COMP 790-088

Networked and Distributed Systems

Inter-domain Routing

Jasleen Kaur

September 23, 2009

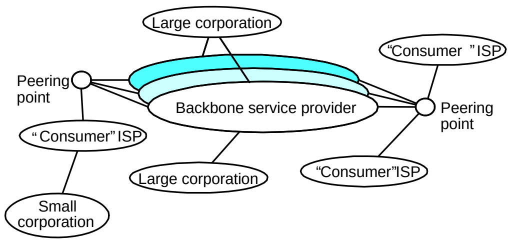
Internet-scale Routing

Approaches

- DV and link-state protocols do not scale to the global Internet
  - » How to make routing scalable?
- Exploit the notion of autonomous systems and divide routing into two parts
  - » Intra-domain routing: Routing within an autonomous system
    - eg: RIP (distance-vector type), OSPF (link-state type)
  - » Inter-domain routing: Routing between autonomous systems
    - Hierarchically aggregate routing information
- Route propagation (“know a smarter router” policy):
  - » Hosts know local (default) router
  - » Local routers know site routers
  - » Site routers know core (backbone) routers
  - » Core routers know everything

Internet AS-level Architecture

Properties



- Tiered Internet service providing
- Multi-homed stub networks
- Peering relations
  - » Points of presence (POPs)

Internet-scale Routing

Challenges

- Matter of scale!
  - » Backbone routers must provide a match for any valid IP address
  - » Even with CIDR, still needs to maintain O(100,000) prefixes
- Autonomous nature of domains:
  - » Each domain runs own interior routing protocol and link-cost assignment scheme
    - ❖ Impossible to calculate meaningful path costs for paths that cross multiple domains

    Therefore, inter-domain routing advertises only reachability information

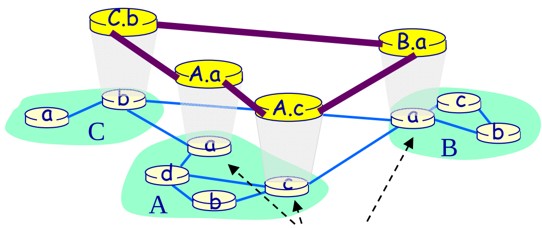
    Find any path that is loop-free (optimality not a consideration)

Internet-scale Routing
Challenges

- Issue of trust:
  - » Provider A may be unwilling to believe route advertisements from provider B
  - » Misconfigured routers, insufficient capacity to carry traffic, malicious intent
- Need to support flexible routing policies:
  - » Prevention of transit traffic
    - ❖ Multi-homed corporations may not wish to carry traffic between the two providers
  - » Provider A may want to implement special policies:
    - ❖ Use provider B only to reach these addresses
    - ❖ Use the path that crosses the fewest ASes
    - ❖ Use AS *x* in preference to AS *y*
    - ❖ Early-exit policy !

Border Gateway Protocol (BGP)
Architectural Components

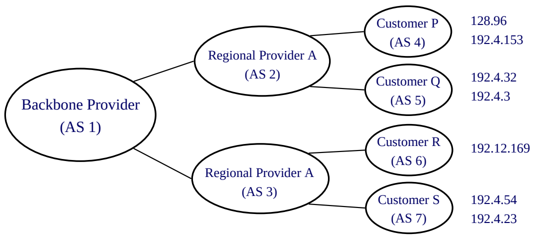
- Each AS has:
  - » At least one BGP speaker (spokesperson for entire AS)
    - ❖ Establish BGP sessions to speakers in other ASes
    - ❖ Exchange reachability information among ASes
  - » One or more Border Gateways (through which packets enter/leave the AS)
    - ❖ Routers charged with the task of forwarding packets between ASes



Border Gateway Protocol (BGP)

Basic Idea

- BGP is a Path-vector protocol:
  - » Advertises complete path for reaching a particular destination
    - ❖ AS 2 advertises: networks 128.96, 192.4.153, 192.4.32, 192.4.3 can be reached directly from AS 2
    - ❖ Backbone AS advertises:
      - 128.96, 192.4.153, 192.4.32, 192.4.3 can be reached along path: (AS1, AS2)
      - 192.12.169, 192.4.54, 192.4.23 can be reached along path: (AS1, AS3)



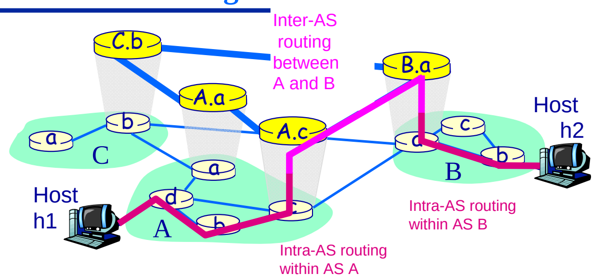
BGP Advertisements

Implementing Policies

- Complete AS path helps implement loop-free routing
  - » If AS finds itself in an advertisement, ignores it
- An AS will advertise only those routes that it considers good enough for itself
  - » And these are the routes that it will actually use for forwarding data
- BGP speakers need not advertise routes, even if they know of one
  - » Helps implement non-transit policy for multi-homed stub networks
    - ❖ If X does not want to route traffic to Z, then X will not advertise any routes to Z
  - » Helps implement cost-related or business-related policies
    - ❖ Don’t advertise routes via competitor’s network (even if competitor has advertised routes to you)
    - ❖ Don’t advertise routes through peers that charge you for bytes routed through them

Putting It All Together

Intra-AS & Inter-AS Routing



- Stub networks send to only border router (if single-homed)
- Provider AS:
  - » Border router injects information into the intra-domain routing protocol
    - “I have a link to customer-prefix *Y* of cost *X*”
  - » All internal routers send packets for this destination to this border router
- Backbone AS:
  - » Use Interior-BGP (IBGP) to distribute info learned by BGP speakers to all routers
    - Enables each router to learn best border router to use for a given prefix

The Internet AS Hierarchy

Why different intra- and inter-AS routing?

- Policy:
  - » Inter-AS: administration wants control over how its traffic routed and who routes through its network
  - » Intra-AS: single administration, so no “policy” decisions needed
- Scale:
  - » Hierarchical routing saves table size, reduced update traffic
- Performance:
  - » Intra-AS: can focus on performance
  - » Inter-AS: policy may dominate over performance

BGP Performance

Path Recovery

- 2-year study of routing updates by the Routeviews project
- Observations:
  - » Delay in Internet inter-domain path failovers averages 3 minutes
  - » Some last 15 minutes
- Cause:
  - » Mostly unforeseen interaction of protocol timers with specific vendor implementation decisions
- User-Impact: Failovers affect end-to-end performance significantly
  - » Measured packet losses grow by 30 times
  - » Latency grows by 4 times

BGP Performance

Misconfigurations

- Observations made in 2001 study:
  - » Each day, 200-1200 prefixes (1% global BGP table) suffer misconfigurations
  - » 2% of the time, these increase routing update load by at least 10%
    - ❖ One observation ***doubled*** load across ***all*** vantage points
  - » 3-4 new prefixes seen everyday result from misconfigurations
- Causes:
  - » Involuntary slips by network operators
  - » Router initialization bugs
  - » Poor understanding of configuration semantics by operators
- User-impact: connectivity is robust
  - » Only 4% of bad announcements disrupt connectivity

BGP Performance

Path Inflation

- 2002 study observed fairly inflated paths
- Causes:
  - » Many paths that use “early-exit” are inflated (longer RTTs)
  - » Topology-insensitive load balancing can cause significant path inflation
  - » Peering points between ISPs may not be on the “shortest path” for two end-hosts
  - » Non-early exit policies
    - ❖ To avoid a congested peering point
  - » Not all ISPs are directly connected to each other