Implementation

Major Internet routing protocols:

- RIP (v1 and v2): intra-domain, Bellman-Ford
  → also called “distance vector”
  → metric: hop count
  → UDP
  → nearest neighbor advertisement
  → popular in small intra-domain networks

- OSPF (v1 and v2): intra-domain, Dijkstra
  → also called “link state”
  → metric: average delay
  → directly over IP: protocol number 89
  → broadcasting via flooding
  → popular in larger intra-domain networks
• IS-IS: intra-domain, Dijkstra
  → “link state”
  → directly over link layer (e.g., Ethernet)
  → more recently: also available over IP
  → flooding
  → popular in larger intra-domain networks

• Source routing: packet specifies path
  → implemented in various link layer protocols
  → ATM call set-up: circuit-switching
  → IPv4/v6: option field
  → mostly disabled
  → large ISPs: sometimes used internally for diagnosis
BGP (Border Gateway Protocol):

• Inter-domain routing
  → border routers vs. backbone routers

→ “peering” between two AS’s
→ includes customer-provider relationship
→ exchanges: peering between multiple AS’s
• CIDR addressing
  → i.e., $a.b.c.d/x$
  → Purdue: 128.10.0.0/16, 128.210.0.0/16, 204.52.32.0/20
  → check at www.iana.org (e.g., ARIN for US)
• Route table look-up: maximum prefix matching
  → e.g., entries: 128.10.0.0/16 and 128.10.27.0/24
  → destination address 128.10.27.20 matches 128.10.27.0/24 best
• Metric: policy
  → e.g., shortest-path, trust, pricing
  → meaning of “shortest”: delay, router hop, AS hop
  → route amplification: shortest AS path $\neq$ shortest router path
  → mechanism: path vector routing
  → BPG update message
BGP route update:

\[ \rightarrow \quad \text{BGP update message propagation} \]

BGP update message:

\[ \text{ASN}_k \rightarrow \cdots \rightarrow \text{ASN}_2 \rightarrow \text{ASN}_1; a.b.c.d/x \]

Meaning: ASN $A_1$ (with CIDR address a.b.c.d/x) can be reached through indicated path

\[ \rightarrow \quad \text{“path vector”} \]

\[ \rightarrow \quad \text{called AS-PATH} \]

Some AS numbers:

- Purdue: 17
- BBN: 1
- UUNET: 701
- Level3: 3356
- Abilene (aka “Internet2”): 11537
Policy:

- if multiple AS-PATHs to target AS are known, choose one based on policy
  
  → e.g., shortest AS path length, cheapest, least worrisome

- advertise to neighbors target AS’s reachability
  
  → also subject to policy

  → no obligation to advertise

  → specifics depend on bilateral contract (SLA)

SLA (service level agreement):

  → bandwidth (e.g., 10 Gbps, 1 Gbps

  → delay (e.g., avrg. 25ms US), loss (e.g., 0.05%) 

  → pricing (e.g., 1 Mbps: below $100)

  → availability (e.g., 99.999%)

  → etc.
BGP-update procedure:

Upon receiving BGP update message from neighbor to target AS $A$

1. Store AS-PATH reachability info for target $A$
   $\rightarrow$ AdjIn table (one per neighbor)

2. Determine if new path to $A$ should be adopted
   $\rightarrow$ policy
   $\rightarrow$ path should be unique
   $\rightarrow$ BPG table ($\text{locRIB}$) & IP routing table update
   $\rightarrow$ inter-domain: IP table update from BGP

3. Determine who to advertise reachability for target $A$
   $\rightarrow$ selective advertisement

Note: if shortest-path then same as Dijkstra in-reverse
BGP-withdrawal:

1. Use BGP keep-alive message to sense neighbor
   → timeout

2. If keep-alive does not arrive within timeout, assume node is down

3. Send BGP withdraw message for neighbor who is deemed down if no alternative path exists; else send BGP update message
   → may trigger further updates

Other BGP features:

• BGP runs over TCP
  → port number 179
  → i.e., “application layer” protocol

• BPG-4 (1995); secure BGP
  → S-BGP: not implemented yet ("BBN vs. Cisco")
Performance

Route update frequency:

→ routing table stability vs. responsiveness
→ rule: not too frequently
→ 30 seconds
→ stability wins
→ hard lesson learned from the past (sub-second)
→ legacy: TTL

Other factors for route instability:

→ selfishness (e.g., fluttering)
→ BGP’s vector path routing: inherently unstable
→ more common: slow convergence
→ target of denial-of-service (DoS) attack
Route amplification:

→ shortest AS path ≠ shortest router path
→ e.g., may be several router hops longer
→ AS graph vs. router graph
→ inter- vs. intra-domain routing: separate subsystems
→ policy: company in Denmark

Route asymmetry:

→ routes are not symmetric
→ estimate: > 50%
→ mainly artifact of inter-domain policy routing
→ various performance implications
→ source traceback
Black holes:

\[ \rightarrow \] persistent unreachable destination prefixes
\[ \rightarrow \] BGP routing problems
\[ \rightarrow \] further aggravated by DNS
\[ \rightarrow \] purely application layer: end system problem
Topology:

\[\rightarrow \text{ who is connected to whom}\]

\[\rightarrow \text{ Internet AS graph (segment of Jan. 2002)}\]
Contrast with random graph: same number of nodes and edges

- random graph: choose each link with prob. $p$
- independently: prob. of $k$ neighbors is $p^k$
Ex.: Delta Airlines route map

→ by design: hub and backbone architecture
→ mixture of centralized/decentralized design
→ small system: centralized is good
→ large system: decentralization necessary
Small system with centralized design:

→ star topology

→ e.g., Southwest Airlines

→ essentially two conjoined star topologies

→ a matter of load balancing

→ backbone topology: trivial