Today’s Lecture

I. ICMP Overview
II. ICMP Error Reporting
III. ICMP Query / Response Messages
IV. ICMP Message Processing

ICMP OVERVIEW

ICMP (RFC 792)
• Communicates…
  – network-level errors
  – information about unexpected circumstances
  – information about the network, in response to queries

What Layer is ICMP?

RFC-792 – “ICMP uses the basic support of IP as if it were a higher level protocol, however, ICMP is actually an integral part of IP, and must be implemented by every IP module.”
ICMP Message Format

- Checksum over entire ICMP message
- ICMP Data usually contains:
  - IP header (including Options, but normally = 20 bytes) of datagram that caused error
  - at least 8 bytes of data from this datagram (usually includes fields needed to identify the cause of the error)

--- Rest of ICMP Header ---

Code

ICMP Message Checksum

Type

ICMP Data (Original IP Header + 8 bytes datagram)

ICMP Message Types

- Error Reporting
- Query & Response

Why ICMP for Reporting Errors?

- Protocol-specific messages?
- For what protocols or functions are ICMP error messages appropriate?

When Not to Send ICMP Error Messages

- An ICMP error message is never generated in response to:
  1. an ICMP error message
  2. a datagram whose source address does not define a single host (address cannot be zero, loopback, broadcast, multicast)
  3. A datagram whose destination address is an IP broadcast address
  4. a datagram sent as a link-layer broadcast
  5. a fragment other than the first one of a datagram
- For each of the above, why?

#1 Destination Unreachable Msgs

- Upon failure to forward/deliver, router sends ICMP message to source before "dropping" datagram
  - IP is best-effort delivery, but discarding datagrams should not be taken lightly
- Several reasons for failure (next slide), but...
  - not all errors can be diagnosed properly (e.g., host IP address changes)
Reasons for Destination Unreachable Messages

- Network unreachable (reason?)
- Host unreachable (reason?)
- Protocol (TCP, UDP) not enabled
- Port not bound to a service
- Fragmentation needed, but DF Flag set
- Source route failed

Path MTU Discovery (RFC 1191)

- Host sets DF Flag and transmits a large datagram
- If datagram size exceeds MTU on some link, the router discards datagram and sends back ICMP Destination Unreachable message
  - message includes size of Next Hop MTU

<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>ICMP Message Checksum</th>
<th>ICMP Data (Original IP Header + 8 bytes datagram)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>(unused)</td>
<td>Next Hop MTU</td>
</tr>
</tbody>
</table>

Path MTU Discovery (cont’d)

- Host receiving this error message knows to reduce maximum packet size to the Next Hop MTU
- Periodically host will increase the packet size and try again
  - Why?

#2 Time Exceeded ICMP Message

- Sent if router has detected that the hop count (TTL) has reached zero (code 0)
  - usually means a routing error (loop) occurred
  - why would loops occur if routing protocols work right?
- Or, sent if destination host timeout occurred while waiting for fragments to arrive (code 1)
  - normal timeout interval on the order of 60-120s

#3 Router Redirect Messages

- Hosts normally initialize their forwarding table from a (static) configuration file at startup
  - contains minimal info (e.g., address of single default gateway) for simplicity
  - if network topology changes, this info is obsolete
  - how learn of such changes (host don’t run routing protocols)?
- Redirect messages do not solve the problem of propagating routes in a general way
  - dynamic routing protocols are used for this

#3 Router Redirect Messages (cont’d)

1. When router detects a host using a “suboptimal” route…
2. send ICMP Redirect message to the host requesting that it change its forwarding table
3. forwards original datagram towards its destination
- How to detect a suboptimal route?
  - if router forwards packet out the same interface it came in on
- Which routes should be updated; only for this specific destination?
Example

- Packet from A to B should go through R2, but A sends to R1 first (i.e., A is misconfigured)

Redirect Message Format

- 3 addresses needed
  - IP address that caused redirect (in "Original IP Datagram" header)
  - IP address of router that sent redirect (in IP header of ICMP Message datagram)
  - correct router IP address (in Redirect message)

Restrictions on Redirection

- Redirect messages sent only by “first hop” router
- No Redirect message if Source Routing Option present
- + a few more restrictions (not covered here)

#4 Congestion and Datagram Flow Control

- IP is connectionless
  - does not reserve buffer space or bandwidth
  - potential for congestion, resulting in packet dropping by routers
- ICMP Source Quench message was used to report congestion to original source
  - a host receiving this message is expected to slow down
  - no ICMP message exists to reverse the effect of a source quench

#4 Congestion and Datagram Flow Control (cont’d)

- Not used any more
  - tends to create rather than solve congestion (why?)
  - congestion control in the Internet is now done mostly in the transport layer

#5 “Parameter Problem” Message

- Some error was detected in the IP header
- Pointer indicates byte offset from start of IP header to the “offending” parameter

<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>ICMP Message Checksum</th>
<th>Pointer</th>
<th>ICMP Data (Original IP Header + 8 bytes datagram)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Data</td>
</tr>
</tbody>
</table>
A Clever Program: traceroute

- Allows us to see the path taken by the packet
  - why not just use IP record route option?

1. Send UDP datagram with TTL=1
   - first router decrements TTL, notices it is 0, sends ICMP Time Exceeded error message back to sender
   - this error message has IP address of the incoming interface of the router generating the error – now we know the first hop!

2. Now send UDP datagram with TTL=2
   - second router sends back “time exceeded” message, with its IP address

3. Etc...

4. How tell when the destination is reached?
   - the UDP datagram is addressed to an “unlikely” port (>30,000)
   - error message sent back by destination is Destination Unreachable (“port not bound to a service”) ICMP error message

Example Traceroute Output

```
1. > tracert www.netf.org
2. Querying the network with icmp
3. 1.2.3.4
   Time to live is set to 30
   Request timer is 4 seconds
   Querying www.netf.org [130.155.6.72] over a maximum of 30 hops
   1. 127.0.0.1 [127.0.0.1] queried at 0 ms [0.000 ms]
   2. 192.168.0.1 [192.168.0.1] queried at 2 ms [2.000 ms]
   3. 10.0.0.1 [10.0.0.1] queried at 5 ms [5.000 ms]
   4. 192.168.100.1 [192.168.100.1] queried at 8 ms [8.000 ms]
   6. 192.168.300.1 [192.168.300.1] queried at 14 ms [14.000 ms]
   7. 192.168.400.1 [192.168.400.1] queried at 17 ms [17.000 ms]
   8. 192.168.500.1 [192.168.500.1] queried at 20 ms [20.000 ms]
   9. 192.168.600.1 [192.168.600.1] queried at 23 ms [23.000 ms]
   10. 192.168.700.1 [192.168.700.1] queried at 26 ms [26.000 ms]
   11. 192.168.800.1 [192.168.800.1] queried at 29 ms [29.000 ms]
   12. 192.168.900.1 [192.168.900.1] queried at 32 ms [32.000 ms]
   13. 192.168.1000.1 [192.168.1000.1] queried at 35 ms [35.000 ms]
   14. 192.168.1100.1 [192.168.1100.1] queried at 38 ms [38.000 ms]
   15. 192.168.1200.1 [192.168.1200.1] queried at 41 ms [41.000 ms]
   16. 192.168.1300.1 [192.168.1300.1] queried at 44 ms [44.000 ms]
   17. 192.168.1400.1 [192.168.1400.1] queried at 47 ms [47.000 ms]
   18. 192.168.1500.1 [192.168.1500.1] queried at 50 ms [50.000 ms]
   19. www.netf.org [130.155.6.72] queried at 53 ms [53.000 ms]

Trace complete.
```

Example Traceroute Output

- “tracert” on Windows machines

QUERYING THE NETWORK WITH ICMP

#1 Echo Request and Reply Messages

- Used to see if destination interface is reachable and functioning

```
Type 8 or 0  Code 0  ICMP Message Checksum
Identifier  Sequence Number
Data
```

- Echo Request
  - Contains Identifier and Sequence Numbers to help match Replies with Requests

- Echo Reply
  - is not mandated! reasons for not sending Echo Reply?
  - data sent by Request must be returned in Reply

Program Using Echo Request: ping

- Even if you can’t ping a host, it might be reachable (i.e., ping is disabled on that host but other services are not)

- Identifier = process number of application sending the ping

- Sequence Number starts at 0 and is incremented by each successive Request
  - can tell if replies are missing, duplicated, or reordered

- Round-trip time can be calculated
  - client puts sending time into Request, subtracts from receiving time when Reply comes back
ping Example

```plaintext
ping -s kronos.csc.ncsu.edu
Pinging kronos.csc.ncsu.edu [130.207.8.17] with 56 bytes of data:
64 bytes from kronos.csc.ncsu.edu [130.207.8.17]: icmp_seq=0 time=47 ms
64 bytes from kronos.csc.ncsu.edu [130.207.8.17]: icmp_seq=2 time=47 ms
64 bytes from kronos.csc.ncsu.edu [130.207.8.17]: icmp_seq=3 time=48 ms
64 bytes from kronos.csc.ncsu.edu [130.207.8.17]: icmp_seq=4 time=38 ms
```

Clock Synchronization

- Each machine maintains its own notion of the current time
  - clocks that differ widely can confuse users of distributed system software
- To synchronize clocks, you need an estimate of round-trip delay
  - simplest technique: ICMP Timestamp Request & Reply messages

Clock Synchronization (cont’d)

- Reported in milliseconds since midnight, coordinated universal time (UTC)
- Sending time = request received – request transmitted
- Receiving time = response received – reply transmitted
- RTT = Sending time + Receiving time
  - not affected by synchronization problems (why not?)

#3 Router Discovery (RFC 1256)

- Routers advertise their presence to hosts
  - using either limited broadcast, or a special multicast address
- Preference level indicates the “desirability” as a default gateway
- Router Advertisement message

```
<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>ICMP Message Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```

#2 Timestamp Request/Reply Messages

- Reported in milliseconds since midnight, coordinated universal time (UTC)
- Sending time = request received – request transmitted
- Receiving time = response received – reply transmitted
- RTT = Sending time + Receiving time
  - not affected by synchronization problems (why not?)

RTT Estimation Problems

- Accurate estimation of round-trip delay can be difficult
  - round-trip delays over Internet may have high variance
  - datagrams can be dropped, delivered out of order \( \rightarrow \) taking many measurements may not guarantee consistency
- Alternative 1: Network Time Protocol (RFC 1305)
  - much more sophisticated (and complicated)
  - ms accuracy in LAN/WAN
- Alternative 2: GPS receivers at every node
  - µs accuracy, but cost and other limitations?

#2 Timestamp Request/Reply Messages (cont’d)

- Reported in milliseconds since midnight, coordinated universal time (UTC)
- Sending time = request received – request transmitted
- Receiving time = response received – reply transmitted
- RTT = Sending time + Receiving time
  - not affected by synchronization problems (why not?)

#3 Router Discovery (RFC 1256)

- Routers advertise their presence to hosts
  - using either limited broadcast, or a special multicast address
- Preference level indicates the “desirability” as a default gateway
- Router Advertisement message

```
<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>ICMP Message Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```
#3 Router Discovery (RFC 1256) (cont’d)

- Does not indicate what route a host should use to reach a specific destination!

- Routers periodically broadcast (or multicast) this information
  - time between advertisements roughly every 10 minutes
  - default lifetime is 30 minutes
  - disabling a router interface: advertise with a lifetime of 0

- Hosts can request this information
  - on bootup, host broadcasts a Router Solicitation message

#4 Address Mask Request / Reply (RFC 950)

- Subnet masks needed for classless addressing / routing (we will discuss this later)

- Host sends Subnet Mask Request to its gateway

- ICMP Subnet Mask Reply message contains the 32 bit mask for the subnet from which the request was received

```
<table>
<thead>
<tr>
<th>Type (0-30)</th>
<th>Code (0-15)</th>
<th>Description</th>
<th>Query / Error</th>
<th>Result / Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>Source quench</td>
<td>E</td>
<td>reduction in TCP send rate</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Redirect for host</td>
<td>E</td>
<td>updates routing table</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>Echo request</td>
<td>Q</td>
<td>send a reply</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>Router advertisement</td>
<td>R</td>
<td>updates routing table</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>Router solicitation</td>
<td>Q</td>
<td>send a advertisement</td>
</tr>
</tbody>
</table>
```

Processing of ICMP Messages

- ICMP covers a wide range of conditions

- Each message handled differently, e.g…
  - ignored (source quench to UDP)
  - handled by kernel (redirect, source quench to TCP)
  - passed to user process (time exceeded, echo/timestamp reply)
  - discarded (if no user processes have registered with the kernel to receive ICMP messages)
  - …

Processing of ICMP Messages (cont’d)

```
<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>Description</th>
<th>Query / Error</th>
<th>Result / Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Echo reply</td>
<td>R</td>
<td>(used by ping)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Network unreachable</td>
<td>E</td>
<td>application request fails</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Host unreachable</td>
<td>E</td>
<td>application request fails</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Protocol unreachable</td>
<td>E</td>
<td>application request fails</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>Fragmentation needed but DF Flag set</td>
<td>E</td>
<td>reduce packet size</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>Source route failed</td>
<td>E</td>
<td>respectively route</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>Other reasons</td>
<td>E</td>
<td>-</td>
</tr>
</tbody>
</table>
```
### Processing of ICMP Messages (cont'd)

<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>Description</th>
<th>Q / R / E</th>
<th>Result / Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>0</td>
<td>Time exceeded (TTL=0)</td>
<td>E</td>
<td>application request fails</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>IP header bad</td>
<td>E</td>
<td>?</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>Timestamp request</td>
<td>Q</td>
<td>send a reply</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>Timestamp reply</td>
<td>R</td>
<td>application calculates RTT</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>Address mask request</td>
<td>Q</td>
<td>send a reply</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>Address mask reply</td>
<td>R</td>
<td>update mask for interface</td>
</tr>
</tbody>
</table>

### Summary

1. ICMP is a “swiss army knife” for lots of problems and small functions
   - common protocol for reporting error conditions
   - also used to query network conditions
2. Some older ICMP functions have been superseded by more powerful, and specialized, protocols
3. ICMP continues to be extended for new purposes

### Next Lecture

- User Datagram Protocol (UDP)