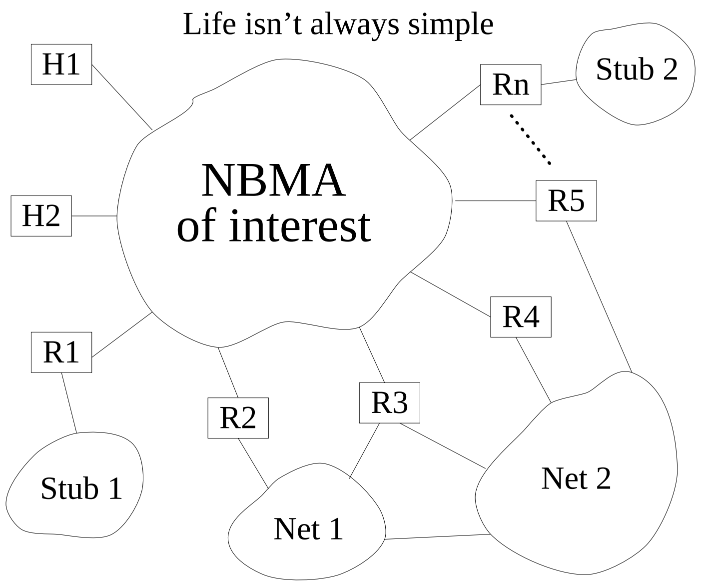
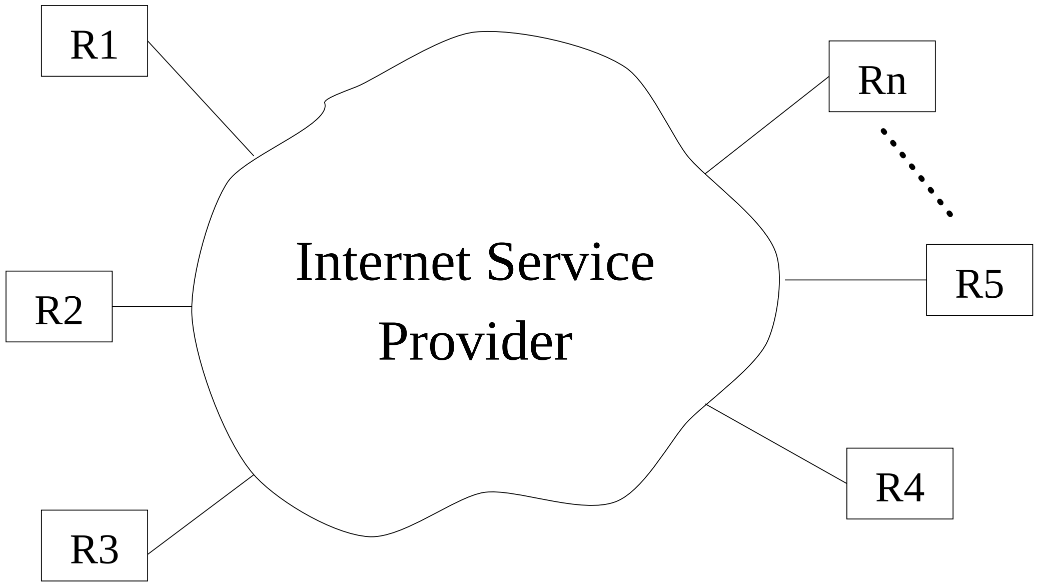
Factors Affecting Route Scaling



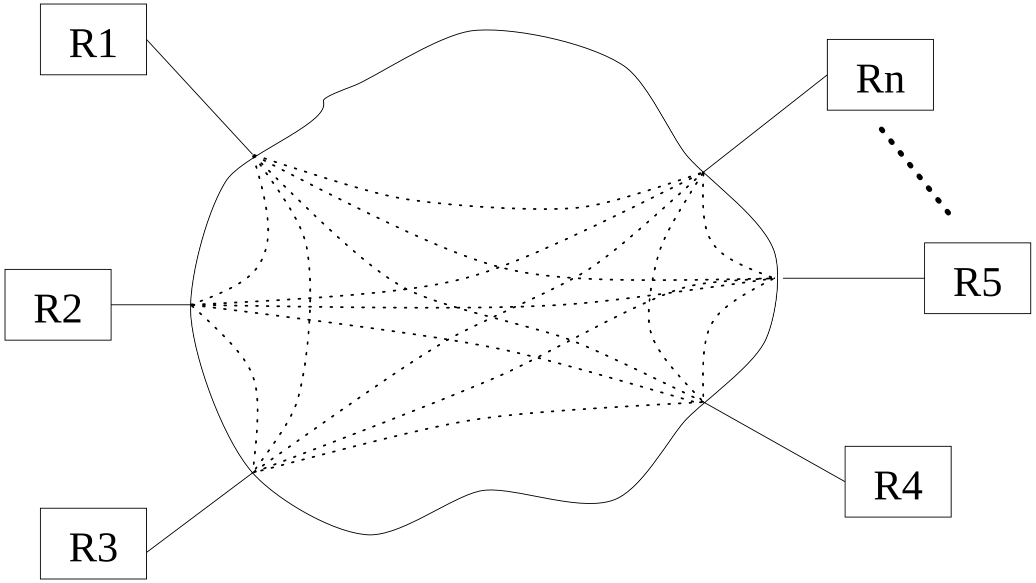
- Number of routing adjacencies
- Number of routes (or amount of state required)
- Volume of Route Change

View of Internet Routing from an NSP



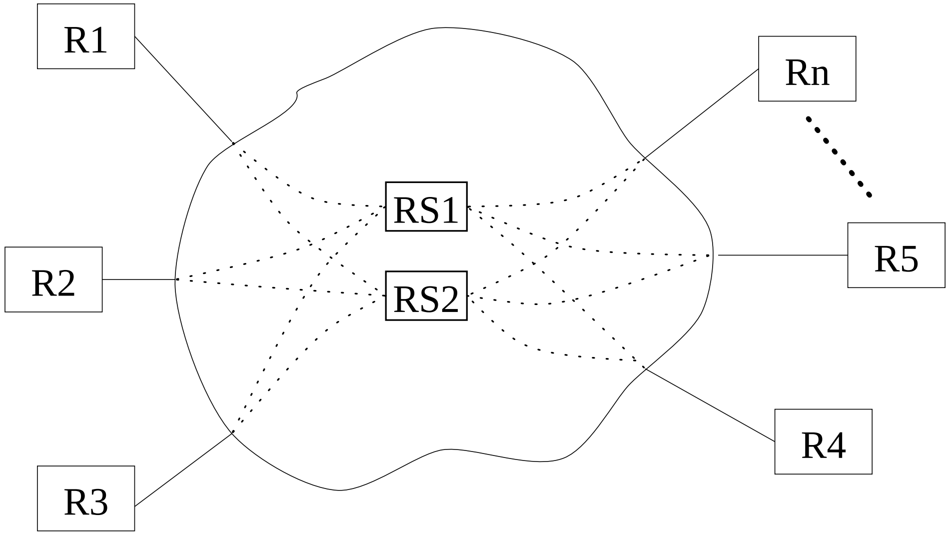
- N_1 through N_n are sets of NLRI announced in
- For any N_i and N_j there may be a non zero intersection
- On average there are 2.5 paths per NLRI

Full Mesh IBGP



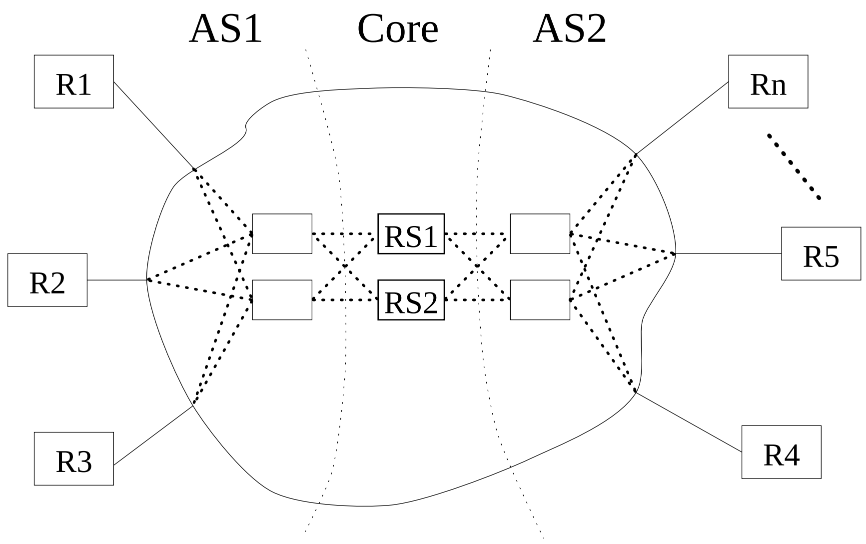
- Obvious scaling problems:
  - $N * (N - 1)$ BGP adjacencies
  - $(N - 1)$ BGP adjacencies per router

Single AS Route Servers



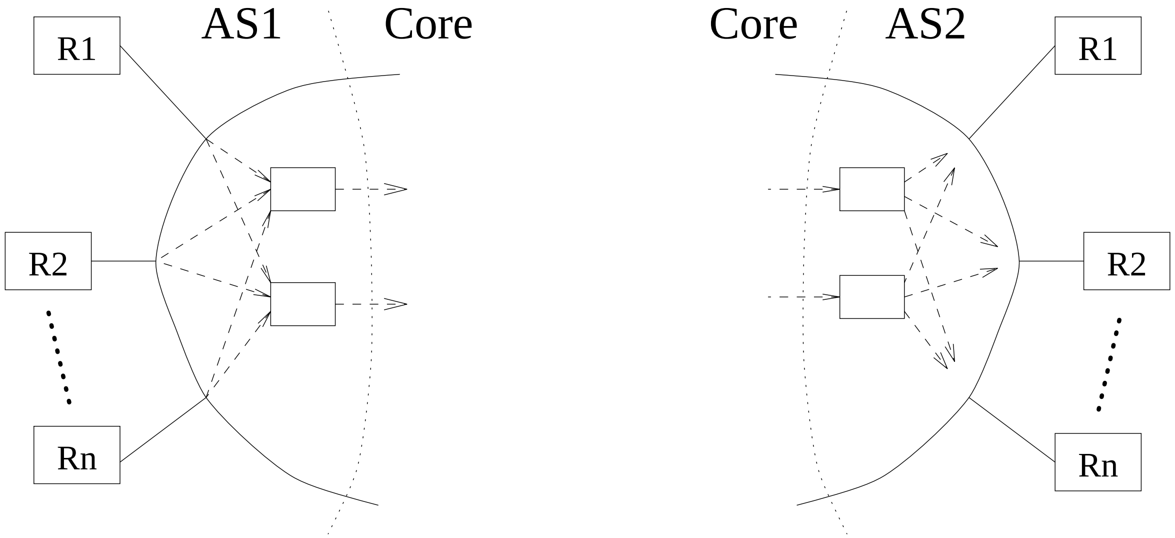
- Simple router servers - one policy for everyone
- N_{RS} BGP adjacencies per router
- N BGP adjacencies per router server
- concentrates difficulty on a small number of machines

Multiple AS with Route Servers



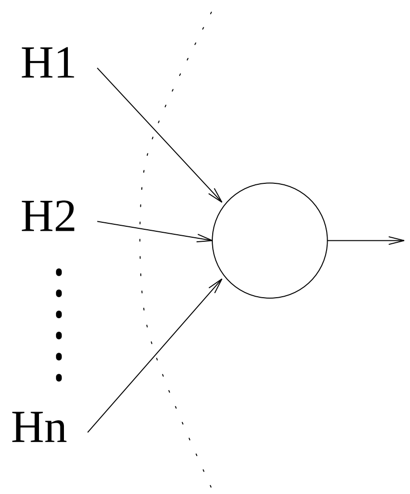
- Multi-View router servers - multiple policies
- N_{RS} BGP adjacencies per router
- N/N_{AS} BGP adjacencies per AS border router
- $N_{RS} * N_{AS}$ BGP adjacencies per route server

Inbound and Outbound Aggregations



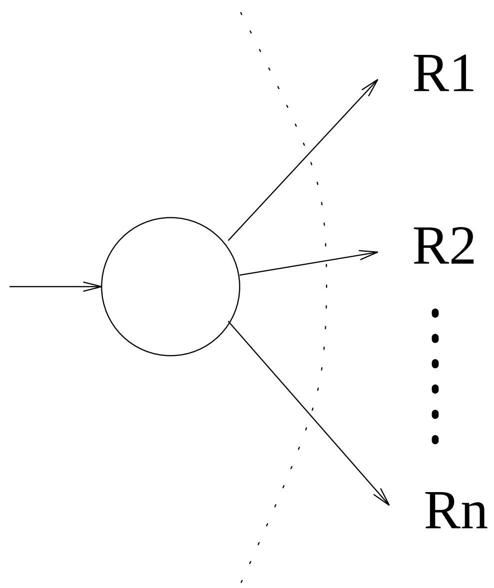
- RIFs support limited controlled deaggregation
- aggregation can be closer to inbound or outbound side

Scaling: Inbound Aggregation of Hosts



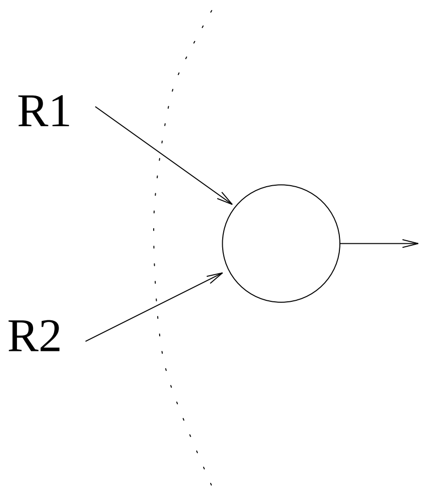
- For example, consider 10,000 hosts, where 1% have direct connections elsewhere at any one given time
- Approx 100 BGP adjacencies at the aggregator
- NLRI can be aggregated to /18 or better.

Scaling: Outbound Aggregation to Hosts and Routers



- The worst case is that the aggregator will have to keep an adjacency for every major router and some percentage of the hosts.

Scaling: Inbound Aggregation of Routers

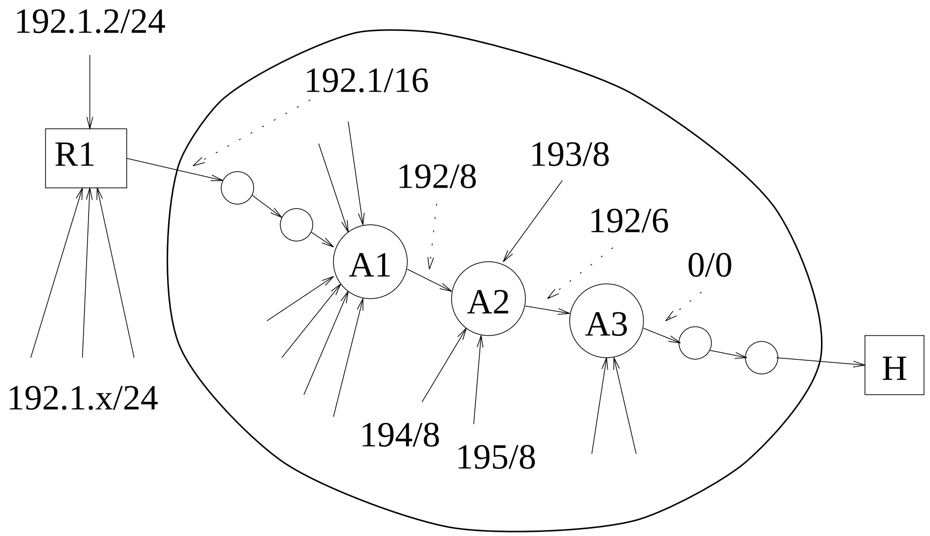


- The aggregator may have to support a BGP adjacency for every other router or host actively reaching a destination behind R1 and R2.
- This could be a very large number of adjacencies
- Recommendation: DON'T DO THIS!

Action Taken for Deaggregate using RIfs

1. Lookup the route, determine the Next HOP (NH)
2. If NH has the "SameNBMA" attribute, open a BGP adjacency with the aggregator and send a RIF with destination
3. If a more specific route is returned with "SameNBMA", repeat step #2.
4. When most specific route for the destination no longer has "SameNBMA", stop and use that next hop.

Example: Action Taken by a Host



- RIF request for 192.1.2.1 to A3, returns 192/6 (A2)
- RIF request for 192.1.2.1 to A2, returns 192/8 (A1)
- RIF request for 192.1.2.1 to A1, returns NH = R.

Remaining Issues (next draft revision)

- Host and Router actions need to be described in detail
- BGP attribute needed for originator's media address
- BGP attribute needed to identify RIF response
- BGP attribute needed to specify a missing set of routes

Host and Router Actions

- RIF responses should never be readvertised BGP routes
- RIF responses should not be placed in the LocalRIB
- RIF responses should be placed only in the FIB
- The draft would benefit from a clear description of:
  - how to determine when a next hop can be further refined
  - what to put in the RIF request
  - the actions taken by an aggregator responding to RIFs
  - the actions taken when RIF responses arrive
  - (maybe) description of how BSD implementation would work as an example
- Thanks to John Garrett for valuable comments

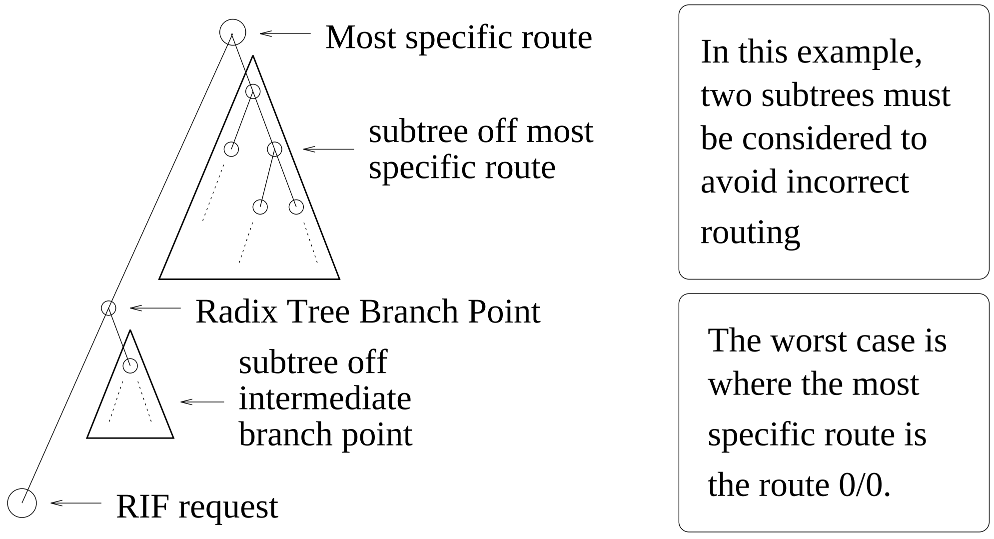
BGP Attribute: Originator's Media Address

- The peer media address is already known
- The originator media address can be held in an attribute
- When aggregating, the aggregator media address is used
- IDRP already has such an attribute

BGP Attribute: Identify RIF Responses

- RIF responses are treated differently from ordinary routes
- The aggregator may be one BGP hop away
- A second BGP session can be set up for RIF requests and responses
- A better solution would be to identify RIF responses

BGP Attribute: Specify Missing Routes



- Failure to provide more specific routes in a RIF response would result in incorrect routing
- In the worst case, the correct response to a RIF request is the full routing table
- The ability to identify missing subtrees solves this