Advanced Unix System Administration

Lecture 11
March 31, 2008

Steven Luo
<sluo+decal@OCF.Berkeley.EDU>

Ethernet

- What is Ethernet?
  - Ken Thompson: “Robert Metcalfe says that if something comes along to replace Ethernet, it will be called 'Ethernet'”
  - The physical layer has varied enormously over the last 30 years: 10 Mbit coaxial cable, 100 Mbit twisted pair, 10 Gbit fiber-optic
  - Actual electrical (optical) standard has varied over the years

Ethernet

- Ethernet topology
  - Classic Ethernet is a bus topology – devices share one medium
    - Only one device can transmit at one time
    - Devices use CSMA/CD to handle access to the medium
    - All devices examine all transmissions to determine whether or not it's intended for them
    - Devices identified by MAC address
    - Doesn't scale well with large networks
  - Switched Ethernet only forwards transmissions to the wanted device

Ethernet

- The Ethernet frame
  - Each Ethernet transmission is packaged into a “frame”
  - Several different Ethernet frame types, which can coexist on the same network
  - “Ethernet II” is most common
    - 14-byte header containing source, destination MAC addresses, and an EtherType
    - Payload; 46-1500 bytes (hence the 1500 byte MTU)
    - 4 byte CRC checksum

Ethernet

- The EtherType
  - Supposedly was originally a length field
  - In Ethernet II, is a 16-bit value greater than 0x0600 used to indicate the network layer protocol contained in the payload
  - Common values:
    - 0x0800 – IPv4
    - 0x0806 – ARP
    - 0x86DD – IPv6

IP: the Internet Protocol

- Packet-switched data protocol for use on unreliable networks
- Responsible for identifying computers and getting packets from A to B
- Addressing
  - 32-bit addresses (4.2 billion possible unique addresses)
  - Usually written in four decimal quads of 8 bits each, for convenience
    - This form is frequently misleading!

IP: the Internet Protocol

- Address resolution
  - There has to be a way of mapping IP addresses to hardware devices
  - The Address Resolution Protocol (ARP) deals with this
    - Devices broadcast an ARP packet: either a request for the physical address of a particular IP address
    - The matching device broadcasts a reply
    - Devices may also broadcast “gratuitous” ARP announcements
    - Address mappings are (or should be) cached by devices

IP: the Internet Protocol

- Networking
  - The IP address is divided into a network portion (n high bits) and a host portion (remaining low bits)
  - Hosts only speak directly to other hosts with the same network ID
  - Network mask (bitmask ANDed with the IP to obtain network ID) is usually written in dotted-quad or in CIDR notation (/n, where n is the number of bits)

IP: the Internet Protocol

- An IP(v4) packet
  - 4-bit version field, 4-bit header length field, 8-bit TOS field (now used for DiffServ and ECN) are first
  - Total length – at least 20 bytes, but can be much larger
  - IPID field – supposed to be used for identifying fragments of an IP packet
  - Flags – 3 bits for fragment control (DF, MF); the high bit is unused (but see RFC 3514)

IP: the Internet Protocol

- An IP(v4) packet con't
  - Fragment offset – 13 bits describing the offset of a fragment in the payload, in 8-byte blocks
  - TTL – 8 bit hop count to ensure that traffic either arrives at its destination or dies
  - Protocol – 8 bits identifying transport layer protocol in the payload
    - Common protocol IDs: 1 (ICMP), 6 (TCP), 17 (UDP), 41 (IPv6 in IPv4)
    - Look in /etc/protocols for a complete list

IP: the Internet Protocol

- An IP(v4) packet con't
  - Checksum – 16 bits for error checking of the header
  - Source address, destination address – 32 bits each
  - Options (optional header fields), padded to an integral multiple of 32 bits
  - Data (on Ethernet, up to 1480 bytes)

IP: the Internet Protocol

- Internet Control Message Protocol (ICMP)
  - Encapsulated in IP, but not usually considered network layer
  - Used for transmitting status information
  - Packet contents: type, code, checksum, variable header information, up to 64 bytes of original datagram
  - Common types: 3 (destination unreachable), 8 (echo), 0 (echo reply)

IP: the Internet Protocol

- Routing
  - IP packets are routed by their network ID (or possibly only the higher-order bits of it)
  - Kernel maintains a routing table of networks that it knows about and where to send packets bound for them
  - Where multiple routing table entries match, the “best” one (whether by a metric, or by prefix length match) is used
  - A “default” route catches all packets not matched by other routes

IP: the Internet Protocol

- Maintaining a routing table
  - Simplest: keep a static routing table
    - Good for small networks, especially where a small number of routers handle all outbound traffic
    - Not practical on the routers themselves, or as networks get larger
  - Dynamic routing tables
    - Use in-band protocols to discover the best routes
    - Two types: interior gateway protocols (IGP) for routing inside big networks, and exterior gateway protocols for inter-network routing

IP: the Internet Protocol

- Routing Information Protocol (RIP)
  - Each router transmits its routing table to others every 30 seconds
  - Bellman-Ford algorithm with hop count as metric used to calculate best route to host
  - Hop count metric severely limits scalability
- Open Shortest Path First (OSPF)
  - Link-state protocol using Dijkstra's algorithm
  - Hierarchical distribution of link state updates
  - Used for big networks

IP: the Internet Protocol

- Border Gateway Protocol (BGP)
  - Used for routing between networks on the Internet – “network” defined as routing organization with an AS number
  - Routers manually configured to talk to others to share routing table prefixes and state information
  - Routing is affected by configured policy, not just by cost metrics
  - Detailed discussion beyond the scope of this course