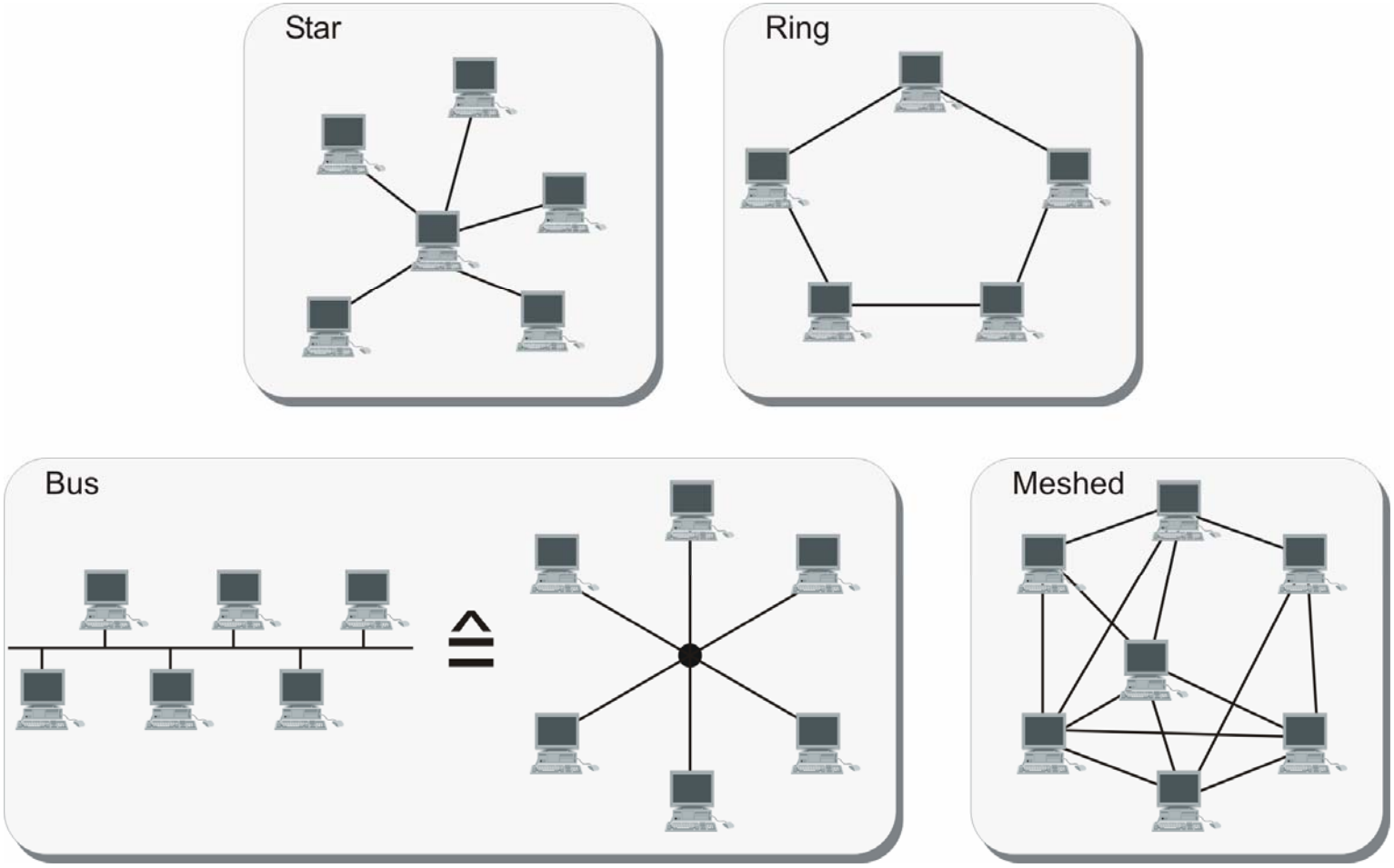
Logical Structure of Networks



Logical Structure of Networks

The logical structure of a network describes in which way the different terminals in a network are connected together.

Star

All terminals are connected to a central node; a direct communication between terminals is not possible. One terminal can operate at a time. Communication control can be managed in a simple way by polling the terminals or through interrupt request by a terminal. Note, in case of central node failure no communication at all is possible.

Ring

There is no central node, each terminal is connected to its two neighbor stations and the information flow passes all terminals. Again only one terminal can operate at a time. In case of one link failure all terminals are still reachable.

Bus

All terminals are connected to one common transmission cable, so the link is shared between all terminals and only one terminal can transmit at a time. In case of a terminal failure the other terminals will not be affected. But in case of a link failure only some of the terminals are still connected together.

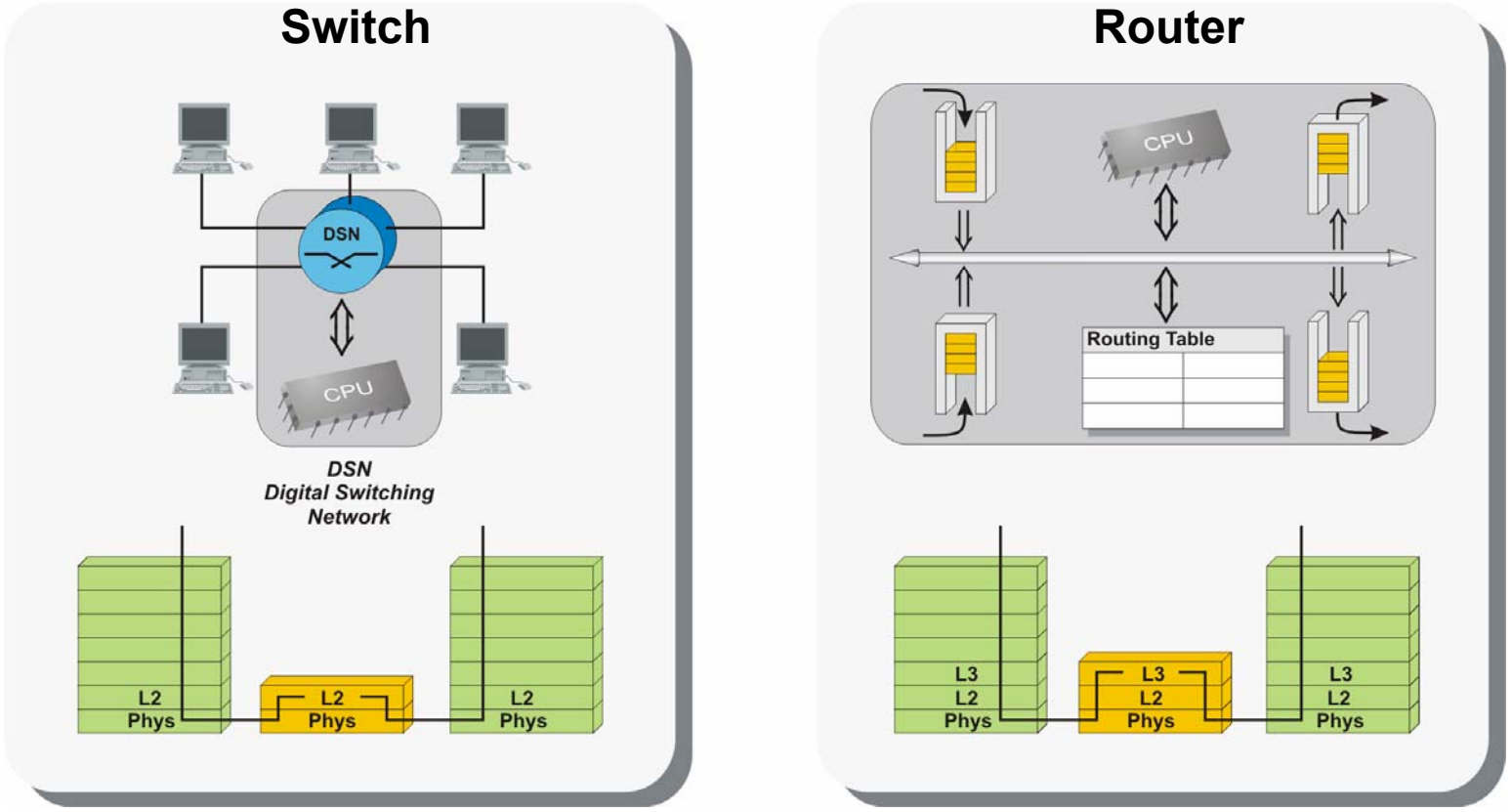
Note, from the physical/cabling point of view a bus structure may look like a star, however it is logical still a bus.

Meshed

Each terminal is connected to several others; no central node exists, but normally several independent transmission links between the terminals. In case no direct connection between two terminals exists the transmission has to pass (multiple) intermediate terminals on the way to its destination.

In real networks several topologies may be combined, e.g. bus and star or bus with other busses which leads to a tree type topology structure. Mainly in Wide Area Networks (WAN) meshed structures can be found where redundant links help to increase network availability.

(2) Network Entities



(2) Network Entities

Switch

From a functional point of view a switch is comparable to a bridge. The difference is that in a switch several connections may be handled in parallel and each connection can handle the full bit rate. Thus switches increase the throughput rate strongly and sometimes a switch is also called multi-port-bridge. A switch is also transparent for all layers above layer 2.
Due to multiple input and output ports the maximum efficiency of a switch requires a suitably network topology. E.g. a local network may be broken into subnets each being connected via a hub to a switch port. So the port capacity of the switch will be shared by only a subgroup of terminals. Other subnets will use other ports with the same throughput capacity. In this way the bus structure of a LAN may be changed into a bus/star structure (bus = subnet; star = hub connections to the switch).
Switches learn step by step which terminals are connected at which port, so they don't need to be configured.

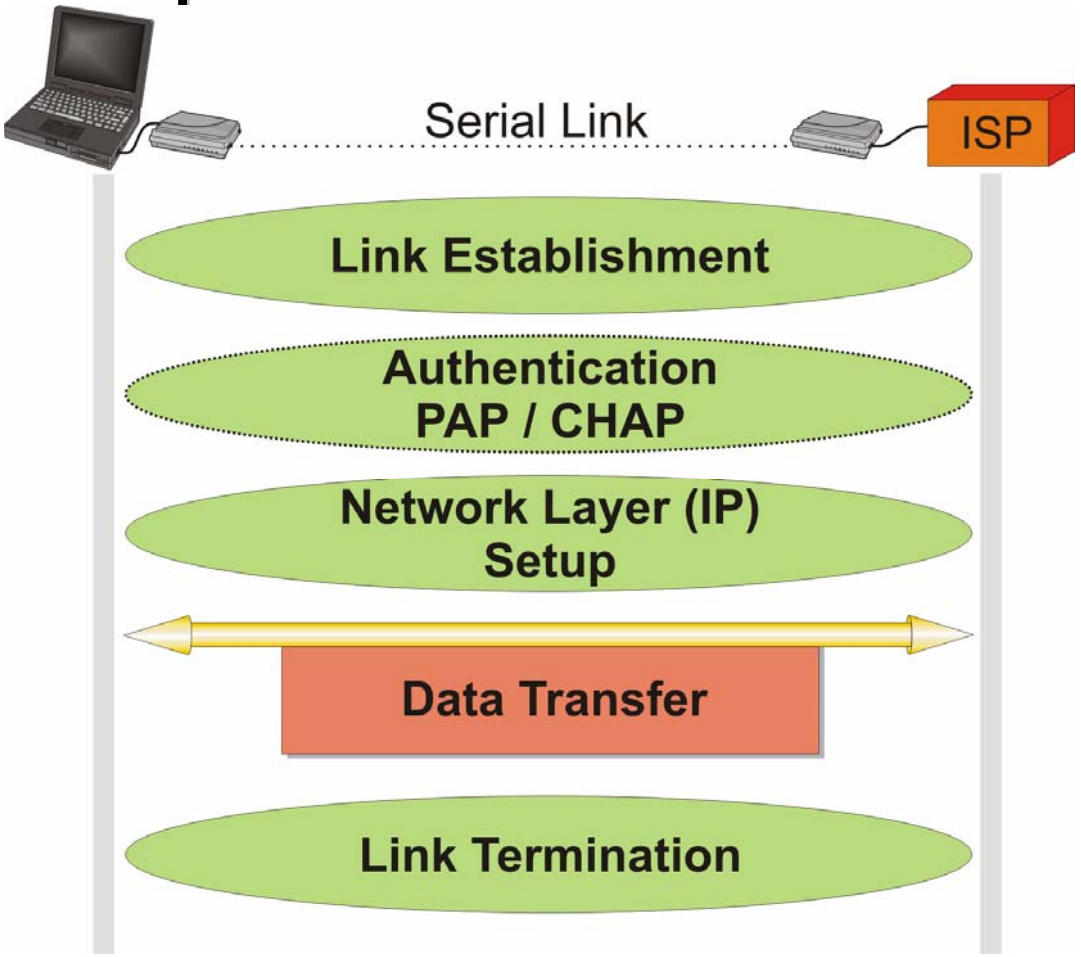
In real system two different switching methods are applied:

- ⇒ Cut-Through (or On The Fly): The switch will not read the whole packet but will start forwarding immediately when the destination address has been read. In this way very short delay times can be realized. However the packet will not be error checked, so errors may be reproduced.
- ⇒ Store and Forward the complete packet has to be received first before error control can start, of course, this causes higher delays.

Router

Routers connect networks with different topologies where the lower protocol layers (layer 2) are different. Routers allow to route different network types and protocols and by this an efficient traffic and network load control is possible. Routers forward data packets based on layer 3 parameters.
In an IP based network routers check the header of every packet for the destination address and decide to which next router the packet has to be sent. Each router contains a routing table which includes the next hop for the data packet. This routing table has to be created first and for this task specific routing protocols are available which will create and maintain the table(s). Of course, routers will not store a complete routing table for the total internet but only the address to the next router (for the next hop) on the way to the destination.
Because the router will interpret the layer 3 (IP) address some processing time is needed which will increase the latency time.
In principle a router contains several network adaptor cards for the different connected network types (e.g. Ethernet, FDDI or ATM), a processing unit and the routing table.
For all higher protocol layers a router is transparent.

Operation of Dial Up Network Access

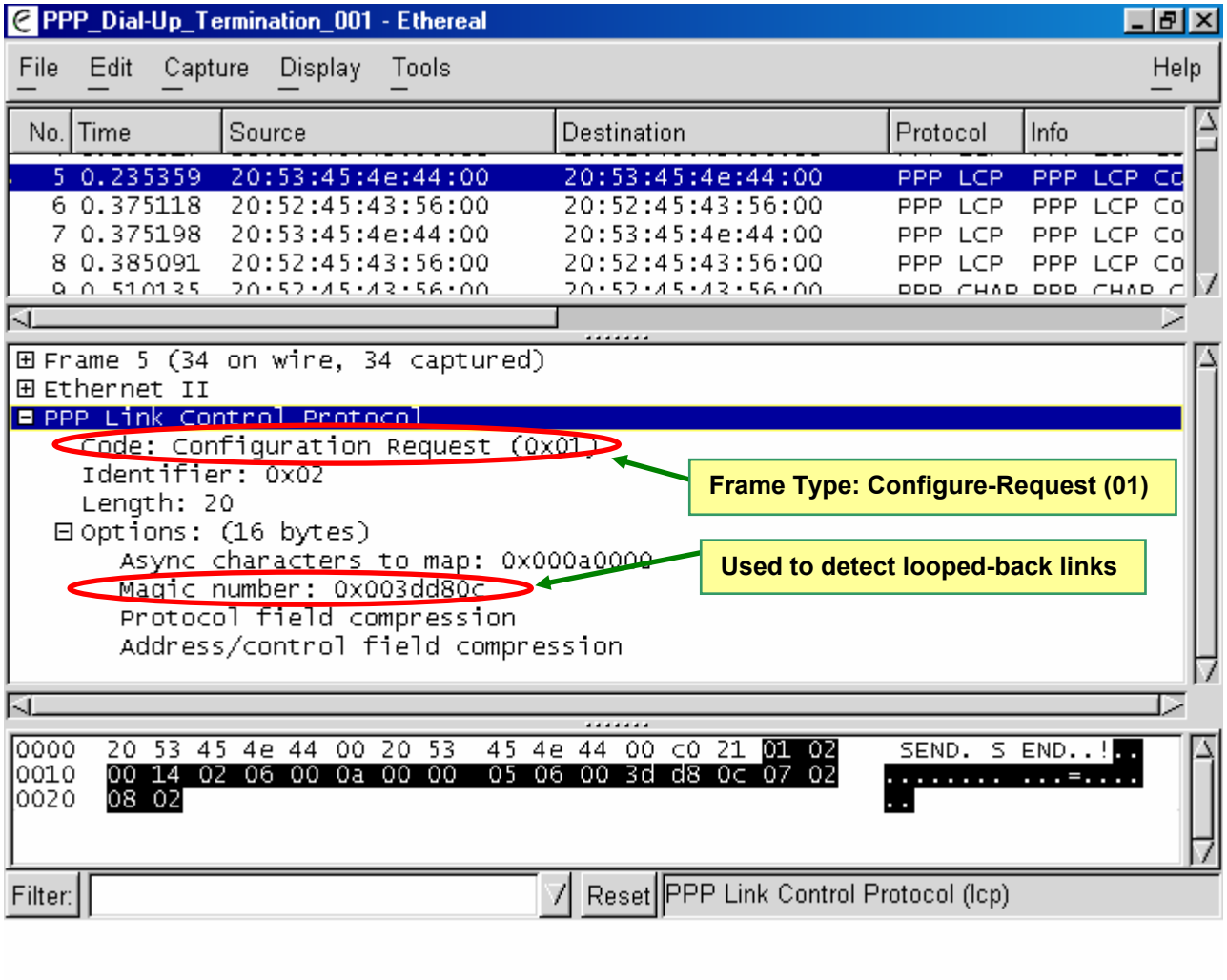


Operation of Dial Up Network Access

Each dial up access using the PPP consists of the following parts:

- **Link Establishment Phase**
  In this phase, the parameters of the connection will be negotiated using the LCP protocol:
  ⇒ The Maximum Receive Unit (MRU) – the maximum size of a data packet that can be processed
  ⇒ Authentication – if authentication shall take place and with which method (PAP / CHAP)
  ⇒ Specifies, if a quality protocol shall be applied
  ⇒ Magic Number to detect loop back links
  ⇒ Compression of PPP fields
- **Authentication (optional)**
  Exchange of username and password in either or both directions. 2 possible methods exist:
  ⇒ Password Authentication Protocol (PAP): Password and Username un-encrypted.
  ⇒ Challenge Authentication Protocol (CHAP): Password and Username are encrypted using challenge handshake between the peers (based on MD5 algorithm)
- **Network Layer Setup Phase for IP**
  Negotiation of further parameters at the network layer. Several control protocols exist related to the different network protocols, PPP could convey. An example is the Internet Protocol Control Protocol (IPCP).
- **Data Transfer Phase**
  HDLC type framing over the data link layer
- **Link Termination Phase**
  Termination of a connection using LCP protocol. It is possible to only disconnect the network layer and keep the link layer open under LCP control.
  PPP does not specify inactivity timeouts, so it is up to provider / system administrator to initiate a termination.

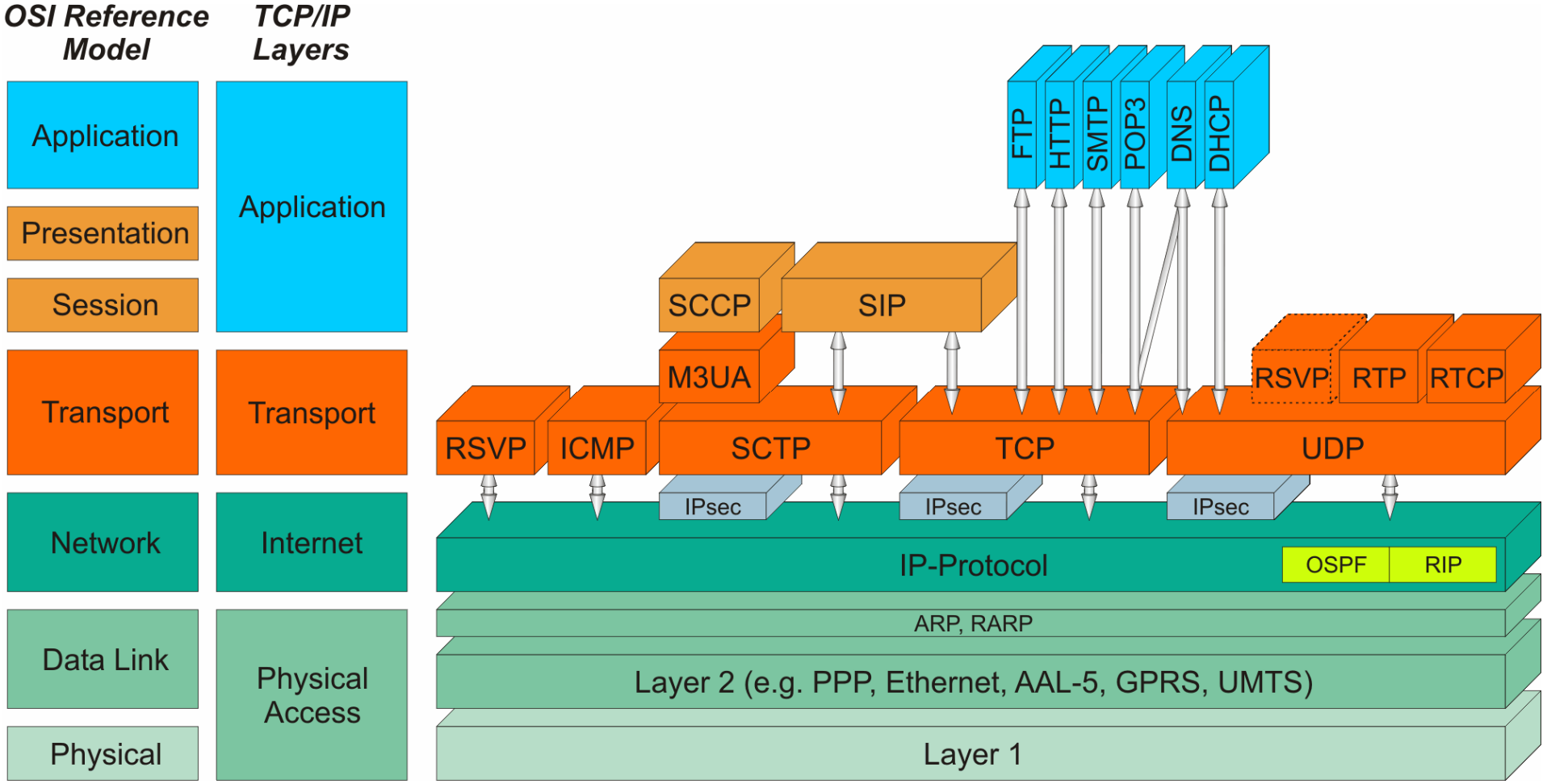
(1) Example for Dial-Up Network Access using the PPP



(1) Example for Dial-Up Network Access using the PPP

Intentionally left blank

Introducing the IP-Protocol Stack



Introducing the IP-Protocol Stack

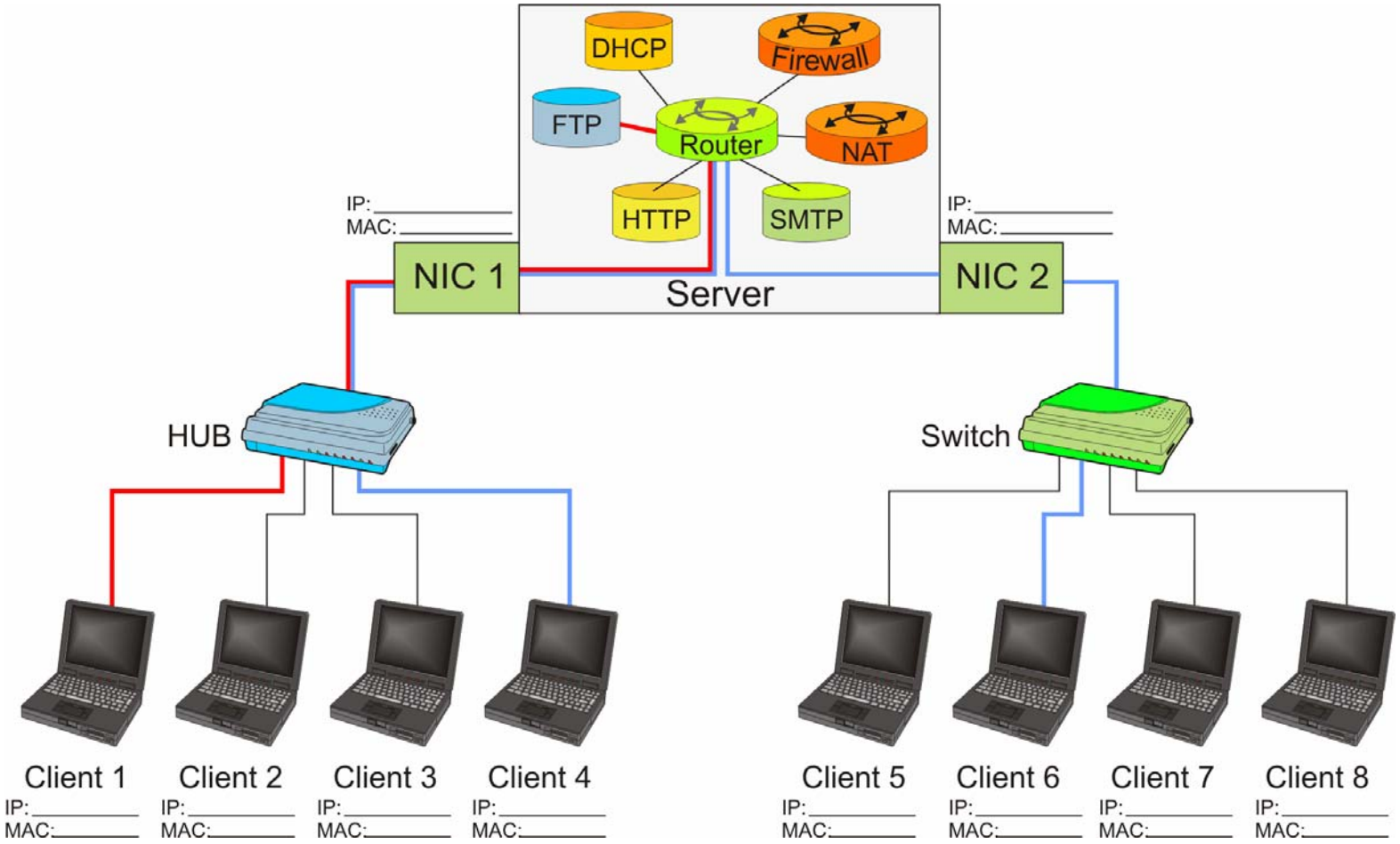
A simplified view of the IP-protocol stack is provided in the figure. The Internet Protocol (IP) itself can be located on top of almost any available layer 2-protocol like for example:

⇒ PPP (⇔ Point-to-Point Protocol / RFC 1661)
⇒ Ethernet (⇔ IEEE 802.3)
⇒ AAL-5 (⇔ ATM / ITU-T I.363.5 (6), Q.2110 (4)
⇒ or even on mobile bearers like GPRS or UMTS (⇔ 3GPP recommendations).

As it can be seen from the slide the term IP-protocol stack relates to IP and the higher layers which make use of the IP's networking capabilities. On top of IP several transport protocols like TCP or UDP are located, itself as carrier for different application protocols like HTTP, FTP or RTP.

Thus from the nested protocol structure one can expect a sequence of protocol headers that have to be included in the complete data packet.

Practical Exercise – Add the LINUX Server to our Network(s)



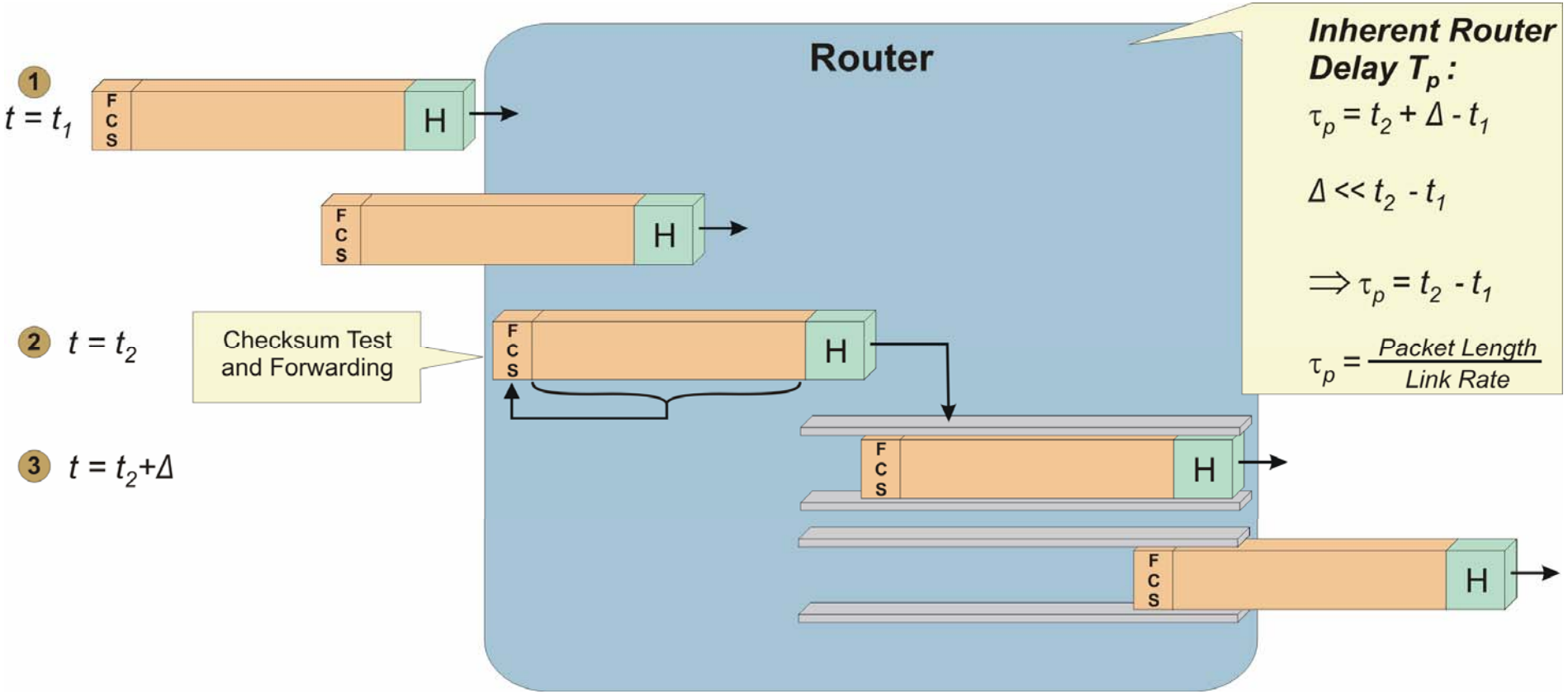
Add the LINUX Server to our Network(s):

- **Connect both workgroup sub-networks to the LINUX Server**
  The Server shall provide the Router, DHCP and NAT function. It will also implement a Mail Server, HTTP Server and Firewall function.
- **Follow the configuration of the Server from the Administrator PC: Router, DHCP, NAT, Firewall …**
  Basics about our LINUX server, Introduction of specific tools and utilities and short roadmap on the PC by the Administrator
- **Re-configure the system such that the IP addresses are provided automatically by the DHCP server.**
- **Modify your Client PC configuration accordingly to enable automatic IP address allocation.**
- **Determine any new MAC addresses and relate all MAC addresses to the IP addresses.**

Questions:

- **Review MAC and IP Addresses – what has changed (if any) and why?**
- **What is the function of the "Standard Gateway"?**

(1) Inherent Delay of Routers



(1) Inherent Delay of Routers

Depending on the layer 2 handling of incoming packets in a network node different delay parameters can be achieved. In cut through mode no processing will be done and the packet will be forwarded directly to layer 3 of a router for routing decision.
In the store and forward mode – which is the mostly used layer 2 processing technique – a received packet will be checked for transmission errors and erroneous packets will be discarded.

Step 1

The first bit of a received packet will arrive at time t_1 at the router input. This packet must be received completely before error checking can start.

Step 2

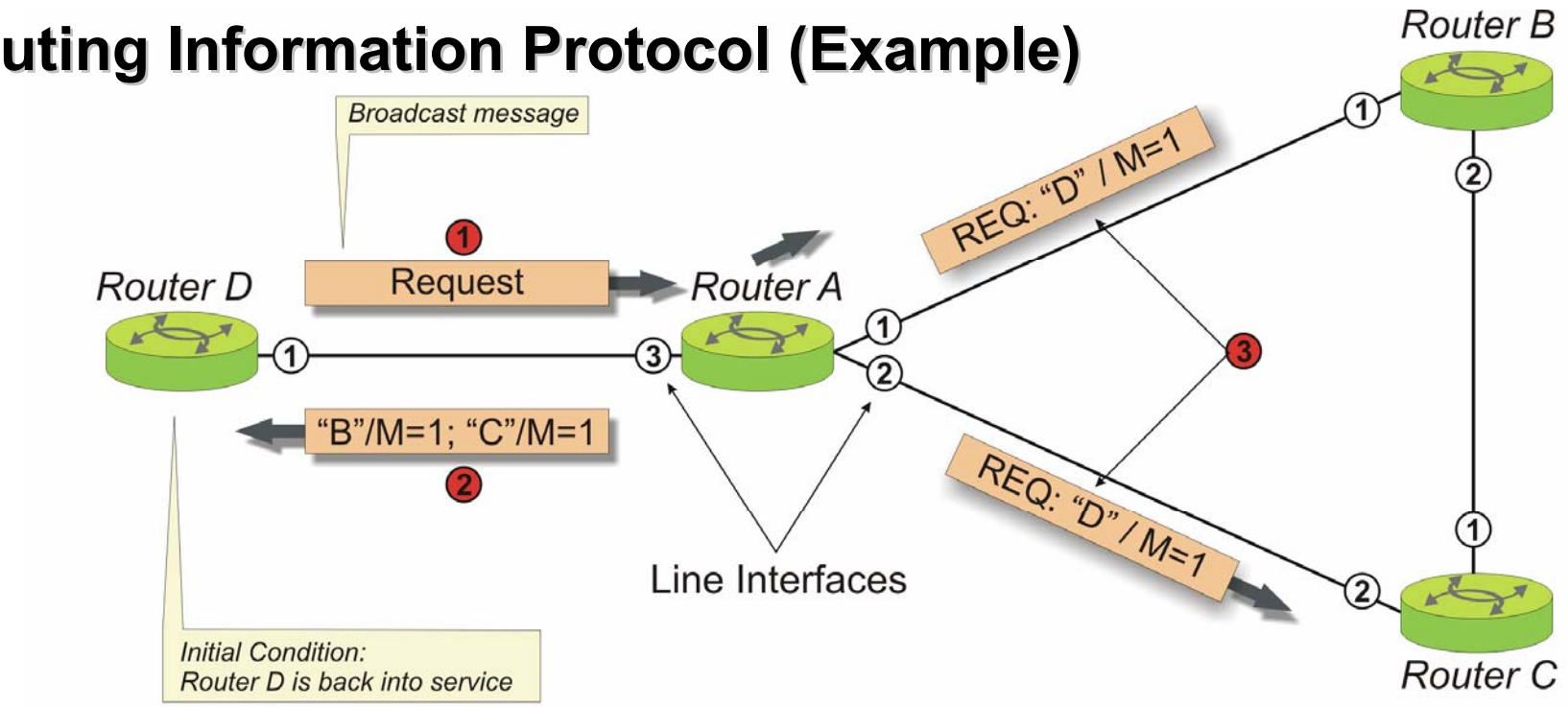
When the last bit of the packet has been received the error check process will start and after positive result the packet will be forwarded to layer 3.

Step 3

After layer 3 processing (address decoding, packet classification) the packet is scheduled for the output port according to the routing table entry and placed in the output queue. This step requires an additional processing time which can be summarized together with the layer 2 processing time to value Δ. Compared to the overall delay time this value can be neglected.

The inherent delay τ_p of a router is determined when the first bit arrives at the router input and when this bit is transmitted at the router output. This delay value depends only on the packet length and the link transmission rate.

Routing Information Protocol (Example)



Router A

Router	Metric	I/f
B	1	1
C	1	2
(2) D	1	3

Router B

Router	Metric	I/f
A	1	1
C	1	2
(4) D	2	1

Router C

Router	Metric	I/f
A	1	2
B	1	1
(4) D	2	2

Router D

Router	Metric	I/f
A	1	1
(2) B	2	1
(2) C	2	1

Routing Information Protocol (Example)

A relatively simple routing protocol is the Routing Information Protocol (RIP) which is widely distributed due to its simplicity but it is only well suited for smaller networks. However, there are also some limitations in this protocol mainly in the sense that a simplistic metric leads to non-optimal routing tables. The decision for a route is based on a metric for which link cost, distance or other parameters may be used. Very often only the shortest distance is derived from the smallest number of hops.

A network with routers A, B and C is in operation and the routing tables have been setup. The routing tables include the connected routers, their addresses, the number of hops towards these routers and the outgoing interface nomination.

Step 1

Router D is put into service (e.g. after a failure) and will inform all its neighbor routers with a RIP: Broadcast Request message about its availability in the network. A request message will be answered immediately with a reply message from all connected neighboring routers.

Step 2

Router A will immediately update its routing table to include router D with one hop away. Further router A will reply towards router D that it can reach routers B and C with one hop.

Step 3

Router A sends a request message to router B and router C to inform that A can reach router D with one hop.

Step 4

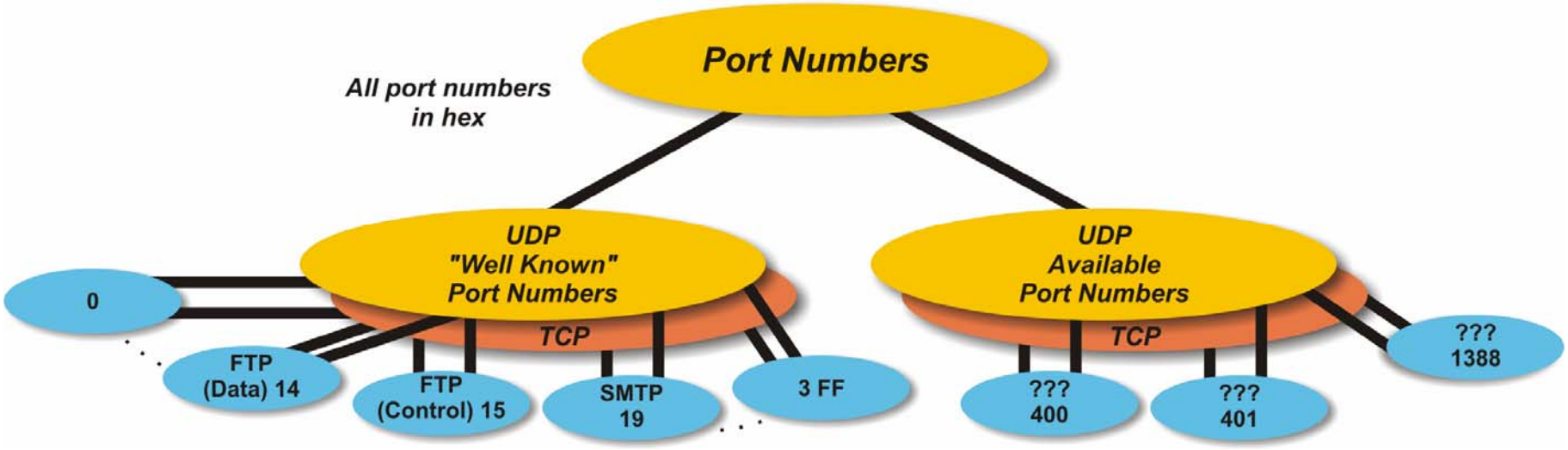
Router B and C will update their routing tables that they can now reach router D with two hops (via router A) and will further inform their next neighbors.

To detect network changes, e.g. in case of failure, each router is supposed to send update messages to all its neighbors every 30 seconds. In case one router has an entry in its routing table that it can reach router X with one hop and there have no update messages been received within 6 update periods (180 seconds) the router will mark the entry as invalid.

In case there is no change in the network topology or network load there will be no routing table changes and all packets to one destination will use the same path.

[IETF RFC 1058, 2453]

Port Numbers



Port Numbers

Both transport layers of IP, which are UDP and TCP, use port numbers for the identification of the application process. As the figure illustrates, there are the so called “well known port numbers” which are assigned and administrated by ICANN and there are port numbers which are randomly chosen by a host for identification purposes.

“Well known” Port Numbers

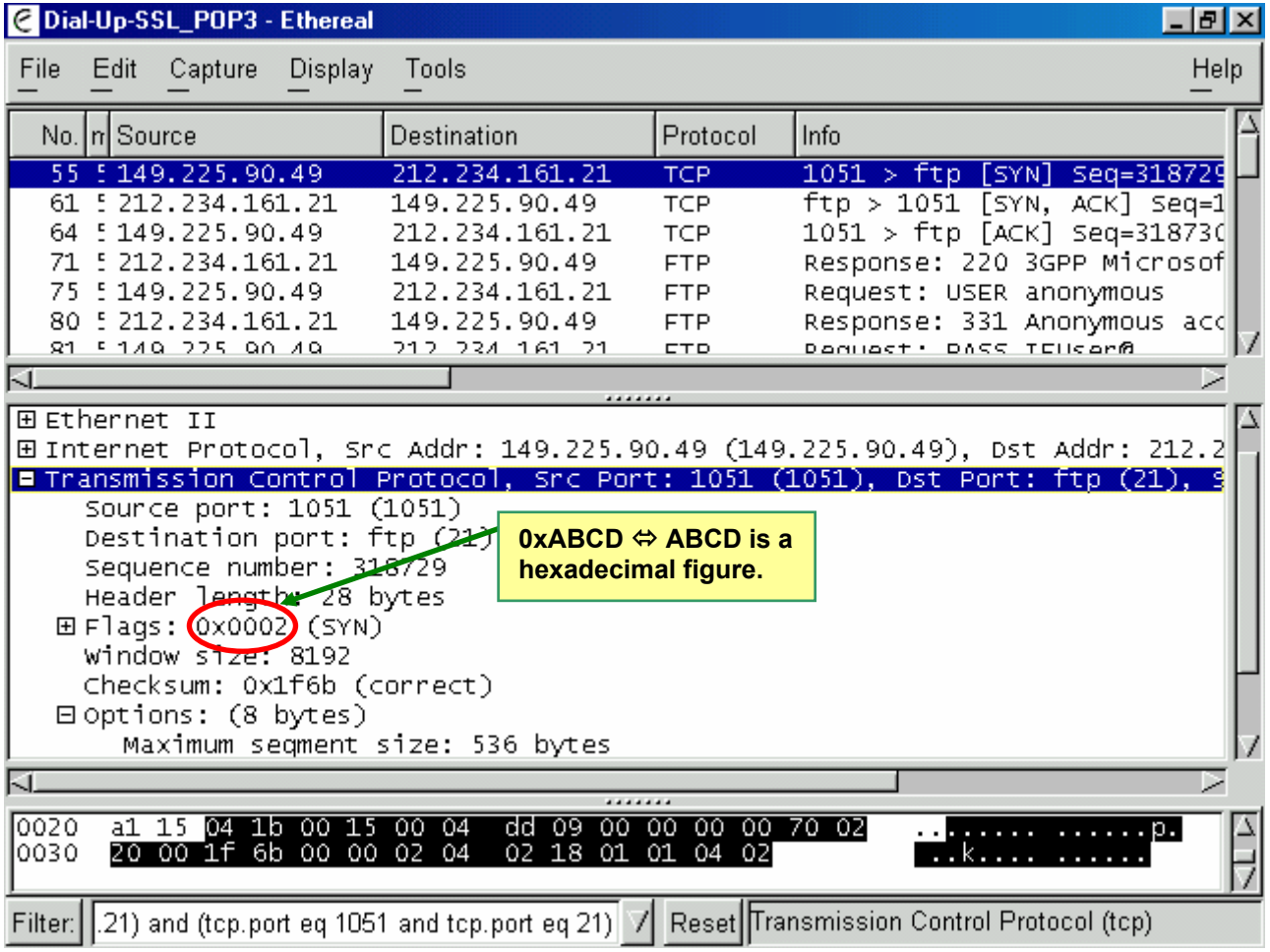
The ICANN has allocated specific port numbers to common applications which need to be known by all IP-implementations. A host will identify the application process it wants to access in a peer node by means of its “well known” port number.

Available Port Numbers

Available port numbers are selected by an initiating host to identify the source process whenever contacting a certain application in another peer. Example: When a file transfer shall be conducted, the calling application in the initiating host will identify itself through some port number between 400_{hex} ($\Leftrightarrow 1024_{dec}$) and 1388_{hex} ($\Leftrightarrow 5000_{dec}$) but it will call on FTP (15_{hex} ($\Leftrightarrow 21_{dec}$)) in its peer.

TCP port numbers are independent from UDP port numbers because the two protocols are independent from each other. However, there is quite some overlap of the well known port numbers between TCP and UDP.

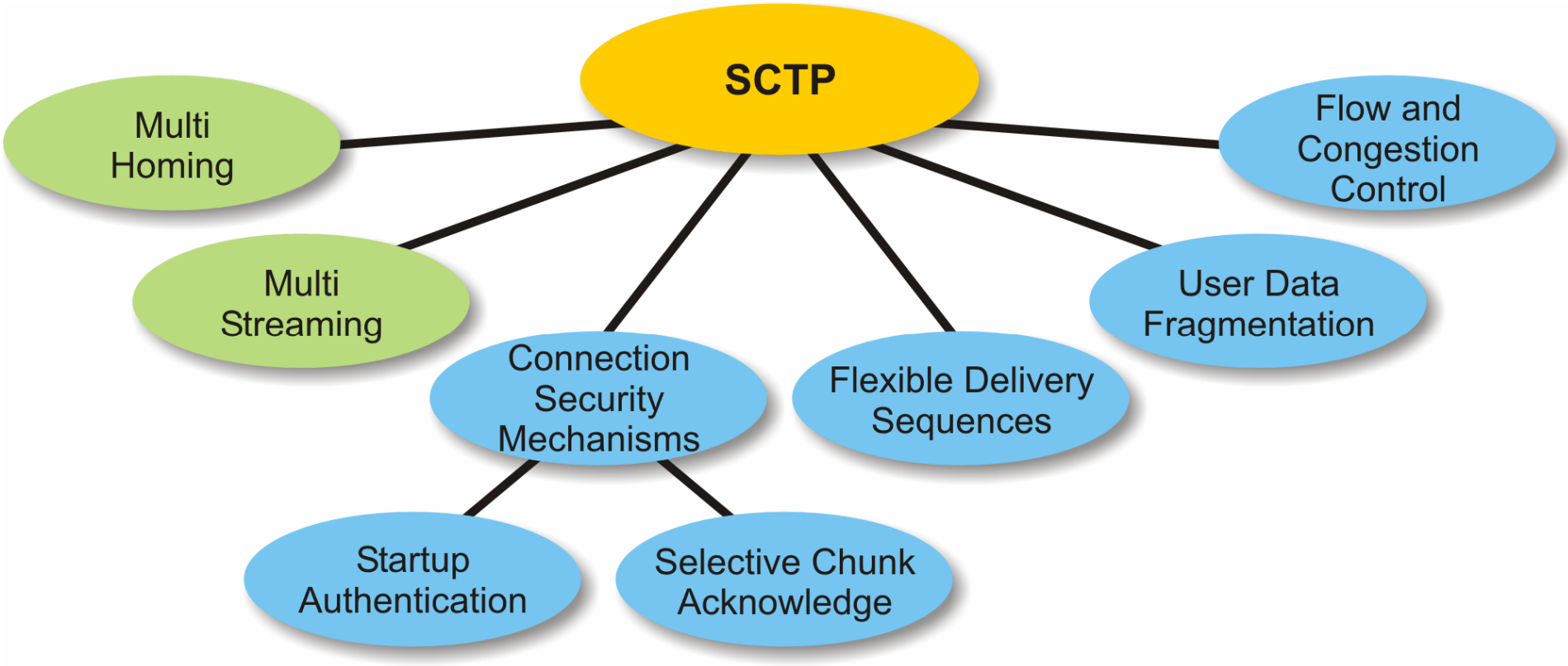
(1) Example for TCP Connection Establishment



(1) Example for TCP Connection Establishment

- **The figure illustrates the initial "active open" TCP-frame which is sent from host A to host B.**
  - ⇒ The application has selected the "free" port number $1{,}051_{dec}$ for this connection
  - ⇒ Note that the <SYN>-flag is set to indicate connection establishment
  - ⇒ The destination port number is the well known port number 21_{dec} (FTP)
  - ⇒ The ISN is selected with $318{,}729_{dec}$
  - ⇒ The window size is set to $8{,}192_{dec}$
  - ⇒ Host A is prepared to receive segments up to 536_{dec} octets (⇔ default size)

Details of the Stream Control Transmission Protocol (SCTP)

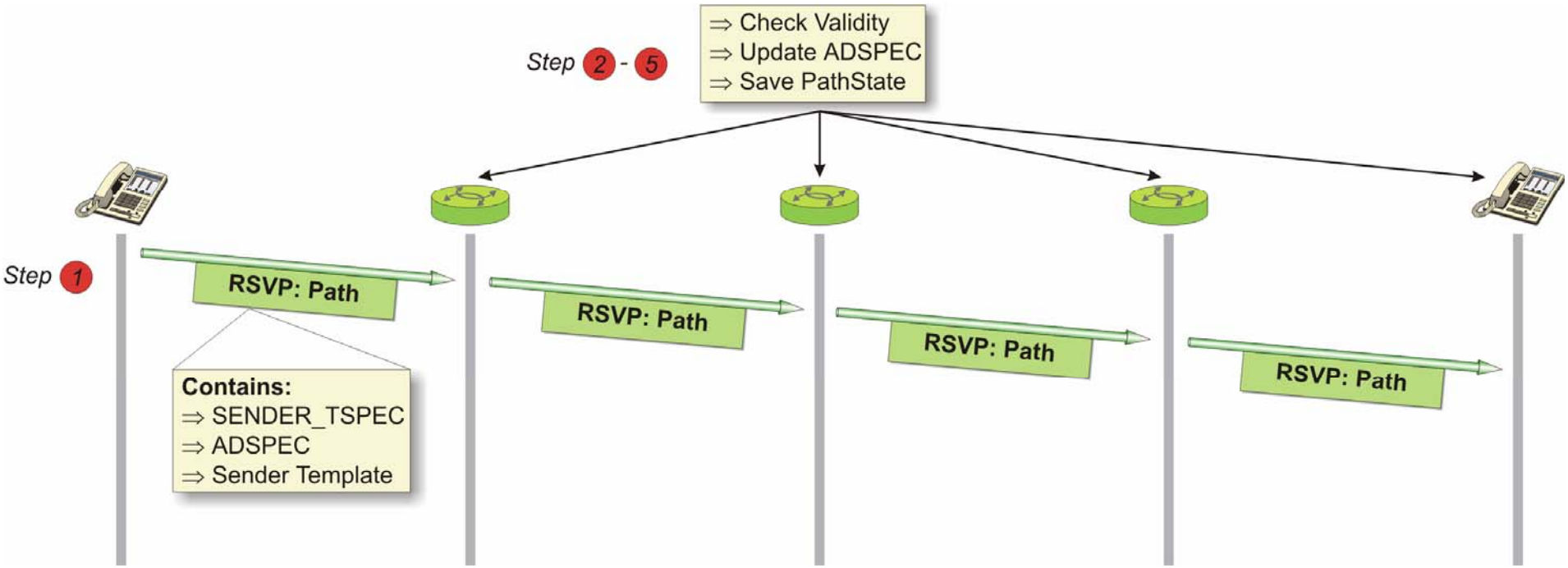


Details of the Stream Control Transmission Protocol (SCTP)

SCTP is an end to end, connection oriented transport protocol, transporting data in independent, sequenced streams. A stream is a sequence of messages (chunks), not octets as in TCP.

- **Multi-Homing**
  SCTP endpoints support multiple IP addresses which are exchanged during initiation of an association. In case of a network failure, traffic could be routed to an alternative IP taking advantage of the interface redundancy.
- **Multi-Streaming**
  User data is separated and transmitted over multiple SCTP streams providing independent, sequenced delivery. Packet loss in a particular stream (in TCP known as "head of the queue blocking") will not affect other streams. RE-transmitted and high priority messages can therefore bypass less significant messages.
- **Connection Security**
  SCTP implements features to provide connection security
  ⇒ At Startup
  SCTP uses a four-way handshake with cookie exchange to defend the endpoints against SYN-type attacks. Also, uses a verification TAG as protection against blind masquerade (denial of service) attacks.
  ⇒ During Data Transmission
  Chunk bundling allows to multiplex data chunks with control chunks in the same stream. The peer endpoint will acknowledge the receipt of all data chunks with selective acknowledges (SACK). SACK chunks carry transmission sequence numbers (TSN) to reveal gaps in the sequence of data chunks. SCTP packets also carry stream sequence numbers (SSN), which identify the sequence of data delivery within each independent stream. If the peer endpoint detects gaps in the SSN, the message will only be delivered, when these gaps are filled.
  ⇒ At Shutdown
  ⇒ SCTP implements a graceful shutdown with a 3-way message exchange. "Half-open" connections as in TCP are not possible.
- **Flexible Delivery Sequences**
  SCTP supports in-order and FIFO delivery towards the application on a per datagram basis.
- **User Data Fragmentation**
  As in TCP
- **Flow and Congestion Control**
  As in TCP

(1) Operation of the Resource Reservation Protocol (RSVP)



(1) Operation of the Resource Reservation Protocol (RSVP)

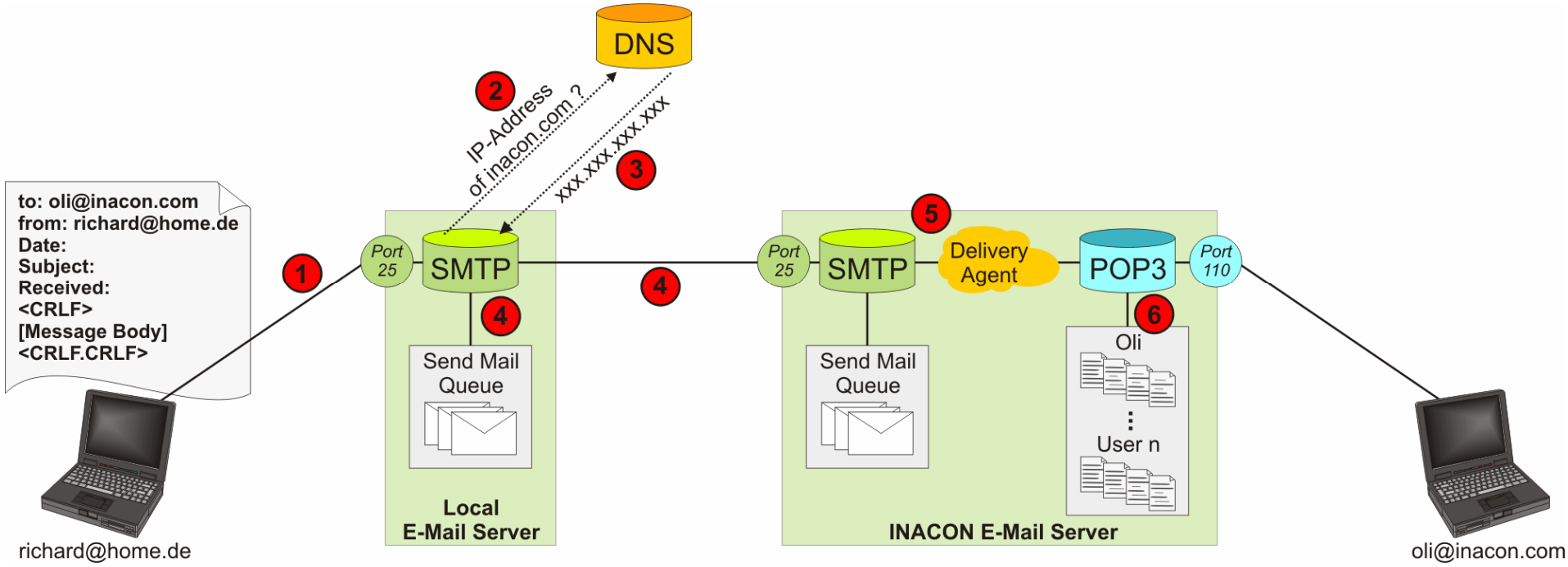
Sending of RSVP: Path-Messages

- **Step 1: Situation at the Original Transmitter**
  - ⇒ Whenever a peer intends to establish a real-time flow with additional QoS-parameters towards one or more other peers, it will construct an RSVP: Path-message and send it to the appropriate router.
  - ⇒ The PATH-message most importantly contains an identification of the sending peer (⇔ Sender Template consisting of IP-address / transport protocol / flow label (IPv6) or port number (IPv4)) and a specification of the traffic characteristics (⇔ SENDER_TSPEC consisting of packet size information, expected data rate and buffer sizes to be reserved along the way).
  - ⇒ Note that the final destination is identified through its IP-address / transport protocol / flow label (IPv6) or port number (IPv4) within the bearing IP-frame.
  - ⇒ Most likely, the original Path-message also contains the ADPEC-object which will be explained under the next bullet.
- **Step 2 - 5: Situation at Intermediate Routers**
  - ⇒ Each *Integrated Services*-aware router along the path towards the destination(s) will intercept an RSVP: Path-message and will perform admission control and policy control.
  - ⇒ That is, the router will check whether or not the requested resources can be reserved (load) and whether the requesting user is authorized to perform this reservation. If either check is negative, the router will reply an RSVP: PathErr-message to the previous hop router. This message is forwarded backwards until it finally reaches the original transmitter. Within the original transmitter, the application will be informed that no sufficient resources are available in the network.
  - ⇒ If these checks are positive, the router will take a number of actions:
    1. It will store the content of the Path-message in a PSB (Path State Block). This PSB most importantly contains the routing information (I/O).
    2. It will update the ADSPEC-object with respect to router-specific QoS-parameters like delay time, MTU and buffer sizes. Note that each router along the way will accordingly update the ADSPEC-object to allow the receiver to estimate the achievable QoS and to relay this information back to the original transmitter.
  - ⇒ Finally, the Path-message is received by the peer user. The RSVP-implementation will evaluate the received information and especially the ADSPEC-object.

While the flow is active, RSVP: Path-messages need to be periodically repeated (Period ≈ 30 s) to avoid an automatic teardown of an established flow.

[RFC 2205, RFC 2209, IEEE Communications Magazine May 1997]

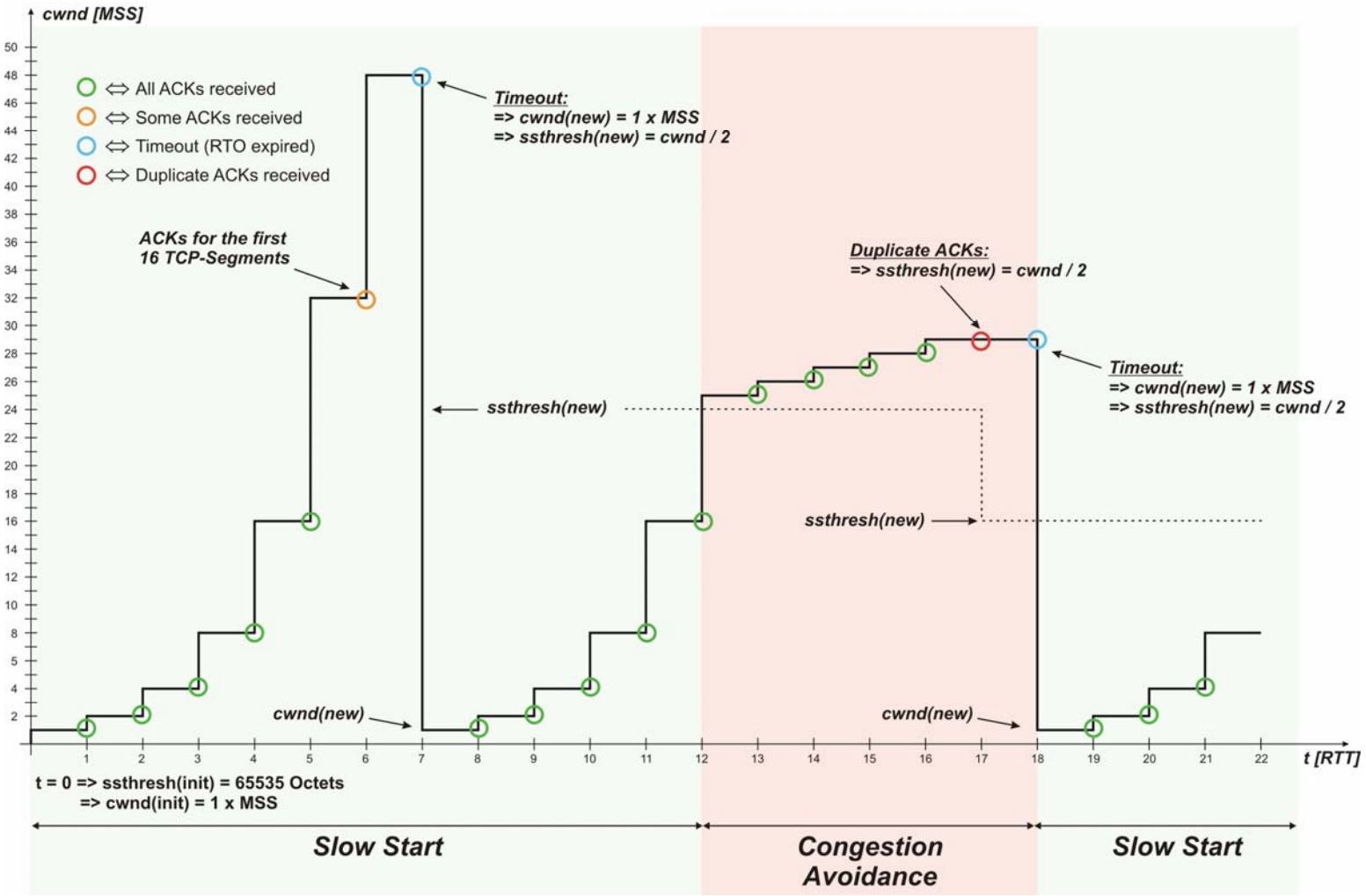
The Simple Mail Transfer Protocol (SMTP)



The Simple Mail Transfer Protocol (SMTP)

- **SMTP is used to send (upload) E-Mails**
- **Local E-Mail Server with the SMTP Server function may need to contact a DNS Server for address resolution and may need to contact an external SMTP Server at the destination to deliver the mail.**
- **SMTP Server maintains a local Send Mail Queue in case a message cannot be delivered to the destination Mail Server.**
- **At the destination Mail Server, a Delivery Agent program will pass the message to the POP3 Server for storage in the inbox system.**

Slow Start and Congestion Avoidance in Operation



Slow Start and Congestion Avoidance in Operation

Slow start and congestion avoidance operate together as follows:

⇒ Initially, the sender operates in slow start mode: It will only send a single TCP-segment to its peer, waiting for the acknowledgement. When this acknowledgement is received, cwnd is incremented to 2 segments.

⇒ Accordingly, the sender sends these two segments again waiting for the acknowledgement. If the acknowledgement for these two segments is received (possibly as a single acknowledgement for both segments), cwnd is incremented to 4 segments. For every acked TCP-segment, cwnd is incremented by 1 segment.

Slow start therefore provides for an exponential opening of the transmit window.

⇒ Eventually, the number of injected TCP-segments will congest the transmission line, resulting in timeouts. If this occurs, half of the current value of cwnd shall be stored in ssthresh and cwnd is reduced to 1 MSS.

A timeout is an indication for the sending TCP that the capacity of the internet is exceeded and that an intermediate router has discarded the TCP-segments for which the RTO expired.

⇒ What we see is another time slow start in operation, at least as long as cwnd ≤ ssthresh. As soon as cwnd > ssthresh, TCP will change to congestion avoidance operation mode.

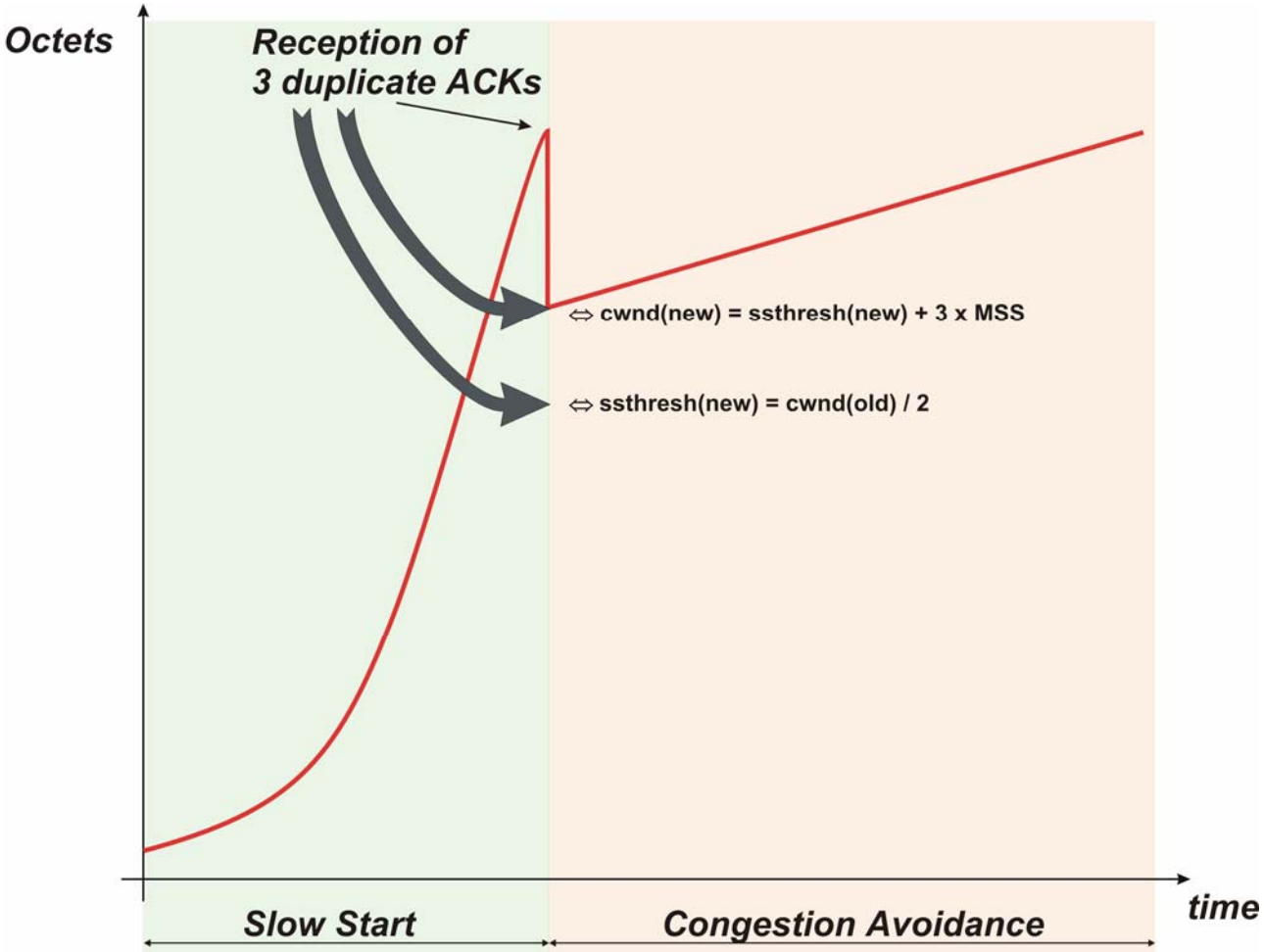
⇒ In congestion avoidance operation mode, the value of cwnd is still incremented but only by a maximum of 1 segment per roundtrip time. As a matter of fact, the increase of cwnd is controlled by the following formula: $cwnd_{new} = cwnd_{old} + (MSS^2 / cwnd_{old}.) \times$ (No of received Acks).

⇒ When duplicate acknowledgements are received (⇔ less than 3), TCP shall adjust the value for ssthresh but not for cwnd.

The reception of duplicate acknowledgements is an indication for the sending TCP that TCP-segments have been received out of order by the peer. Still, the transmission line appears to be still open. Therefore, cwnd shall not be reduced but ssthresh is adjusted.

[RFC 2001]

The Fast Recovery Algorithm



The Fast Recovery Algorithm

The fast recovery algorithm is really an amendment of the slow start and congestion avoidance algorithms. Whenever the fast retransmit algorithm is triggered by the reception of a sufficient number of duplicate acknowledgements, fast recovery will provide that the TCP-connection consecutively operates in congestion avoidance mode rather than in slow start.

- $\Rightarrow$ $ssthresh_{new} = cwnd_{old}/2$
- $\Rightarrow$ $cwnd_{new} = ssthresh_{new} + 3 \times MSS$

[RFC 2001]