Interdomain traffic engineering with BGP

L. Swinnen, S. Tandel, B. Quoitin and O. Bonaventure

URL: http://www.info.ucl.ac.be/people/OBO
Outline

- BGP basics
- BGP-based traffic engineering techniques
- The provider selection problem
- Evaluation of AS-Path prepending
Intradomain versus Interdomain routing

• Intradomain routing (IGP)
  • Objective
    ✗ select the best path towards each destination based on some metrics (e.g. Delay, bandwidth) used in domain
  • Issues
    ✗ IGP should react quickly to changes in topology

• Interdomain routing (EGP)
  • Objective
    ✗ select the best path towards each destination that is compatible with the routing policies of the transit domains without knowing their topology
  • Issues
    ✗ Each domain can define its own routing policy
    ✗ EGP should be scalable (14k domains, 130k routes)
The Border Gateway Protocol

- **Objective**
  - Distribute interdomain routes in a scalable manner while supporting routing policies

- **Principles**
  - Path-vector routing protocol
  - BGP routers exchange routing tables
    - BGP session is established over TCP connection
    - No periodic advertisement of routes as with RIP
      - Routes are first advertised when BGP session is established
      - Routes are updated when they change
      - Routes are removed when they stop being reachable
  - BGP routers use policies to filter and rank the routes sent or received
Routing policies: customer-provider

- Principle of **customer-provider peering**
  - ASc is a smaller ISP than ASp
  - ASc buys transit service from ASp
    - ASp agrees to transmit packets from ASc towards any destination
    - ASp agrees to announce the routes received from ASc
Routing policies: shared-cost

- Principle of shared-cost peering
  - usually used on links between Ases of same size
  - ASx (ASy) agrees to receive from ASy (ASx) packets sent towards ASx or its direct customers
    - ASx (ASy) does not provide transit to ASy (ASx)
The Internet today

- **Tier-1 ISPs**
  - About 20
  - Full-mesh

- **Tier-2 ISPs**
  - About 200
  - Customers of T1

- **Tier-3 ISPs**
  - About 12000
  - Enterprise networks
  - Customers of T1, T2

---

Séminaire CERTI - 6 Décembre 2002 page 7  
O. Bonaventure, 2002
The Border Gateway Protocol (2)

- The two variants of BGP
  - eBGP between border routers of distinct AS
  - iBGP between BGP routers inside AS
BGP route advertisements

- Content
  - IP prefixes
    - List of reachable IP prefixes
  - AS-PATH
    - the list of AS through which the announcement passed
  - NEXT-HOP
    - the IP address of the router that advertised the route
  - LOCAL-PREF
    - can be used for traffic control purposes
  - MULTI-EXIT-DISCERNIMinator
    - can be used for traffic control between neighbors
  - ORIGIN
    - how the route was learned (IGP, EGP,Incomplete)
  - ATOMIC AGGREGATE and AGGREGATOR
    - rarely used
Organization of a BGP router

A BGP router can filter the routes received from each peer

A BGP router can filter the routes advertised to each peer

BGP Decision process
- Highest local-pref
- Shortest AS-Path
- Lowest MED
- eBGP over iBGP
- nearest IGP neighbor
- Tie breaking rules

Forwarding table
Outline

- BGP basics
- BGP-based traffic engineering techniques
- The provider selection problem
- Evaluation of AS-Path prepending
Control of the outgoing traffic

- Objectives
  - balance traffic among external links
  - prefer some links over others for any reason

- How to achieve this control?
  - AS administrator tunes the BGP decision process of all its border routers to influence the selection of the best-path on each router
    - rely on input filters to ignore/change some routes learned
    - add local-pref to the routes advertised via iBGP inside the AS to influence the other border routers of the AS
    - Rely on IGP to select closest next-hop
Control of the outgoing traffic
Sample network

- Routing without tuning BGP decision process

R11's routing table
- 4/8:AS3:AS4 via R32 (eBGP,best)
- 4/8:AS2:AS4 via R12 (iBGP)

R12's routing table
- 4/8:AS2:AS4 via R22 (eBGP,best)
- 4/8:AS3:AS4 via R11 (iBGP)
- 4/8:AS3:AS4 via R32 (eBGP)
Control of the outgoing traffic local-pref

- **Principle**
- set local-pref to influence all routers of the AS

R11's routing table
- 4/8:AS3:AS4 via R31 (eBGP, best)
- 4/8:AS2:AS4 via R12 (iBGP), pref

R12's routing table
- 4/8:AS2:AS4 via R22 (eBGP, best), pref
- 4/8:AS3:AS4 via R11 (iBGP)
- 4/8:AS3:AS4 via R32 (eBGP)
Control of the incoming traffic

- Objectives
  - balance traffic among external links
  - prefer some links over others for any reason

- How to achieve this control ?
  - AS administrator needs to send different BGP advertisements on different links to influence the BGP decision process of routers in distant AS

- But routers in distant AS and transit AS can also tune their outgoing traffic ...
Control of the incoming traffic
Sample network

- Routing without tuning the announcements
  - packet flow towards AS1 will depend on the tuning used on the decision process of AS2, AS3 and AS4
Control of the incoming traffic
AS-Path prepending

- Principle
- Artificially prepend own AS number on some routes

R31's routing table
- 10/7:AS1 via R11 (eBGP, best)
- 10/7:AS1:AS1:AS1:AS1 via R32 (iBGP)

R32's routing table
- 10/7:AS1 via R31 (iBGP, best)
- 10/7:AS1:AS1:AS1:AS1 via R12 (eBGP)
Control of the incoming traffic
Selective announcements

- Principle
- Advertise some prefixes only on some links

- Drawbacks
  - splitting a prefix increases size of all BGP routing tables
  - No redundancy in case of link failure

Séminaire CERTI - 6 Décembre 2002 page 18
O. Bonaventure, 2002
Control of the incoming traffic
AS-Path prepending (2)

- AS-Path prepending can be combined with more specific prefixes

R31's routing table
- 10/8:AS1 via R11 (eBGP, best)
- 10/8:AS1:AS1:AS1 via R32 (iBGP)
- 11/8:AS1:AS1:AS1 via R11 (eBGP)
- 11/8:AS1 via R32 (iBGP, best)

R32's routing table
- 10/8:AS1 via R31 (iBGP, best)
- 10/8:AS1:AS1:AS1 via R12 (eBGP)
- 11/8:AS1:AS1:AS1 via R31 (iBGP)
- 11/8:AS1 via R12 (eBGP, best)
Outline

- BGP basics
- BGP-based traffic engineering techniques
- The provider selection problem
- Evaluation of AS-Path prepending
The provider selection problem

- How does an ISP select a provider?

- Economical criteria
  - Cost of link
  - Cost of traffic

- Quality of the BGP routes announced by provider
  - Number of routes announced by provider
  - Length of the routes announced by provider

- Often, ISPs have two upstream providers for technical and economical redundancy reasons
An experiment in provider selection

- **Principle**
  - Obtain BGP routing tables from several providers
    - 20 providers from www.routeviews.org
  - Simulate the connection of an ISP to 2 providers
    - Rank providers based on the routes selected by the BGP decision process of the simulated ISP
Selection among 20 providers

![Graph showing comparison between average routes in LOC-RIB, average common routes, as-path shorter 1, and (avg) as-path shorter. The x-axis represents AS peers, and the y-axis represents the number of routes.](image)

Séminaire CERTI - 6 Décembre 2002 page 23

O. Bonaventure, 2002
Selection among the 12 largest providers

![Diagram showing selection among 12 largest providers](image-url)
Outline

- BGP basics
- BGP-based traffic engineering techniques
- The provider selection problem
- Evaluation of AS-Path prepending
The network model

- Each AS modelled as a single BGP router advertising a single IP prefix
  - Core ASes provide transit service
  - Each Stub AS attached to two providers with preferential attachment
The network core

- Degree-based
  - Albert-Barabasi model

  - 200 core ASes
  - 400 stub ASes

- Generate nodes sequentially
  - New AS attached to 2 existing nodes
  - Inter-AS links chosen with a probability (AS connectivity), highly connected Ases are preferred

- Consequence
  - ASes with lowID have a better connectivity than ASes with highID
Distribution of the interdomain paths

Without prepending
Distribution of the interdomain paths with prepending once on lowID link

![Graph showing the distribution of interdomain paths with and without prepending.](image-url)
Distribution of the interdomain paths with prepending twice on lowID link

![Graph showing the distribution of incoming traffic on LowID with and without prepending. The graph compares the percentage of traffic on LowID with ordered stubs.]
Distribution of the interdomain paths with prepending once highID link

![Graph showing distribution of incoming traffic on LowID](Image)
Impact of BGP-based techniques that control the outgoing traffic

- How is the distribution of the interdomain paths affected by the utilization of local-pref by stub ASes?

- Simple simulation study

  - Configure local-pref on each stub to force it to send all its packets via its lowID provider

  - Configure local-pref on each stub to force it to send all its packets via its highID provider
Distribution of interdomain paths
Stubs send only via lowID provider

Order of incoming traffic on LowID

- Without prepending, without localpref
- Without prepending, localpref-lowID
Distribution of interdomain paths
Stubs send only via highID provider

Without prepending, without localpref
Without prepending, localpref-highID

Percentage of incoming traffic on LowID vs. Ordered stubs
Findings and further work

Findings
- AS-Path is not sufficient to rank BGP routes
  - Internet is becoming more and more flat
  - Need new performance-oriented attributes
- AS Path prepending is not sufficient to control incoming traffic
- Outgoing and incoming BGP-based traffic control techniques seem decoupled

Further work
- Better techniques to control traffic
  - Evolutionary algorithms developed by Steve Uhlig
    - Automatic selection of local-pref values
    - Automatic selection of redistribution communities
  - Simulations with more realistic and larger models
  - Internet measurements together with Belnet, BT and Skynet