Two types of Internet Routing Protocols

- Internet consists of roughly 19,000 Autonomous Systems
- What is an Autonomous system (AS)?
  - A network belonging to single administrative entity
  - With unified routing policies
- Intradomain routing protocol: within an Autonomous System
  - Distance Vector, e.g., RIP
  - Link State, e.g., OSPF, IS-IS
- Interdomain routing protocol: between Autonomous Systems
  - Border Gateway Protocol (BGP-4)
  - Path vector protocol

Intra-domain Routing Protocols

- Link state vs. distance vector
  - Uses unreliable datagram delivery
    - Flooding at layer 2
  - Distance vector
    - Routing Information Protocol (RIP), Bellman-Ford based
    - Each router periodically exchange reachability information with its neighbors
    - Minimal communication overhead
    - Takes long to converge, i.e., in proportion to the maximum path length
    - Has count to infinity problems
  - Link state
    - Open Shortest Path First Protocol (OSPF), based on Dijkstra
    - Each router periodically floods immediate reachability information to other routers
    - Fast convergence
    - High communication and computation overhead

Internet Routing Instability

- Methodology
  - Collect routing messages from five public exchange points
- Problems caused by routing instability
  - Increased delays, packet loss and reordering, time for routes to converge (small-scale route changes)
- Relevant BGP information
  - AS-Path
  - Next hop: Next hop to reach a network
  - Two routes are the same if they have the same AS-Path and Next hop
- Other attributes (e.g., MED, communities) ignored for now
Measurement methodology

AS-Path

- Sequence of AS's a route traverses
- Used for loop detection and to apply policy

```
AS-3
193.168.194

AS-2

AS-1
10.0.0.16 AS-1 AS-2 AS-3
10.0.0.16 AS-1 AS-2
10.0.0.16 AS-1 AS-2
```

BGP Information Exchange

- Announcements: a router has either
  - Learned of a new route, or
  - Made a policy decision that it prefers a new route
- Withdrawals: a router concludes that a network is no longer reachable
  - Explicit: associated to the withdrawal message
  - Implicit: (in effect announcement) when a route is replaced as a result of an announcement message
- In steady state BGP updates should be only the result of infrequent policy changes
  - BGP is stateful requires no refreshes
  - Update rate: indication of network stability

Example of delayed convergence

```
0: stage node msg queued
2: [3] [4][1][41] 1
3: [3] [4][1][241] 1
4: [3] [3] 1

1 2 3 4

Assuming node 1 has a route to a destination, and it withdraws the route:

Stage (msg processed) Msg queued
0: 1 -> [2,3,4]W
1: 1 -> [3,4][41] [2,3,4]A[431]
2: 2 -> [3,4][41] [2,3,4]A[341]
3: 3 -> [3,4][41] [2,3,4]A[241]
4: 4 -> [2,3,4][41] [2,3,4]A[431]

MinRouteAdver timer expires: 4 -> [2,3,4]W, 3 -> [2,4][431], 2 -> [3,4][431]

... (omitted)

9: 3 -> [2,4]W
```

Note: In response to a withdrawal from 1, node 3 sends out 3 messages:

3: [2,4][431], 3: [2,4][431], 3: [2,4][431]

Routing Successive Events (Instability)

- WADiff: a route is explicitly withdrawn as it becomes unreachable, and is later replaced with an alternative route (forwarding instability)
- AADiff: a route is implicitly withdrawn and replaced by an alternative route as the original route becomes unavailable or a new preferred route becomes available (forwarding instability)
- WADup: a route is explicitly withdrawn, and reannounced later (forwarding instability or pathological behavior)

Types of Inter-domain Routing Updates

- Forwarding instability:
  - may reflect topology changes
- Policy fluctuations (Routing instability):
  - may reflect changes in routing policy information
- Pathological updates:
  - redundant updates that are neither routing nor forwarding instability
- Instability:
  - forwarding instability and policy fluctuation \( \rightarrow \) change forwarding path
Routing Successive Events (Pathological Instability)

- **AADup:**
  - A route is implicitly withdrawn and replaced with a duplicate of the original route (pathological behavior or policy fluctuation)
- **WWDup:**
  - The repeated transmission of BGP withdrawals for a prefix that is currently unreachable (pathological behavior)

Findings

- BGP updates more than one order of magnitude larger than expected
- Routing information dominated by pathological updates
  - Implementation problems:
    - Routers do not maintain the history of the announcements sent to neighbors
    - When a router gets topological changes they just send these announcements to all neighbors, irrespective of whether the router sent previous announcements about that route to a neighbor or not
    - Self-synchronization – BGP routers exchange information simultaneously may lead to periodic link/router failures
    - Unconstrained routing policies may lead to persistent route oscillations

Findings

- Instability and redundant updates exhibits strong correlation with load (30 seconds, 24 hours and seven days periods)
  - Overloaded routers fail to respond an their neighbors withdrawn them
- Instability usually exhibits high frequency
- Pathological updates exhibits both high and low frequencies
- No single AS dominates instability statistics
- No correlation between size of AS and its impact on instability statistics
- There is no small set of paths that dominate instability statistics

Conclusions

- Routing in the Internet exhibits many undesirable behaviors
  - Instability over a wide range of time scales
  - Asymmetric routes
  - Network outages
  - Problem seems to worsen
- Many problems are due to software bugs or inefficient router architectures

Lessons

- Even after decades of experience routing in the Internet is not a solved problem
- This attests the difficulty and complexity of building distributed algorithm in the Internet, i.e., in a heterogeneous environment with products from various vendors
- Simple protocols may increase the chance to be understood
- Implemented right

Beacons [2003], Motivation:
Better understanding of BGP dynamics

- Border Gateway Protocol (BGP)
  - Internet interdomain routing protocol
- Difficult to understand BGP’s dynamic behavior
  - Multiple administrative domains
  - Unknown information (policies, topologies)
  - Unknown operational practices
  - Ambiguous protocol specs

What is a BGP Beacon?

- An unused, globally visible prefix with known Announce/Withdrawal schedule
  - For long-term, public use
- For research purposes to study BGP dynamics
  - To calibrate and interpret BGP updates
  - To study convergence behavior
  - To analyze routing and data plane interaction
- Useful to network operators
  - Serve to debug reachability problems
  - Test effects of configuration changes:
    - e.g., flap damping setting

Related work

- Differences from Labovitz’s “BGP fault-injector”
  - Long-term, publicly documented
  - Varying advertisement schedule
  - Beacon sequence number (AGG field)
  - Enabler for many research in routing dynamics
- RIPE RIS Beacons
  - Set up at 9 exchange points

Active measurement infrastructure

Many Observation points:

- Send route update
- Stub AS
- Upstream Provider

BGP Beacon #1
198.133.206.0/24

1: Oregon RouteViews
2: RIPE
3: AT&T
4: Verio
5: MIT
6: Berkeley

Beacon terminology

- Input signal:
  - Beacon-injected change
  - 3:00:00 GMT: Announce (A0)
  - 5:00:00 GMT: Withdrawal (W)
- Output signal:
  - 5:00:10 A1
  - 5:00:40 W
- Signal length:
  - number of updates in output signal
  - (3 updates)
- Signal duration:
  - Time between first and last update in the signal
  - (5:00:10 -- 5:01:10
  - 60 seconds)
- Inter-arrival time:
  - Time between consecutive updates

Deployed PSG Beacons

<table>
<thead>
<tr>
<th>Prefix</th>
<th>AS</th>
<th>Start date</th>
<th>Upstream provider AS</th>
<th>Beacon period</th>
<th>Beacon location</th>
</tr>
</thead>
<tbody>
<tr>
<td>198.133.206.0/24</td>
<td>47</td>
<td>9/4/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.83.230.0/24</td>
<td>4</td>
<td>8/10/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>198.32.7.0/24</td>
<td>6</td>
<td>8/10/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.135.183.0/24</td>
<td>37</td>
<td>9/4/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.133.206.0/24</td>
<td>2</td>
<td>9/4/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.135.183.0/24</td>
<td>37</td>
<td>9/4/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>198.32.7.0/24</td>
<td>6</td>
<td>8/10/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>198.133.206.0/24</td>
<td>2</td>
<td>9/4/02</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Beacon 5 schedule

- Live host behind the beacon for data analysis
- Study fail-over behavior for multi-homed customers

Beacon 5 schedule diagram
How to process Beacon data?

- How to identify output signals, ignore external events?
  - Data cleaning
  - Anchor prefix as reference
    - Same origin AS as beacon prefix
    - Statically nailed down
- How to minimize interference between consecutive input signals?
  - Beacon period is set to 2 hours
- Time stamp and sequence number
  - Attach additional information in the BGP updates
    - Make use of a transitive attribute: Aggregator fields

Beacon data cleaning process

- Goal
  - Clearly identify updates associated with injected routing changes
  - Discard beacon events influenced by external routing changes

Beacon example analysis

- BGP implementation impact:
  - Cisco vs. Juniper
- Route flap damping analysis
- Convergence analysis
- Inter-arrival time analysis

Cisco vs. Juniper update rate-limiting

- Known last-hop Cisco and Juniper routers from the same AS and location
- Average signal length: Average number of updates observed for a single beacon-injected change

“Cisco-like” last-hop routers

- Linear increase in signal duration with respect to signal length
- Slope = 30 sec
- Due to Cisco’s default rate-limiting setting

“Juniper-like” last-hop routers

- Signal duration relatively stable with respect to signal length
- Shorter signal duration compared to “Cisco-like” last-hops
What is route flap damping?

- A mechanism to punish unstable routes by suppressing them
- RFC2439 [Villamizar et al. 1998]
  - Supported by all major router vendors
  - Believed to be widely deployed [AT&T, Verio]
- Goals:
  - Reduce router processing load due to instability
  - Prevent sustained routing oscillations
  - Do not sacrifice convergence times for well-behaved routes
- There is conjecture a single announcement can cause route suppression.

Exponentially decayed

Scope
- Inbound external routes
- Per neighbor, per destination

Penalty
- Flap: route change
  - Increases for each flap
  - Decays exponentially

Cisco default setting:
- Scope
  - Inbound external routes
  - Per neighbor, per destination
- Penalty
  - Flap: route change
    - Increases for each flap
    - Decays exponentially

\[ P(t) = P(0)e^{-\lambda t} \]

Strong evidence for withdrawal- and announcement-triggered suppression.

Route flap damping analysis

Distinguish between announcement and withdrawal

Summary:
- WD-triggered sup more likely than ANN-triggered sup
- Cisco: overall more likely trigger sup than Juniper
- AAAS-pattern
- Juniper: more aggressive for WAWS pattern

Convergence analysis

Summary:
- Withdrawals converge slower than announcements
- Most beacon events converge within 3 minutes

Output signal duration
Beacon 1’s upstream change

Beacon 1 at Router Hex2, signal detection marks

Inter-arrival time analysis
Cisco-like last-hop routers

Complementary cumulative distribution plot

Inter-arrival time modeling

Inter-arrival time modeling

Beacon conclusion

Beacon conclusion

Routing stability in congested networks (Shaikh 2000)

Routing stability in congested networks (Shaikh 2000)

Network configuration

Network configuration
Methodology

- Mean-Time-to-Flap (U2D)
- Mean-Time-to-Recover (D2U)
- Overload factors: 25-400%
- Data packet size: 64, 256, 1500 bytes
- Buffer size at HR: 4MB, 16MB

Analytical models

- Assumptions:
  - The overload factor remains constant
  - Every packet has the same probability of being dropped depending on the overload factor
  - Packet dropping probability is independent for each packet
- Markov chains to find expected values of U2D and D2U for OSPF and BGP

Conclusions

- Developed detailed analytical models
- OSPF's behavior depends only on traffic overload factor
  - Independent of buffer size, packet dropping policy
- BGP's behavior depends on overload factor and RTT
- BGP's resilience to congestion decreases as RTT increases
- There is a need to isolate routing messages from data traffic
  - Through scheduling and buffer management

Lecture summary

- Internet routing is still not well-understood
  - For example, difficult to interpret BGP update messages
  - Holy grail: root cause analysis of BGP updates, need to correlate intradomain and interdomain changes
  - Measurement is useful for understanding routing stability
- Effect of congestion on routing protocols
  - Is TCP the right transport for BGP?
  - How should router treat routing messages differently?