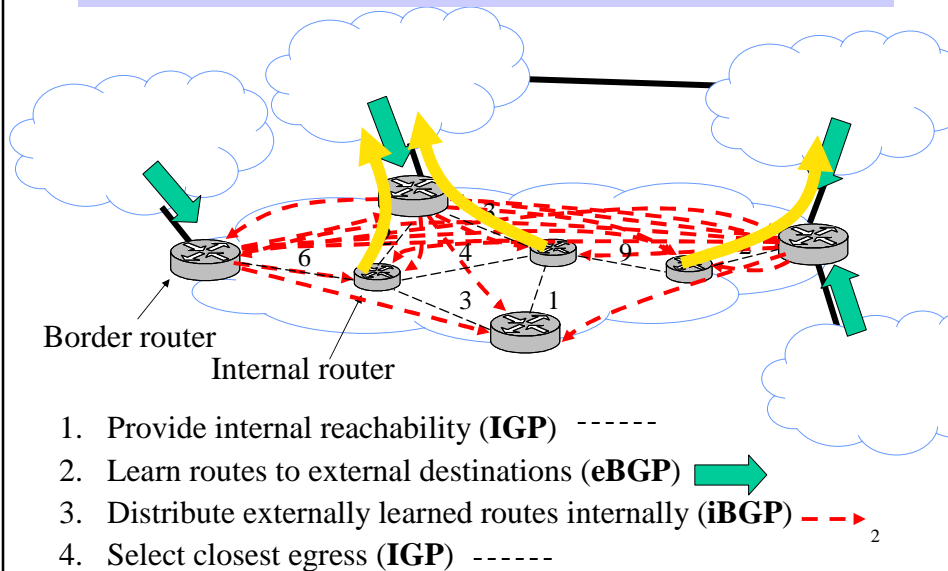


Design and implementation of a Routing Control Platform

Matthew Caesar, Donald Caldwell,
Nick Feamster, Jennifer Rexford,
Aman Shaikh, Jacobus van der Merwe

1

How ISPs route



What's wrong with Internet routing?

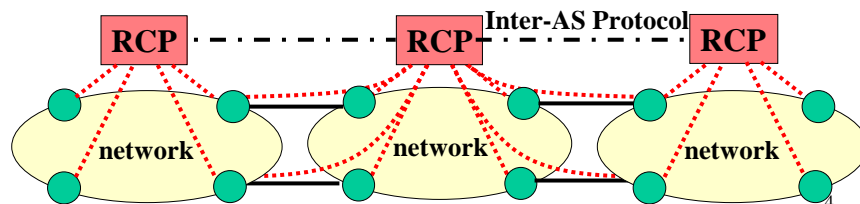
- **Full-mesh iBGP doesn't scale**
 - # sessions, control traffic, router memory/cpu
 - Route-reflectors help by introducing hierarchy
 - but introduce configuration complexity, protocol oscillations/loops
- **Hard to manage**
 - Many highly configurable mechanisms
 - Difficult to model effects of configuration changes
 - Hard to diagnose when things go wrong
- **Hard to evolve**
 - Hard to provide new services, improve upon protocols

3

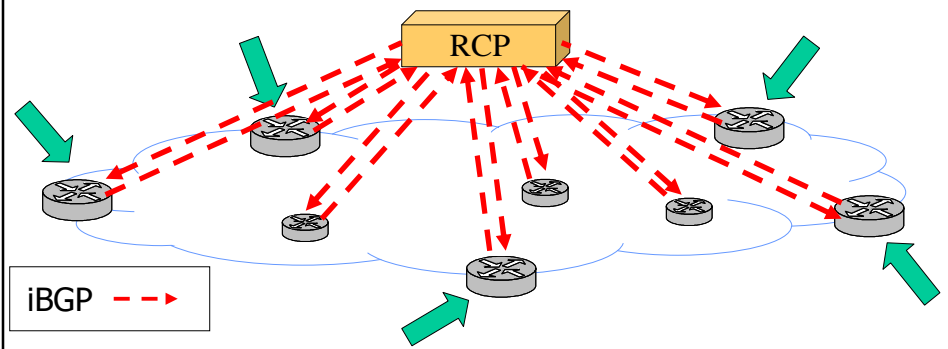
Routing Control Platform

- **What's causing these problems?**
 - Each router has limited visibility of IGP and BGP
 - No central point of control/observation
 - Resource limitations on legacy routers

Solution: compute routes from central point, remove protocols from routers



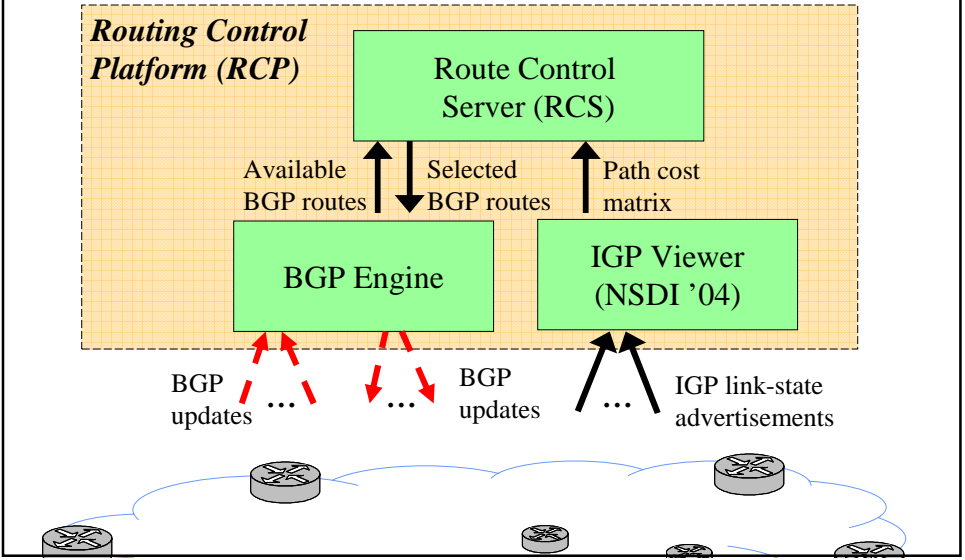
RCP in a single ISP



- **Better scalability:** reduces load on routers
- **Easier management:** configuration from a single point
- **Easier evolvability:** freedom from router software

5

RCP architecture



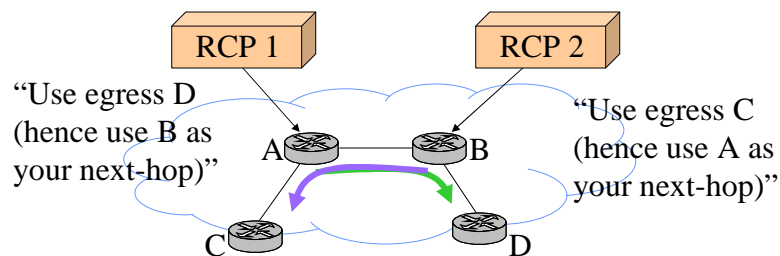
Challenges and contributions

- Reliability
 - **Problem**: single point of failure
 - **Contribution**: simple replication of RCP components
- Consistency
 - **Problem**: inconsistent decisions by replicas
 - **Contribution**: guaranteed consistency without inter-replica protocol
- Scalability
 - **Problem**: storing all routes increases cpu/memory usage
 - **Contribution** : can support large ISP in one computer

→ Building this system is feasible

7

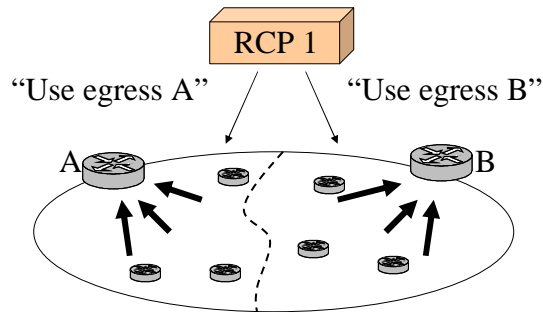
Potential consistency problem



- Need to ensure routes are consistently assigned
 - Even in presence of failures/partitions

8

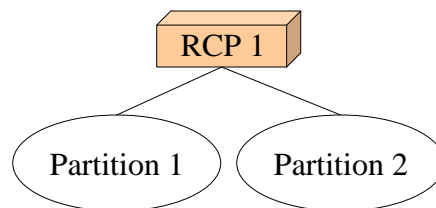
Consistent assignment Single RCP, single partition



- **Solution:** Assign all routers along the shortest IGP path the same exit router
 - Ensures forwarding loops don't arise

9

Consistent assignment Single RCP, multiple partitions

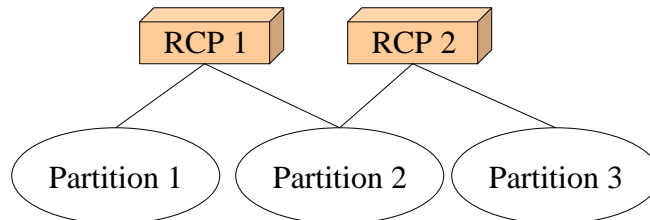


Forwarding plane is partitioned (RCP not in forwarding plane!)

- **Solution:** Only use state from router's partition in assigning its routes
 - Ensures next hop is reachable

10

Consistent assignment Multiple RCPs, multiple partitions



- **Solution:** RCPs receive same IGP/BGP state from each partition they can reach
 - IGP provides complete visibility and connectivity
 - RCS only acts on partition if it has complete state for it
 - Conservative → may affect liveness (but not correctness)

→ No consistency protocol needed to guarantee consistency in steady state

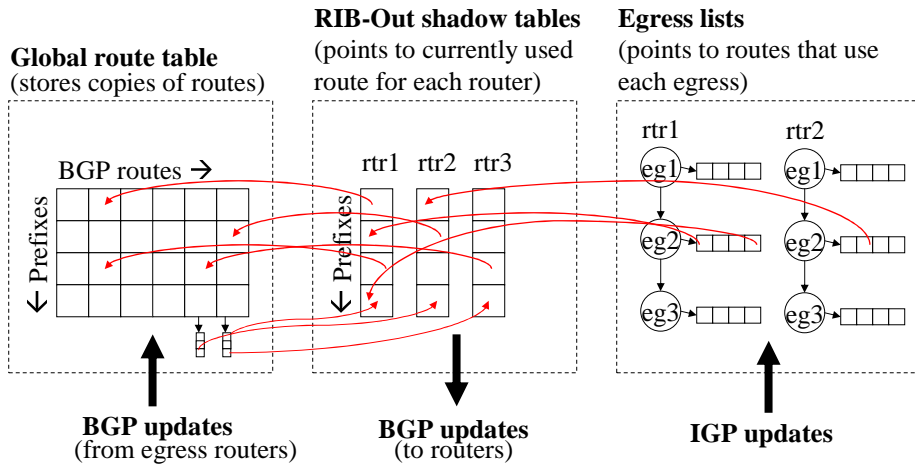
11

Scalability solution

- **Eliminate redundancy**
 - Store only a single copy of each BGP route
- **Accelerate lookup**
 - Quickly find routers whose routes changed
- **Avoid recomputation**
 - Compute routes once for groups of routers
 - Don't recompute if relative ranking of egress routers unchanged

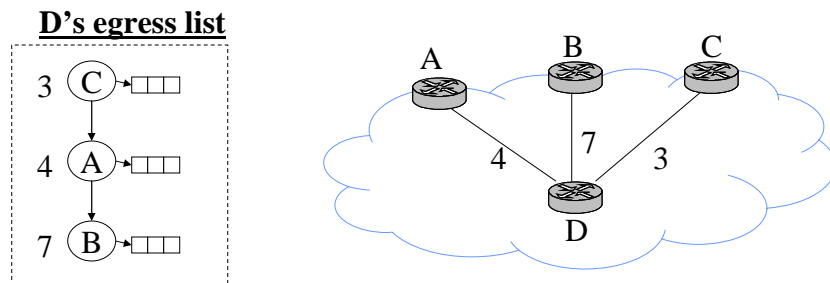
12

RCS data structures



13

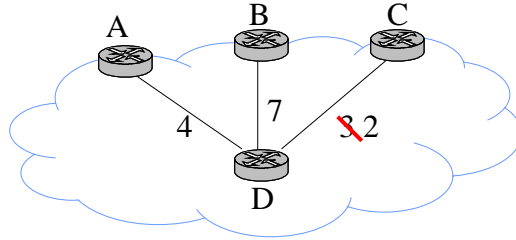
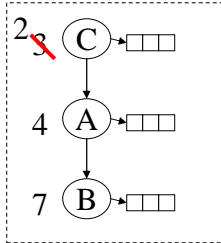
Example of egress list operation



14

Example of egress list operation

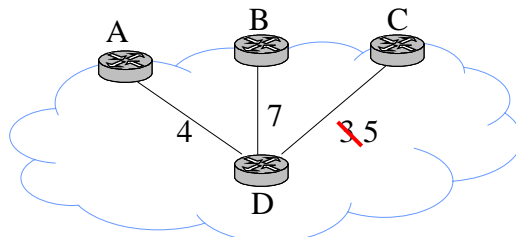
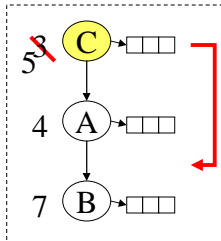
D's egress list



15

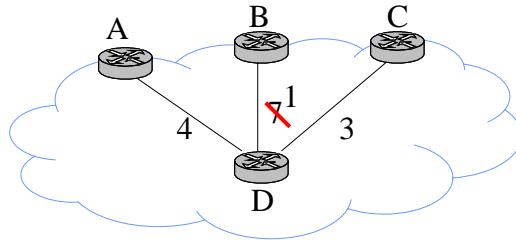
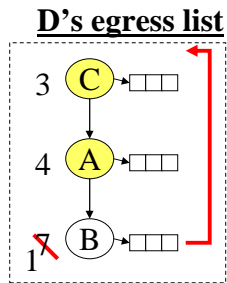
Example of egress list operation

D's egress list



16

Example of egress list operation



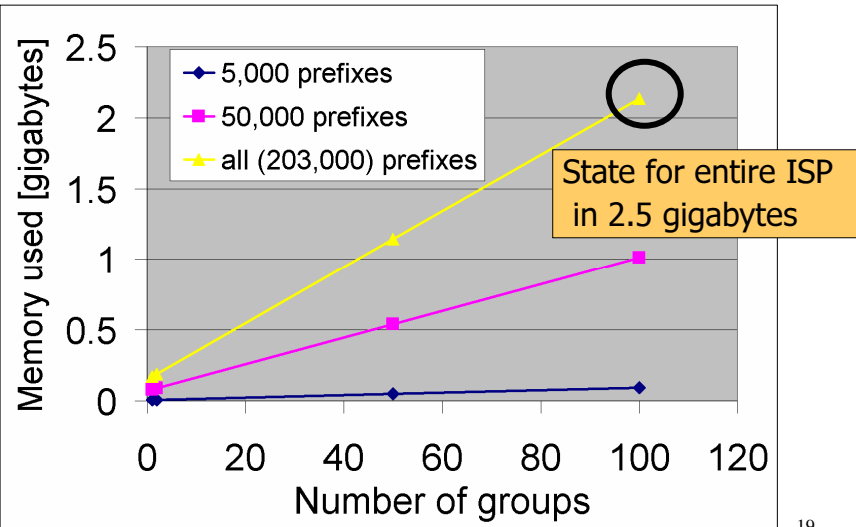
17

Performance evaluation

- **BGP and OSPF logs from Tier-1 ISP backbone**
 - collected on Aug 1 2004, ~500 routers
- **Metrics:** memory usage, update processing time
- **Measurement techniques:**
 - **Whitebox** (instrument code with timers)
 - **Blackbox** (workload generator on separate machine)
 - no-queuing (one update at a time)
 - real-time (allow updates to queue)
- **3.2 Ghz P4, 4GB memory, Linux 2.6.5**

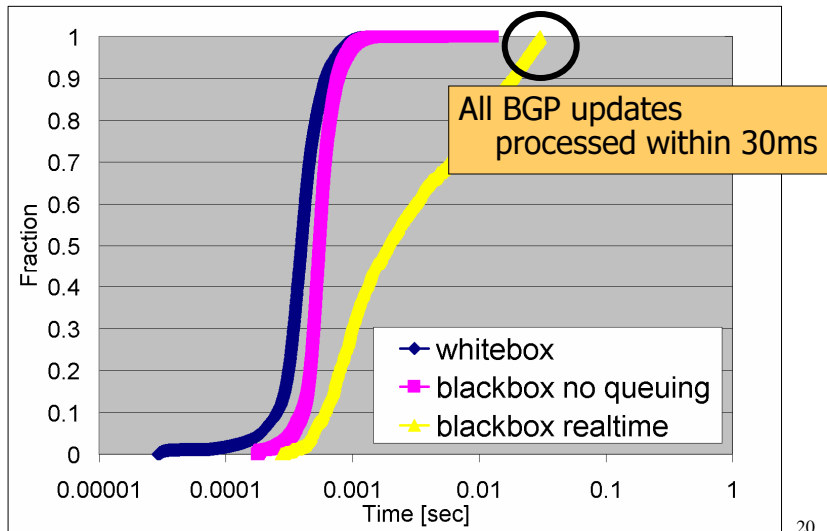
18

Results: RCS memory usage



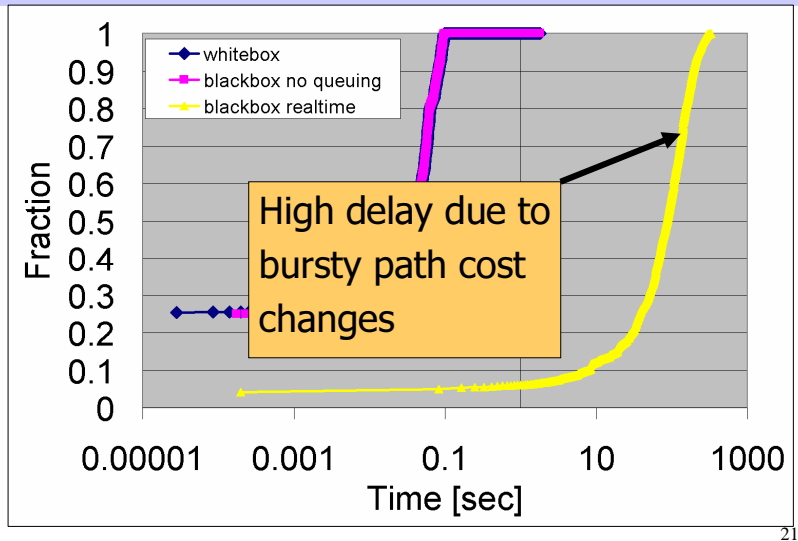
19

BGP change processing time

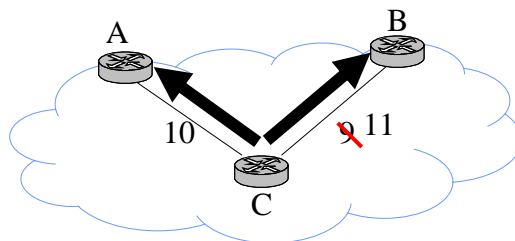


20

IGP change processing time



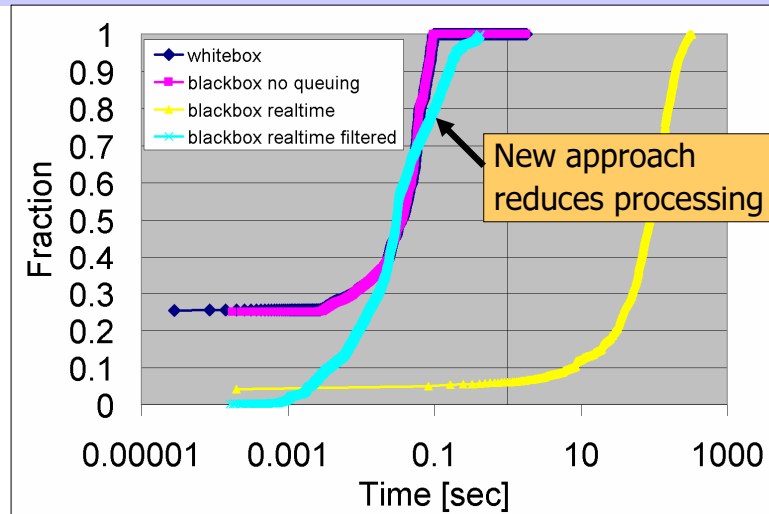
Towards decoupling BGP from IGP



- **Problem:** Single link change can affect many paths
 - Transient delay/loss, traffic shift, and eBGP updates
- **Solution:** Decouple egress point ranking and cost
 - Experiment: process only reachability-affecting events

22

IGP change processing time



23

Conclusions

- **RCP improves routing**
 - Correct, scalable route distribution
 - Eases management and evolvability
- **RCP is feasible**
 - Reliability, scalability, deployability, consistency
- **Many open problems:**
 - How to simplify network management
 - How to enable new services
 - RCP cooperation between ISPs

24