Routing

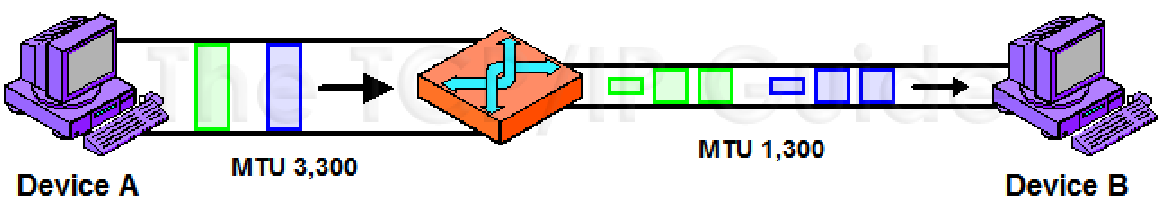
Routing / Indirect Delivery

NAT

Internet Control Messages Protocol (ICMP)

Fragmentation and Reassembly

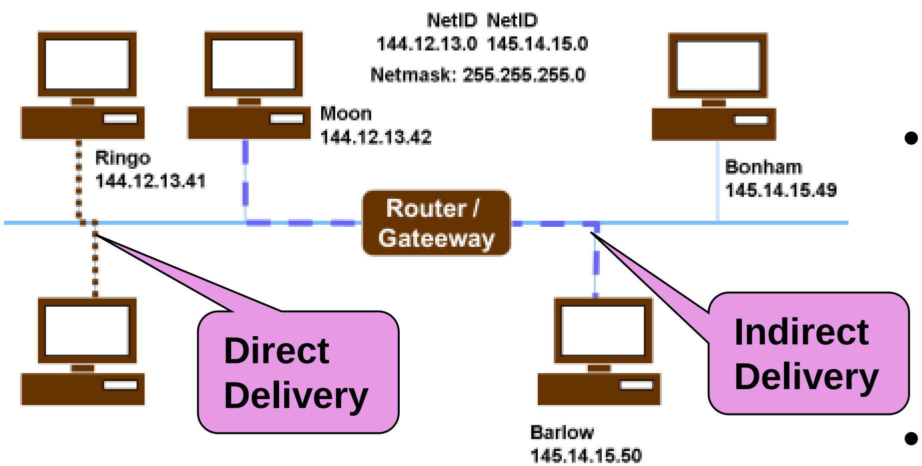
- The IP layer splits data from the higher layer into datagrams that are suitable for transmission over a physical network
- The largest datagram that can be transmitted over a physical network is called the *Maximum Transmission Unit* (MTU)
- Different types of physical networks have different limits on the amount of data that can be sent in one frame, eg Ethernet 1500 bytes, Fibre ~4500 bytes



MTU

- If a datagram fragment is re-fragmented into smaller fragments, the fragments are not reassembled until the final destination is reached
- There are ways to work out the smallest MTU size in a path and use that size from the start. This avoids re-fragmentation.

Routing – Direct Delivery



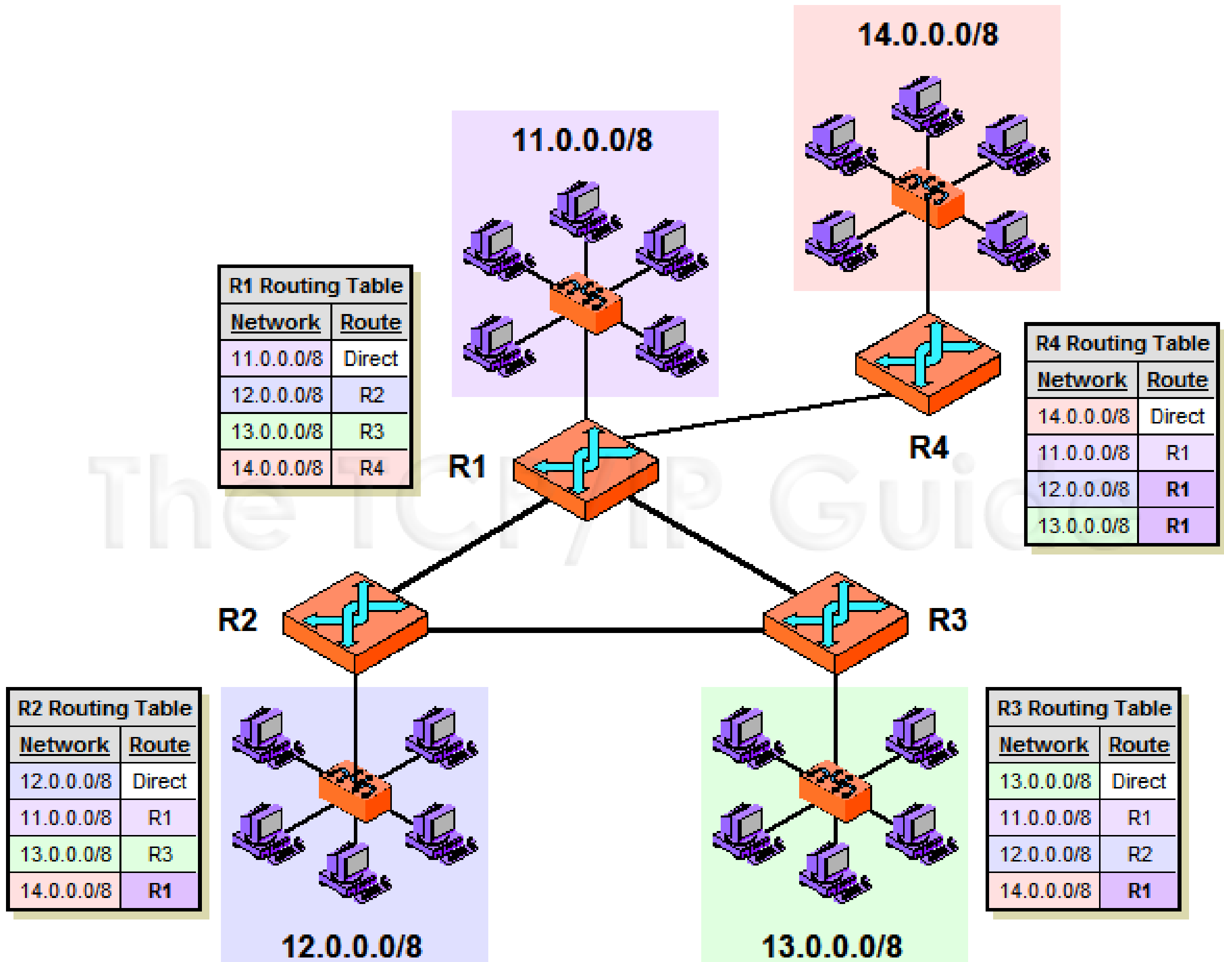
Simplest Indirect Delivery Model

- Direct delivery occurs if the final destination is on the same local network and is relatively straight forward
- Indirect delivery relies on the internet to know how to reach the destination
- The protocol can work out which type by looking at the IP address, specifically, the network address component (refer back to IP addressing)
- If the network address is the same as the current network, then direct delivery is required

Indirect Delivery

- Indirect delivery is implemented by ***<u>“next hop” routing</u>***. Each router knows which routers it is directly connected to.
- For any destination the router only knows the next router it needs to send the datagram to.
- This information is maintained in the routing table
- The routing table information is set up and maintained by means of the Internet Control Message Protocol (ICMP) – discussed later

Routing Table example



IP Network Address Translation (IP NAT)

- Because of the continual growth in the number of devices connected to the internet, there has been a shortage of global IP addresses since the 1990s

- A new IP protocol called IPv6 has been developed, which amongst other things has longer IP numbers (128 bits instead of 32, or about 3×10^{38} compared to 4×10^{9} addresses)
- However this is still not widely used, so in the meant time, NAT has been developed to overcome the shortage of IP addresses

Inside & Outside, Local & Global

- NAT uses the concepts of
  - Inside and Outside addresses
  - Local and Global addresses
- ***Inside address*** are those used to refer to devices within an organization or local area network
- ***Outside addresses*** are public IP addresses

- ***Local addresses*** are addresses used inside a local network, whether they are inside or outside addresses
- ***Global addresses*** are those used in datagrams outside the local network, whether they are inside or outside addresses

Example

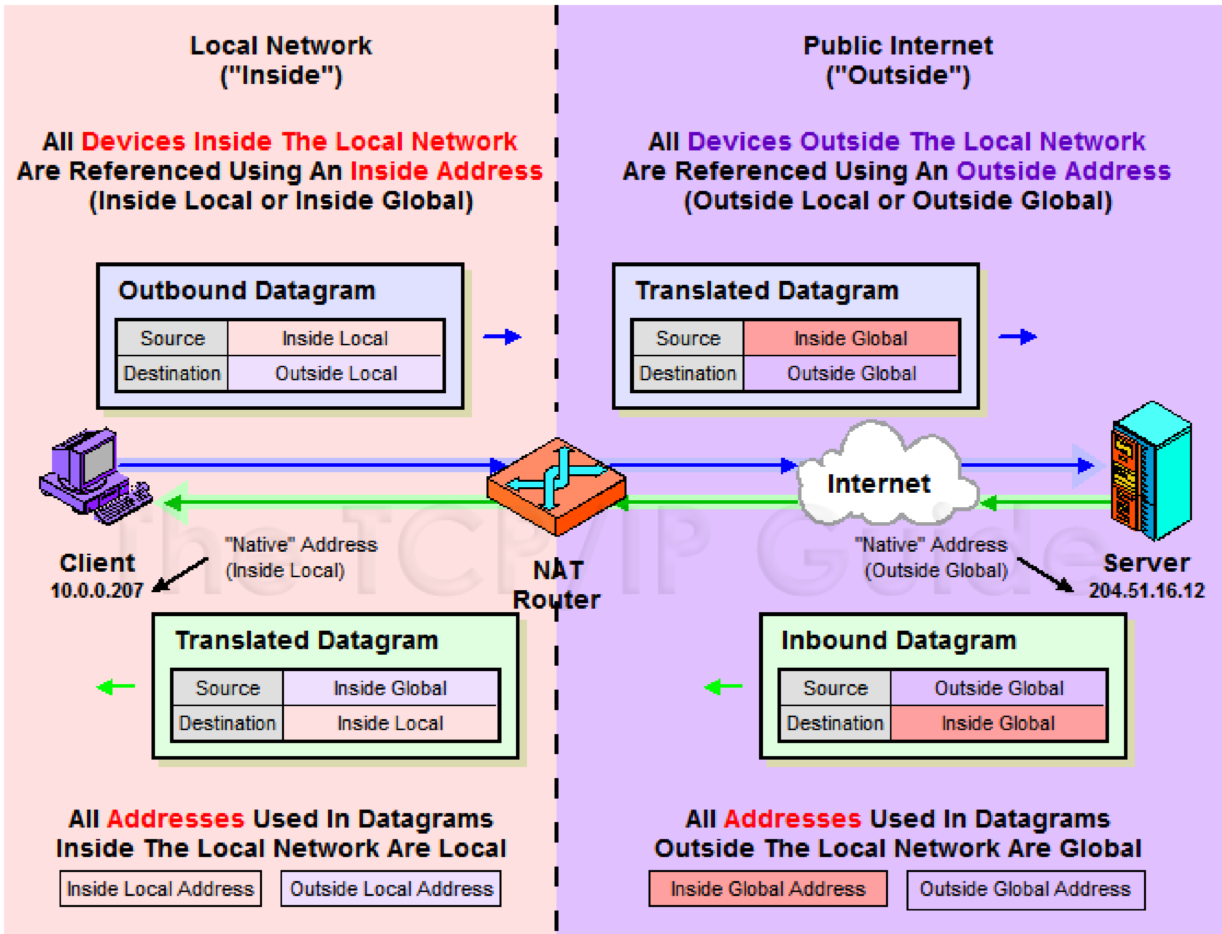
- An organization might have local network with addresses 10.0.0.0 to 10.0.0.255
- To communicate with the outside world it will have a smaller pool of global public addresses, one of which might be 204.51.16.12
- When a local device (eg 10.0.0.207) wants to communicate with a public server, it will use one of the networks public addresses - 204.51.16.12 as the return address. When this address appears in a datagram that is within the organizations local network it is referred to a local outside address

- When the datagram travels outside the local network and into the internet the IP address is referred to as an outside global address.

- Inside the local network the datagram may at various times have either the local or global IP address, but because it is inside the local network the address is considered an inside address

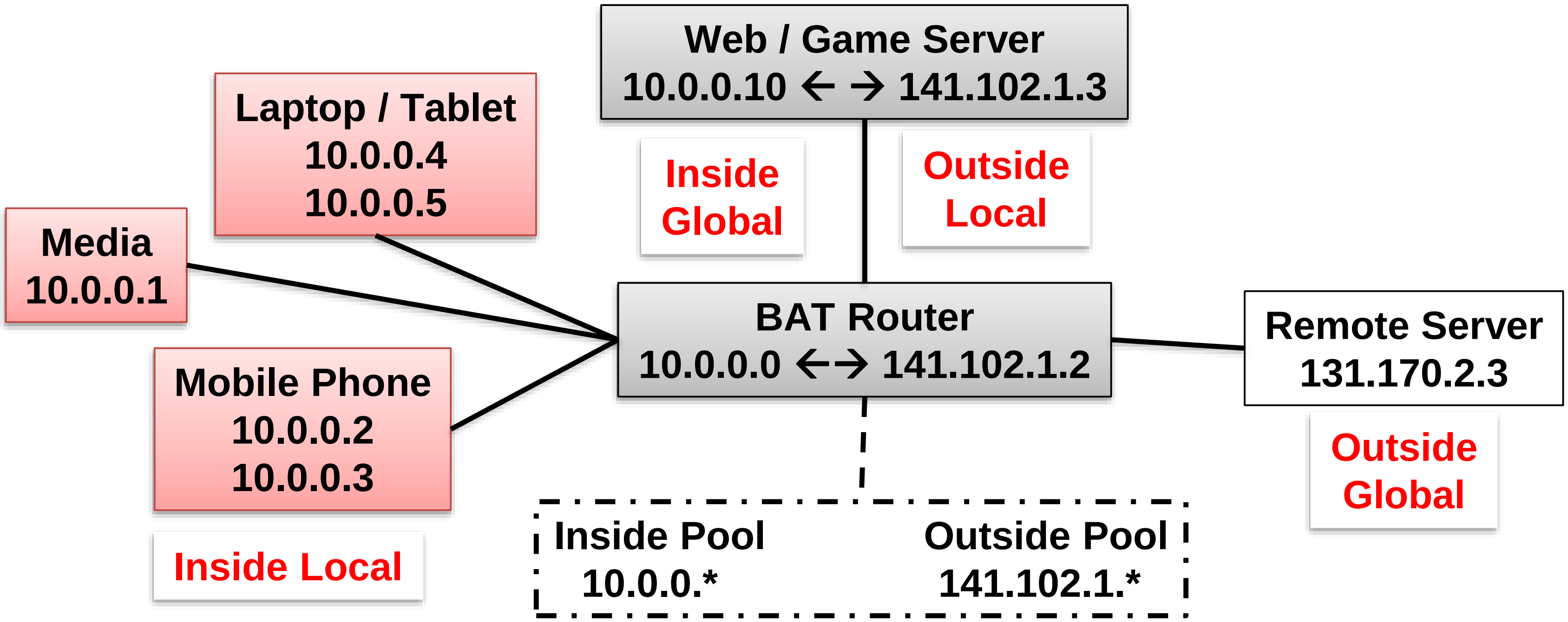
Similarly, when the frame travels into the public internet the IP address is considered to be a global address.

NAT Overview



NAT Summary

- Translation from local to global IP addresses is done by the NAT router, which amongst other things maintains a table of local and global addresses.
- **Global addresses are allocated dynamically from the organizations pool**



Issues

- NAT is actually an inelegant fix and there are many issues that need to be dealt with
- Some higher level applications imbed IP addresses. NAT must watch for these and handle them FTP is one example
- Every time an IP address substitution is made the header checksum must be recalculated
- The ICMP protocol uses IP addresses so NAT needs to understand this and know which IP addresses should not be changed

Internet Control Message (ICMP)

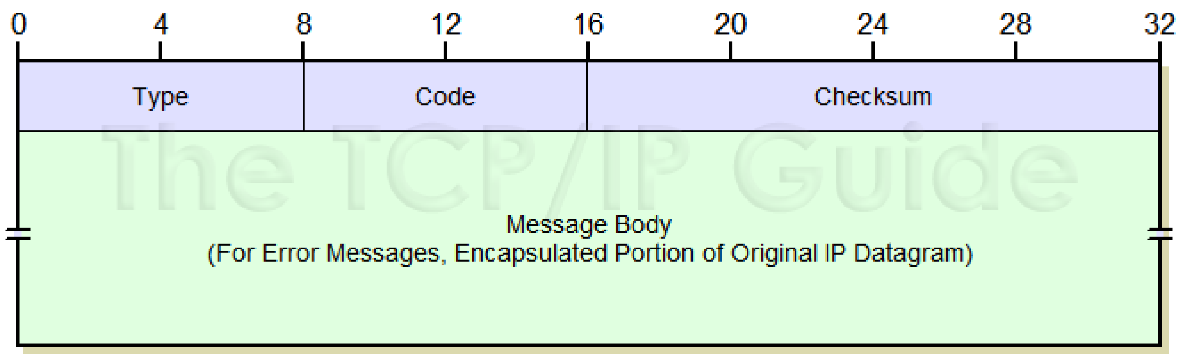
- The ICMP protocol is a layer 3 protocol which is a companion or helper protocol to the IP protocol. ICMP provides the ability to report errors and to provide information, such as the location and address of routers

- Functions
  - Error Reporting
  - Information Messages

- 255 message types each with 256 sub types

ICMP message format

- 1 byte for type
- 1 byte for sub type
- 2 byte checksum
- Body of message



ICMP Message types

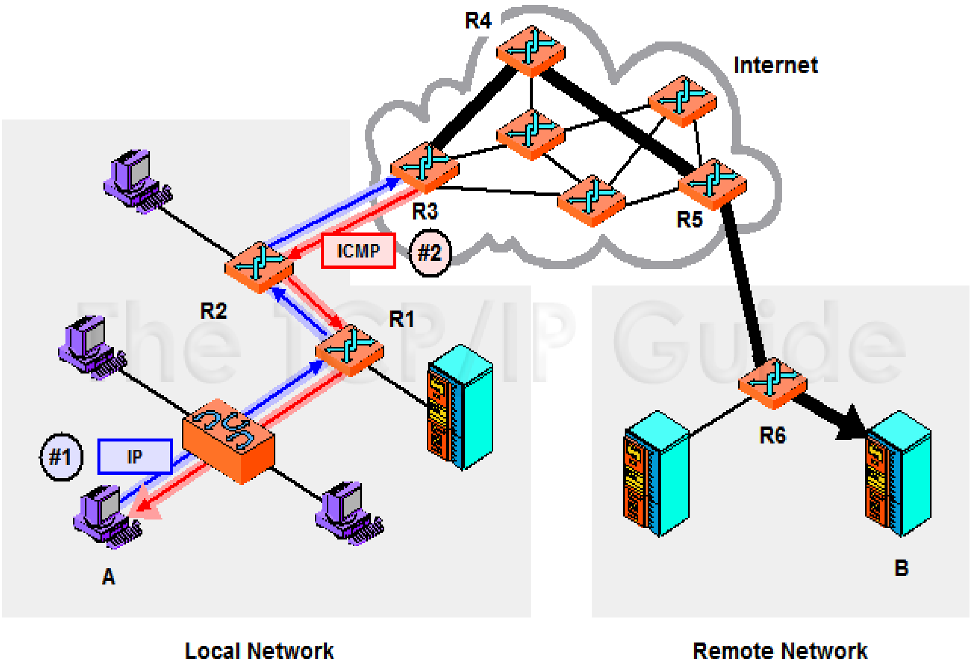
Message Class	Type Value	Message Name	Summary Description of Message Type	Defining RFC Number
ICMPv4 Error Messages	**3**	***Destination Unreachable***	Indicates that a datagram could not be delivered to its destination. The *Code* value provides more information on the nature of the error.	792
	4	***Source Quench***	Lets a congested IP device tell a device that is sending it datagrams to slow down the rate at which it is sending them.	792
	5	***Redirect***	Allows a router to inform a host of a better route to use for sending datagrams.	792
	11	***Time Exceeded***	Sent when a datagram has been discarded prior to delivery due to expiration of its *Time To Live* field.	792
	12	***Parameter Problem***	Indicates a miscellaneous problem (specified by the *Code* value) in delivering a datagram.	792
ICMPv4 Informational Messages (part 1 of 2)	**0**	***Echo Reply***	Sent in reply to an *Echo (Request)* message; used for testing connectivity.	792
	8	***Echo (Request)***	Sent by a device to test connectivity to another device on the internetwork. The word “Request” sometimes appears in the message name.	792
	9	***Router Advertisement***	Used by routers to tell hosts of their existence and capabilities.	1256
	10	***Router Solicitation***	Used by hosts to prompt any listening routers to send a *Router Advertisement*.	1256
	13	***Timestamp (Request)***	Sent by a device to request that another send it a timestamp value for propagation time calculation and clock synchronization. The word “Request” sometimes appear in the message name.	792

IPV4 part 2 Informational messages

14	***Timestamp Reply***	Sent in response to a *Timestamp (Request)* to provide time calculation and clock synchronization information.	792
15	***Information Request***	Originally used to request configuration information from another device. Now obsolete.	792
16	***Information Reply***	Originally used to provide configuration information in response to an *Information Request* message. Now obsolete.	792
17	***Address Mask Request***	Used to request that a device send a subnet mask.	950
18	***Address Mask Reply***	Contains a subnet mask sent in reply to an *Address Mask Request*.	950
30	***Traceroute***	Used to implement the experimental “enhanced” *traceroute* utility.	1393

Error reporting

- Error messages are sent back to the source to correct the problem, not intermediate routers



- This may be problematic if the problem was caused by a router, but will eventually get corrected

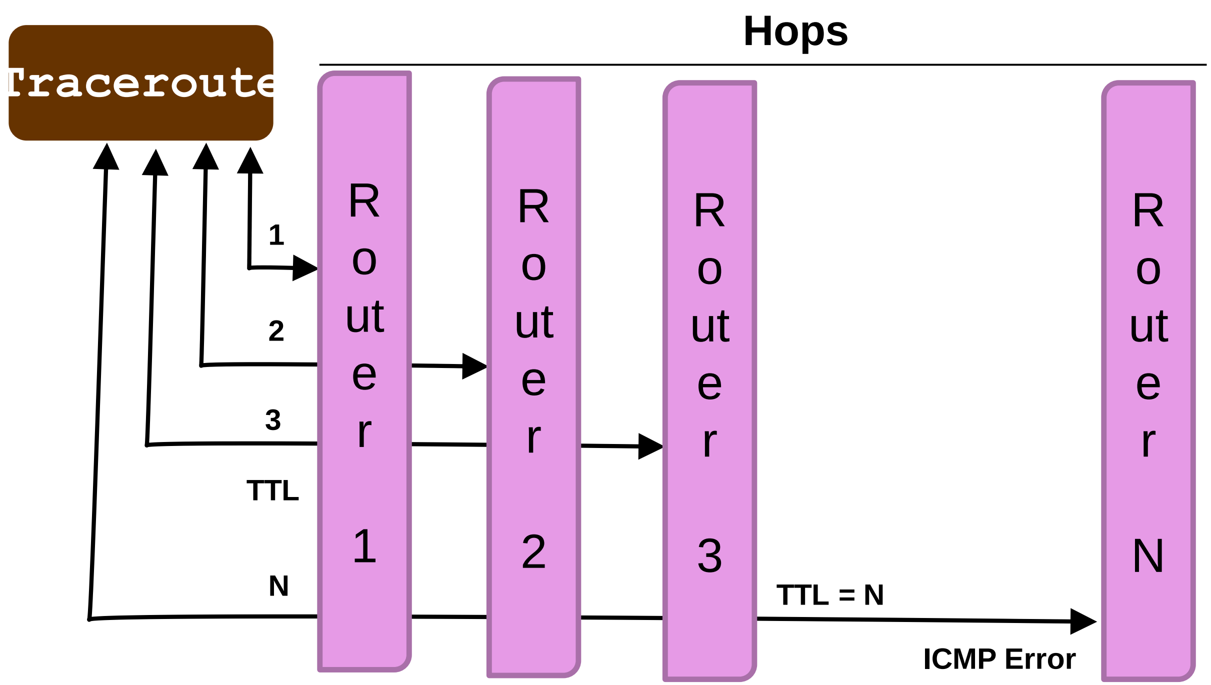
- In error messages, the first part of the body is the header of the datagram that caused the error, which contain the source and destination address, amongst other things.

- It is also possible that because of some other error the ICMP datagram that is reporting an error may not reach its destination. This needs to be handled by higher level protocols

Router advertisement and solicitation

- Two ICMP message types which allow hosts and routers to learn about nearby routers
- Router Advertisement
  - Tells listening devices that the router exists and contains information about its address(es)
  - Usually sent every 7 – 10 minutes by the router, which can be a problem for new host – solved by:
- Router Solicitation
  - Sent by hosts
  - Broadcast on the local network asking any devices that are routers to respond with a router advertisement message

Using ICMP – TraceRoute



- Traceroute successively sends out packets to the destination with a 'time to live' (TTL) starting at 1 and increasing until the destination is reached. Each router decrements the TTL and when it gets to 0, that particular router sends back an ***ICMP*** error identifying itself and its time and date.

- ***Ping*** works in a similar manner, but does not manipulate TTL. It also uses ICMP packets. Ping is often blocked at the host for security reasons.

Using ICMP – TraceRoute

```
traceroute to Zen.pc.helsinki.fi (128.214.62.27), 30 hops max, 40 byte packets
1 gw131.gw.mds.rmit.edu.au (131.170.70.254) 0.375 ms 0.307 ms 0.321 ms
2 b5-vnr.rmit.edu.au (131.170.13.254) 1.417 ms 1.305 ms 1.24 ms
3 b5-xlr.rmit.edu.au (131.170.5.254) 6.873 ms 3.719 ms 3.233 ms
4 131.170.4.254 (131.170.4.254) 1.968 ms * 1.763 ms
5 131.170.75.254 (131.170.75.254) 2.936 ms 3.03 ms 2.909 ms
6 203.21.130.129 (203.21.130.129) 3.2 ms 3.484 ms 5.076 ms
7 192.65.88.193 (192.65.88.193) 18.626 ms 18.305 ms 39.386 ms
8 GigabitEthernet9-0-0.rr1.optus.net.au (202.139.1.197) 76.489 ms 262.656 ms
255.714 ms
9 Hssi4-0-0.sf1.optus.net.au (192.65.89.230) 285.375 ms 337.396 ms 564.301 ms
10 205.174.74.165 (205.174.74.165) 264.398 ms 338.822 ms 322.563 ms
11 s2-0-0.paloalto-cr18.bbnplanet.net (4.1.142.253) 577.819 ms 717.478 ms
12 p3-2.paloalto-nbr2.bbnplanet.net (4.0.3.85) 617.377 ms 648.684 ms *
13 p4-0.paloalto-nbr1.bbnplanet.net (4.0.5.65) 651.288 ms 826.48 ms 873.934 ms
14 p1-0.paix-bi2.bbnplanet.net (4.0.6.98) 598.376 ms * 603.963 ms
15 p7-0.paix-bi1.bbnplanet.net (4.0.3.141) 616.007 ms 808.929 ms 668.49 ms
16 4.0.3.217 (4.0.3.217) 638.838 ms 907.203 ms 687.141 ms
17 if-8-0.core1.NewYork.Teleglobe.net (207.45.222.178) 884.507 ms 662.875 ms *
18 207.45.202.22 (207.45.202.22) 841.245 ms 860.958 ms 839.009 ms
19 sw-gw.nordu.net (193.10.252.185) 943.884 ms 973.134 ms 852.775 ms
20 e-gw.nordu.net (193.10.252.209) 859.486 ms 933.297 ms 963.172 ms
21 fi-gw.nordu.net (193.10.252.202) 859.348 ms * 896.598 ms
22 funet3-a3001-funet1.funet.fi (193.166.187.174) 839.12 ms 809.535 ms 888.153 ms
23 helsinki1-a203-funet3.funet.fi (193.166.187.134) 985.698 ms * *
24 128.214.231.10 (128.214.231.10) 912.567 ms 970.315 ms *
25 128.214.62.27 (128.214.62.27) 922.567 ms 980.315 ms *
```

Three packets sent.
ms = reply delay for each
* = no response

rraceroute in Linux in 2010

```
traceroute to Zen.pc.helsinki.fi (128.214.62.27), 30 hops max, 40 byte packets
1 gw131.gw.mds.rmit.edu.au (131.170.70.254) 0.375 ms 0.307 ms 0.321 ms
2 b5-vnr.rmit.edu.au (131.170.13.254) 1.417 ms 1.305 ms 1.24 ms
3 b5-xlr.rmit.edu.au (131.170.5.254) 6.873 ms 3.719 ms 3.233 ms
4 131.170.4.254 (131.170.4.254) 1.968 ms * 1.763 ms
5 131.170.75.254 (131.170.75.254) 2.936 ms 3.03 ms 2.909 ms
6 203.21.130.129 (203.21.130.129) 3.2
7 192.65.88.193 (192.65.88.193) 18.62
8 GigabitEthernet9-0-0.rr1.optus.net.
255.714 ms
9 Hssi4-0-0.sf1.optus.net.au (192.65.
10 205.174.74.165 (205.174.74.165) 26
11 s2-0-0.paloalto-cr18.bbnplanet.net
12 p3-2.paloalto-nbr2.bbnplanet.net (
13 p4-0.paloalto-nbr1.bbnplanet.net (
14 p1-0.paix-bi2.bbnplanet.net (4.0.6
15 p7-0.paix-bi1.bbnplanet.net (4.0.3
16 4.0.3.217 (4.0.3.217) 638.838 ms 9
17 if-8-0.core1.NewYork.Teleglobe.net
18 207.45.202.22 (207.45.202.22) 841.
19 sw-gw.nordu.net (193.10.252.185) 9
20 e-gw.nordu.net (193.10.252.209) 859.486 ms 933.297 ms 963.172 ms
21 fi-gw.nordu.net (193.10.252.202) 859.348 ms * 896.598 ms
22 funet3-a3001-funet1.funet.fi (193.166.187.174) 839.12 ms 809.535 ms 888.153 ms
23 helsinki1-a203-funet3.funet.fi (193.166.187.134) 985.698 ms * *
24 128.214.231.10 (128.214.231.10) 912.567 ms 970.315 ms *
25 128.214.62.27 (128.214.62.27) 922.567 ms 980.315 ms *
```

Question?

- What is the circumference of the earth?

  c = 40,000 km = 40 Mm

  (metric system is based on this!!)

- What is the speed of light?

  v = $3x10^8$ m / s = 0.3 Gm / s

  Actually, though fiber its 66% or $2x10^8$ ms^{-1}

- So how long would it take light to travel halfway around the world?

  d = c / 2 *(eg Australia to Finland as in this example)*

  t = d / v

  = 20,000/200,000 = 2/15

  = 0. 1 s = 100 ms

Three packets sent.
ms = reply delay for each
* = no response

tracert in Windows in 2020

- $ tracert 128.214.62.27
- Tracing route to lh9-570-21104.biosci.helsinki.fi [128.214.62.27]
- over a maximum of 30 hops:
- 1 2 ms 11 ms 11 ms mymodem.modem [192.168.0.1]
- 2 23 ms 20 ms 30 ms gateway.vb05.melbourne.asp.telstra.net [58.162.26.197]
- 3 18 ms 17 ms 16 ms ae255.win-ice301.melbourne.telstra.net [203.50.62.188]
- 4 29 ms 25 ms 28 ms bundle-ether25.win-core10.melbourne.telstra.net [203.50.61.144]
- 5 30 ms 29 ms 29 ms bundle-ether12.ken-core10.sydney.telstra.net [203.50.11.122]
- 6 29 ms 29 ms 31 ms bundle-ether1.pad-gw11.sydney.telstra.net [203.50.6.61]
- 7 32 ms 36 ms 30 ms bundle-ether1.sydp-core04.sydney.reach.com [203.50.13.90]
- 8 168 ms 179 ms 164 ms i-20802.eqnx-core02.telstraglobal.net [202.84.141.25]
- 9 169 ms 166 ms 171 ms i-0-0-0-1.paix-core02.telstraglobal.net [202.84.143.210]
- 10 165 ms 164 ms 164 ms i-92.paix02.telstraglobal.net [202.84.247.41]
- 11 * * * Request timed out.
- 12 * * * Request timed out.
- 13 * * * Request timed out.
- 14 * * * Request timed out.
- 15 229 ms 232 ms 231 ms ae-32.2603.rtsw2.ashb.net.internet2.edu [64.57.21.53]
- 16 228 ms 228 ms 228 ms 64.57.21.54
- 17 305 ms 304 ms 304 ms uk-hex.nordu.net [109.105.97.140]
- 18 330 ms 323 ms 324 ms dk-uni.nordu.net [109.105.97.126]
- 19 335 ms 325 ms 330 ms dk-bal2.nordu.net [109.105.97.222]
- 20 347 ms 346 ms 350 ms se-tug.nordu.net [109.105.102.102]
- 21 342 ms 357 ms 344 ms helsinki3.ip.funet.fi [109.105.102.103]
- 22 352 ms 352 ms 359 ms helsinki1-et-0-1-1-1.ip.funet.fi [86.50.255.246]
- 23 365 ms 354 ms 352 ms r1.helsinki.kampus.funet.fi [193.167.244.245]
- 24 352 ms 356 ms 355 ms funet2-up2.fe.helsinki.fi [128.214.173.185]
- 25 355 ms 351 ms 351 ms up2-funet2.fe.helsinki.fi [128.214.173.186]

In 5 year the transit speed has tripled.

- 30 * * * Request timed out.