Advanced Unix System Administration

Lecture 12
October 27, 2008

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

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

How routing doesn't work:

“[...] the Internet is not something that you just dump something on. It's not a big truck. **It's a series of tubes.** And if you don't understand, those tubes can be filled and if they are filled, when you put your message in, it gets in line and it's going to be delayed by anyone that puts into that tube enormous amounts of material, enormous amounts of material.”

- Sen. Ted Stevens (R-AK)

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

IP: the Internet Protocol

- IPv6
  - IPv4 address exhaustion a looming issue (possibly as early as 2010)
    - Is already an issue in Japan and developing countries, which got less of the address allocation than existing developed countries
  - Complex CIDR network topology means that the IPv4 routing table is growing exponentially
  - IP needs to catch up with advances in routing technology and with the modern security environment

IP: the Internet Protocol

- IPv6 con't
  - Addressing: same idea as IPv4, but 128-bit addresses instead of 32-bit
    - Usual notation: 8 parts in hex separated by colons
  - By convention, high 64 bits of the IP used for the network ID, low 64 bits for the host ID
    - The (excessively) large host ID allows for stateless autoconfiguration via MAC or EUI-64 addresses
    - Router just announces prefix, clients automagically configure themselves
    - DHCPv6 available for more complex setups

IP: the Internet Protocol

- IPv6 con't
  - The original concept: hierarchical allocations and hierarchical routing
    - Regional Internet Registries get very large blocks (/16) from IANA
    - RIRs allocate /32 or larger blocks to ISPs
    - ISPs allocate /48 to organizations, /64 to known small subscribers
    - Prefixes to be allocated to facilitate prefix aggregation; would make the IPv6 routing table much smaller and easier to manage

IP: the Internet Protocol

- IPv6 con't
  - Failure of the hierarchical allocation model
    - Large organizations insisted on provider-independent (PI) address space
      - Renumbering is difficult, so PI addresses make switching ISPs easier
      - Also makes multihoming (connecting to more than one ISP) easier
    - Large endpoints can therefore now get /32 allocations on their own
    - With this allocation model, the IPv6 routing table may get to be just as large – or larger! – than the IPv4 routing table