ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

Communication Systems

ICMP, NAT

University of Freiburg
Computer Science
Computer Networks and Telematics
Prof. Christian Schindelhauer

CoNe Freiburg

IIF
INSTITUT FÜR INFORMATIK FREIBURG

Copyright Warning

- This lecture is already stolen
- If you copy it please ask the author
  - Prof. Dr. Gerhard Schneider
- like I did

Internet Protocol – the Universal Service

- By now: Introduced IPv4 operation and protocol headers
- But spared:
  - Details on packet fragmentation as a central concept in IP (as an universal service)
  - Helper protocol to IP to cope with problems of stateless operation (how to get information on failures)
- Then: Special routing in IPv4 NAT (main issue of the practical part)

IP – Fragmentation of Packets

- Adapting datagram size one of the most important tasks of the Communication Systems protocol:
- IP datagrams itself cannot exceed 64kbyte
- Lower protocol levels report MTU (max. transfer unit)
  - Linux loopback 16384byte
  - Ethernet frames offer max. payload of 1500byte
  - ATM offers 48byte
  - slow modem-ppp connections 296byte packet length
- The tool ifconfig or ip (first practical course) reports MTU of each interface

IP – Fragmentation of Packets

- Fragmentation & Reassembly
  - divide network-layer datagram into multiple link-layer units, all have to be equal or smaller than link MTU size
  - further fragmentation may be needed if MTU is decreased along the path again
  - sometimes it is cleverer to set MTU smaller at source to avoid later fragmentation
  - reconstruct datagram at final station
- Each fragment otherwise acts as a complete, routeable datagram
- Datagrams are identified by the (source, destination, identification) triple
- Concept of fragmentation changes with IPv6

IP – Fragmentation of Packets

- If fragmented, identification triple is copied into each resulting packet
- Also contains (offset, length, more) triple
  - more - boolean indicates is last fragment
  - offset - relative to original datagram
- Relating fragments to original datagram provides:
  - Tolerance to re-ordering and duplication
  - Ability to fragment fragments (!)

IP – Fragmentation of Packets

- IP fragments are re-assembled at final destination before datagram is passed up to transport layer
- Routers do not reassemble fragmented datagrams
  - Allows for independent routing of fragments
  - Reduces complexity (need for CPU and memory) in routers
- Problems with fragmenting:
  - Loss of 1 or more fragments implies loss of datagram at the IP layer
  - IP is best effort, provides no retransmission, will time-out if frag(s) appear to be lost
  - May be interesting for DoS attacks

IP – Fragmentation of Packets

- Avoid fragmentation through computing path MTU
  - Problems if path changes (dynamic routing) and new path has smaller MTU along its way
- Adapting size of packets in the source machine according to the “minimum MTU”: Path MTU Discovery
  - IPv6 uses MTU discovery and assumes standard minimum MTU
- If datagram size is smaller then MTU, no fragmentation needed
- How to do this?
  - Probe network for largest size that will fit
  - If possible, have network tell us this size
  - Operates through ICMP messaging (presented later on)

Internet Control Message Protocol (ICMP)

- Remember IP packet orientated
- It provides no direct way of discovering the fate of a packet
  - Send & forget principle
  - Packets could be delayed for too long or even lost
  - Destination could be unreachable
    - Machine itself (routing broken, machine down, ...)
    - Specific protocol or port (above layer 3)
- Upper layer protocols or application may implement time out or helper protocol on network layer could be introduced ...

Internet Control Message Protocol (ICMP)

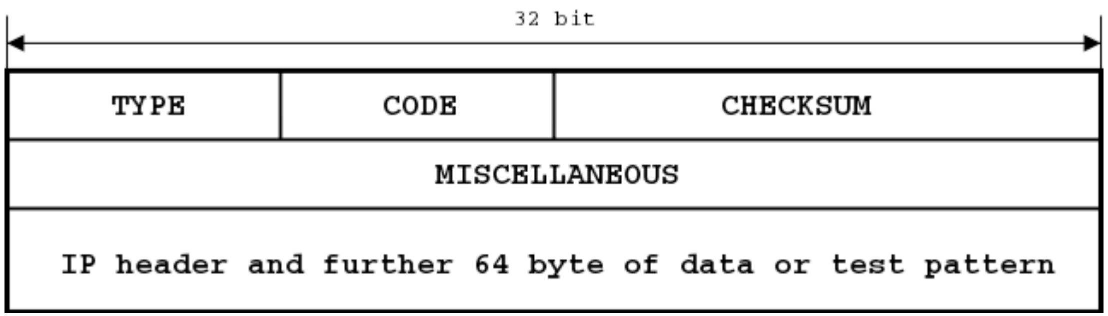
- Want a mechanism for error reporting and information exchange
  - ICMP Protocol defines “extensions” to the unreliable IP
  - Logically part of IP module, but is actually encapsulated within IP
  - Provides IP module to IP module message delivery
  - Error and information reporting only
  - Queries: client/server info request/response
  - Errors: reports of error conditions
  - Restrictions are placed on the generation of ICMP messages to avoid cascades

ICMP

- Restrictions for use of ICMP messages
- ICMP messages are not allowed to be sent in response to:
  - an ICMP error message (ok for queries)
  - datagrams failing header validation tests
  - broadcast or multicast IP datagrams
  - link-layer broadcast or multicast frames
  - invalid source address or zero network prefix
  - any fragment other than the first

ICMP Header

- Encapsulated as IP payload, common header:
  - Type field is 1 of 15 message types
  - Code indicates subtypes
  - Checksum covers entire ICMP message



ICMP Error Message Data

- Historically, ICMP errors returned the offending IP header and 1st 8 data bytes

```
▽ Internet Control Message Protocol
    Type: 3 (Destination unreachable)
    Code: 3 (Port unreachable)
    Checksum: 0x1b31 (correct)
  ▽ Internet Protocol, Src Addr: 132.230.9.160 (132.230.9.160), Dst Addr: 132.230.9.124 (132.230.9.124)
      Version: 4
      Header length: 20 bytes
    ▷ Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00)
      Total Length: 334
      Identification: 0x845c (33884)
    ▷ Flags: 0x00
      Fragment offset: 0
      Time to live: 128
      Protocol: UDP (0x11)
      Header checksum: 0x985a (correct)
      Source: 132.230.9.160 (132.230.9.160)
      Destination: 132.230.9.124 (132.230.9.124)
  ▷ User Datagram Protocol, Src Port: bootpc (68), Dst Port: bootps (67)
  ▷ Bootstrap Protocol
```

ICMP Error Message Data

- Test pattern (in hex) could be defined with ping tool (helps for easier identification of packets -> practical course)
- No longer adequate with more complicated headers like IP in IP tunnels
- New rules say should contain as much as original datagram as possible, without the length of ICMP datagram being larger then 576 bytes (standard Internet min size)
- Error Message Types (first header field):
  - 3 = Destination Unreachable, 4 = Source Quench
  - 5 = Redirect, 11 = Time Exceeded, 12 = Parameter Problem

ICMP Query Message Types

- 0 = Echo Reply ("ping response") and 8 = Echo Request ("ping query")
  - Example given last slide
  - Well known from the widely used ping command
  - Should not be blocked, needed for easy network debugging
- 9 = Router Advertisement, 10 = Router Solicitation
- 13 = Time Stamp Request,14 = Time Stamp Reply
- 17 = Address Mask Request,18 = Address Mask Reply
- Most of the ICMP messages named last are blocked because of easy misuse (redirection of routes for packet sniffing, spoofing, ...)

ICMP – Destination Unreachable

- Unreachable entities (codes):
  - 0:network
  - 1:host
  - 2:protocol
  - 3:port
  - Destination in general because of:
  - 4: frag needed, but DF set
  - 5: source route failed
- Network Unreachable generated by router lacking any route to destination

ICMP – Destination Unreachable

- Host Unreachable indicates last hop router cannot contact destination
- Protocol Unreachable: host lacks a layer-4 protocol implementation
- Port Unreachable no process bound to port (usually with UDP)
- Code 4 indicates the datagram required fragmentation but the DF bit was set
- Newer implementations replace (unused) 2nd word of ICMP header with next MTU
- MTU info returned to host, where it can subsequently alter its packet size to avoid fragmentation (process path MTU discovery)

ICMP – Further Messages

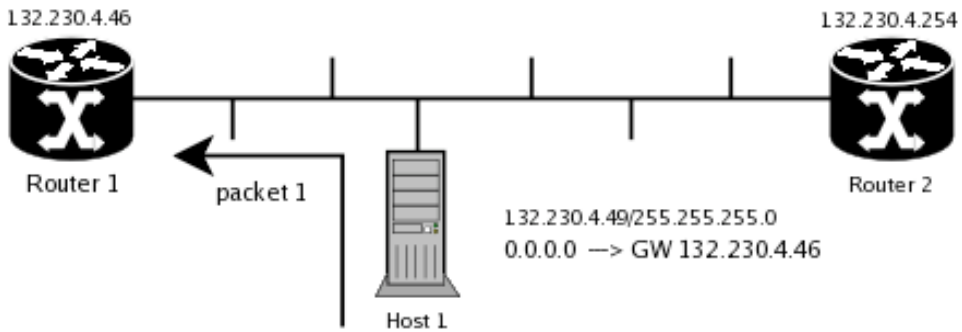
- Source Quench: Initial idea was that routers could generate "slow down" messages
- Problem is generating more traffic during periods of high traffic is not very attractive
- Currently, routers should not generate source quench ICMP messages
  - May generate much additional traffic in already congested networks
  - May interfere with TCP flow control

ICMP – Further Messages

- Time Exceeded (type 11)
- Indicates IP packet's delivery time has been exceeded
- Code field values:
  - 0: TTL exceeded in transit
  - 1: fragment reassembly time exceeded
- Parameter problem (type 12) - General catch-all for any delivery error not otherwise covered
- ICMP Router Solicitation, router advertisement (type 10 – finding nearby routers) is mostly replaced by DHCP which will be discussed next ...

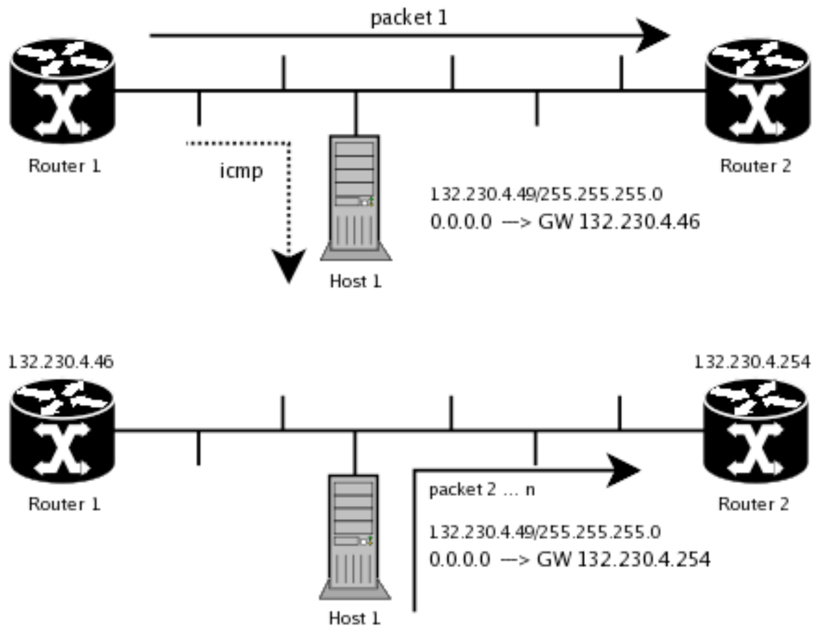
ICMP – Redirect

- Indicates wrong router on network is being used as first hop. Redirect indicates which router to use instead
- Code field values: 0:network, 1:host, 2:TOS & Network, 3:TOS & Host
- May be misused for redirecting traffic from/to a host (sniffing, hijacking packets, ...)



ICMP – Redirect

- Host sends packet to default router (as listed in its routing table)
- Designated router sends ICMP redirect, because default router is in same subnet (one hop could be saved if sent directly)



NAT – Special Routing in IPv4

- Talked of standard concept of IPv4 routing last lecture
- Original idea of IP networking – end-to-end routing (present in the IP header via source and destination address)
- Special requirements, beginning of IPv4 addresses shortage and security considerations introduced NAT
- Network Address Translation (NAT) process of modifying network address information in packet headers while transiting a router
- Idea: Map one address space to an other, typically requiring
  - Rewrite of source and/or destination address in layer 3 IP header
  - And/or rewrite of port numbers in layer 4 headers

NAT – Typology

- Two levels of network address translation.
  - Basic NAT – IP address translation only, rather seldom used e.g. to directly map a routed IP to a machine in a private network
  - Often term Port Address Translation (PAT) or Network Address Port Translation, NAPT – emphasizing the translation of both IP addresses and port numbers
- NAT involving translation of the source IP address and/or source port – source NAT or SNAT
  - Rewriting IP of originating machine, typically the case in masquerading NAT
- NAT involving translation of the destination IP address and/or destination port – destination NAT or SNAT
  - Typical scenario of port forwarding over a NAT router

NAT – IP Masquerading

- DNAT and SNAT often found together in many router setups
- Today: NAT typically synonymous with IP masquerading, where a "private" address space mapped to (single) public IP address(es)
  - Popular from mid-1990's NAT as a tool for alleviating the IPv4 address shortage
  - Especially found in countries with lesser allotted address space than Northern America and Europe
- NAT is not without problems
  - Breaking the concept of end-to-end addressing – the original source of a packet is hidden behind the masquerading gateway
  - Communication does not flow symmetrical any more – 1:n mapping in e.g. masquerading allows uni directional setups of communication channels only

NAT – Problems on Network and Transport Layer

- ICMP problems
  - may or may not correctly parse ICMP packets, depending on whether the payload is interpreted by a host on the "inside" or "outside" of translation
- Checksum recalculation
  - IP header checksum has to be recomputed (changed source and/or destination addresses)
  - Fragmented packets needs to be reassembled to allow higher level checksumming corrected: TCP and UDP use checksums covering their respective headers, the data and a "pseudo-header" with source and destination IP addresses
  - Thus MTU path discovery (RFC 1191, used in IPv6 too) might be a good idea

NAT – Problems on Application Layer

- Special applications
  - FTP in active mode with separate connections for control and data traffic: When requesting a file transfer host behind NAT will fail using its IP address and some port
  - SIP puts IP information (for setup of RTP channels, later lectures) into the application layer headers
  - Application Layer Gateway (ALG) could fix the issue: special software running on a NAT router updating payload data
  - Problem: ALG needed for every affected protocol
  - Another possible solution:
    - NAT traversal techniques like STUN
    - UPnP (Universal Plug and Play) requiring cooperation of the NAT device (security risk)

NAT – Operation Problems

- Stateful NAT tables
  - Router keeps entry for each connection
  - List could grow significantly, slowing down packet processing
  - Typically short living entries in NAT table
    - Failing connections of long living services like SSH
    - Or keep-alive procedures like in SIP could reduce battery saving efforts in mobile devices

ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

Communication Systems

University of Freiburg
Computer Science
Computer Networks and Telematics
Prof. Christian Schindelhauer

CoNe Freiburg

IIF
INSTITUT FÜR
INFORMATIK
FREIBURG