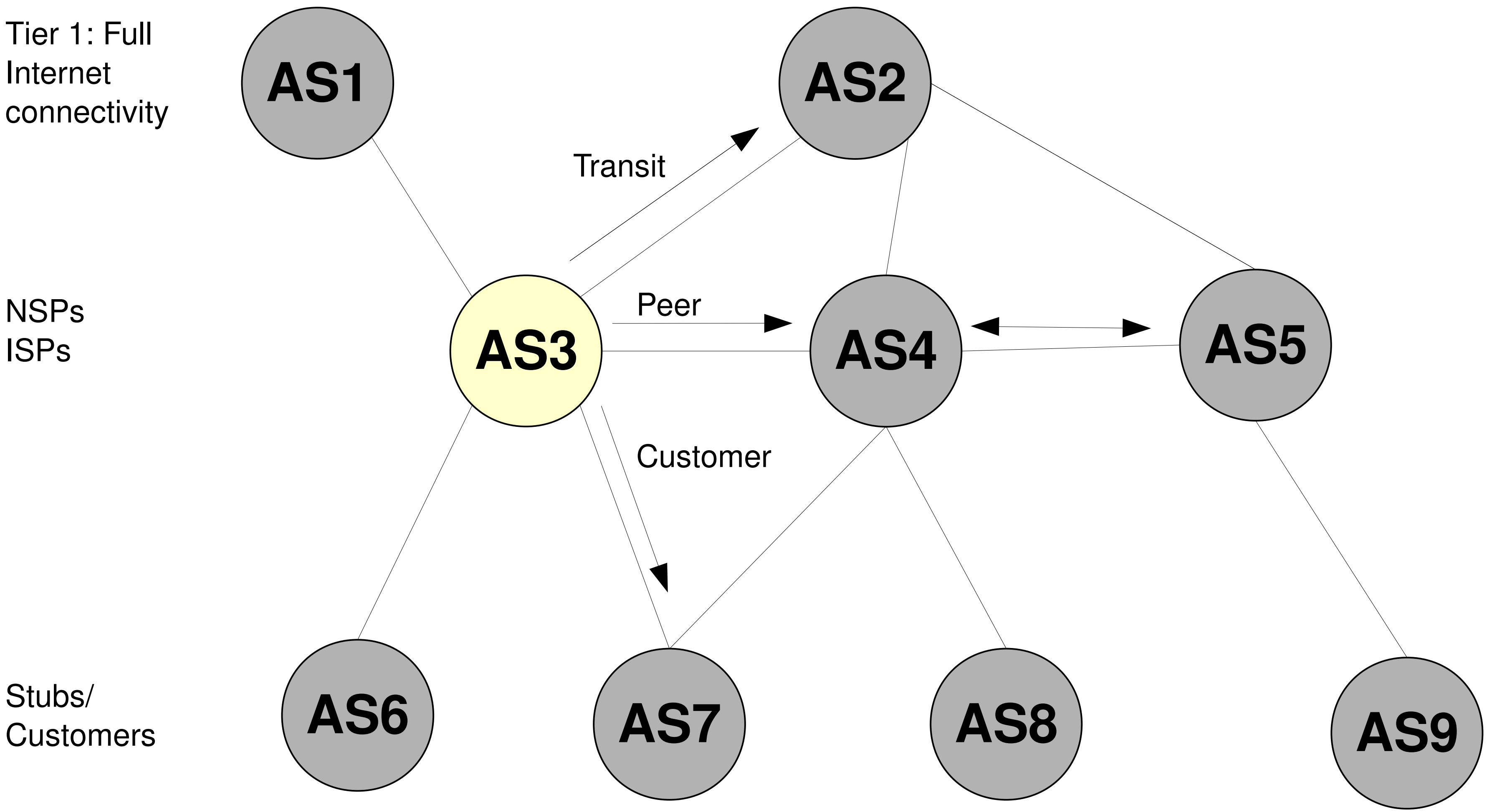
- A set of routers that has a single routing policy, that run under a single technical administration
  - A single network or group of networks
  - University, business, organization, operator
- This is viewed by the outside world as an Autonomous System
  - All interior policies, protocols, etc are hidden within the AS
- Represented in the Internet by an Autonomous System Number (ASN). 0-65535
  - Example: ASN 1653 for SUNET
  - Note: RFC 4893: BGP Support for Four-octet AS Number Space

# Customer / ISP Relations: Stub AS

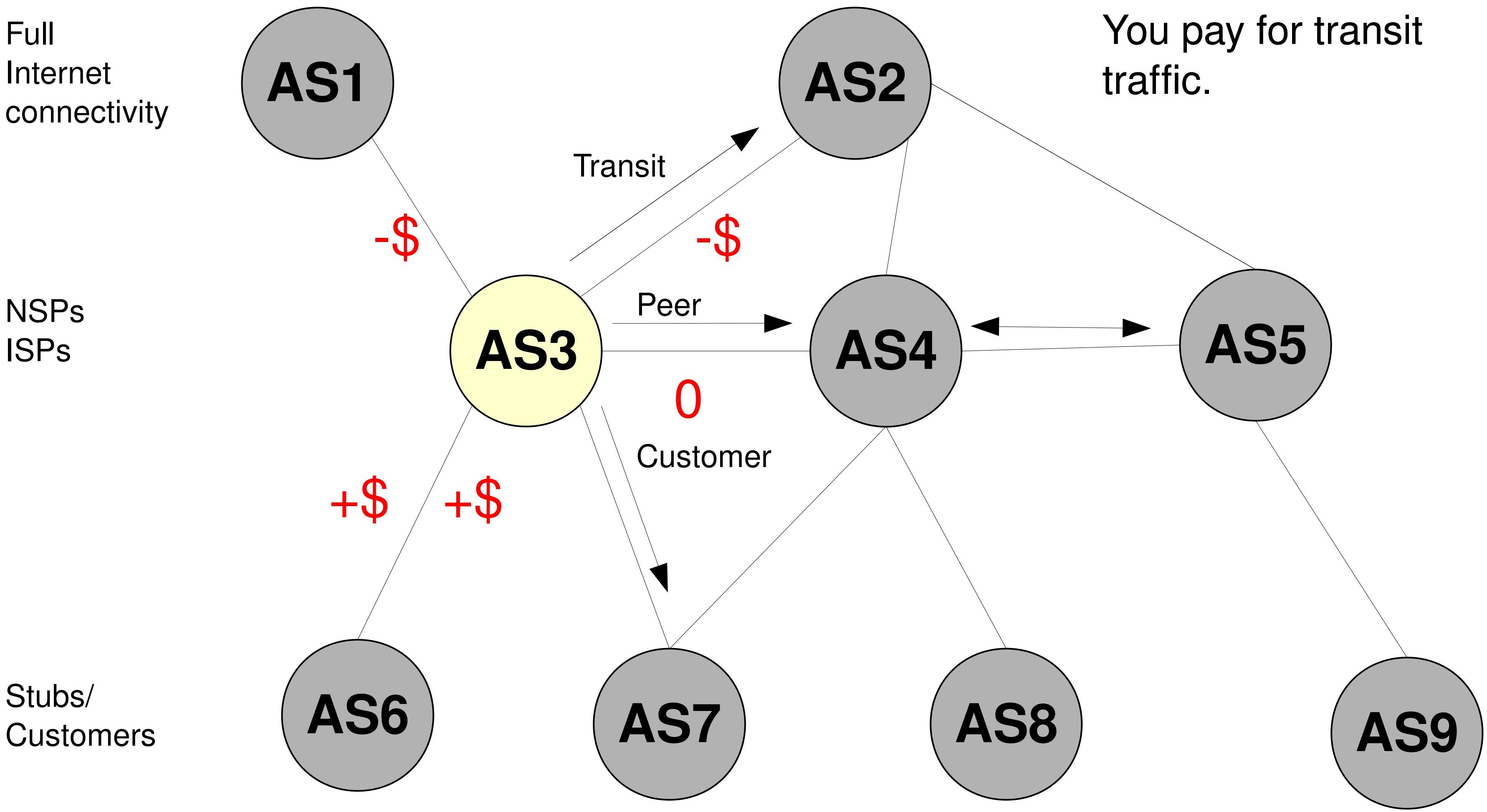



- Typical customer topology
- Can use default route to reach the Provider and Internet
- Customer can use address block of provider
- Customer does not need to be a separate AS
- Typically use static routing but can also use BGP
  - Less common. Use a separate IGP (eg RIP) only to exchange routes between border routers.

# AS graph and peering relations

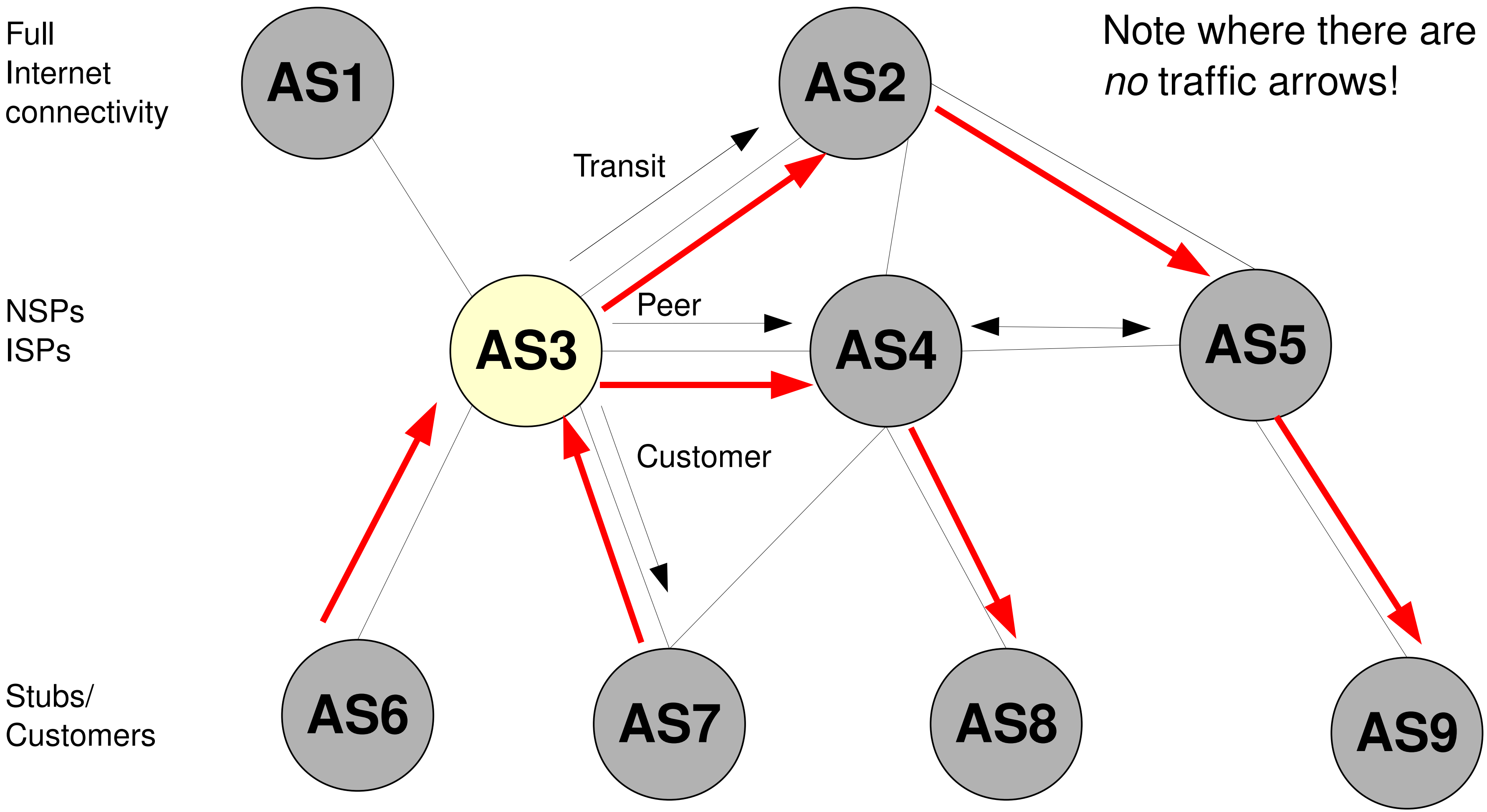


# Cost and peering relations




You pay for transit traffic.

# Traffic patterns



# IP resources

- IP numbers Ipv4 (exhausted when?)
- IP numbers IPv6
- AS-numbers 16 vs 32 bit
    - LIR, AfriNIC or through provider.

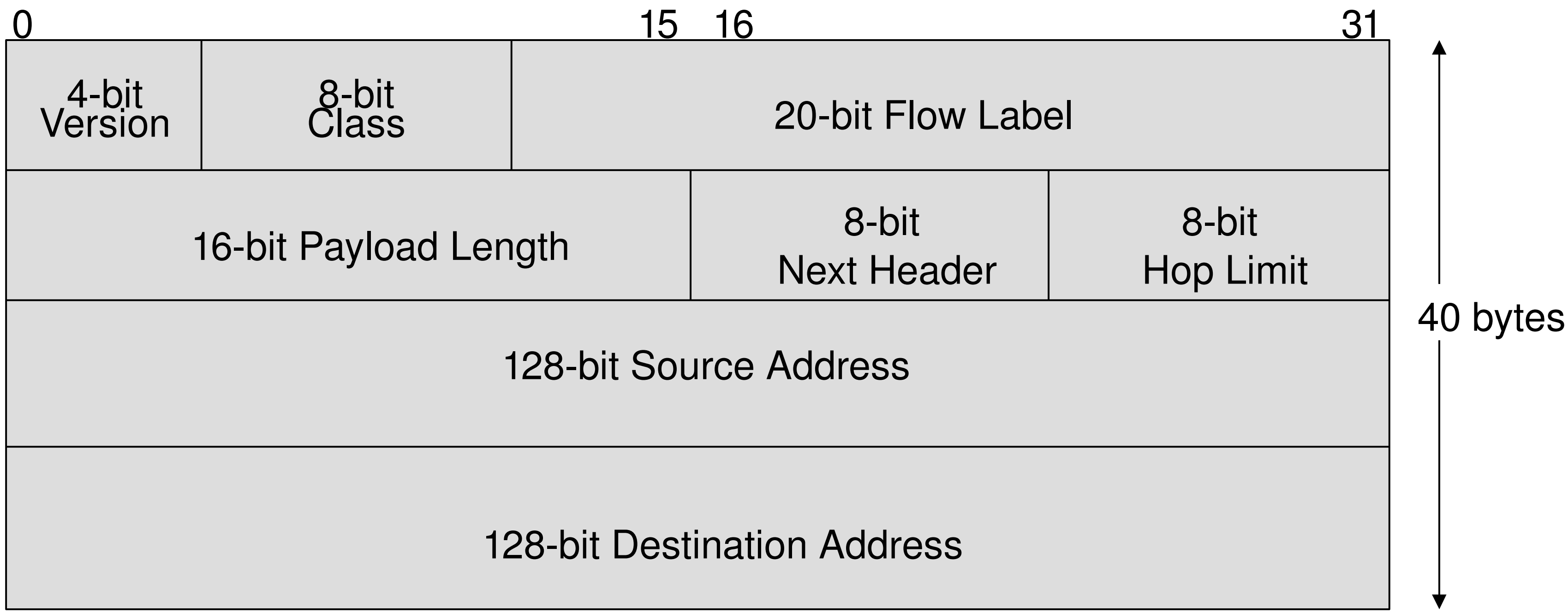
# IPv6: What drives deployment?

- Asia and Europe
  - Smaller pools of IPv4 addresses
  - Faster at adopting new technology
  - Government-driven (ASIA)
  - Wireless (3G in Europe)
- U.S.
  - DoD announced that it will move to IPv6 by 2008
    - Public address assignment simplifies end-to-end security
- IPv6 has been added to  DNS root servers
- SUNET and NorduNET runs IPv6 in core

# IPv6 vs IPv4

- Changes in IPv6 compared to IPv4
  - 128 bit addresses
  - extended address hierarchy
  - simplified header
  - simpler and better support for options
  - possible to extend the protocol
  - support for auto-configuration (plug-and-play)
  - support for QoS treatment
  - host mobility
  - security
  - provider selection
  - no fragmentation in routers

# IPv6 Header Format




- **Version** Only field identical to IPv4. Code is 6 in IPv6
- **Class** New field. Revised concept of priority bits. Facilitates handling of real-time traffic.
- **Flow Label** New field. To distinguish packets requiring the same treatment.
- **Payload Length** Replaces *length* field in IPv4. Gives length of data following IPv6 header
- **Next Header** Replaces *protocol* field in IPv4. Extension headers can be used.
- **Hop Limit** Replaces *TTL* field in IPv4. Hop limit more accurately reflects the use of TTL.
- **Src Address** Revised *source address* field. 128 bits in IPv6 vs 32 bits in IPv4.
- **Dst Address** Revised *destination address* field. 128 bits in IPv6 vs 32 bits in IPv4.

# What wasn't mentioned

- Different services /etc/services
- ICMP Internet Control Messages Protocol
- ARP
- UDP nor TCP
- NAT
- DNS
- Ipsec, SSL
- Filtering

# That's all

Questions?