Interdomain routing with BGP
Issues and challenges

Olivier Bonaventure

Department of Computing Science and Engineering
Université catholique de Louvain (UCL)

Email : Bonaventure@info.ucl.ac.be
URL : http://www.info.ucl.ac.be/people/OBO

Outline

- Routing in the Internet and BGP principles
- Some issues and challenges
  - Scalability of interdomain routing
  - Performance of interdomain routing
  - Security of interdomain routing

Routing in the Internet

- Two different types of routing in Internet

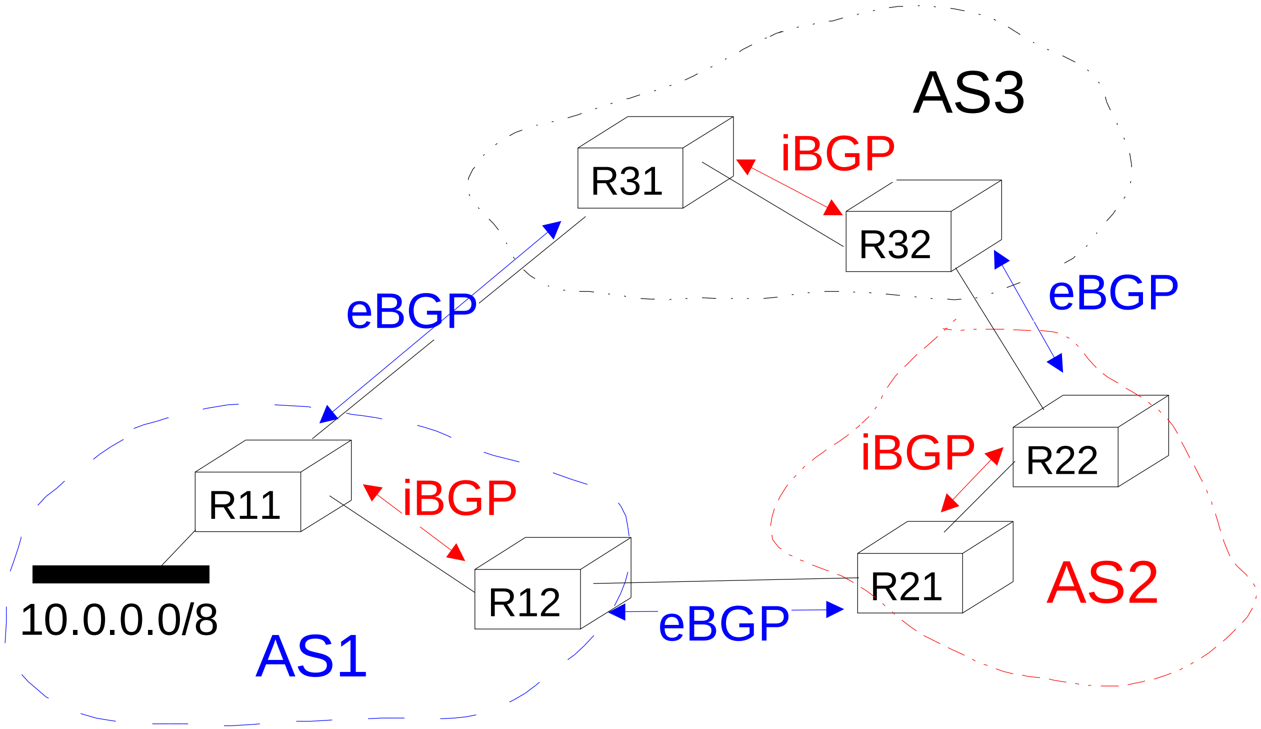
- Intradomain routing (IGP)
  - Objective
    - select the best path towards each destination based on some metrics (e.g. Delay, bandwidth) used inside AS
- Interdomain routing (EGP)
  - Objective
    - select the best path towards each destination that is compatible with the routing policies of the transit ASs without knowing the topology of those transit ASs
  - Issues
    - Each AS is allowed to define its own routing policy
    - EGP should be scalable (13.000 AS, 120.000 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      - routes are withdrawn when they stop being reachable
  - BGP routers use policies to filter and rank the routes sent or received

The Border Gateway Protocol (2)

- The two variants of BGP
  - eBGP between border routers of distinct AS
  - (full-mesh) iBGP between BGP routers inside AS

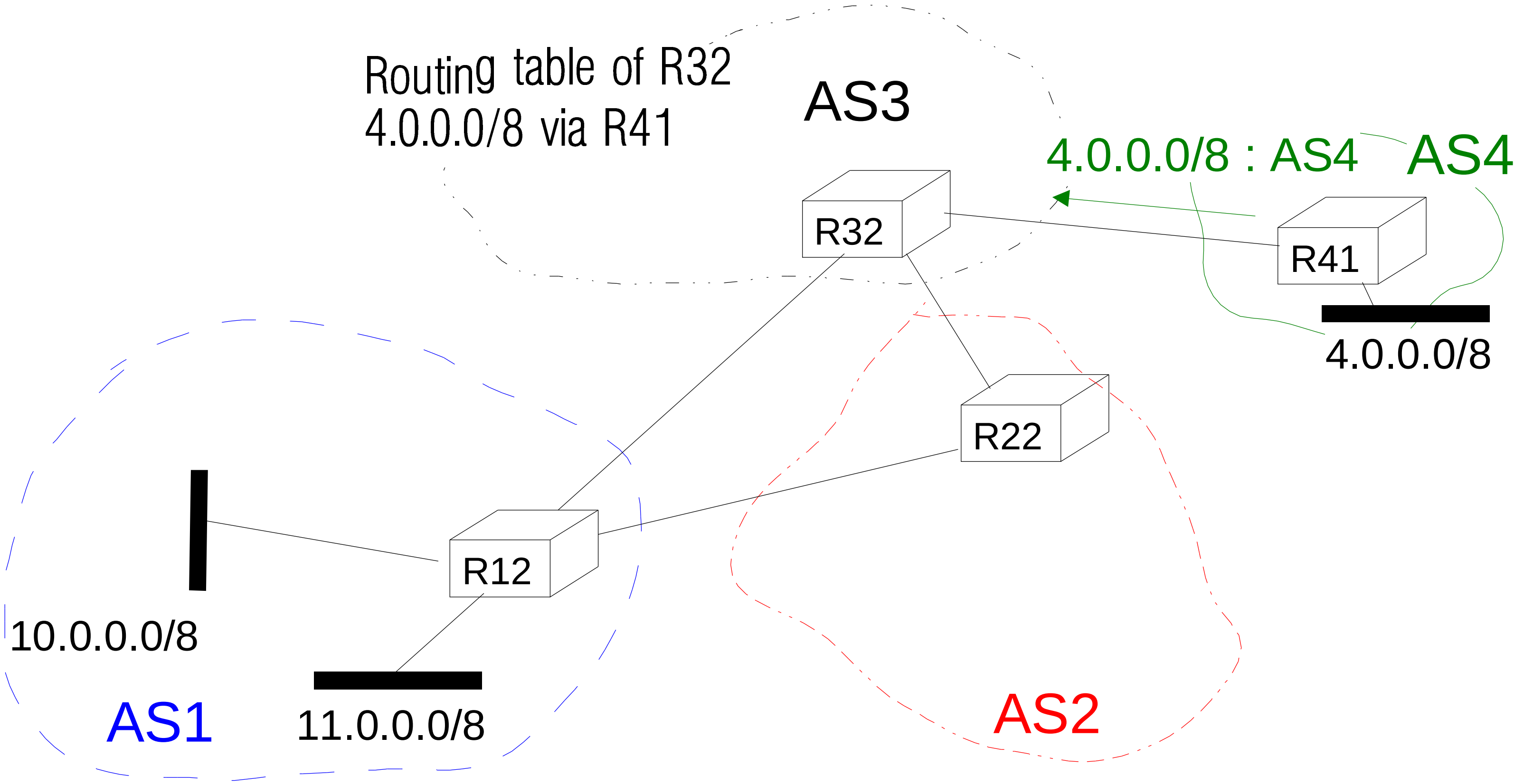


The Border Gateway Protocol (3)

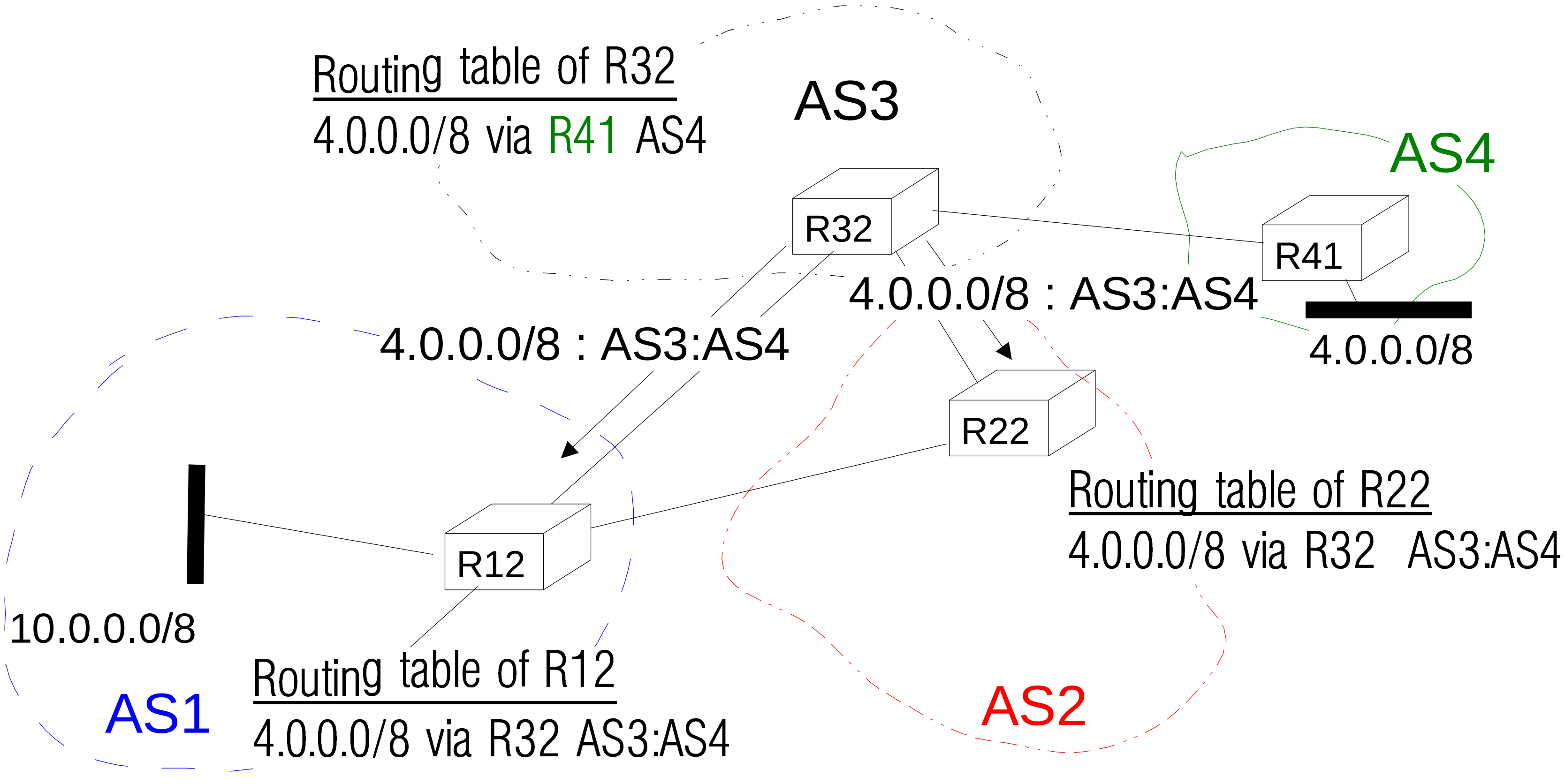
- Routes distributed in UPDATE messages
  - Contents
    - List of reachable IP prefix, List of withdrawn IP prefixes and several attributes (e.g. AS-Path)
- Processing of UPDATE message
  - For each reachable IP prefix in UPDATE
    - Add route to set of known routes towards IP prefix
    - Select the best route among all those routes for forwarding
    - If the best route towards this destination changed readvertise the best route to peers
  - For each withdrawn IP prefix in UPDATE
    - Remove route from set of known routes towards IP prefix
    - Select the best route among remaining routes for forwarding
    - If the best route towards this IP prefix changed readvertise the best route to peers

BGP : example

- Distribution of the route towards 4.0.0.0/8



BGP : example (2)



Routing table of R32

4.0.0.0/8 via R41 AS4

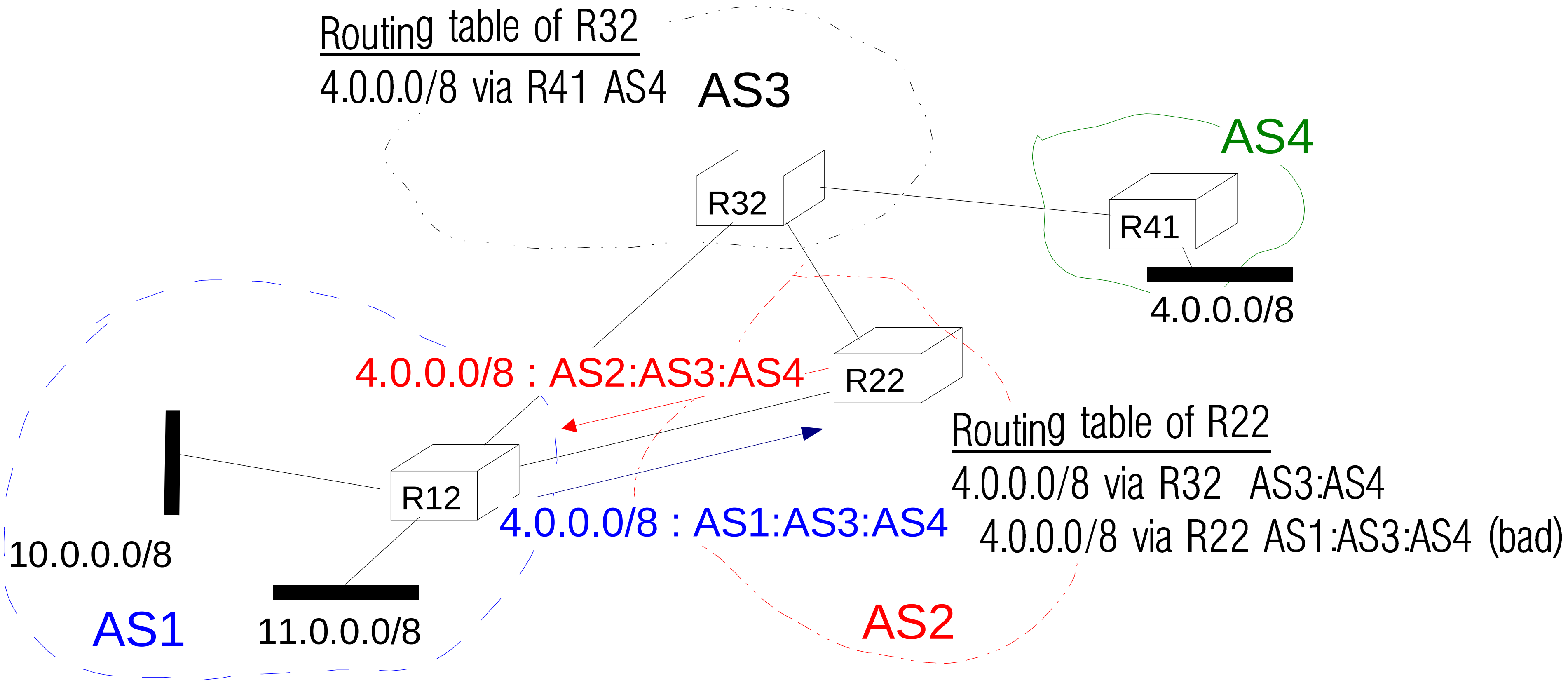
Routing table of R22

4.0.0.0/8 via R32 AS3:AS4

Routing table of R12

4.0.0.0/8 via R32 AS3:AS4

BGP : example (3)



Routing table of R32

4.0.0.0/8 via R41 AS4

Routing table of R22

4.0.0.0/8 via R32 AS3:AS4

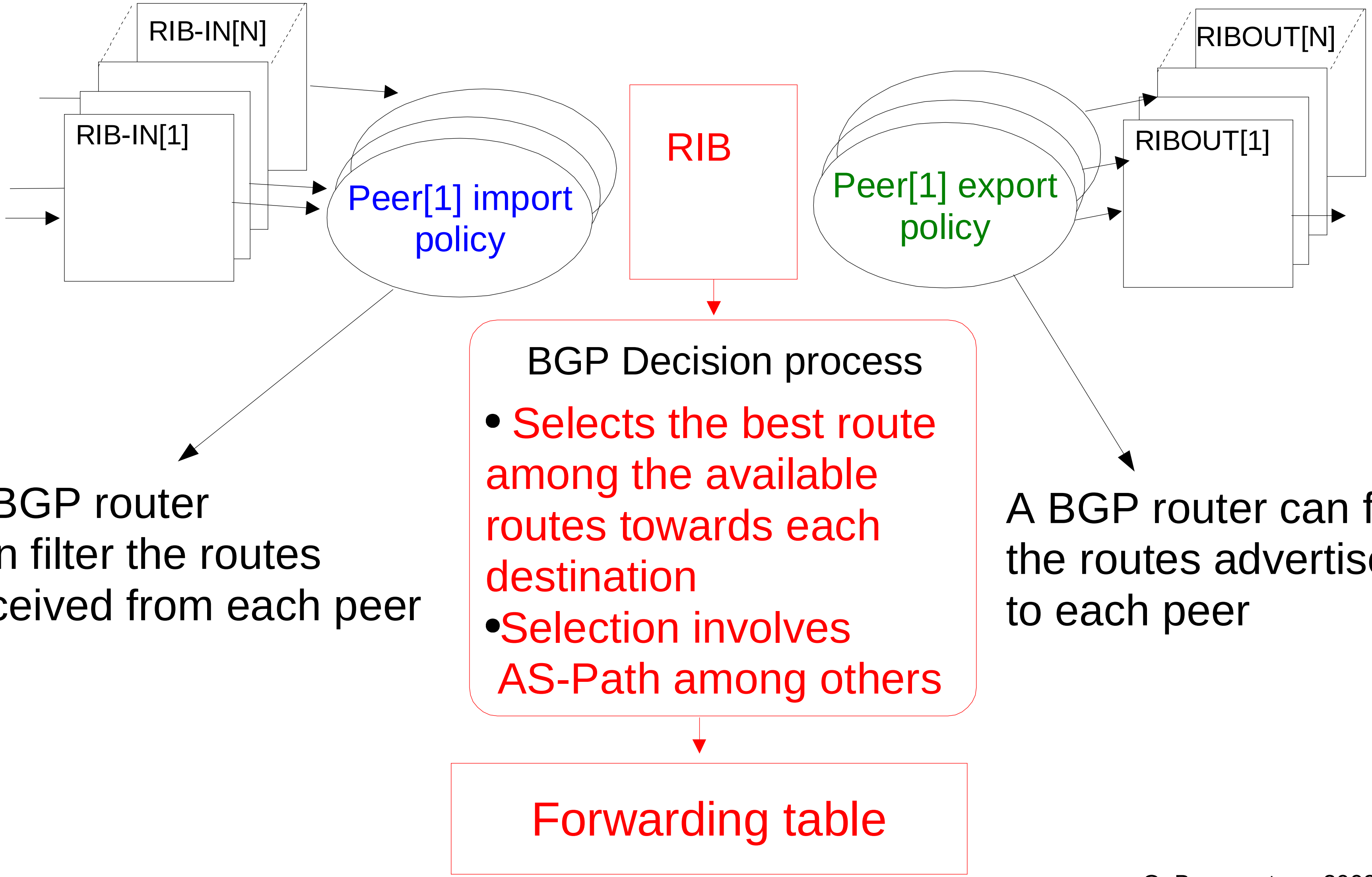
4.0.0.0/8 via R22 AS1:AS3:AS4 (bad)

Routing table of R12

4.0.0.0/8 via R32 AS3:AS4

4.0.0.0/8 via R22 AS2:AS3:AS4 (bad)

Organization of a BGP router



A BGP router can filter the routes received from each peer

BGP Decision process

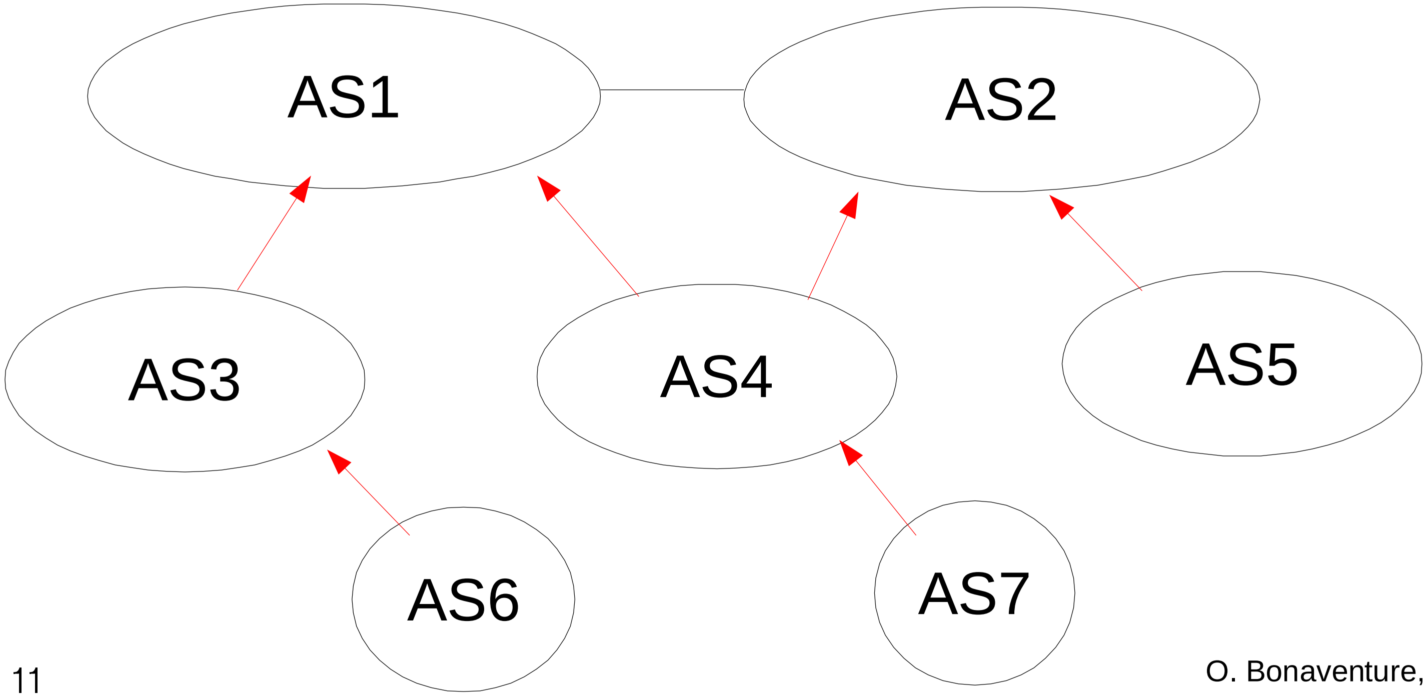
- Selects the best route among the available routes towards each destination
- Selection involves AS-Path among others

Forwarding table

A BGP router can filter the routes advertised to each peer

Routing policies : customer-provider

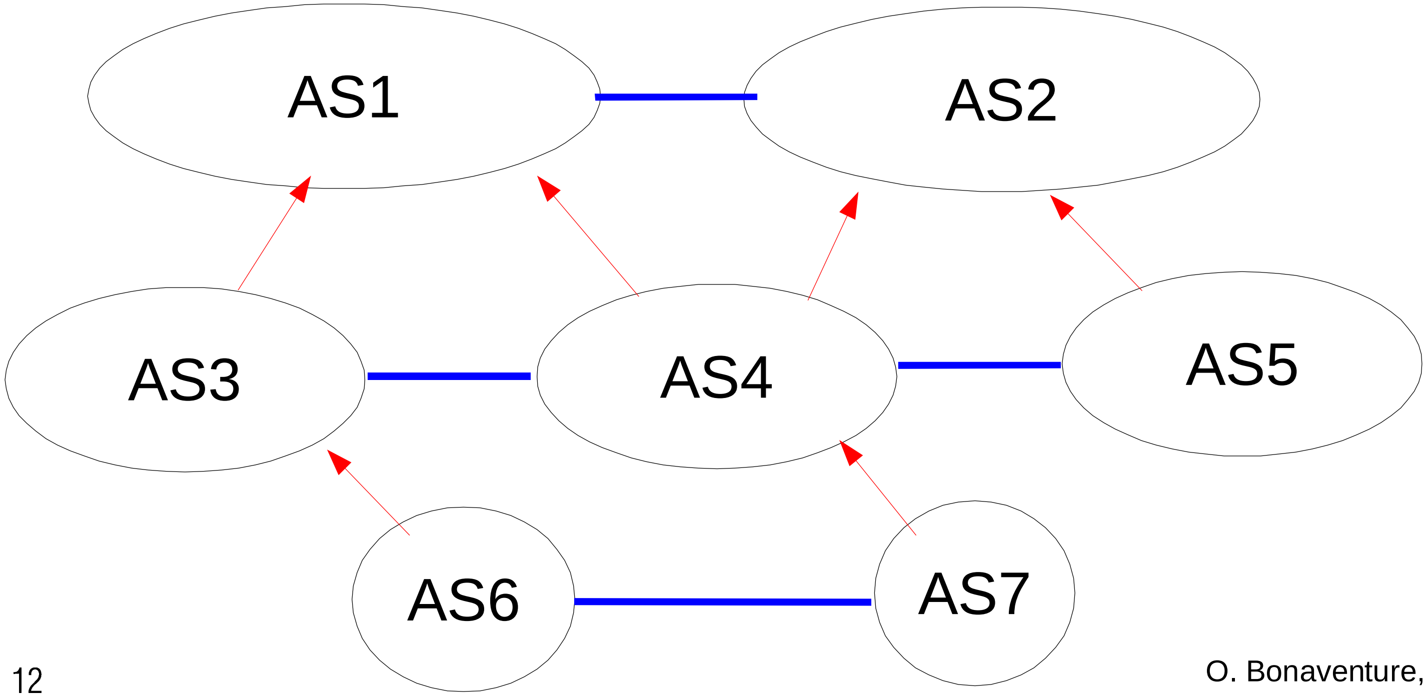
- Principle of customer-provider peering
  - ASc is a smaller ISP than ASp
  - ASc buys transit service from ASp
    - ASp agrees to transmit packets from Asc towards any destination
    - Asp agrees to announce the routes received from ASc



O. Bonaventure, 2002

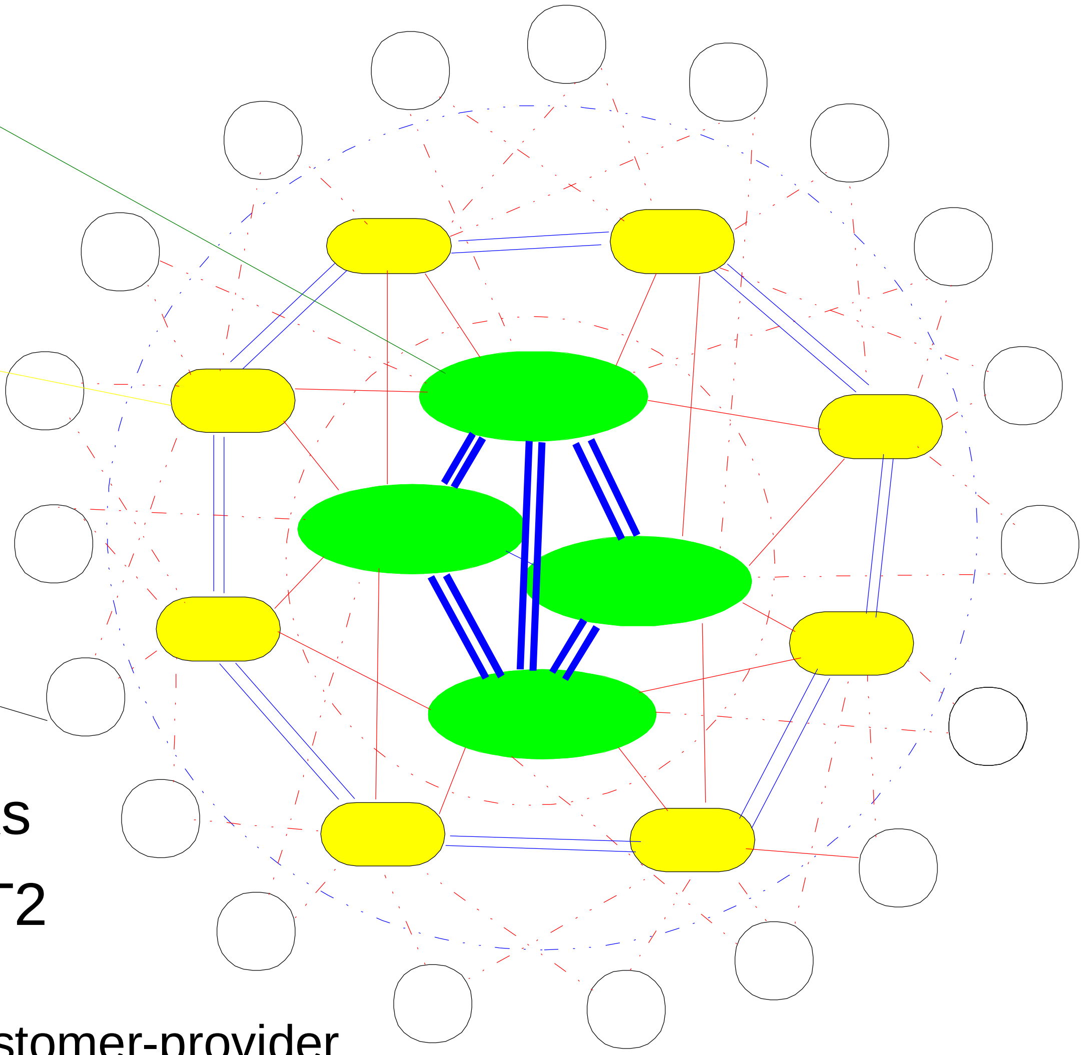
Routing policies : shared-cost

- Principle of shared-cost peering
  - usually used on links between Ass 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



═ Peer-to-peer

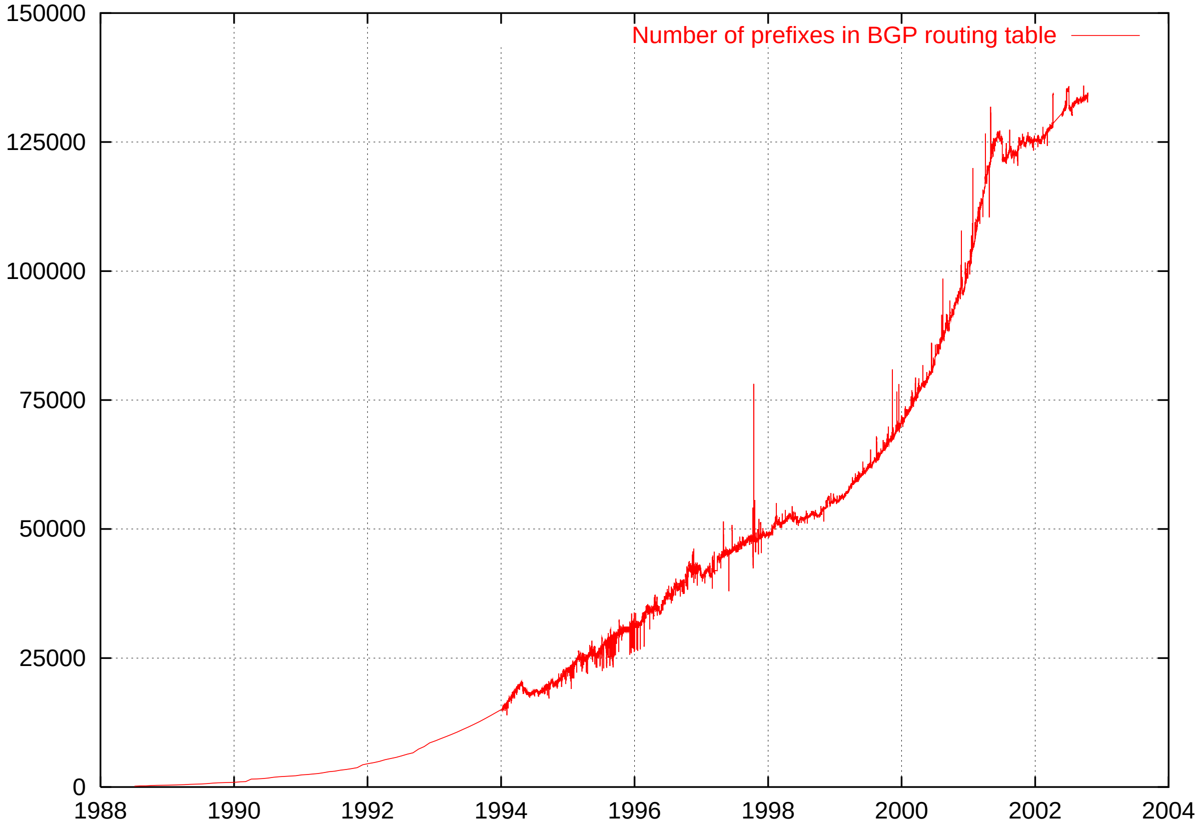
— Customer-provider

Issues and challenges

- How to sustain the growth of the Internet ?
  - In theory anyone can announce its routes with BGP
  - In practice, BGP routing tables cannot be infinite...
- How to support mission critical services in addition to the current best effort service ?
  - BGP should react quickly to link failures
  - An ISP should be able to control the flow of its interdomain traffic
- Security of interdomain routing ?

The growth of the BGP routing tables

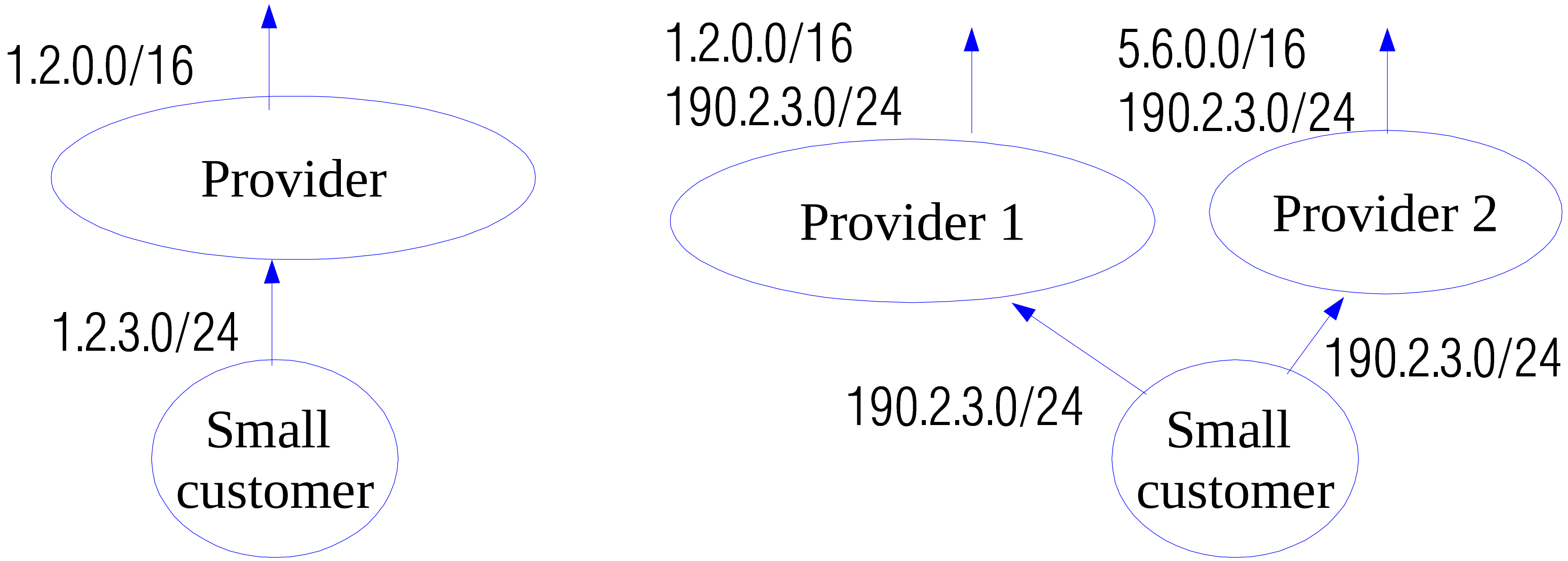
- Evolution of the number of prefixes in BGP routing tables



Source : G. Huston, http://bgp.potaroo.net

The reasons for the growth

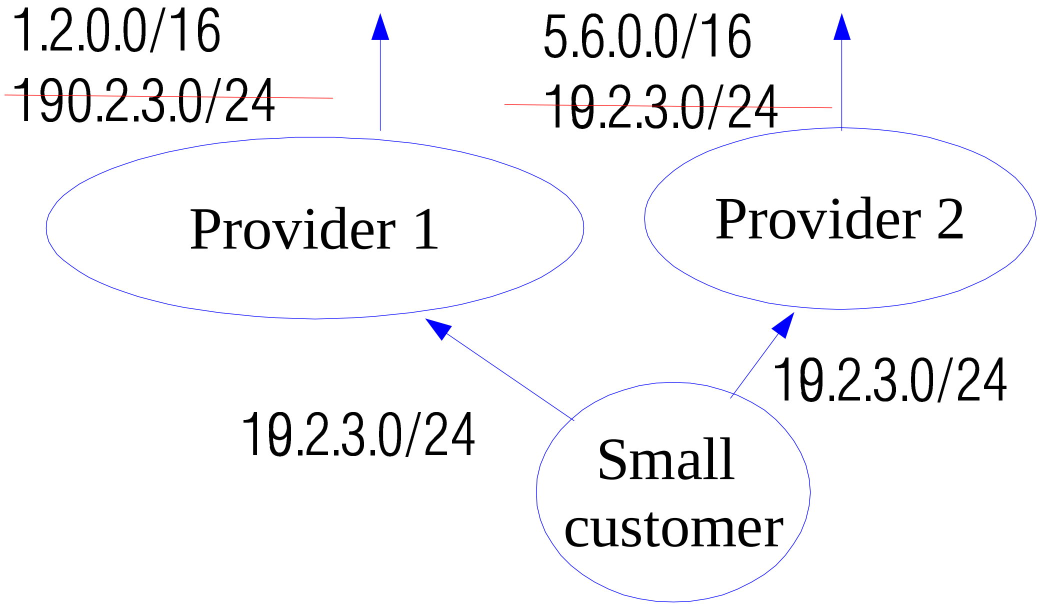
- The Internet is growing
  - The total number of IP addresses advertised increases slowly
- The Internet is more and more fragmented
  - More and more customer networks multi-homed



How to deal with growth of BGP tables ?

- Current « solution » (aka quick hack)
  - Some ISPs filter routes towards too long prefixes
  - Consequence
    - Some routes are not distributed to the global Internet

- Towards a better solution
  - Providers should perform more aggregation

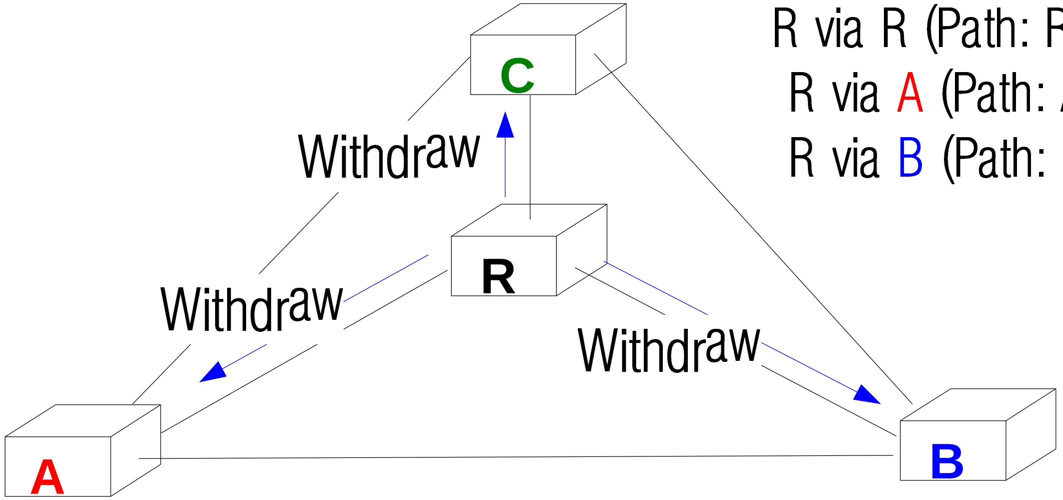


How to support mission critical services ?

- Example services
  - Voice over IP
  - Virtual Private Networks
- When an interdomain link fails, BGP should
  - Quickly announce the failure
  - Quickly distribute a new route to the destination
- Current BGP restoration times on the global Internet
  - From several tens of a second up to a few minutes and sometimes worse...

The reasons for the slow convergence

- The BGP protocol itself



Routing table of C

R via R (Path: R)

R via A (Path: A-R)

R via B (Path: C-R)

Routing table of A

R via R (Path: R)

R via B (Path: B-R)

R via C (Path: C-R)

Routing table of B

R via R (Path: R)

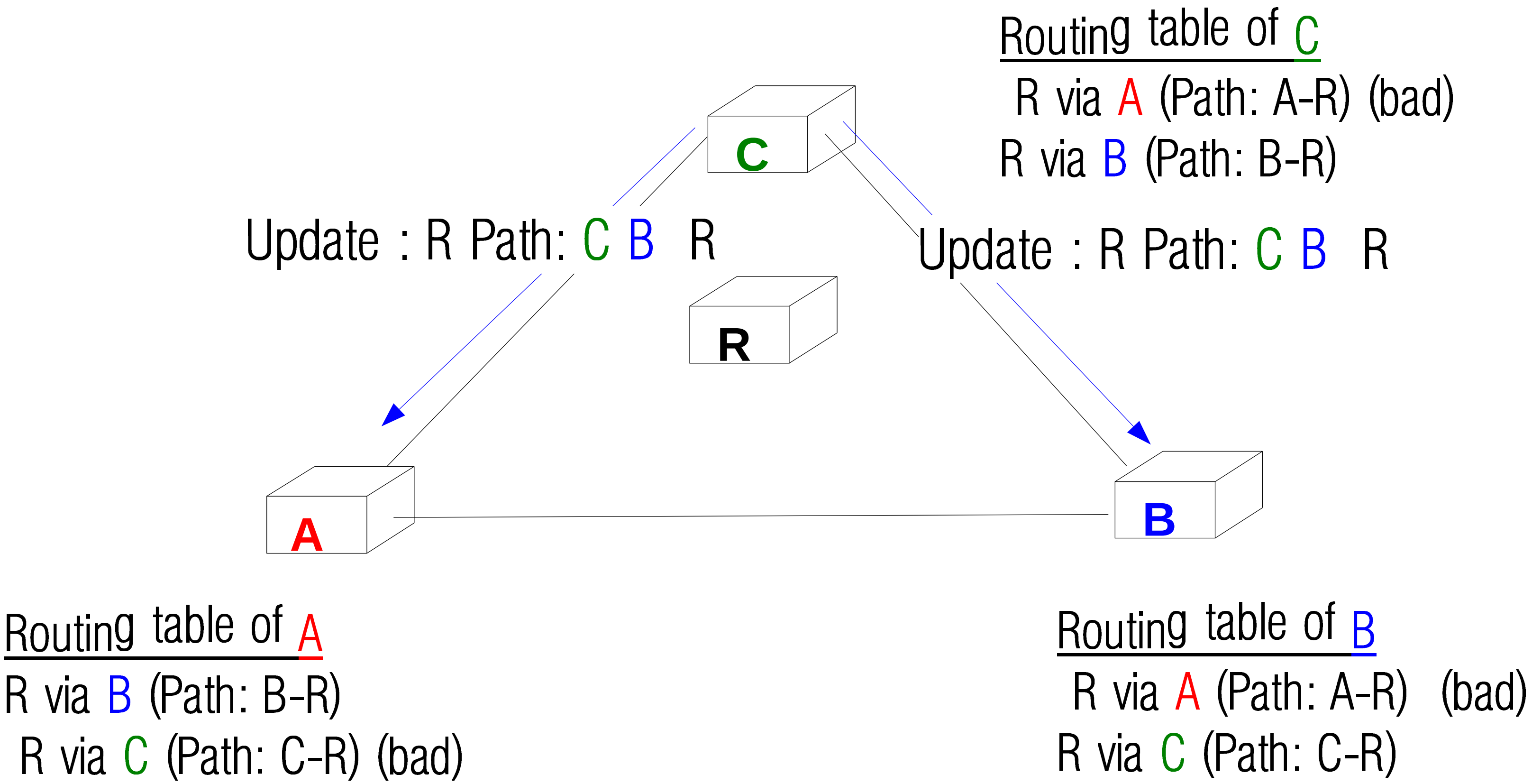
R via A (Path: A-R)

R via C (Path: C-R)

- Routers will process the withdraw and advertise alternate routes

The reasons for the slow convergence (2)

- C sends announcements



Routing table of C

R via A (Path: A-R) (bad)

R via B (Path: B-R)

Update : R Path: C B R

Update : R Path: C B R

Routing table of A

R via B (Path: B-R)

R via C (Path: C-R) (bad)

Routing table of B

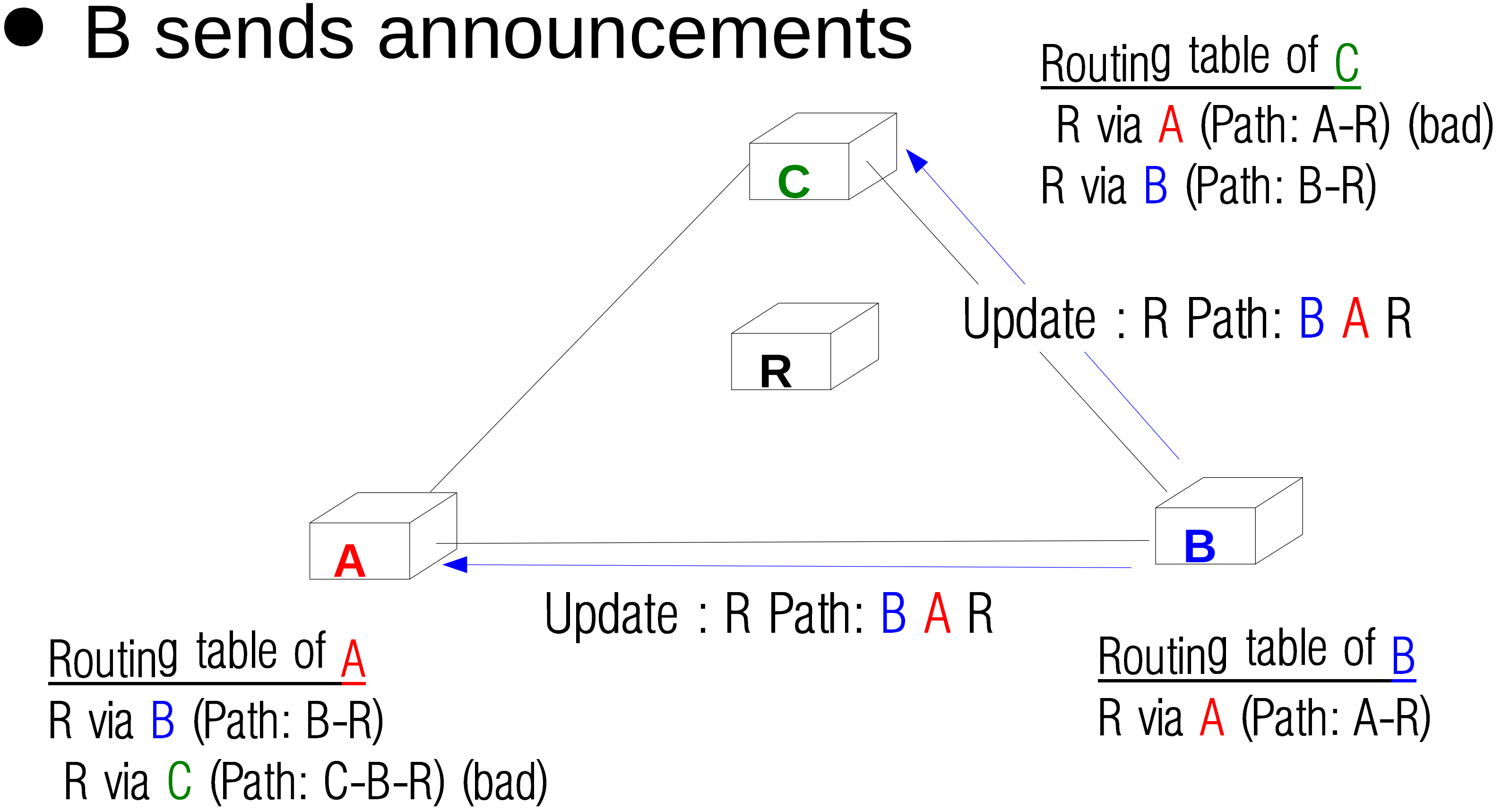
R via A (Path: A-R) (bad)

R via C (Path: C-R)

  - A learns a worse (but valid) route towards R
  - B learns that the route via C is a loop

The reasons for the slow convergence (3)

- B sends announcements



Routing table of C
R via A (Path: A-R) (bad)
R via B (Path: B-R)

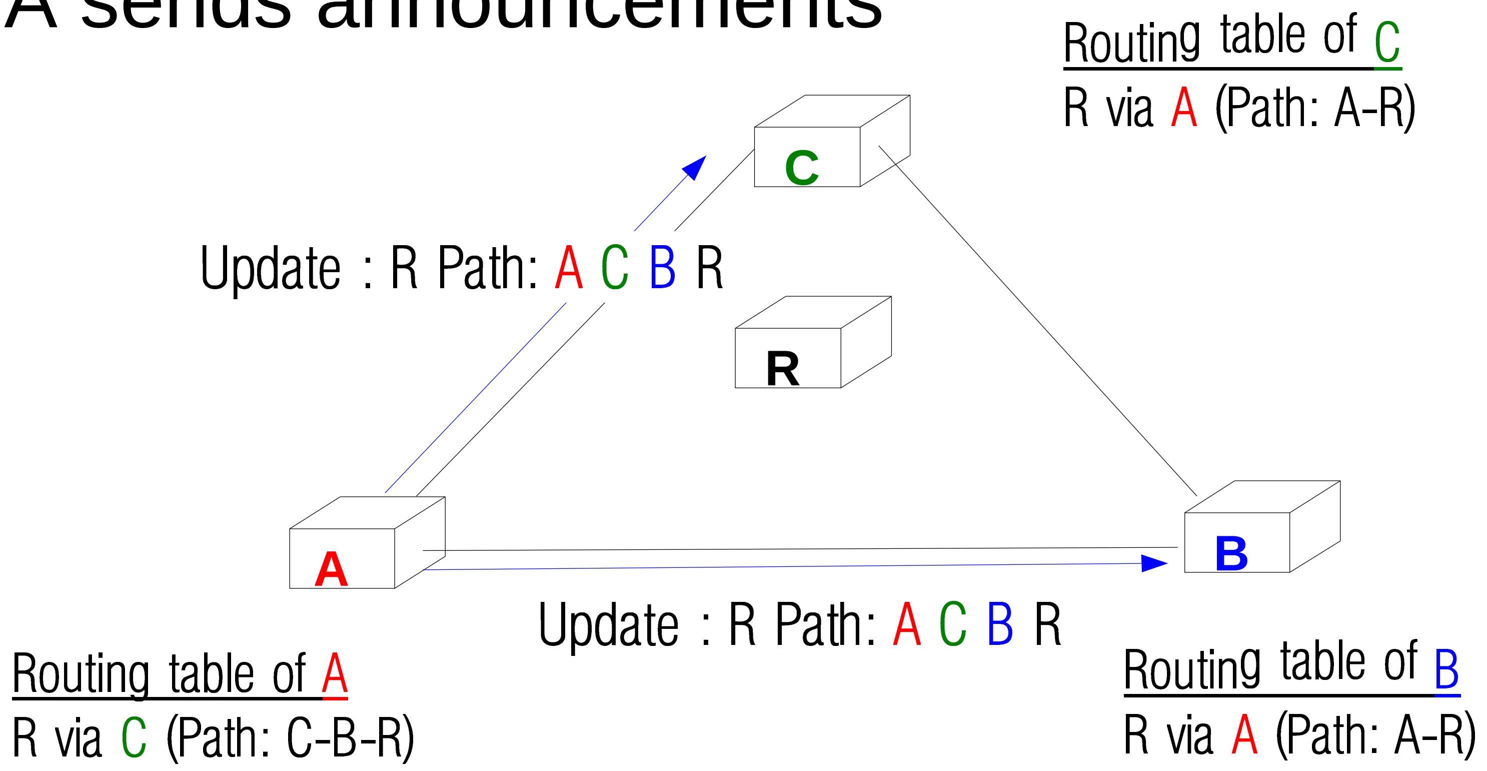
Routing table of A
R via B (Path: B-R)
R via C (Path: C-B-R) (bad)

Routing table of B
R via A (Path: A-R)

  - C learns a longer (but valid) path towards R
  - A learns that the route via B is a loop

The reasons for the slow convergence (4)

- A sends announcements



Routing table of C

R via A (Path: A-R)

Update : R Path: A C B R

Update : R Path: A C B R

Routing table of A

R via C (Path: C-B-R)

Routing table of B

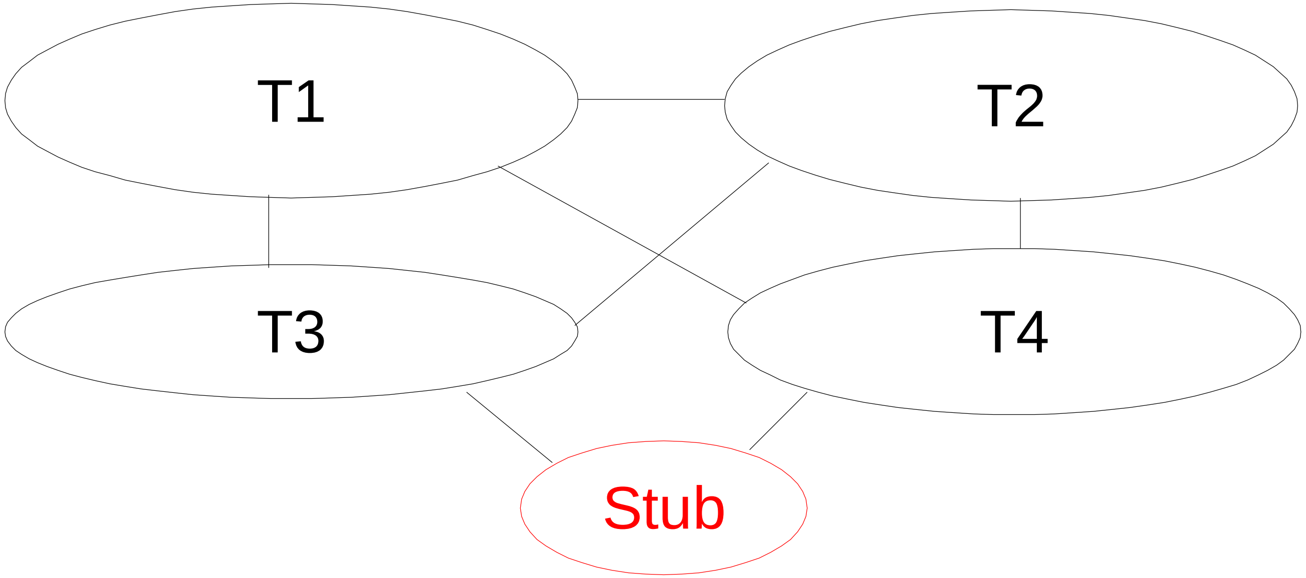
R via A (Path: A-R)

  - C learns that route via A is a loop
    - C will withdraw its route and inform A
  - B learns that route via A is a loop
- Improving the convergence of BGP is not easy

How to control the flow of interdomain traffic ?

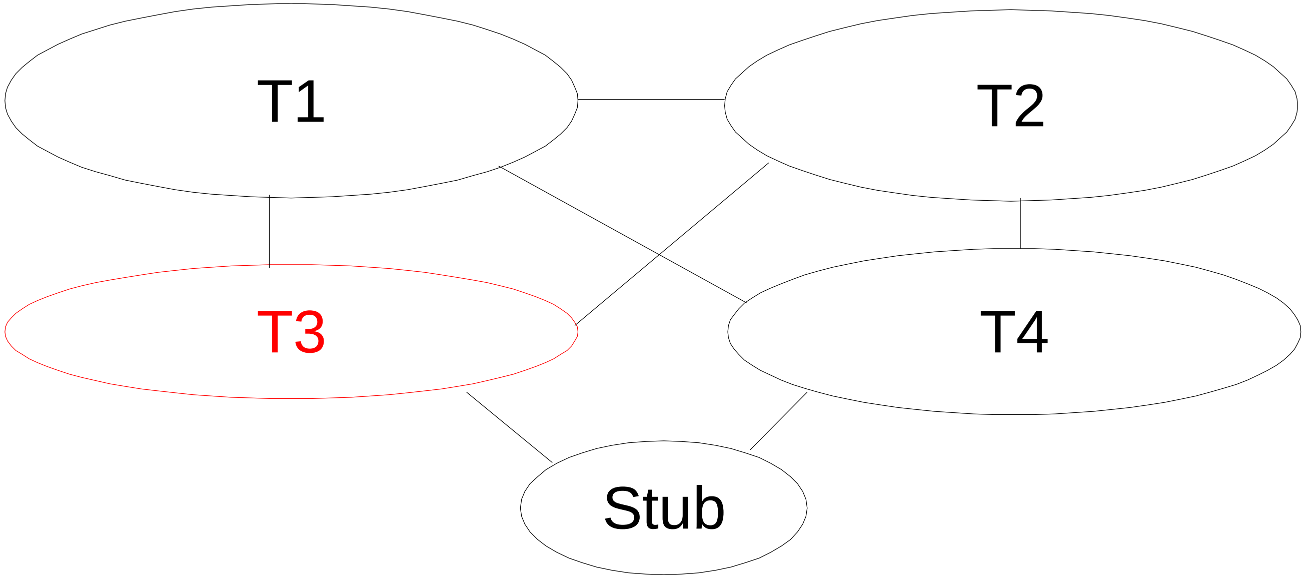
- Principle
  - If router x advertises a route towards destination d on link l, it implicitely agrees to forward to this destination any amount of traffic received on this link
- How to control the interdomain traffic on a link ?
  - 2 cases to consider
    - Stub domain that does not provide transit service
    - Transit domain that provides transit service to others

How to control interdomain traffic
Stub domain



- Control of the outgoing traffic
  - Stub can choose any received route
- Control of the incoming traffic
  - send different route advertisements on different links
    - Only announce part of the routes from stub on a link
    - Announce some routes as « bad » routes on one link
  - Dynamic changes require transmission of new BGP msgs

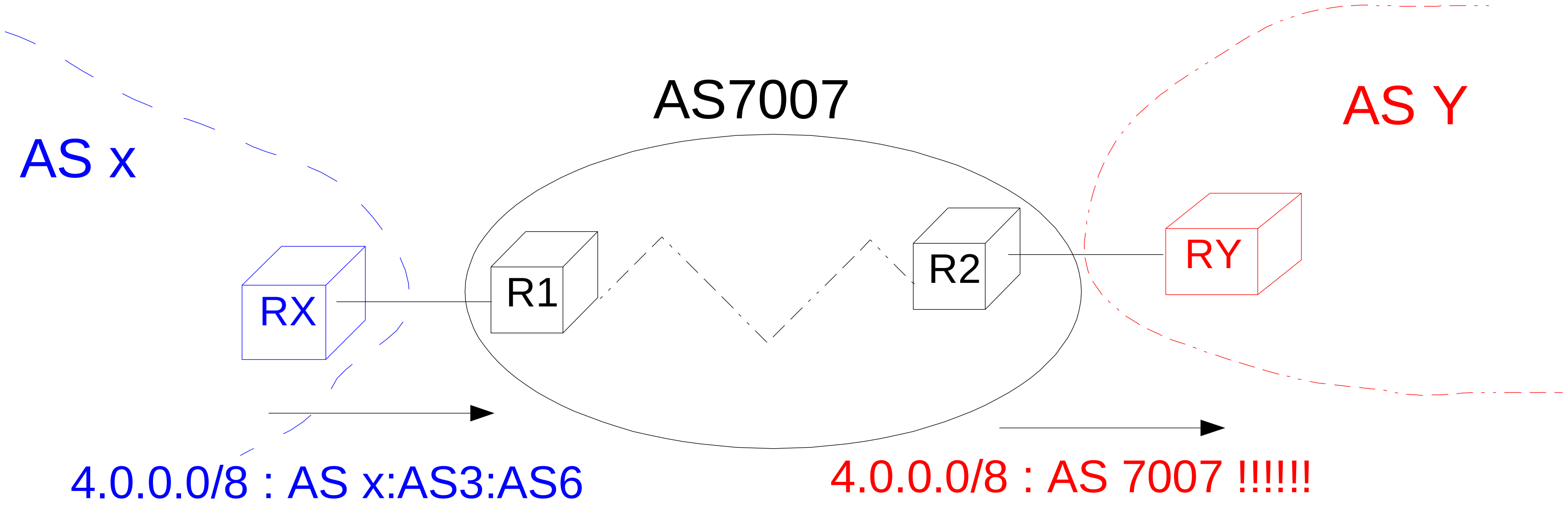
How to control interdomain traffic
Transit domain



- Control of the outgoing traffic
  - BGP must advertise any change in the chosen route
- Control of the incoming traffic
  - send different route advertisements on different links
- Issue
  - BGP messages sent will change in function of traffic load
  - Traffic load will change in function of quality of routes

The (in)security of BGP

- The AS7007 accident



- A single configuration error in two routers
  - Two hours of disruption for large parts of the Internet
- How to deal with this problem ?
  - Filters installed by providers to detect customer errors
  - S-BGP, but requires a non-existing PKI

Research issues on interdomain routing

- How to continue to scale interdomain routing ?
- Is path vector the best technique to distribute interdomain routing information ?
  - Any new proposal should interoperate with BGP
- How to provide a faster convergence ?
  - Is one second a decent target convergence time ?
- How to secure interdomain routing ?
  - The security features must be deployable ...