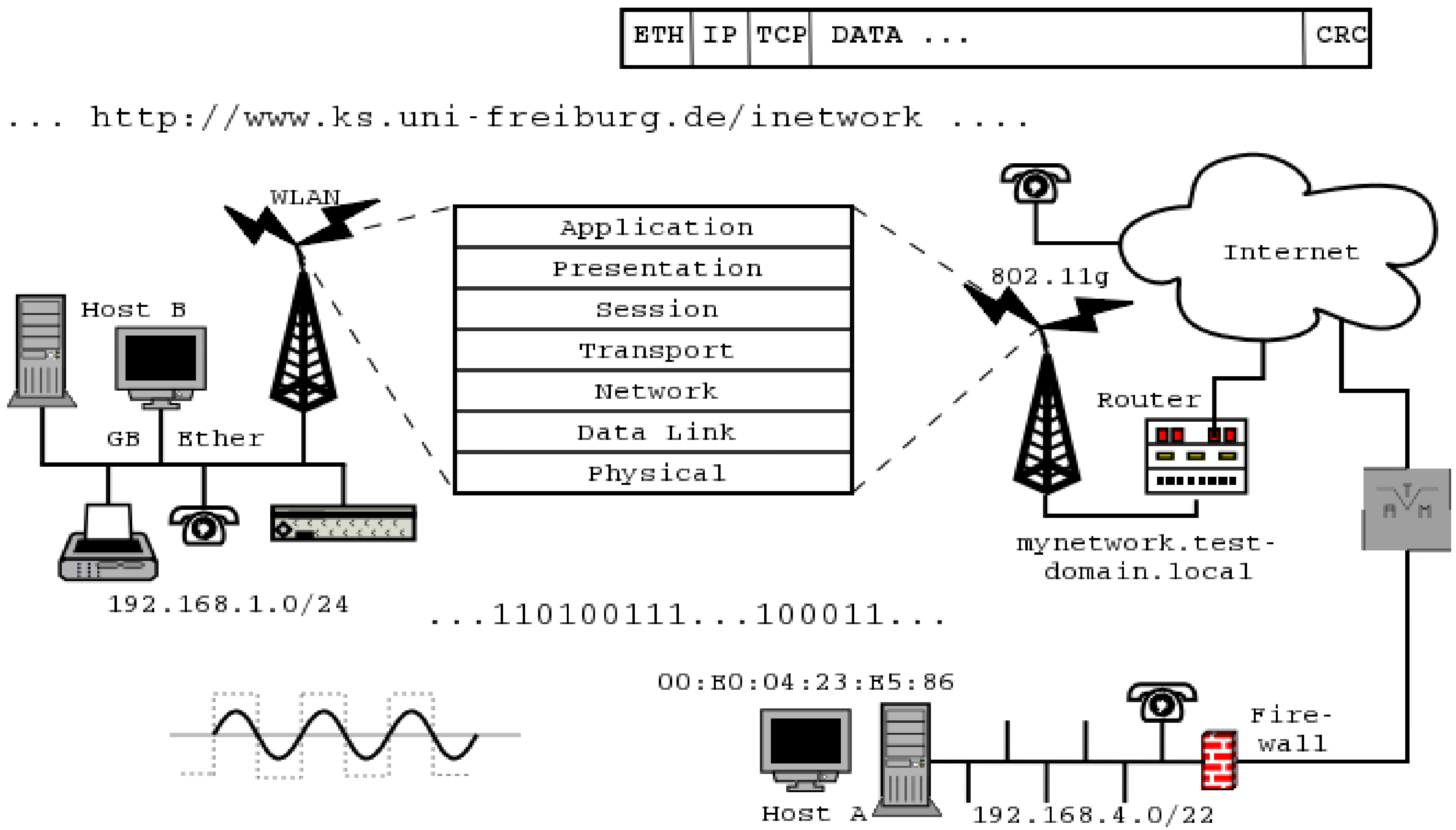
Communication Systems

3rd lecture



Chair of Communication Systems
Department of Applied Sciences
University of Freiburg
Winter Term 2008

Communication Systems
Organization and Q&A

- Any questions from last lecture?
- We were dealing with Ethernets last lecture:
- Why not only Ethernets are used for networking – working plug&play, stations in the net are found automatically!?
- Why it is impossible to use jumbo frames in a mixed Gigabit and Fast Ethernet LAN?
- What is the minimum length of an Ethernet packet, why? Is that really needed in switched Ethernets, why/not?
- Which restriction may apply if a Gigabit Ethernet adapter is plugged into the old-standard PCI?

Communication Systems
10GBit/s Ethernet

- 10 Gigabit Ethernet (10GbE or 10 GigE) standard first published in 2002 as IEEE Std 802.3ae-2002
- Several substandards followed to define different media types (mainly fiber, later several copper variants)
  - To support different PHYs often pluggable PHY modules deployed in switches
- Major difference to earlier standards:
  - Obsoleting half duplex operation and CSMA/CD
  - Only full duplex links connected by switches
  - Introducing a special 10 Gigabit Ethernet WAN standard: 802.3aq-2006 (slightly slower with extra encapsulation)

Communication Systems
10GBit/s Ethernet

- Optical PHYs for single or multi-mode fiber
  - Single mode: 8.3µm – thus difficult to deploy but much longer distances (some km, single path of light travelling)
  - Multi-mode: 50 or 62.5µm – multi path with the problem of differential mode delay (up to 300m, cheaper cable, optics)
- 10 Gigabit Ethernet copper standards followed later because on quite demanding wire characteristics (IEEE 802.3an-2006 for twisted pair copper wire)
  - Allows up to 55 m (180 ft) with existing Cat 6 cabling
  - Typical 100m with partitioned Category 6a cables with reduced crosstalk between UTP cables (alien crosstalk)
  - Wire-level modulation: Tomlinson-Harashima precoded (THP) version of pulse amplitude modulation with 16 discrete levels (PAM16 – PAM5 in 1GbE TP), encoded in a two-dimensional checkerboard pattern (DSQ128)

Communication Systems
100GBit/s Ethernet

- Because of modulation overhead TX copper standard adds rather high latency
- Other copper standards for low latency like for "backplane" wiring as needed in blade computer units (1m distances) or 802.3ak (15m)
- 10 GbE deployed mostly in infrastructure, network adaptors available but not widely used (1GbE much cheaper, other strategies like channel bonding)
- 100 Gigabit Ethernet, first discussions in 2006, standardization begun in 2007 (including standard for 40GbE)
  - Same frame formats and sizes, full duplex only like in 10 GbE
  - Aiming for 10-40km with optical single mode fiber (4 wave length), copper standards too, but for much shorter ranges (~10m)

Communication Systems
Ethernet link aggregation/channel bonding

- Ethernet by now predominant standard in LAN / MAN – alternatives searched for not depending on next generation standard
- Implementations for Ethernet line backup and seemless bandwith upgrades
  - Often impossible to switch directly to the next higher standard – imagine 100% used 1GbE link means 10+% used 10GbE
  - Idea to use more than one link between two switches (proprietary extensions exist to span link aggregation over different switches)
  - Allows network's backbone speed to grow incrementally on demand, without having to replace the whole infrastructure

Communication Systems
Ethernet link aggregation/channel bonding

- IEEE 802.3ad standard for link aggregation (LA – Ethernet, channel bonding – several names exist)
  - Either static setup or implementation with Link Aggregation Control Protocol (LACP)
  - Some proprietary alternatives exists of companies like Cisco or Nortel (offering multi switch bonding / high availability, (split) multi link trunking - SMLT)
  - Common way to balance traffic by using Layer 3 (IPv4, IPv6 address) hashes (Which problems you would expect when using this approach?)

Communication Systems
Ethernet link aggregation/channel bonding

- Mixed port speed aggregation possible (useful for backup but not load balancing)
- Not possible to aggregate links operating mixed in full and half duplex mode
- Some disadvantages
  - Depending on sessions rarely 50/50 distribution (on two lane LA) reached, often distribution like 70/30
  - Imagine 1GbE capable station connected to a switch using 2*100Mbit/s uplink – what is the resulting max. Bandwidth?
  - More advanced switches: Using Layer 4 link hashes for distribution (does this solve the problem named in the previous slide?)

Communication Systems
Ethernet link aggregation/channel bonding

- Higher layer implementations for most common Unix OS (like Linux)
  - Bonding over different vendors Ethernet adaptors possible
  - Several setups
    - High availability and outgoing traffic equalization (independent of switch)
    - HA and link aggregation for both directions (dependent on compatibility to switch link layer aggregation)
- Talked of bandwidth issues; addressing higher layer network design questions by now

Communication Systems
Ethernet extensions

- Large collision domains allowed restricted number of stations
  - Broadcast traffic – even a rather moderate load can reduce Ethernet throughput significantly up to rendering a network unuseable completely
  - Switched operation reduces domains to just one wire/direction (thus no collision handling, half duplex operation standardized in 10GBit/s Ethernet)
- Diameter of Ethernets increased throught switching and optical transmission
  - Whole university campus connected using Ethernet: 100/1000MBit/s to the stations, up to 10GBit/s in backbone
  - Number of connected stations quite large

Communication Systems

Ethernet on Freiburg University campus

- Freiburg University campus: 156 switching centers offering ~29k Ethernet ports of 100 / 1000Mbit/s speed, ~16k ports acive
- Ethernet only infrastructure by now (obsoleting FDDI, TokenRing, ATM)
- Backbone is using Ethernet trunking ??

rechenzentrum universität freiburg

Internet Working
Extensions: VLAN tagging and QoS

- VLANs a means for complete virtualisation of broadcast domains
  - Same characteristics as physical LAN, but allowing end stations to be grouped together independent of network switch location
  - Major advantage: reconfiguration done via switch software configuration instead of physically relocating hardware / changing cabling
  - Segmentation service traditionally provided by routers in a LAN
  - Provides a mean to expedite time-critical network traffic by setting transmission priorities for outgoing frames
  - Bridges and switches filter destination addresses and forward VLAN frames only to ports that serve the VLAN to which the traffic belongs
  - Multiple layer 3 networks within the same physical LAN – we will make extensive use of that in upcoming lectures giving a good opportunity to setup multiple IP routers connected by different links not "seeing" each other

Internet Working

Extensions: VLAN tagging and QoS

- VLAN 802.1q tagging is a MAC option that provides capabilities not previously available to classical Ethernet networks (defined since Fast Ethernet standards)
- VLAN standard deploys internal tagging process modifying the Ethernet frame adding a 4 byte header extension
  - 2-byte tag protocol identifier (TPID) plus 2-byte tag control information (TCI)
  - TPID has a fixed value of 0x8100 indicating the frame is carrying 802.1q/802.1p tag
  - TCI contains:
    - 3-bit user priority
    - 1-bit canonical format indicator (CFI)
    - 12-bit VLAN identifier (VID) – Uniquely identifies the VLAN the frame belongs to

Internet Working

Extensions: VLAN handling

- If the MAC is installed in a switch port, the frame is forwarded according to its priority level to all ports that are associated with the indicated VLAN identifier
- If the MAC is installed in an end station, it removes the 4-byte VLAN header and processes the frame in the same manner as a basic data frame

PRE	SFD	DA	SA	VLAN type ID	Tag control information	Length/Type	Data	Pad	FCS	Extension
7	1	6	6	2	2	4	46 -	1500	4	

Field length in bytes

Internet Working

Extensions: VLAN handling

- VLANs violate the old MTU restriction of 1518 Byte producing packets with a MTU of 1522 Byte
- Application:
  - Static – port based configuration: All machines connected to a port are in the same VLAN (invisible to them), standard scenario in campus setup
  - All VLAN tags added by these stations are silently dropped in switch (thus a reconfiguration of local device was required for producing the proper playground for the exercises)
  - Dynamic – using special software to create VLAN automatically e.g. grouping on source MAC address (e.g. putting all IP telephones in a specific LAN with higher forwarding priority)

Communication Systems
Ethernet and IP

- Flat addressing scheme of physical/data link layer Ethernet
- Why two addresses for a LAN connected host?
- IP addressing – higher layer, to overcome the flat addressing restrictions: routed/hierarchical
  - Changing places of many hosts (e.g. your laptop: connected at home to Ethernet and to different WLANs throughout the day, but the physical addresses of your machine do not change)
  - Manual setup/configuration needed (todays practical)
- How to encapsulate IP datagram within link-layer frame

Communication Systems
Ethernet and IP

- What low level destination (MAC) address to use?
- Helper protocol is needed
  - IP has no feature to do mapping itself
  - Such mapping is not needed in PPP environments
  - This protocol is specific to the underlying hardware / software protocol
- Address Resolution Protocol (ARP) is for address mapping in Ethernets (and TokenRings, ATM, ...)
  - Fairly old protocol around for a while (RFC 826)

rechenzentrum universität freiburg

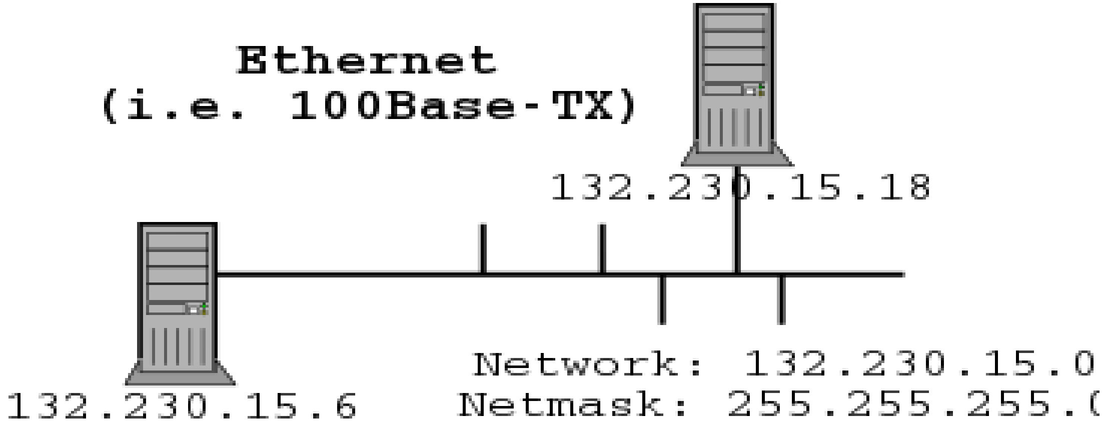
Communication Systems

IP to MAC and vice versa

- Address Mapping: IP to MAC – to get the host where to deliver a given packet locally
- Simple solution could just broadcast everything (and every machine listens to everything)
  - Unnecessary, burdens uninterested stations with others' traffic, congests the network
- IP to MAC address mapping mechanisms
  - Configured by hand [cumbersome]
  - Dynamic [learned by system automatically]
- Address Mapping IP to MAC: Learning

Internet Working
address mapping in broadcast nets

- But what to do in broadcast nets with many connected hosts?
  - In broadcast nets every host gets every packet sent out in the segment (switching may reduce traffic, but for some services packets to all are inevitable)
- For local delivery, need to map network-layer address to link-layer address:
  - Consider the machines 132.230.15.6 and 132.230.15.18 (netmask e.g. 255.255.255.0) ... [on same network/subnet]



Communication Systems
address resolution protocol (ARP)

- Dynamic approach
  - Each station runs **Address Resolution Protocol** (ARP)
  - Client/server architecture, each station is both client and server , routers have to implement the same mechanism too
  - Cache lookups with timeouts on each resolution
- Introduction of an intermediate protocol – operating between layer 2 & 3
- Address Resolution Protocol is basically address independent (at both network & link layer)
- Protocol is specialized for each particular network/link address pairing

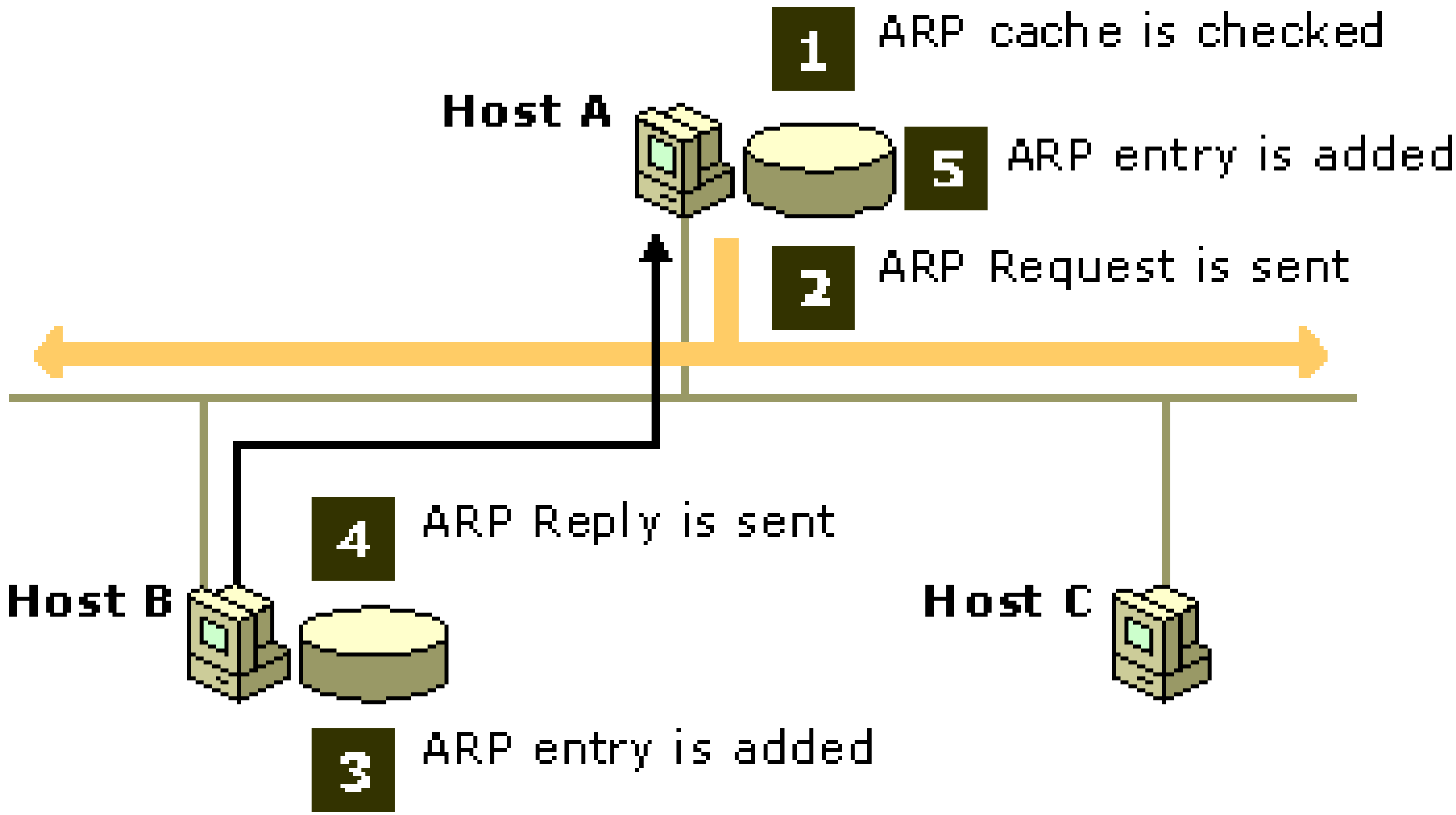
Communication Systems
address resolution cont.

- The term **address resolution** refers to the process of finding an address of a computer in a network
- Address is "resolved" using a protocol in which a piece of information is sent by a client process executing on the local computer to a server process executing on a remote computer
- The information received by the server allows the server to uniquely identify the network system for which the address was required and therefore to provide the required address
- Procedure is completed when the client receives a response from the server containing the required address

Communication Systems
ARP operation

- Step-by-Step operation
  - 0 - Requesting station A has IP address I, wants the associated MAC address M
  - 1 – Check the own ARP cache
  - 2 - A broadcasts the query: who has I? tell A
  - 3 – B adds MAC for A to its cache
  - 4 - Machine assigned address I responds directly to A with its MAC address M
  - 5 - A adds the (I,M) entry to its ARP cache

Communication Systems
ARP operation cont.



Communication Systems
ARP on ethernet with IP payload

- Common example is Ethernet/IPv4 (look at Ethernet/IPv6 in *wireshark* in upcoming practical)
- Ethernet MAC: 6Byte (48bit), IPv4: 4Byte (32bit), IPv6: 16Byte (128bit)

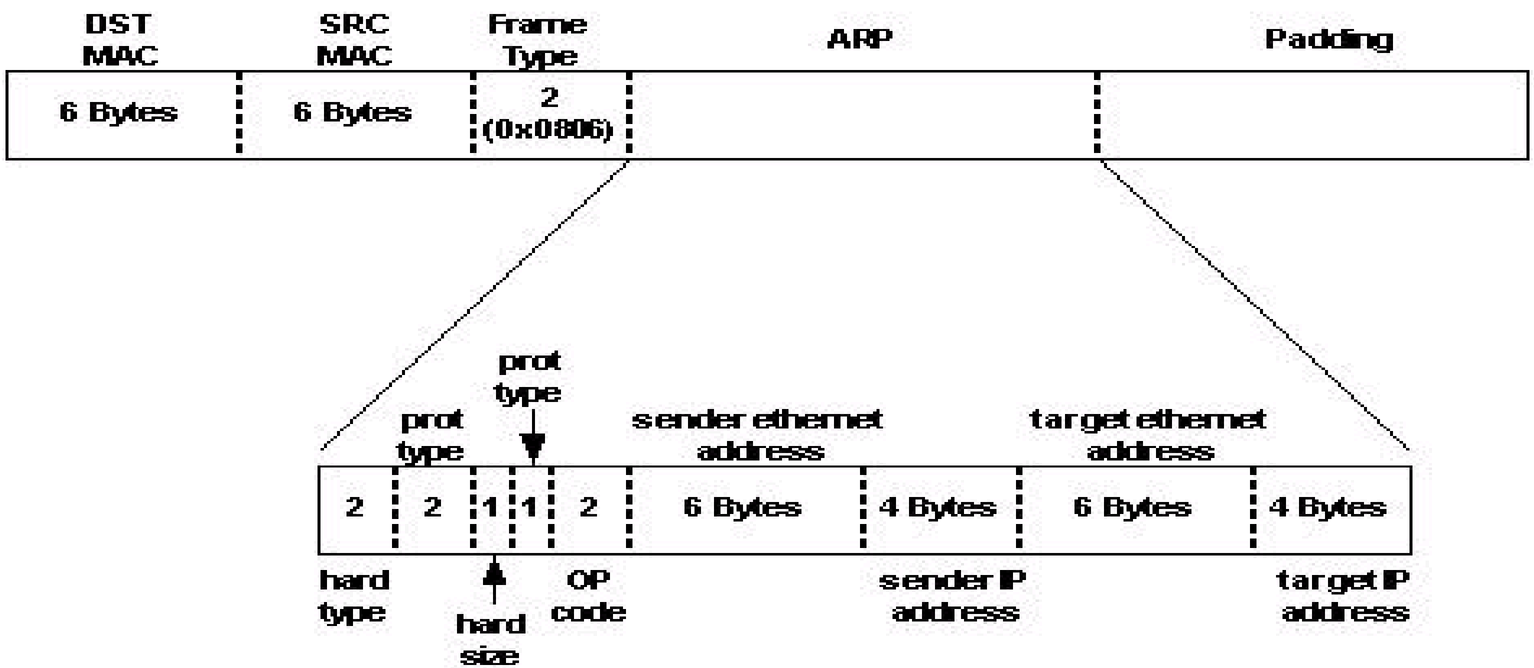
32 bit

HARDWARE ADDRESS TYPE		PROTOCOL ADDRESS TYPE
HADDR LENGHT	PADDR LENGHT	OPERATION
SENDER HARDWARE ADDRESS (first 4 Byte)		
SENDER HADDR (last 2 Byte)		TARGET HADDR (first 2 Byte)
TARGET HARDWARE ADDRESS (last 4 Byte)		
TARGET PROTOCOL ADDRESS (all 4 Byte)		

Communication Systems

ARP on ethernet with IP payload cont.

- ARP frames marked with Frame Type 0x0806
- IPv4 frames marked 0x0800 (*wireshark* exercise)
- Ethernet frame on wire with all headers and ARP payload



rechenzentrum universität freiburg

Communication Systems
ARP cache table (example)

- Contains hostname or IP address, hardware type, MAC, flag (c for cached), interface (use of *arp* or *ip neighbor* commands presented in experimental part of the course)

```
$
$ /sbin/arp
Address                  HWtype  HWaddress           Flags Mask            Iface
npserv2.ruf.uni-freibur  ether   00:09:6B:00:41:78   C                     eth0
fs1.ruf.uni-freiburg.de  ether   00:02:B3:4C:57:23   C                     eth0
mbone.ruf.uni-freiburg.  ether   08:00:20:88:96:76   C                     eth0
b1-9.ruf.uni-freiburg.d  ether   00:09:6B:00:40:DA   C                     eth0
npserv4.ruf.uni-freibur  ether   00:09:6B:00:41:BC   C                     eth0
b1-8.ruf.uni-freiburg.d  ether   00:09:6B:00:26:CF   C                     eth0
login9.ruf.uni-freiburg  ether   00:09:6B:00:3F:8D   C                     eth0
npserv1.ruf.uni-freibur  ether   00:02:B3:4C:57:37   C                     eth0
132.230.1.254            ether   00:09:97:30:3A:14   C                     eth0
mawa.ruf.uni-freiburg.d  ether   00:02:B3:B5:04:9A   C                     eth0
b2-7.ruf.uni-freiburg.d  ether   00:09:6B:00:41:72   C                     eth0
b2-6.ruf.uni-freiburg.d  ether   00:09:6B:00:40:4E   C                     eth0
b2-8.ruf.uni-freiburg.d  ether   00:09:6B:00:3B:08   C                     eth0
b1-6.ruf.uni-freiburg.d  ether   00:09:6B:00:41:88   C                     eth0
b1-7.ruf.uni-freiburg.d  ether   00:09:6B:00:54:B6   C                     eth0
ldap1.ruf.uni-freiburg.  ether   00:02:B3:4C:4D:5B   C                     eth0
$
```

Communication Systems
literature list & course organization

- You should have read Kurose&Ross (or similar text books on Ethernet and ARP) chapter 5 up to 5.7
- Read section 5.7 for upcoming lecture and on PPPoE
- Read on ARP insecurities and how to exploit them
- Next lecture is **Friday, 7th November** 9am (sharp!) same lecture pool room: -100/-101

Communication Systems
The exercise environment

- You should have grabbed practical exercise sheet #3 passed by
- This exercise block will require some more setup than the last one
  - Updated special Linux image which will run in a virtual machine (VMware/Player) using a virtual bridge to the physical Ethernet of the host machine
  - Start the pool system machine – select "**Kursraum Entwicklung**" entry in the list (just the one below the highlighted entry, otherwise the command for the virtual environment would not be found)
  - Log-on to the pool system machine with your computer center ID (choose a session like "KDE3" or "Gnome" at this moment) or
  - Choose "KDE3" or "Gnome" after login (do not select the "Communication Systems" or any Windows course session)
  - Same access procedure like last time, run *run-vmware.sh /var/lib/vmware/vmconfigs/linux-comsys-ws08.xml*

Communication Systems

The exercise environment

- No DHCP is running this time – manual IP address assignment required for network access
  - The Linux within virtual machine should not have an IP address assigned
  - Use the proper default gateway – use the commands of last practical course
  - New commands of this exercise: *brctl* and *vconfig*
  - In ping tests try to use IP addresses as the resolver might not be configured properly or the name server is unreachable
- Work with your neighbors and coordinate your actions!
- Homework: Practice the use of VLANs – create VLANs e.g. in a host-only Ethernet setup in VMware/Player