Advanced Unix System Administration

Lecture 10
March 1, 2007

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

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

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
  - 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 – IPv6 routing table will always be much shorter than the IPv4 one

IP: the Internet Protocol

- IPv6 con't
  - IPv4 and IPv6 can coexist, but v6 is not backwards-compatible
  - Transition mechanisms:
    - IPv6-in-IPv4 tunnels
    - Mapping the IPv4 space into the IPv6 space
    - Protocol translation mechanisms
  - Transition is going slowly, but Catch-22 involved: no one will deploy until commonly used, no one will use until commonly deployed