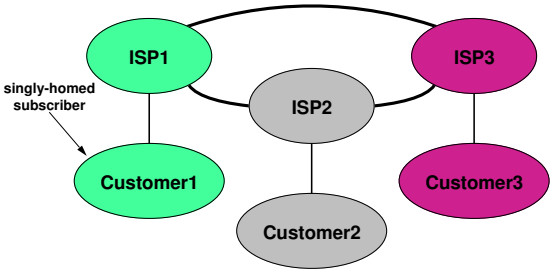
CS551
Multi-homing in BGP

Bill Cheng

http://merlot.usc.edu/cs551-f12

Transit vs. Nontransit Services



- ISP*n* provides transit service to Customer*n*
- ISP*n* provides non-transit service from another ISP for traffic for its customer

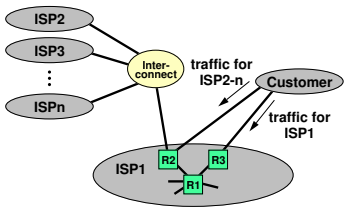
Multi-homing

- With multi-homing, a single network has more than one connections to the Internet
- Improves reliability and performance:
  - can accommodate link failure
  - bandwidth is sum of links to Internet
- Multiple connections provide ***load sharing*** but *not* load balancing
  - BGP cannot do load balancing

Issues With Multi-homing

- Symmetric routing
  - while conventional wisdom prefers symmetric paths, many are asymmetric
- Packet re-ordering
  - may trigger TCP's fast retransmit algorithm
- Other concerns:
  - addressing, DNS, aggregation
- *Note:* using BGP in multi-homing situation is not an off-the-shelf use of the protocol

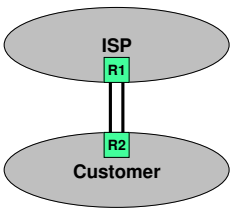
Static Routing May Not Work



- Static routing may send traffic to ISPs 2-n from customer over one link and traffic to ISP1 over the other link
- Lacks flexibility (especially when ISP1 grows and shrink)

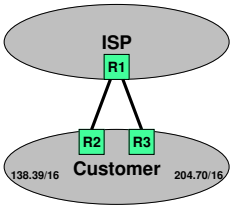
Multi-homing to a Single Provider: Case 1

- Easy solution:
  - no BGP, but use IMUX or Multilink PPP
- Harder solution:
  - use BGP
  - makes assumptions about traffic (same amount of prefixes can be reached from both links)



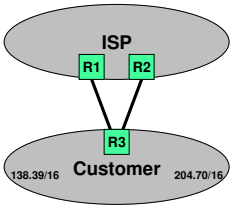
Multi-homing to a Single Provider: Case 2

- For ISP-> Customer traffic:
  - use MED in Customer or LOCAL-PREF in ISP
- For Customer->ISP traffic:
  - break-down prefix and advertise different prefixes over different links with default routes
- Good if traffic load to/from prefixes is equal
  - if single prefix in Customer, only 1 link will be used for ISP->Customer traffic



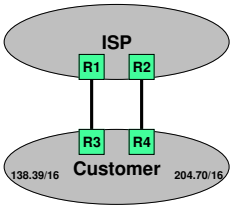
Multi-homing to a Single Provider: Case 3

- For ISP->Customer traffic, same as before:
  - use MED in Customer or LOCAL-PREF in ISP
- For Customer->ISP traffic:
  - R3 alternates links (reordering?)
  - Customer learns full BGP routes and load-shares
- Good if traffic load to/from prefixes is equal



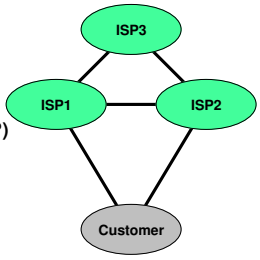
Multi-homing to a Single Provider: Case 4

- Most reliable approach
  - no equipment sharing
- ISP -> Customer:
  - same as case 3
- Customer -> ISP:
  - same as case 2



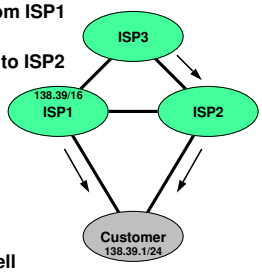
Multi-homing to Multiple Providers

- Major issues:
  - addressing
  - aggregation
- Customer address space: (what are the advantages and disadvantages of each approach?)
  - delegated by ISP1
  - delegated by ISP2
  - delegated by ISP1 and ISP2
  - obtained independently



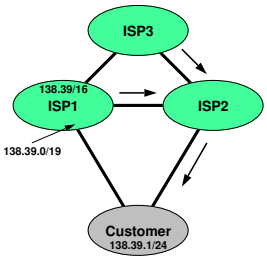
Case 1: Customer Uses Address Space From One ISP (1 or 2)

- Customer uses address space from ISP1
  - ISP1 advertises /16 aggregate
  - Customer advertises /24 route to ISP2
  - ISP2 relays route to ISP1 and ISP3
  - ISP2-3 use /24 route
  - ISP1 routes directly
  - Problems with traffic load? (longest prefix becomes a *"traffic magnet"*)
  - *Note:* this can actually work well if the relative sizes of the providers have a good match



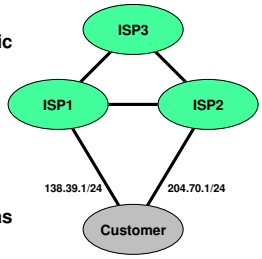
Pitfalls

- ISP1 aggregates to a /19 at border router to reduce internal tables
  - ISP1 still announces /16
  - ISP1 hears /24 from ISP2
  - ISP1 routes packets for customer to ISP2!
  - *Workaround:* ISP1 must inject /24 into I-BGP



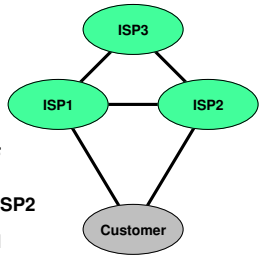
Case 2: Customer Uses Address Space From Both ISPs

- ISP1 and ISP2 continue to announce aggregates
  - Load sharing depends on traffic to two prefixes
  - Lack of reliability: if ISP1 link goes down, part of customer becomes inaccessible
  - Customer may announce prefixes to both ISPs, but still problems with longest match as in case 1



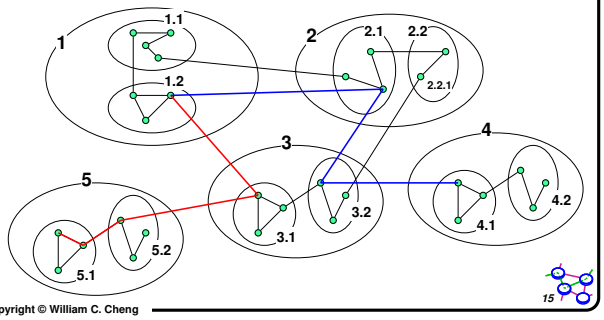
Case 3: Customer Uses Its Own Address Space

- Offers the most control, but at the cost of aggregation
- Still need to control paths:
  - suppose ISP1 large, ISP2-3 small
  - customer advertises long path to ISP1, but LOCAL-PREF attribute used to override
  - ISP3 learns shorter path from ISP2
- ***Bottom line:*** no good and general solution for multi-homing to multiple providers



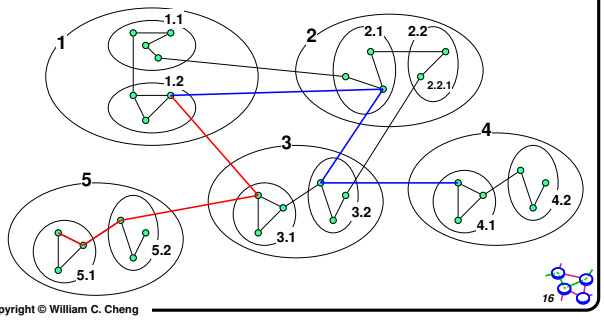
How Can BGP Express the Following Policies:

- 2 will not act as transit to 3
- 1 will use the blue path for packets destined to 4 and the red for packets destined to 5
- 2 will not accept packets sourced in 1



How Can BGP Express the Following Policies:

- 2 will not act as transit to 3 *(do not tell anyone about 3)*
- 1 will use the blue path for packets destined to 4 and the red for packets destined to 5 *(cannot control path, just first hop)*
- 2 will not accept packets sourced in 1 *(no way)*

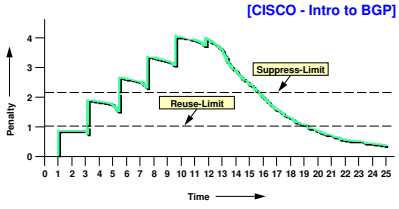


Route Flap Dampening

- Problem: route flap when a flaky link constantly goes up and down:
  - BGP sessions disappear and reappear
  - routes are withdrawn and re-advertised
  - global effects (does the flap of a butterfly's wing in Brazil set off a tornado in Texas?)
- BGP was extended to dampen route flaps
- Associate a penalty with each route
  - increase when route flaps
  - exponentially decay penalty with time
- When penalty reaches threshold, suppress route
  - must never forget routes

Route Flap Dampening (Cont...)

[CISCO - Intro to BGP]

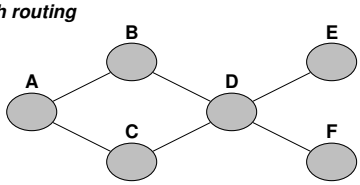


Tricky Issues

- "Synchronizing" intra and inter-domain routing
- Getting packets to the right exit router without introducing too much flux into intra-domain routing
- Multi-homing
  - interaction with aggregation
- How much policy *should* we actually be able to support???

BGP Limitations: Policy

- Ex: *fish routing*



  - E would like A to send traffic to E via B while F would like A to send traffic to F via C
  - but how?

Other BGP-related Issues

- Convergence Time [Labovitz00a]
- Router synchronization [Floyd94b]
- Congestion [Shaikh00a]
- Policy and convergence [Gao00a, Tangmunarunkit01a]
- Misconfiguration [Mahajan02a]
- Other other issues
  - routing arbiter - central DB of policies
  - robustness in the face of router resource exhaustion [Chang, Govindan, Heidemann]

Some BGP Stats (as of 30-Jan-2003)

[data from Japan (bgp-stats@lists.apnic.net)]

- BGP routing table entries: 120,000
  - prefixes after max aggregation: 76,596
- Addresses announced: 1,180,368,745
  - 31.8% of available address space announced
  - 57.9% of the allocated address space announced
  - 55.0% of available address space allocated
- AS's in Internet routing table: 14,513
  - origin-only AS's: 12,615
    - origin-only AS's with only 1 prefix: 5,690
  - transit AS's: 1,898
- AS path length
  - mean: 5.3
  - maximum seen: 17