EE 122: Interdomain Routing Protocol (BGP)

Ion Stoica
TAs: Junda Liu, DK Moon, David Zats
http://inst.eecs.berkeley.edu/~ee122/fa09
(Materials with thanks to Vern Paxson, Jennifer Rexford, and colleagues at UC Berkeley)

Outline

- Why does BGP exist?
  - What is interdomain routing and why do we need it?
  - Why does BGP look the way it does?

- How does BGP work?
  - Path vector algorithm
  - ...and many boring details

  *pay more attention to the “why” than the “how”*

BGP Routing

- Provides paths between networks

- Previous lecture presented two routing designs
  - link-state
  - distance vector

- Previous lecture assumed single domain
  - all routers have same routing metric (shortest path)
  - no privacy issues, no policy issues

Internet is more complicated.....

- Internet not just unstructured collection of networks

- Internet is comprised of a set of “autonomous systems” (ASes)
  - independently run networks, some are commercial ISPs
  - currently around 30,000 ASes

- ASes are sometimes called “domains”
  - hence “interdomain routing”
Internet: a large number of ASes

- Large ISP
- Large ISP
- Small ISP
- Access Network

Who speaks BGP?

- Two types of routers
  - Border router (Edge)
  - Internal router (Core)

Purpose of BGP

- Share connectivity information across ASes

I-BGP and E-BGP

- IGP: Intradomain routing
  - Example: OSPF

- E-BGP
  - R2
  - R3
  - R4
  - R5
  - AS3
  - AS1
  - AS2
  - AS1

- Announce B

- R1
  - AS1
  - AS2
  - AS3
  - AS4
  - AS5

- Border router
- Internal router
In more detail

1. Provide internal reachability (IGP)
2. Learn routes to external destinations (eBGP)
3. Distribute externally learned routes internally (iBGP)
4. Select closest egress (IGP)

Rest of lecture...

- Motivate why BGP is the way it is
- Discuss some problems with interdomain routing
- Discuss (briefly!) what a new BGP might look like

1. ASes are autonomous

- Want to choose their own internal routing protocol
  - different algorithms and metrics
- Want freedom to route based on policy
  - “my traffic can’t be carried over my competitor’s network”
  - “I don’t want to carry transit traffic through my network”
  - not expressible as Internet-wide “shortest path”!
- Want to keep their connections and policies private
  - would reveal business relationships, network structure

5 Minute Break

Questions Before We Proceed?
2. ASes have business relationships

Relations between ASes
- Provider
- Customer
- Peer

Business Implications
- Customer pay provider
- Peers don’t pay each other

AS-level topology
- Destinations are IP prefixes (e.g., 12.0.0.0/8)
- Nodes are Autonomous Systems (ASes)
  - Internals are hidden
- Links are connections & business relationships

What routing algorithm can we use?
- Key issues are policy and privacy
- Can’t use shortest path
  - Domains don’t have any shared metric
  - Policy choices might not be shortest path
- Can’t use link state
  - Would have to flood policy preferences and topology
  - Would violate privacy
What about distance vector?
- Does not reveal any connectivity information
- But is designed to compute shortest paths
- Extend distance vector to allow policy choices?

Path-Vector Routing
- Extension of distance-vector routing
- Support flexible routing policies
- Faster loop detection (no count-to-infinity)
- Key idea: advertise the entire path
- Distance vector: send distance metric per dest d
- Path vector: send the entire path for each dest d

Faster Loop Detection
- Node can easily detect a loop
  - Look for its own node identifier in the path
  - E.g., node 1 sees itself in the path “3, 2, 1”
- Node can simply discard paths with loops
  - E.g., node 1 simply discards the advertisement

Flexible Policies
- Each node can apply local policies
  - Path selection: Which path to use?
  - Path export: Which paths to advertise?
- Examples
  - Node 2 may prefer the path “2, 3, 1” over “2, 1”
  - Node 1 may not let node 3 hear the path “1, 2”
Selection vs Export

- Selection policies
  - determines which paths I want my traffic to take

- Export policies
  - determines whose traffic I am willing to carry

- Notes:
  - any traffic I carry will follow the same path my traffic takes, so there is a connection between the two
  - from a protocol perspective, decisions can be arbitrary
    - can depend on entire path (advantage of PV approach)

Examples of Standard Policies

- Transit network:
  - Selection: prefer customer to peer to provider
  - Export:
    - Let customers use any of your routes
    - Let anyone route through you to your customer
    - Block everything else

- Multihomed (nontransit) network:
  - Export: Don’t export routes for other domains
  - Selection: pick primary over backup

Illustration

Route selection: controls traffic out of the network
Export: controls traffic into the network

Issues with Path-vector Policy Routing

- Reachability
- Security
- Performance
- Lack of isolation
- Policy oscillations
Reachability

- In normal routing, if graph is connected then reachability is assured
- With policy routing, this does not always hold
  - Peers don’t provide transit to each other
  - Customer doesn’t provide transit for provider

Security

- An AS can claim to serve a prefix that they actually don’t have a route to (blackholing traffic)
  - problem not specific to policy or path vector
  - important because of AS autonomy
- Even worse: snoop on all traffic to almost any destination
  - without anyone realizing that anything is wrong
- Fixable: make ASes “prove” they have a path
  - but not used in today’s Internet

Performance

- BGP designed for policy not performance
- “Hot Potato” routing common but suboptimal
  - AS wants to hand off the packet as soon as possible
  - Even BGP “shortest paths” are not shortest
    - Fewest AS’s != Fewest number of routers
- 20% of paths inflated by at least 5 router hops
- Not clear this is a significant problem

Performance (example)
Lack of Isolation: dynamics

- If there is a change in the path, the path must be re-advertised to every node using the path

![BGP updates per hour per prefix](Fig. from [Huston])

Lack of isolation: routing table size

- Each BGP router must know path to every other IP prefix
  - but router memory is expensive and thus constrained
- Number of prefixes growing more than linearly
- Subject of current research

![Number of prefixes in BGP table](Fig. from [G. Huston])

Persistent Oscillations due to Policies

Depends on the interactions of policies

- “1” prefers “1 3 0” over “1 0” to reach “0”

![Persistent Oscillations due to Policies](Diagram 1)

Persistent Oscillations due to Policies

Initially: nodes “1”, “2”, and “3” know only shortest path to “0”

![Persistent Oscillations due to Policies](Diagram 2)
"1" advertises its path "1 0" to "2"

"3" advertises its path "3 0" to "1"
Persistent Oscillations due to Policies

“1” withdraws its path “1 0” from “2” since it is no longer using it.

Persistent Oscillations due to Policies

“2” advertises its path “2 0” to “3”.

Persistent Oscillations due to Policies

Persistent Oscillations due to Policies
"3" withdraws its path “3 0” from “1” since is no longer using it.

"1" advertises its path “1 0” to “2”.

 withdrawn 3 0

0 withdraws its path “3 0” from “1” since is no longer using it.

advertises its path “1 0” to “2”.

withdraw: 3 0

Persistent Oscillations due to Policies

Persistent Oscillations due to Policies

Persistent Oscillations due to Policies

Persistent Oscillations due to Policies
Persistent Oscillations due to Policies

“2” withdraws its path “2 0” from “3” since is no longer using it

Policy Oscillations (cont’d)

- Policy autonomy vs network stability
  - focus of much recent research
- Not an easy problem
  - PSPACE-complete to decide whether given policies will eventually converge! [FP2008]
- However, if policies follow normal business practices, stability is guaranteed [GR2000]

Redesigning BGP

- If we keep all the current constraints, not many alternative design options (at high-level)
  - Which constraints might we lift?
- Are most policies really private?
  - could use link-state for some of the routing
- Do ASes really need to see the entire path?
  - could hide some of the path, reducing updates
- Can AS structure be integrated into addressing?
Summary

- BGP is essential to the Internet
  - ties different organizations together

- Poses fundamental challenges....
  - leads to use of path vector approach

- ...and myriad details (not presented in this lecture)

- Reading: Sections K&R 4.6.3