A Model of BGP Routing for Network Engineering

Nick Feamster, *MIT*
Jared Winick, *Lockheed Martin*
Jennifer Rexford, *AT&T Labs--Research*

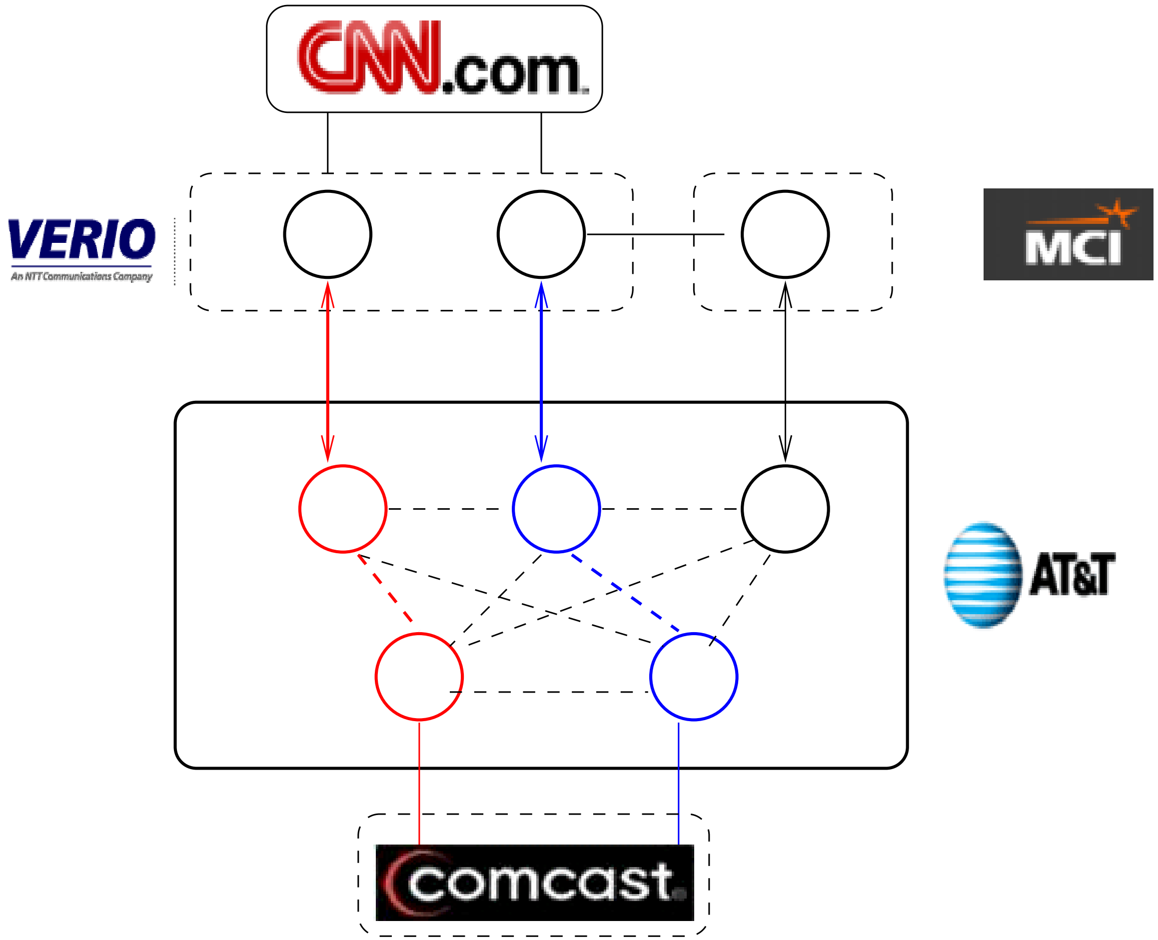
Problem: Network operators must tune routing protocols to provide good performace in the face of changing conditions.

Today: Tweak configuration and pray.

Our Solution: Compute how a configuration change will affect traffic flow before deployment.

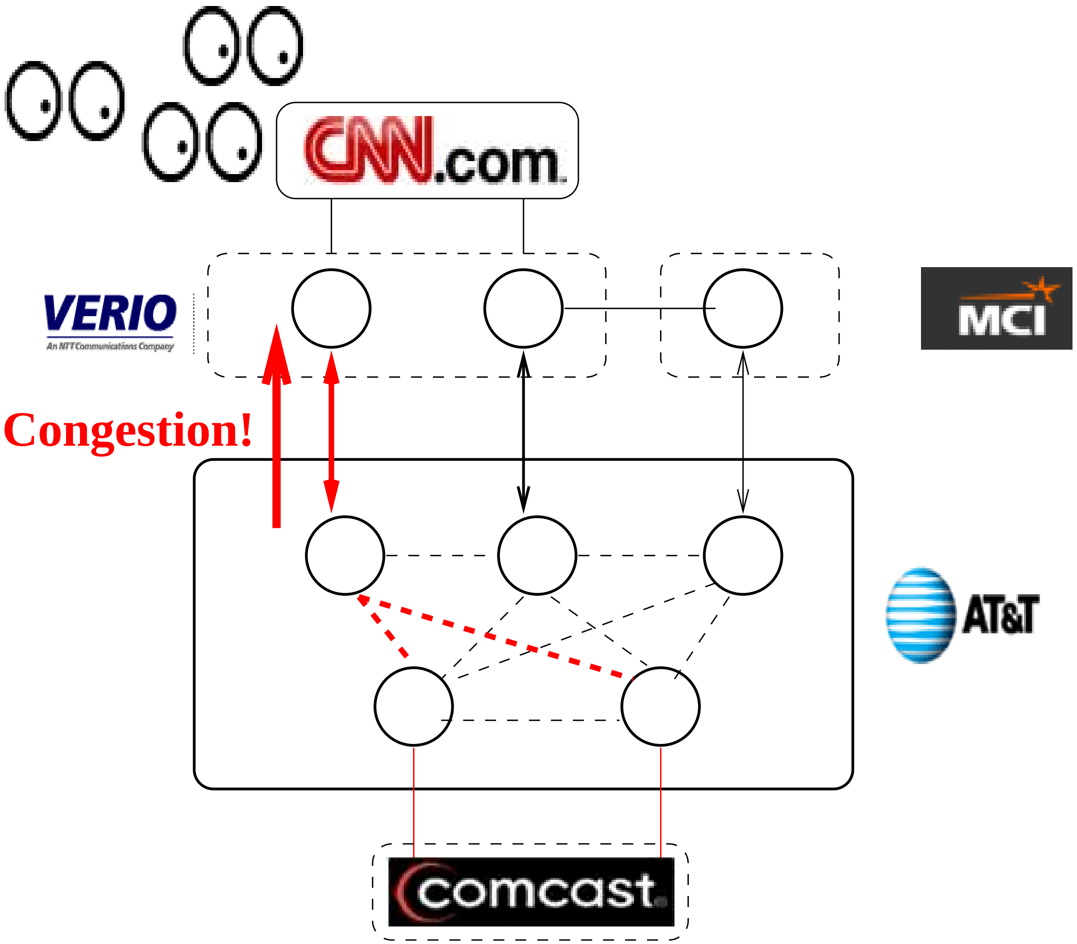
Overview

- Internet composed of autonomous systems (ASes)
- Multiple connections between ASes



Network Operators Must Respond to Events

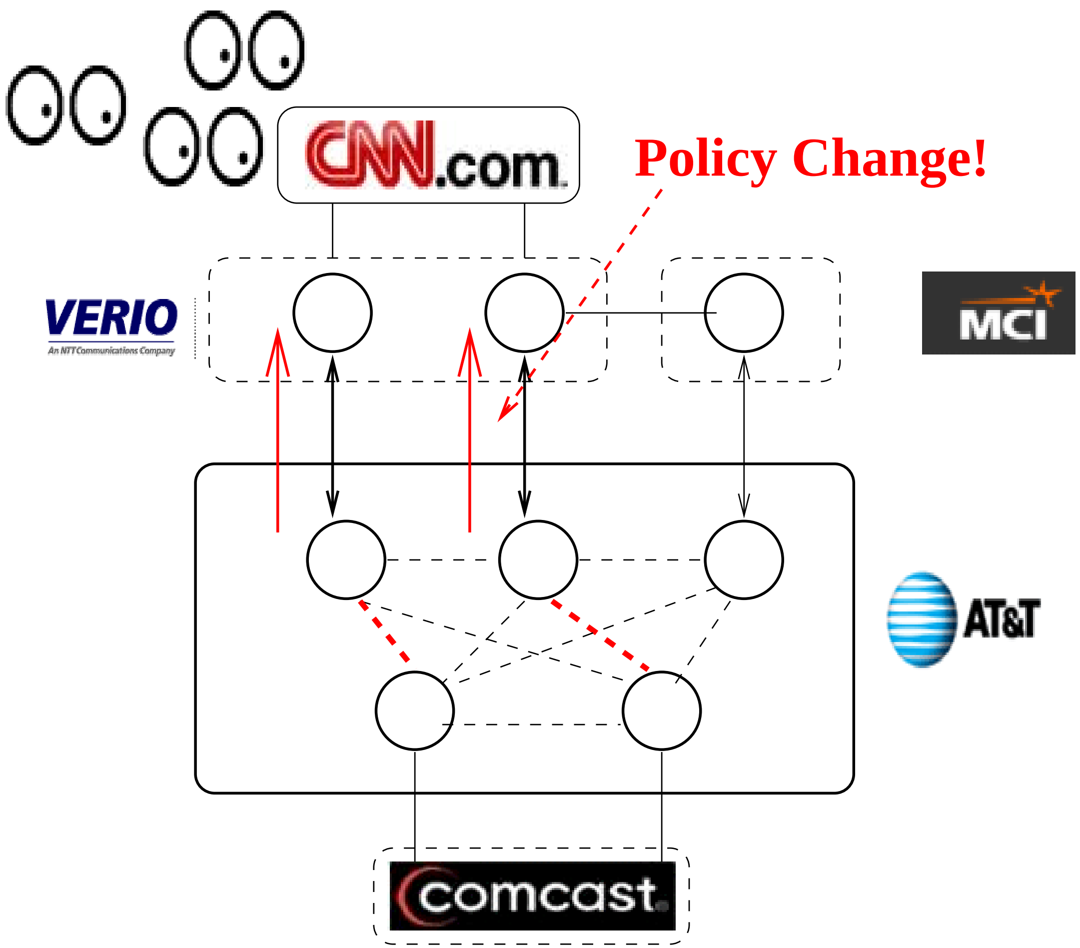
- Changes in **Traffic Volume**



Network Operators Must Respond to Events

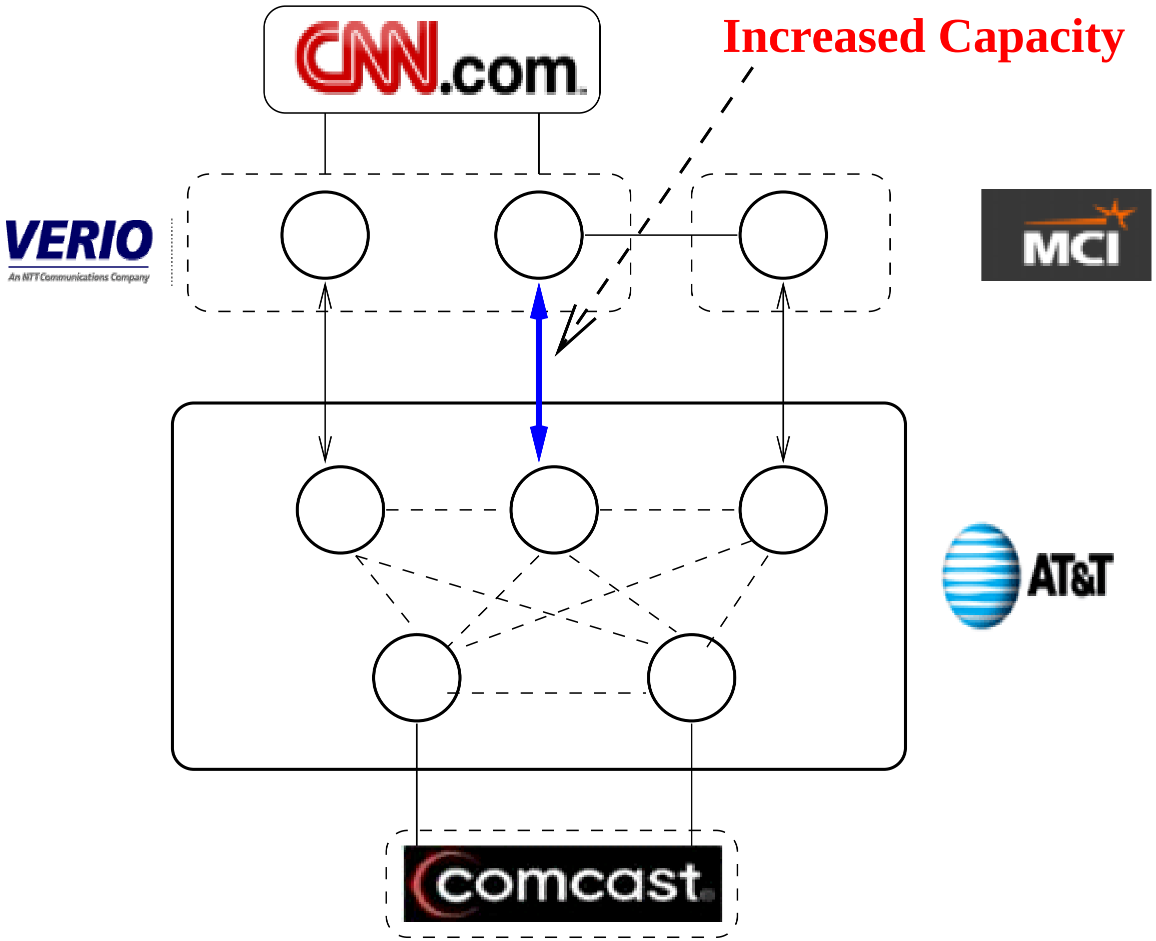
- Changes in **Traffic Volume**

  Note: The network does not adapt automatically!



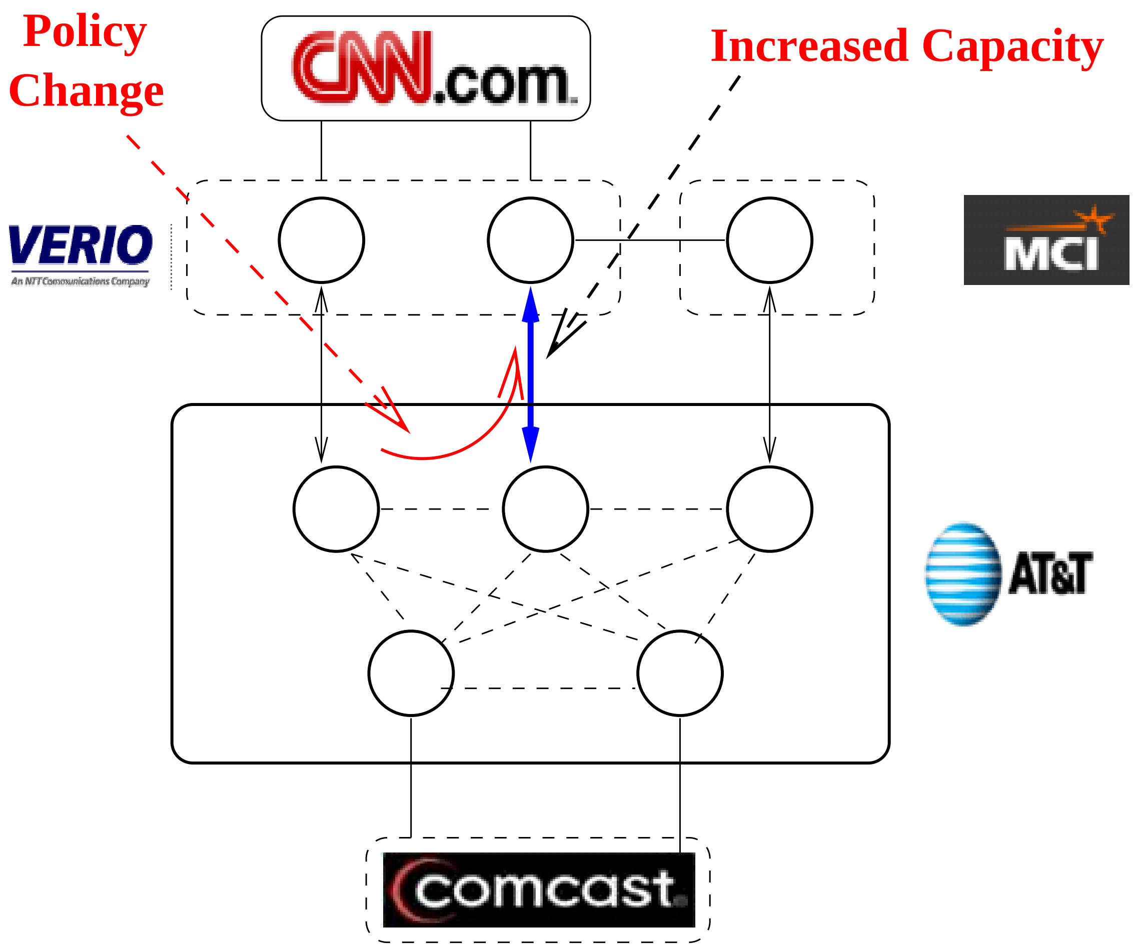
Network Operators Must Respond to Events

- Changes in **Link Capacity**



Network Operators Must Respond to Events

- Changes in **Link Capacity**



Predicting Traffic Flow: Many Requirements

- Knowledge about traffic volumes
- Information about available routes
- Prediction of paths between routers within the AS

Our contribution: Modeling how each router within one autonomous system (AS) will select routes to external destinations.

Implemented in a accurate and fast tool that has been evaluated and tested on the AT&T IP backbone.

Strawman #1: Simulation

Observation:

If a routing system converges to a unique outcome, the outcome is independent of the order that routers exchange messages and select paths.

Advantages

- *Time ordering of messages does not affect outcome.*

  Simulation will arrive at correct answer.

Disadvantages

- Operators must know **outcomes,** not dynamics.
- Many possible message **orderings**: potentially slow.

Simulation: Accurate, but slow.

Strawman #2: Rank Routes, Pick the Best One

Problem

Incorrect answer!

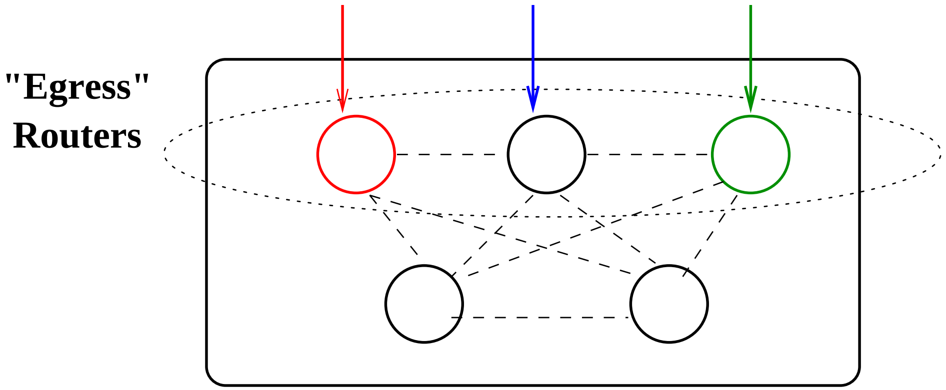
Two Artifacts of Border Gateway Protocol (BGP)

- **Impossible to impose a complete ranking of routes at a single router.**
  - Ranking between two routes can depend on presence (or absence) of other routes.
- **All routes may not be visible at every router.**
  - Set of routes learned at one router depend on route selection at other routers.

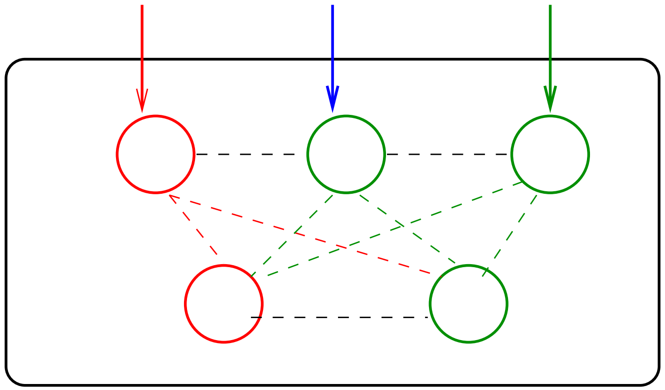
Note: These "artifacts" provide flexibility and scalability!

Instead: Model a Certain Message Ordering

- **Step 1:** Egress routers compute best routes.
  - Outcome: A set of egress routers for each destination.

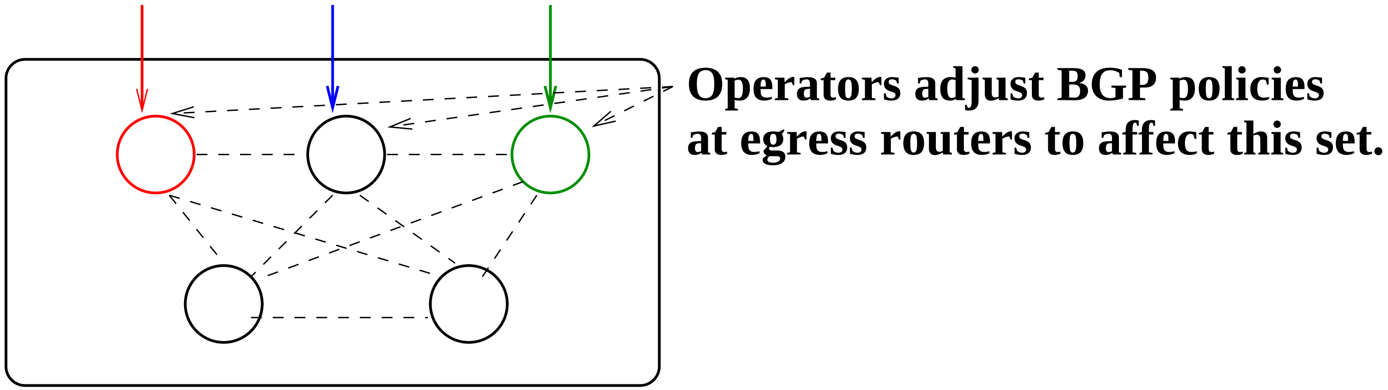


- **Step 2:** Egress routers propagate these routes to other routers within the AS.
  - Outcome: Each router in the AS selects a egress router.

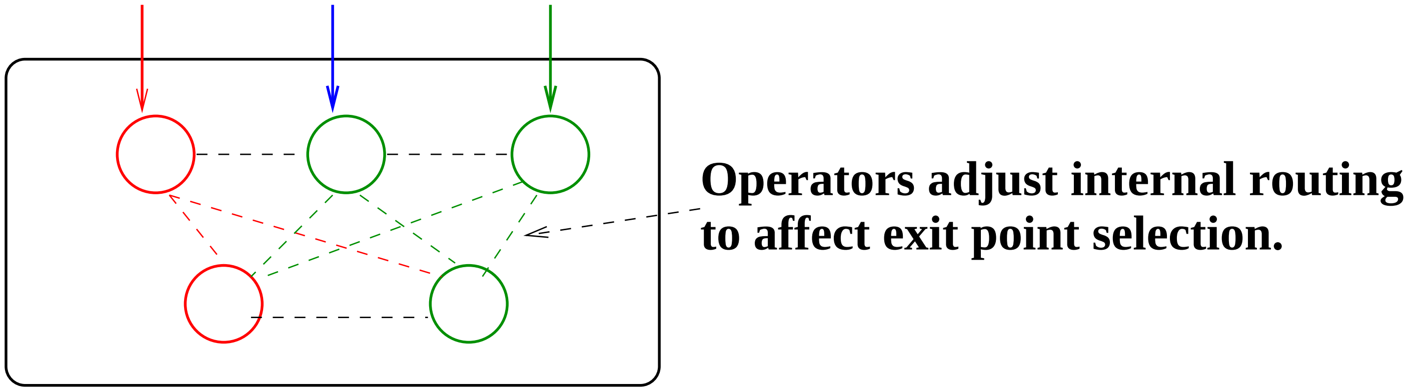


Instead: Model a Certain Message Ordering

- **Step 1:** Egress routers compute best routes.
  - Outcome: A set of egress routers for each destination.

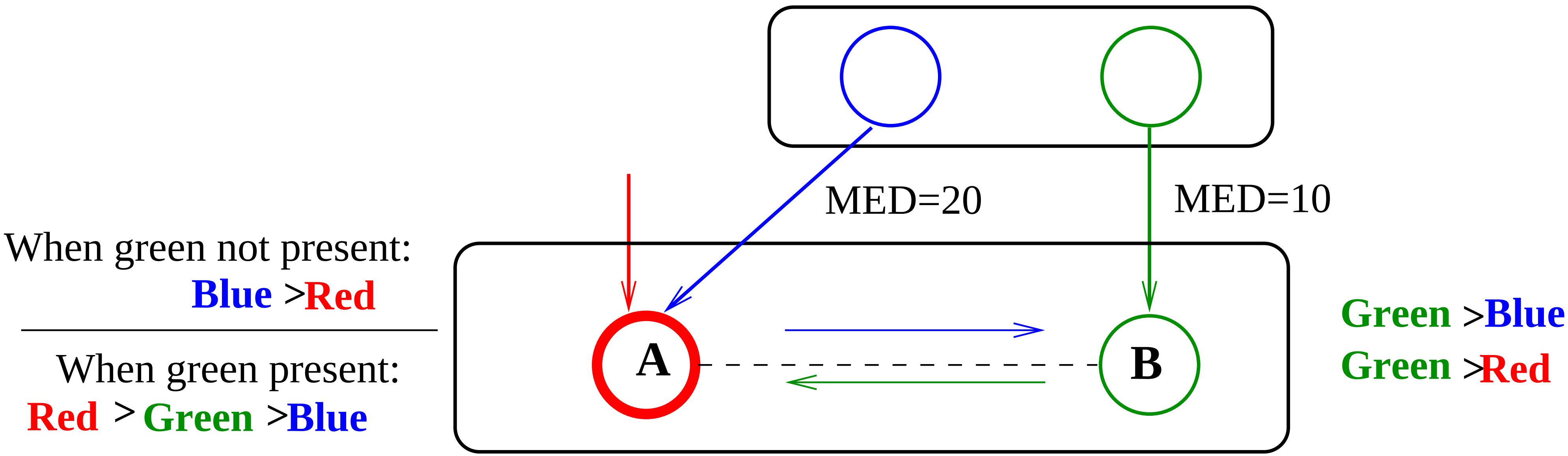


- **Step 2:** Egress routers propagate these routes to other routers within the AS.
  - Outcome: Each router in the AS selects a egress router.



Step 1: Egress Routers Compute Best Routes

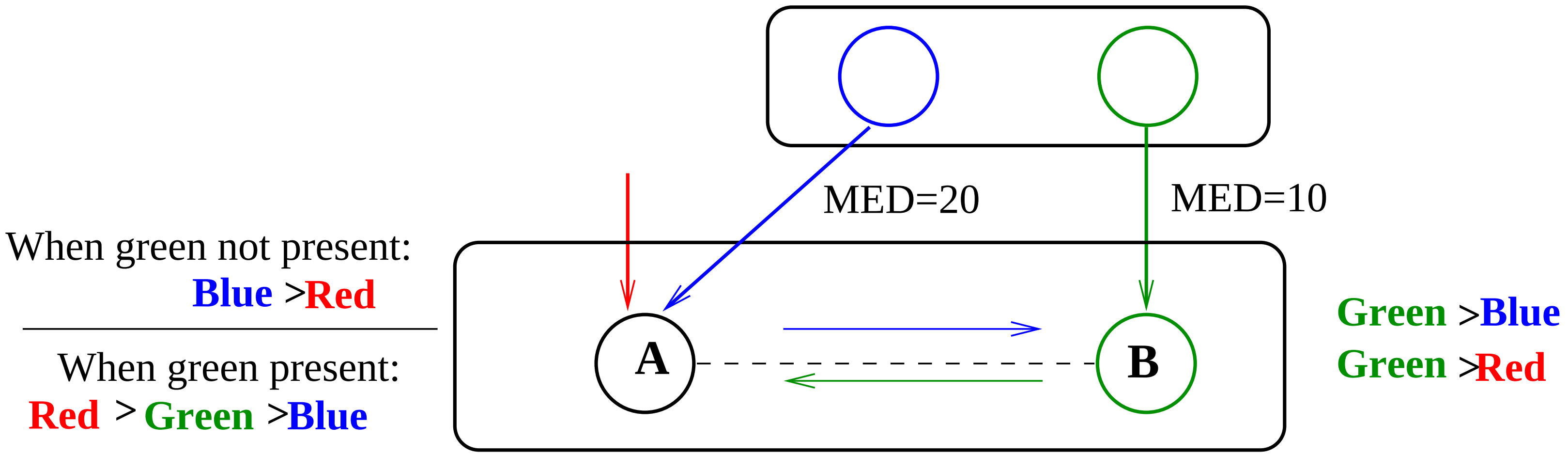
- *Problem:* Ranking of routes at one router can depend on routes learned at other routers.



- *Solution:* Compute best local routes where possible, and propagate the effects.

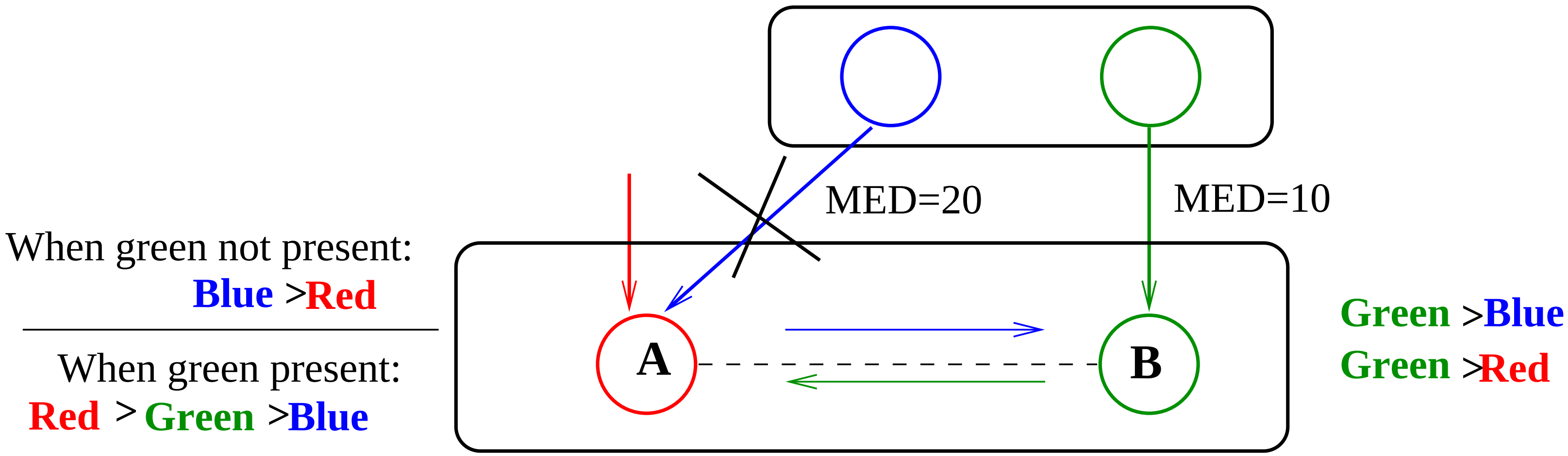
Step 1: Egress Routers Compute Best Routes

- Algorithm:
  1. Consider locally-best route at one router.
  2. Eliminate routes as follows:
     - If the route is worse than the locally-best route at another router, eliminate it.
     - Else, select it, and eliminate all other routes at that router.



Step 1: Egress Routers Compute Best Routes

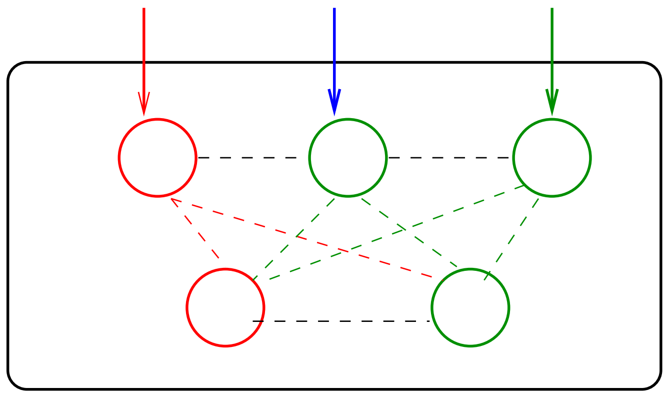
- Algorithm:
  1. Consider locally-best route at one router.
  2. Eliminate routes as follows:
     - If the route is worse than the locally-best route at another router, eliminate it.
     - Else, select it, and eliminate all other routes at that router.



Model a Certain Message Ordering

- **Step 1:** Egress routers compute best routes.
  - Outcome: Each egress router has a route to the destination.

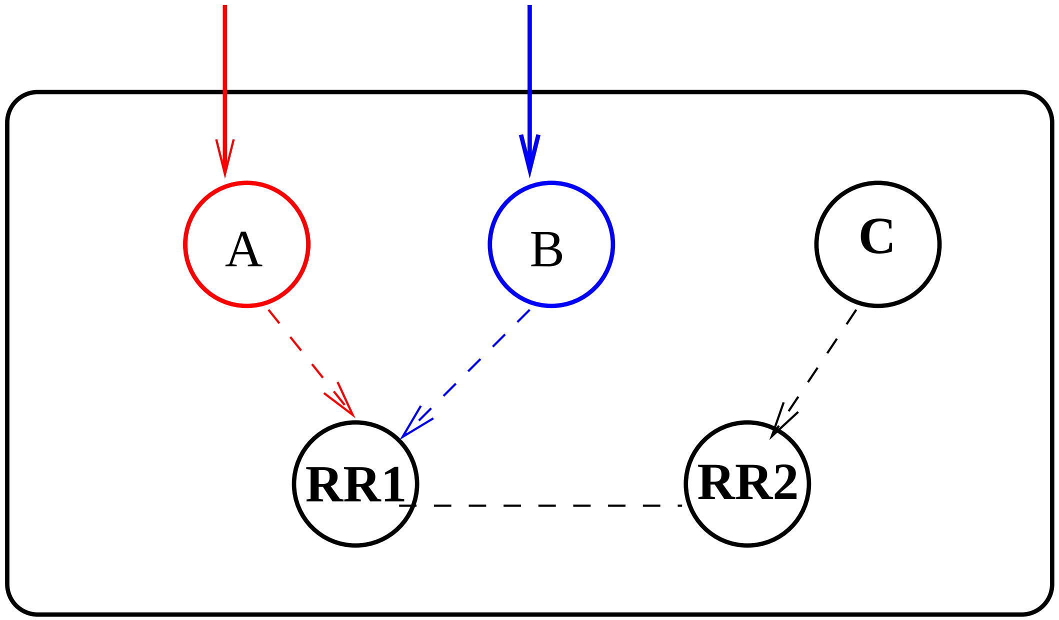
- **Step 2:** Egress routers propagate these routes to other routers within the AS.
  - Outcome: Each router in the AS selects a egress router.



Step 2: Choosing the Best Egress Router

- *Problem:*

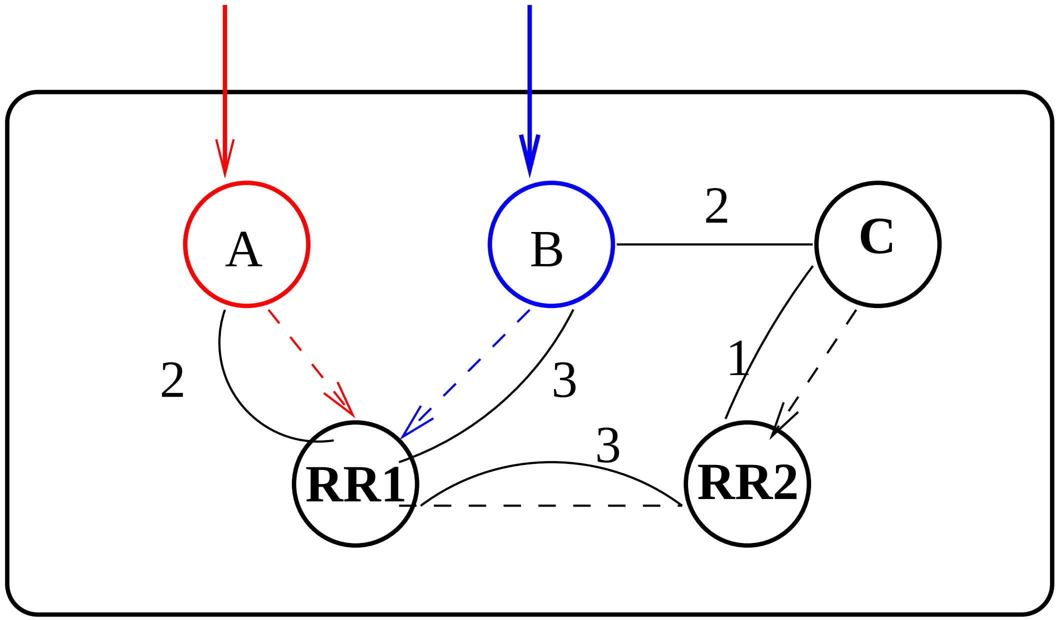
  Route at closest egress may not be visible.



Step 2: Choosing the Best Egress Router

- *Problem:*

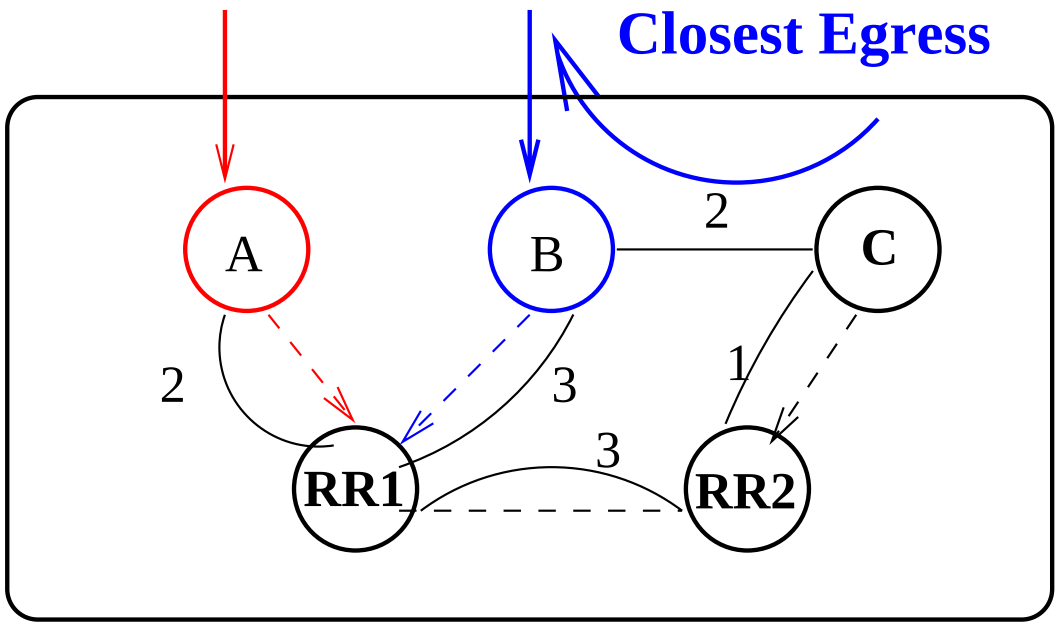
  Route at closest egress may not be visible.



Step 2: Choosing the Best Egress Router

- *Problem:*

  Route at closest egress may not be visible.



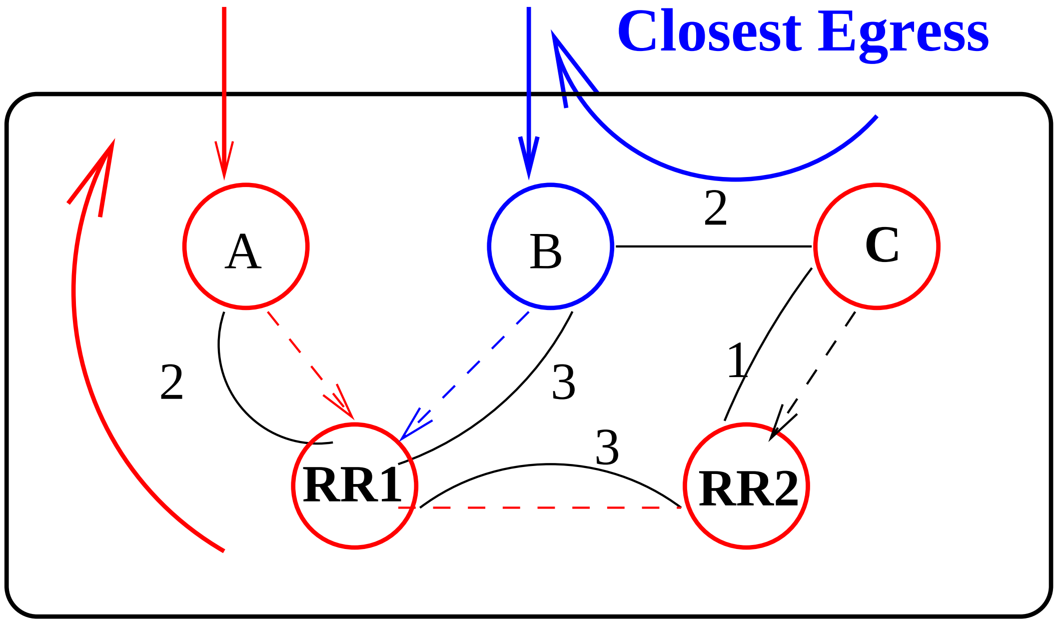
Step 2: Choosing the Best Egress Router

- *Problem:*

  Route at closest egress may not be visible.

RR1's closest egress is A.

RR2 and C select A.



- *Solution:*

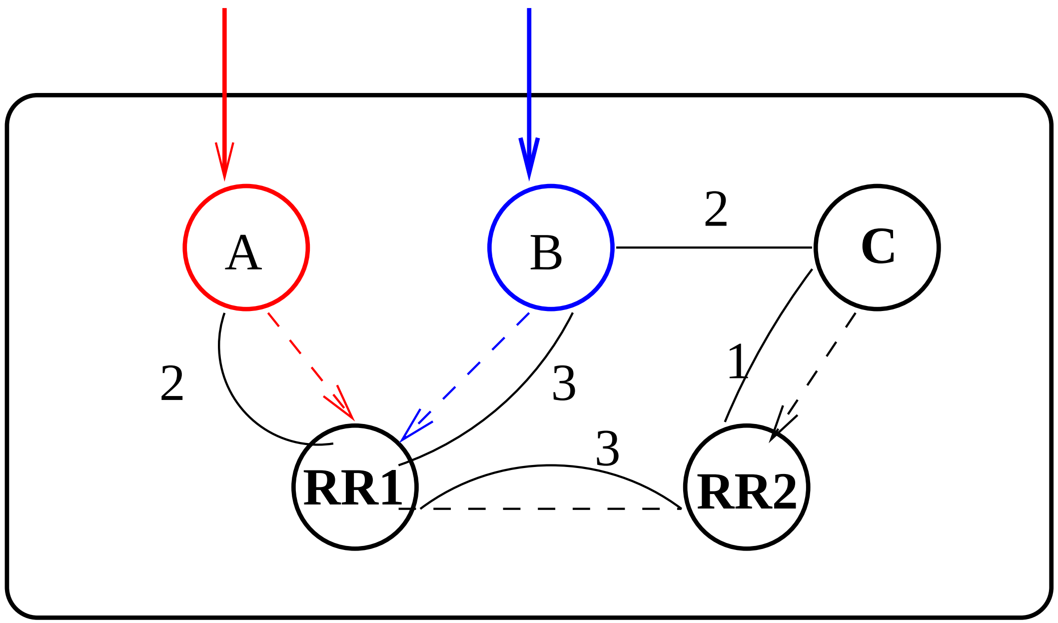
  First, compute routes at routers for which available routes are known. Propagate effects.

Step 2: Choosing the Best Egress Router

Algorithm (Graph Walk):

1. Assign routes to egress routers. *(done in Step 1)*
2. Assign routes to the parents of these routers.

Once at top level of the hierarchy, proceed down.

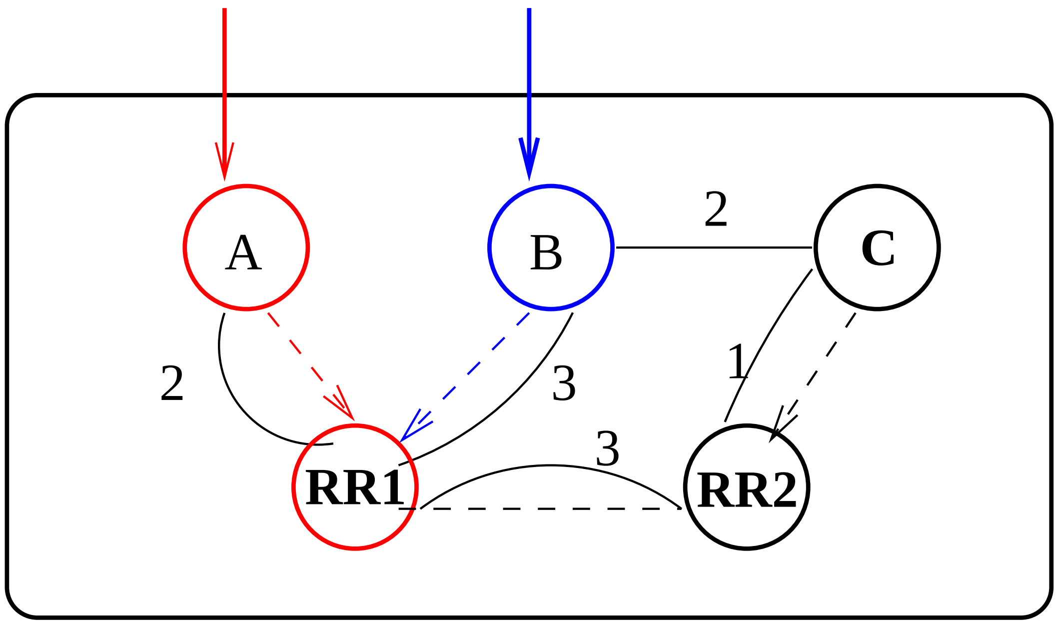


Step 2: Choosing the Best Egress Router

Algorithm (Graph Walk):

1. Assign routes to egress routers. *(done in Step 1)*
2. Assign routes to the parents of these routers.

Once at top level of the hierarchy, proceed down.

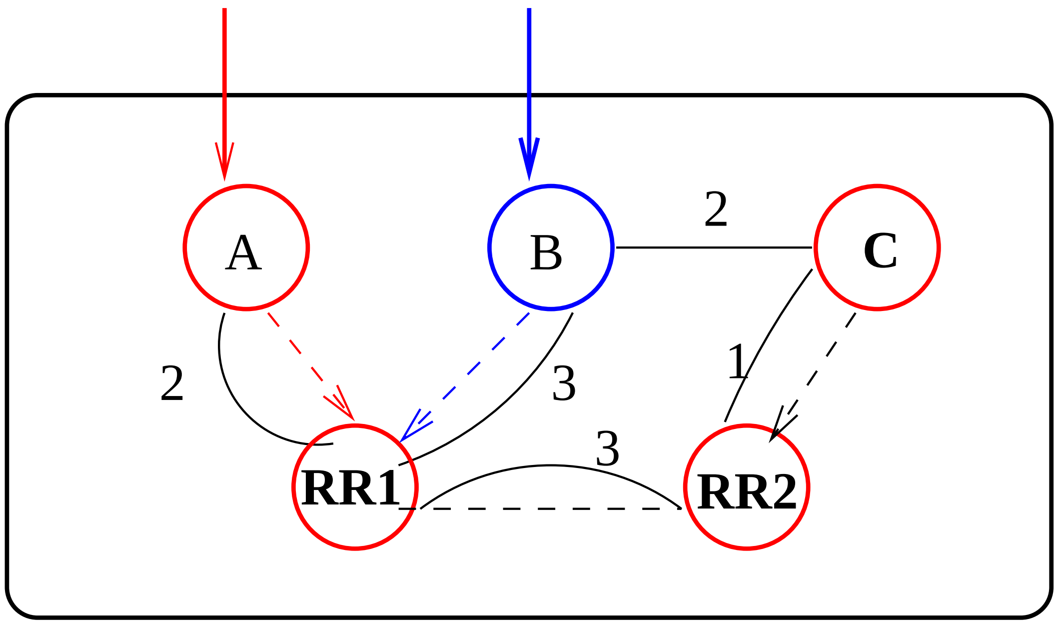


Step 2: Choosing the Best Egress Router

Algorithm (Graph Walk):

1. Assign routes to egress routers. *(done in Step 1)*
2. Assign routes to the parents of these routers.

Once at top level of the hierarchy, proceed down.



Step 2: Choosing the Best Egress Router

- **Constraints:**
  1. **No partitions in the internal BGP graph.**
  2. **Routers are "closer" to clients than non-clients.**
  3. **No cycles in the internal BGP graph.**

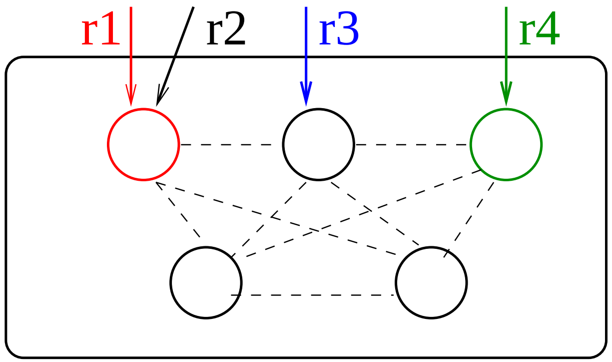
These constraints can be checked with static analysis.

Summary: Model of BGP Route Selection

- **Step 1:** Egress routers compute best routes.
  - Outcome: Each egress router has a route to the destination.

Input:
(r1, r2, r3, r4)

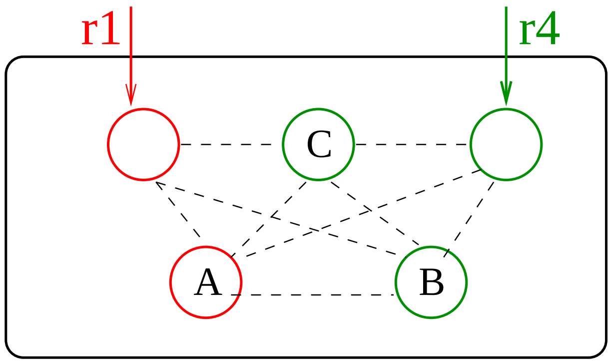
Output:
(r1, r4)



- **Step 2:** Intra-AS propagation.
  - Outcome: Each router in the AS selects a egress router.

Input:
(r1, r4)

Output:
A −> r1
B −> r4
C −> r4

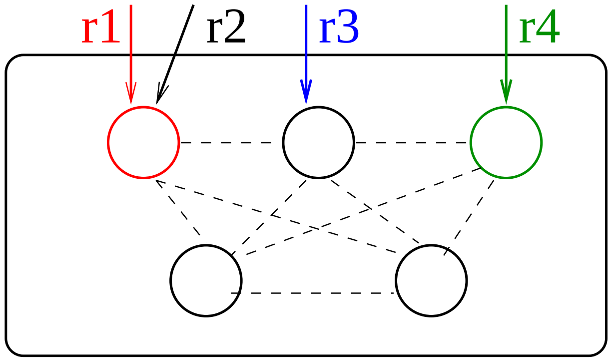


Summary: Model of BGP Route Selection

- **Step 1:** Egress routers compute best routes.
  - Outcome: Each egress router has a route to the destination.

Input:
(r1, r2, r3, r4)

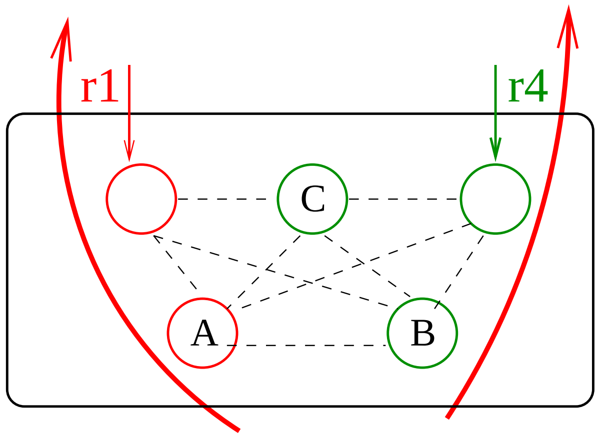
Output:
(r1, r4)



- **Step 2:** Intra-AS propagation.
  - Outcome: Each router in the AS selects a egress router.

Input:
(r1, r4)

Output:
A −> r1
B −> r4
C −> r4



Implementation is Efficient and Accurate

- **< 1 second** to compute effects of a policy change on AT&T network. (100s of routers, ~ 90k destinations)
  - Could be used as the "inner loop" for optimization.
- Model is accurate in more than 99% of all cases.

Router	*# Predictions*	*Total Errors*
RR1	89,343	554 (0.620%)
RR2	88,647	394 (0.444%)
AR1	88,649	391 (0.441%)
AR2	76,733	511 (0.666%)

Conclusion

- Operators must tune routing protocols as network conditions change.
  - Unfortunately, predicting the effects of these changes is difficult.
- We present a model of BGP routing.
  - Useful for offline computation
  - Fast and accurate
- Two artifacts complicate modeling BGP
  - No total ordering (MED attribute)
  - Limited route visibility (route reflection)

In the future, we should design distributed routing protocols in ways that facilitate modeling.