When the cure is worse than the disease: The impact of graceful IGP operations on BGP

Laurent Vanbever

Princeton University &
University of Louvain

IEEE INFOCOM

April 18, 2013

Joint work with Stefano Vissicchio, Luca Cittadini, and Olivier Bonaventure

Interior Gateway Protocols (IGP)
are frequently reconfigured

Interior Gateway Protocols (IGP) are frequently reconfigured

Motivation

Traffic engineering && green networking

Reconfiguration operation

Change link weights

Interior Gateway Protocols (IGP) are frequently reconfigured

Motivation

Reconfiguration operation

Traffic engineering && green networking

Change link weights

Maintenance

Cost-out links and/or routers

Interior Gateway Protocols (IGP) are frequently reconfigured

Motivation	Reconfiguration operation
Traffic engineering && green networking	Change link weights
Maintenance	Cost-out links and/or routers
Service deployment && scaling/performance	Protocol changes, hierarchy deployment

Reconfiguring the IGP can create numerous problems

IGP reconfiguration can lead to

- forwarding loop
- network congestion
- blackhole

or any combination of those

A lot of research has been made to solve these problems

- forwarding loop [Francois05-07], [Alimi08], [Fu08], [Vanbever12]
- network congestion [Raza09], [Shi09]
- blackhole [Alimi08], [Vanbever12]

Most of these research works *exclusively* focus on the IGP

but

BGP routers depend on the underlying IGP

to discriminate between equivalent routes

Most network traffic in an ISP is due to BGP

the IGP is used as a reachability mechanism

Problem

Can *safely* reconfiguring the IGP create BGP anomalies?

Most of these research works *exclusively* focus on the IGP

but

BGP routers depend on the underlying IGP

to discriminate between equivalent routes

Most network traffic in an ISP is due to BGP

the IGP is used as a reachability mechanism

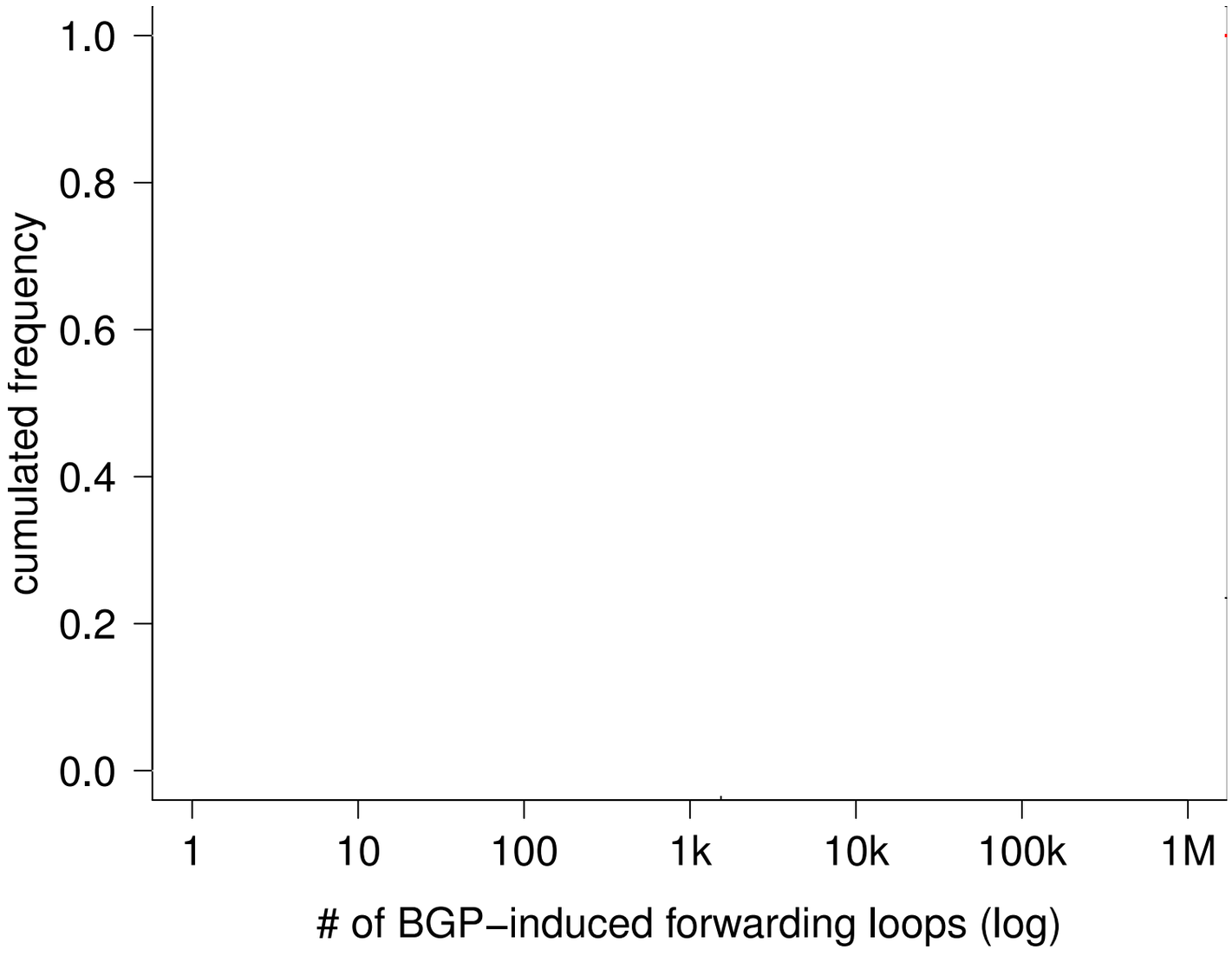
Problem

Can *safely* reconfiguring the IGP create BGP anomalies?

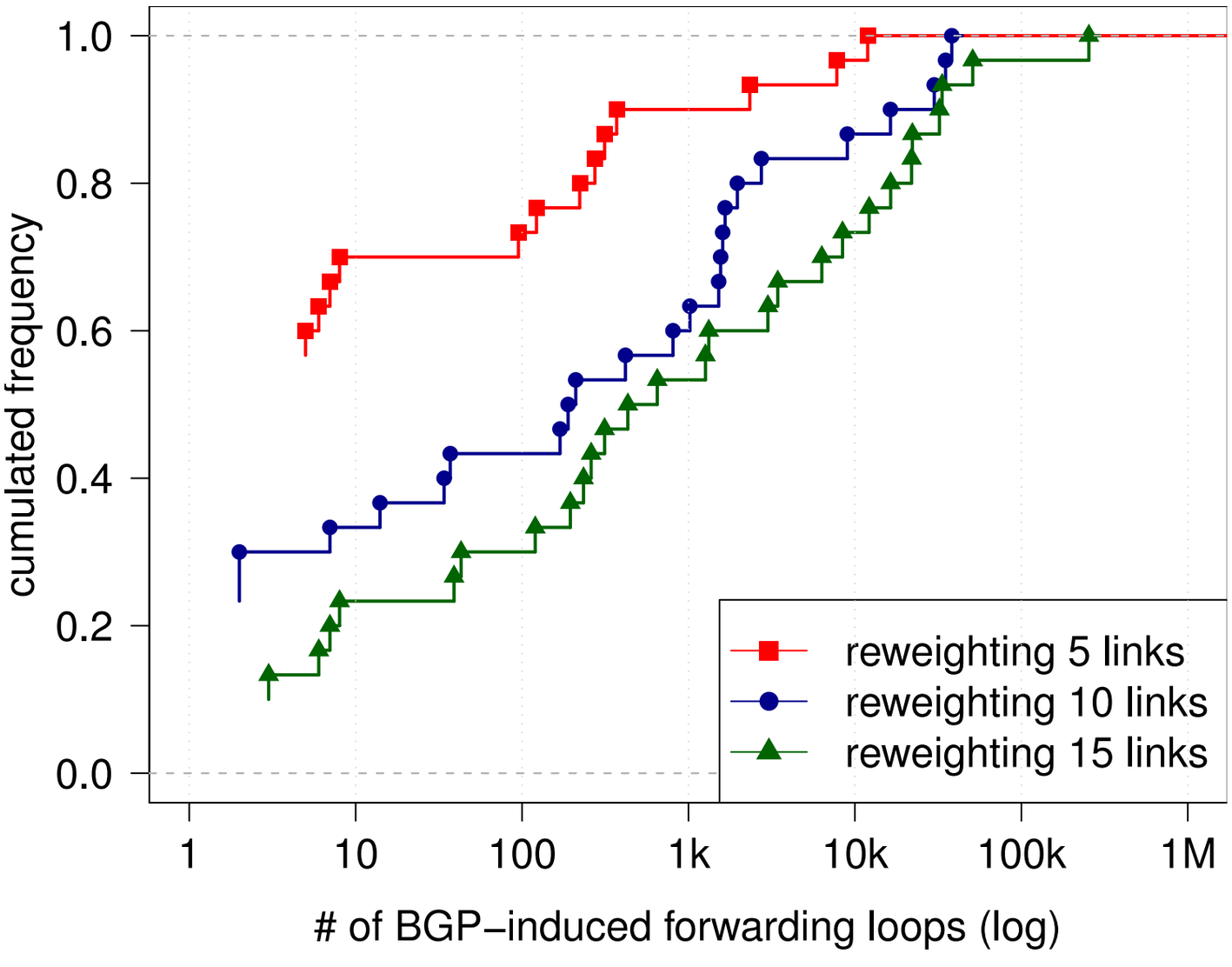
The answer is ... YES!

Safely reconfiguring the IGP
can and do create BGP anomalies

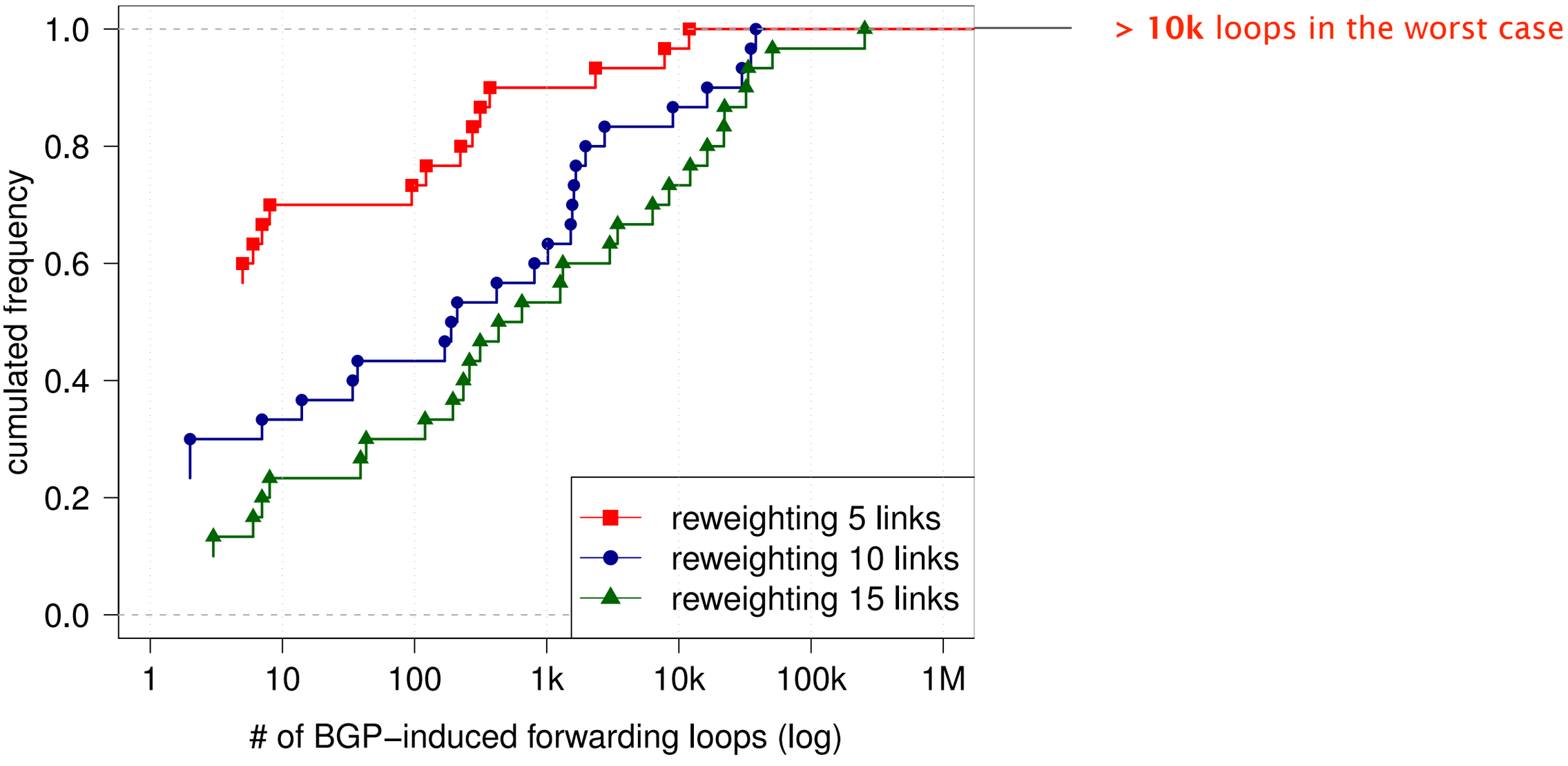
Dataset	IGP and BGP configuration of a Tier1 backbone 100+ routers, 150+ links Representative BGP route feed
Reconfiguration	Randomly reweight 5, 10, 15 links using *provably correct* IGP reconfiguration technique [Vanbever12]
Experiments	Measure the amount of BGP-induced loop


1.0
0.8
0.6
0.4
0.2
0.0
cumulated frequency
1
10
100
1k
10k
100k
1M
# of BGP−induced forwarding loops (log)

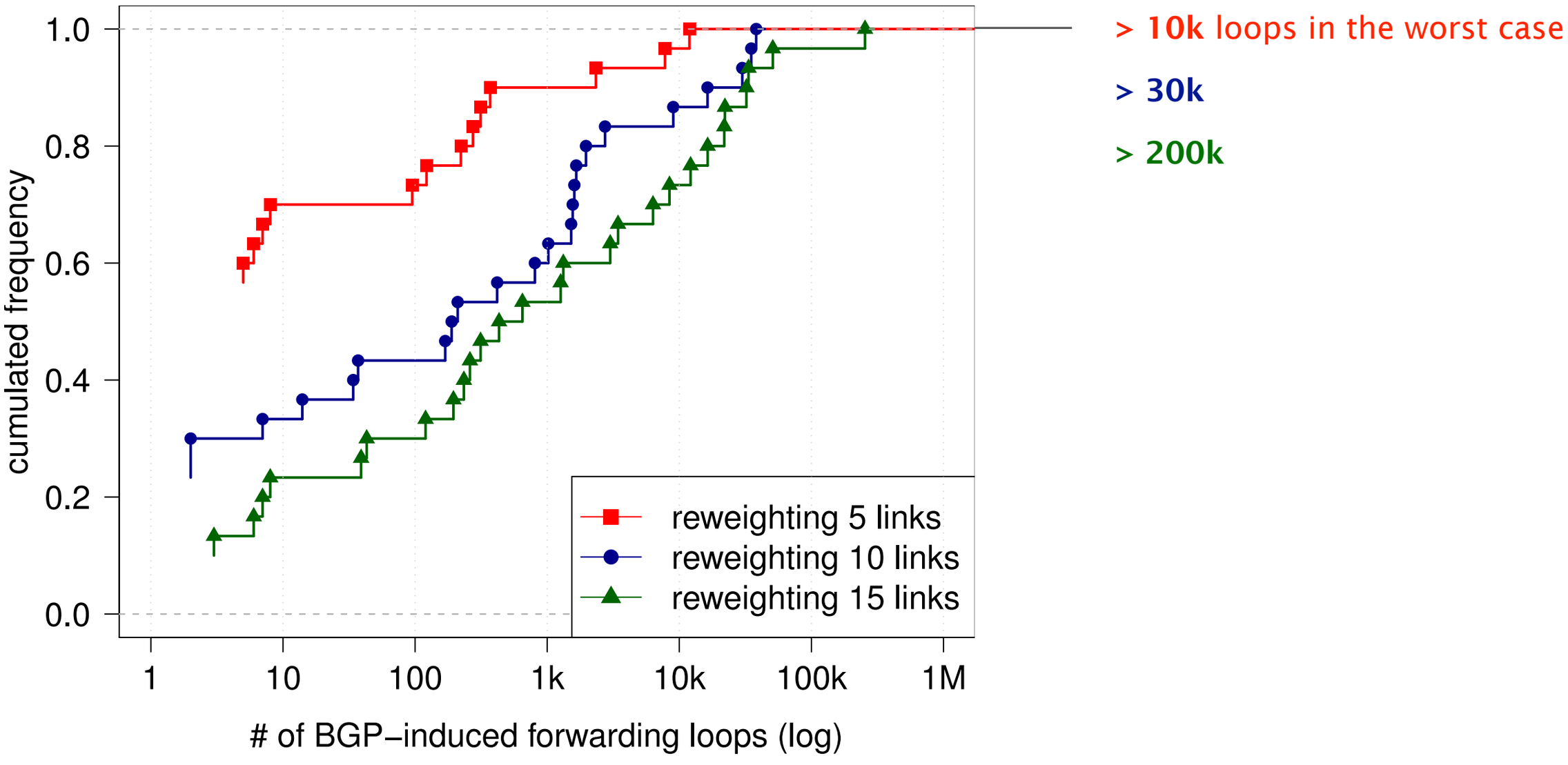
Safely reconfiguring the IGP can create numerous BGP anomalies



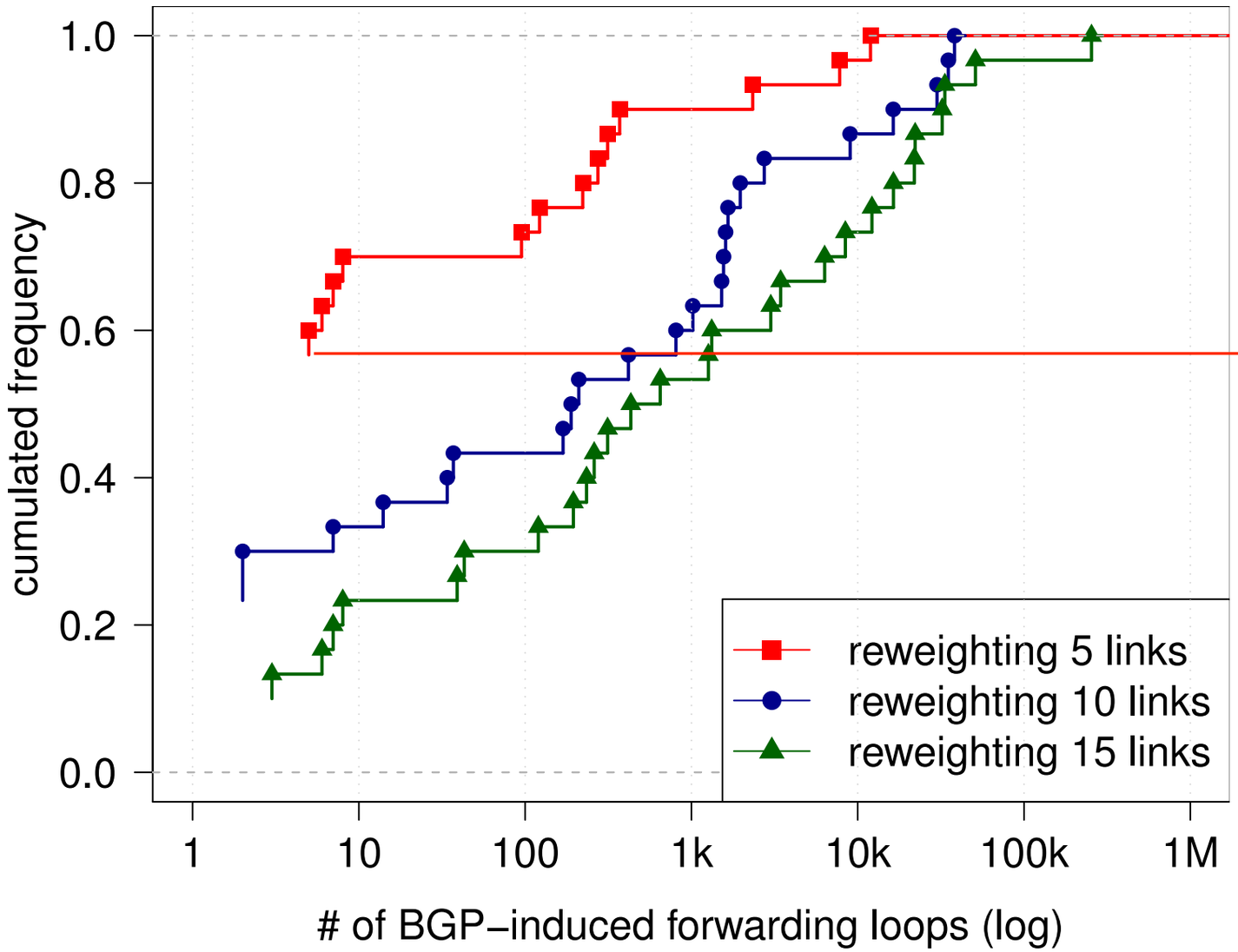
In the worst case, tens of thousands of loops can be created



In the worst case, tens of thousands of loops can be created

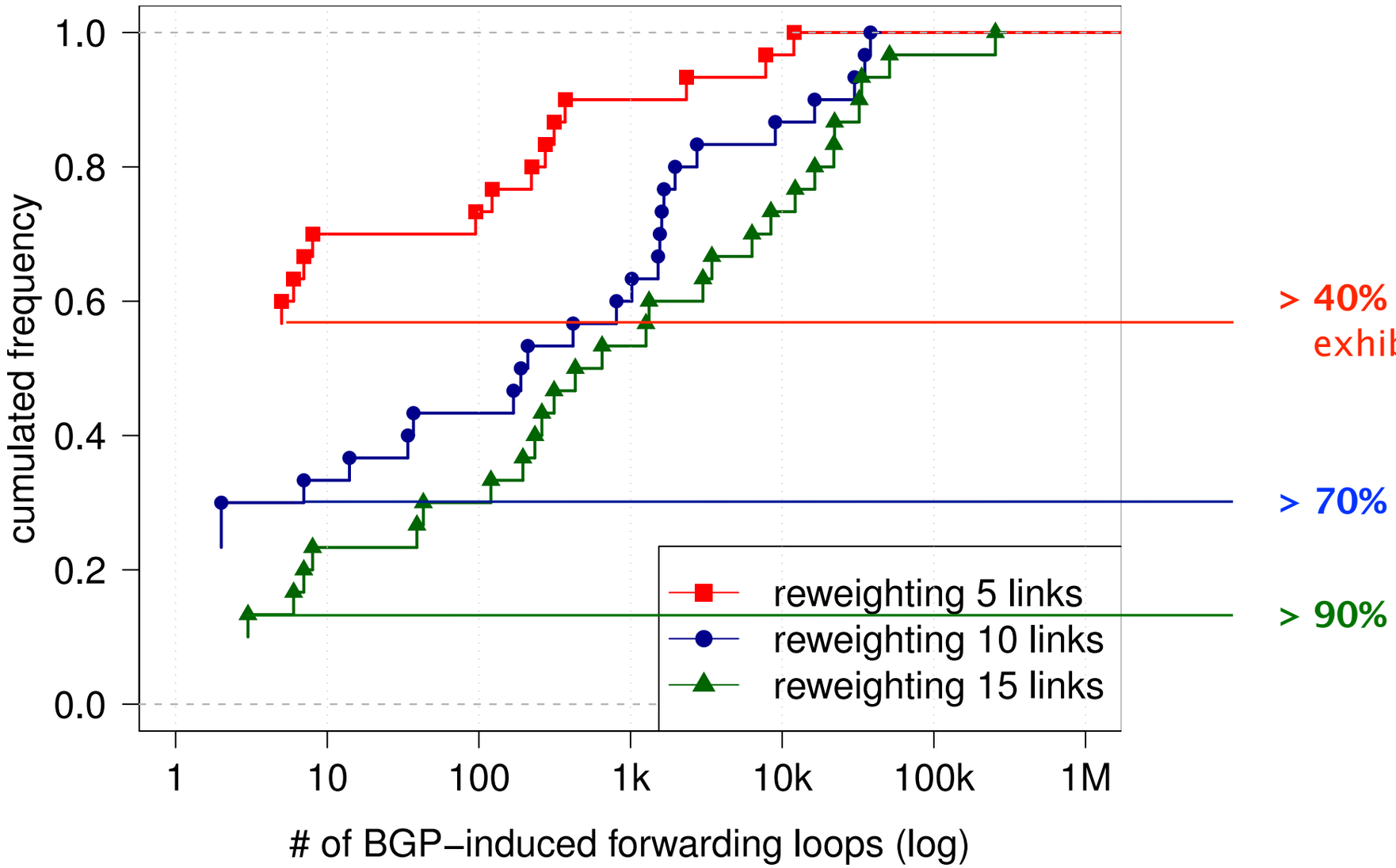


Most IGP reconfiguration triggers BGP-induced loops



> **40%** of experiments exhibit loops, *in the best case*

Most IGP reconfiguration triggers BGP-induced loops



> **40%** of experiments exhibit loops, *in the best case*

> **70%**

> **90%**

Our contributions range from practice, to theory, and back to practice

Theory

Complexity

Guidelines

Our contributions range from practice, to theory, and back to practice

Theory

Reconfiguring IGP can introduce any BGP anomaly
even with state-of-the-art IGP reconfiguration

Complexity

Guidelines

Our contributions range from practice, to theory, and back to practice

Theory

Reconfiguring IGP can introduce any BGP anomaly
even with state-of-the-art IGP reconfiguration

Complexity

Deciding if an anomaly-free IGP reconfiguration triggers BGP anomaly is NP-hard

Guidelines

Our contributions range from practice, to theory, and back to practice

Theory

Reconfiguring IGP can introduce any BGP anomaly
even with state-of-the-art IGP reconfiguration

Complexity

Deciding if an anomaly-free IGP reconfiguration triggers BGP anomaly is NP-hard

Guidelines

Sufficient conditions and configuration guidelines that guarantee the absence of BGP-induced anomalies

When the cure is worse than the disease:
The impact of graceful IGP operations on BGP

The cure

IGP reconfiguration

The side effects

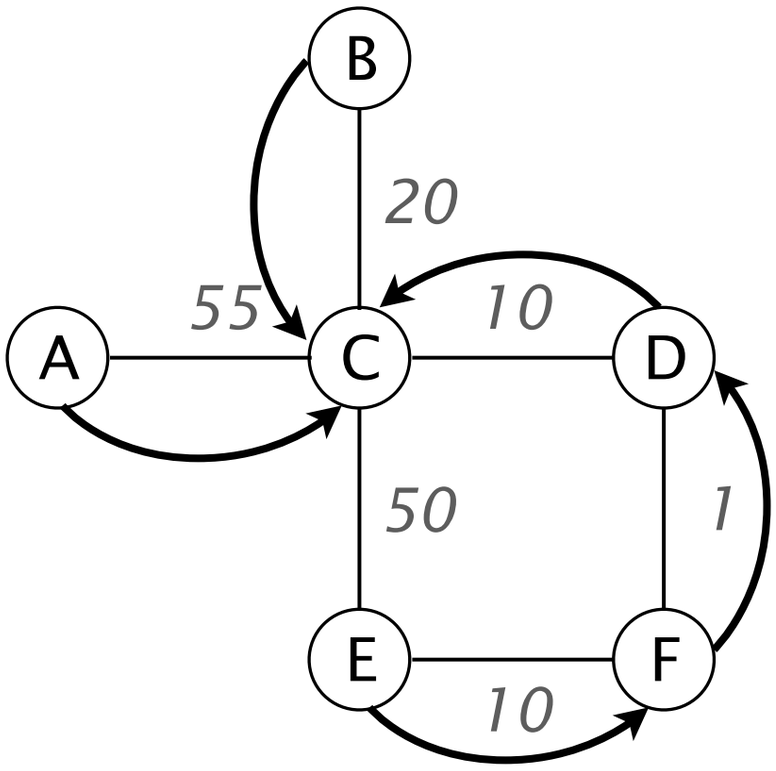
BGP-induced anomalies

The solutions

sufficient conditions

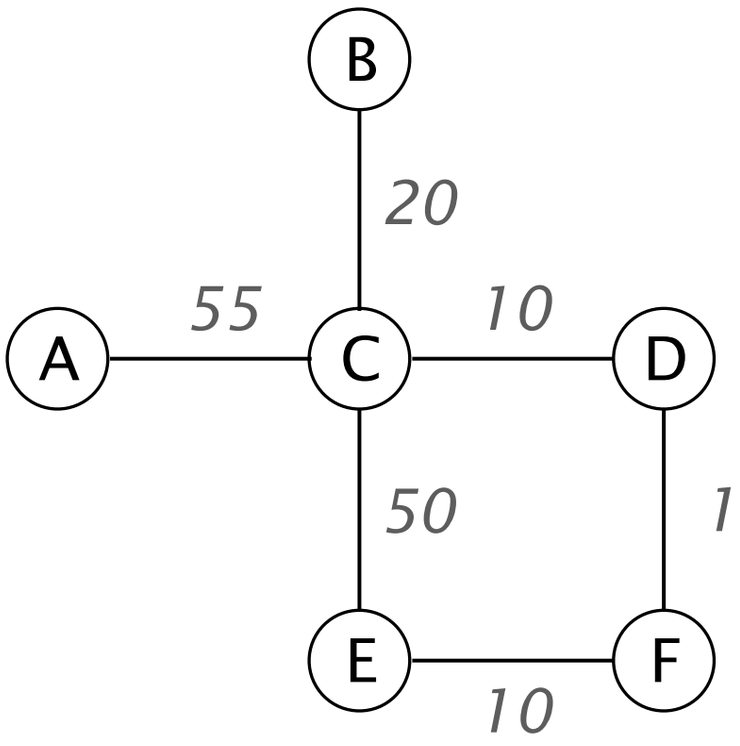
When the cure is worse than the disease: The impact of graceful IGP operations on BGP

Intradomain routing protocols (IGP) rule traffic forwarding within a routing domain

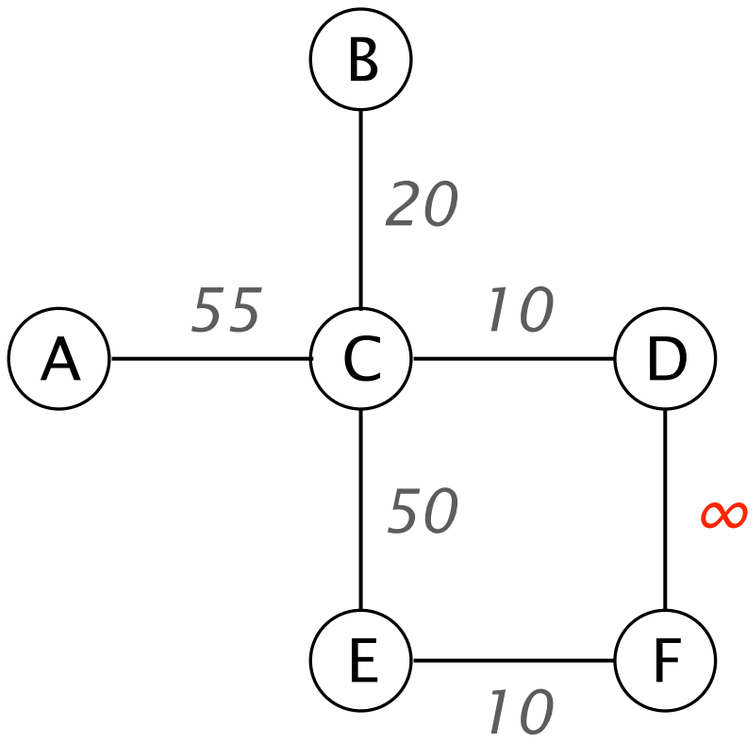


forwarding paths towards C

IGP reconfiguration consists in changing some IGP parameters, such as link weights

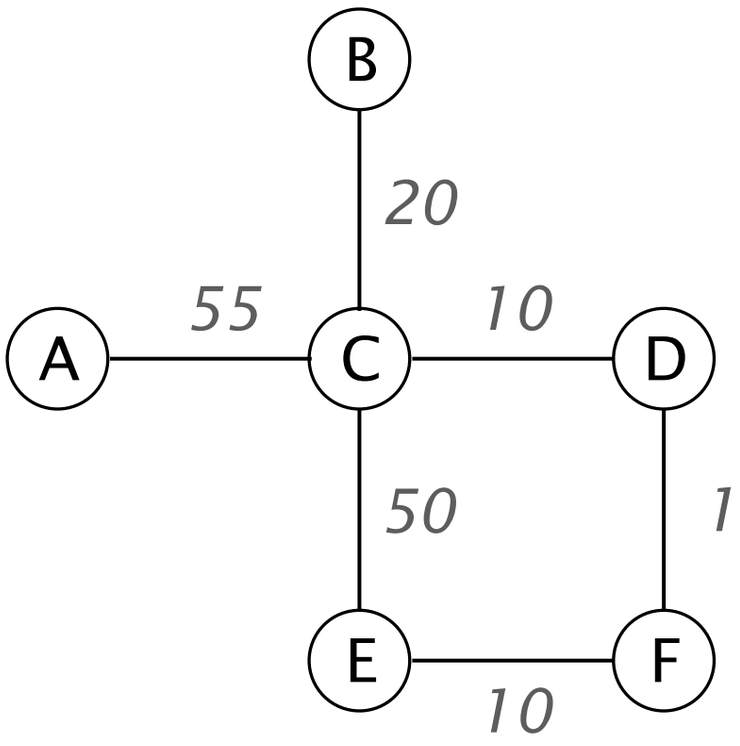


initial IGP

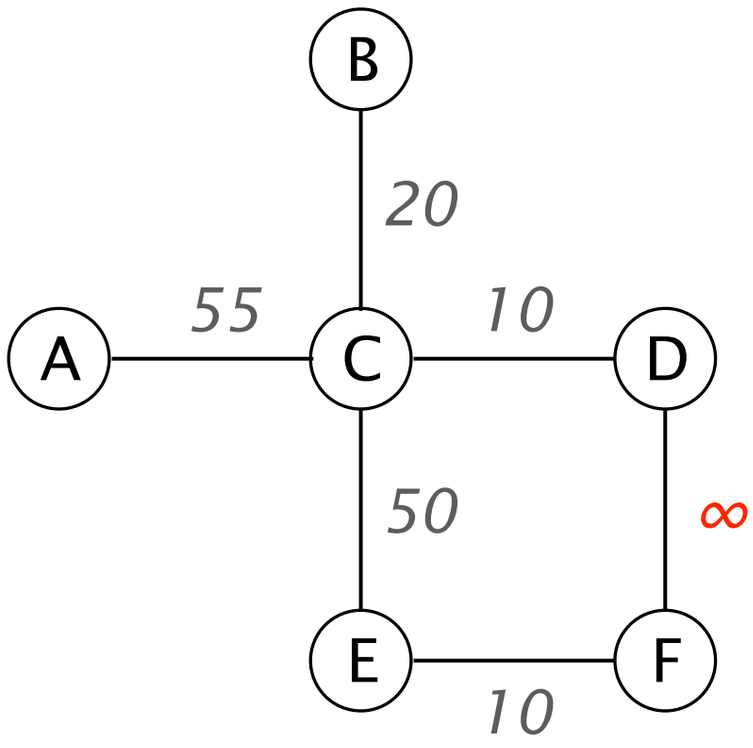


final IGP

IGP reconfiguration can impact the forwarding paths

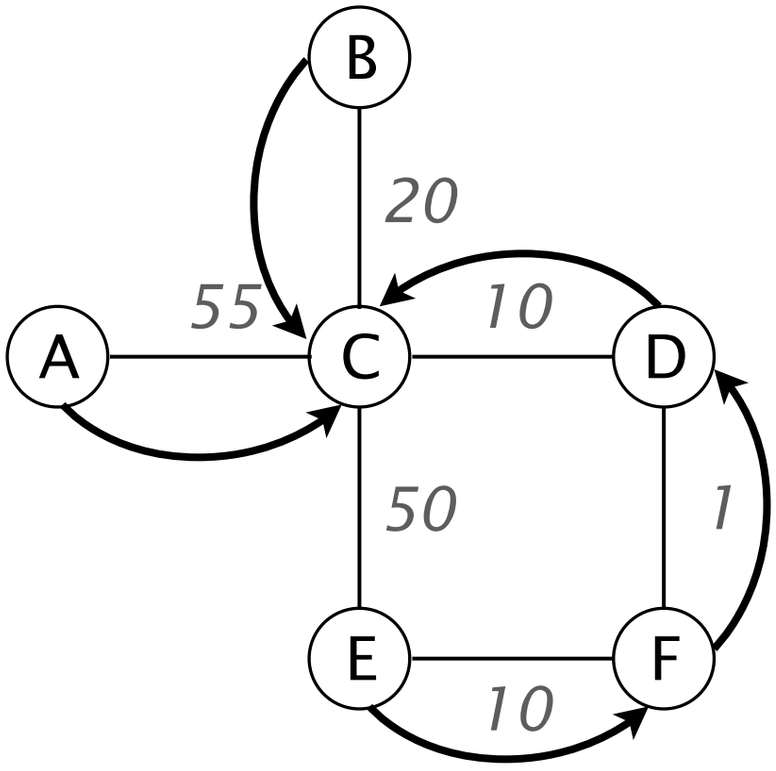


initial IGP

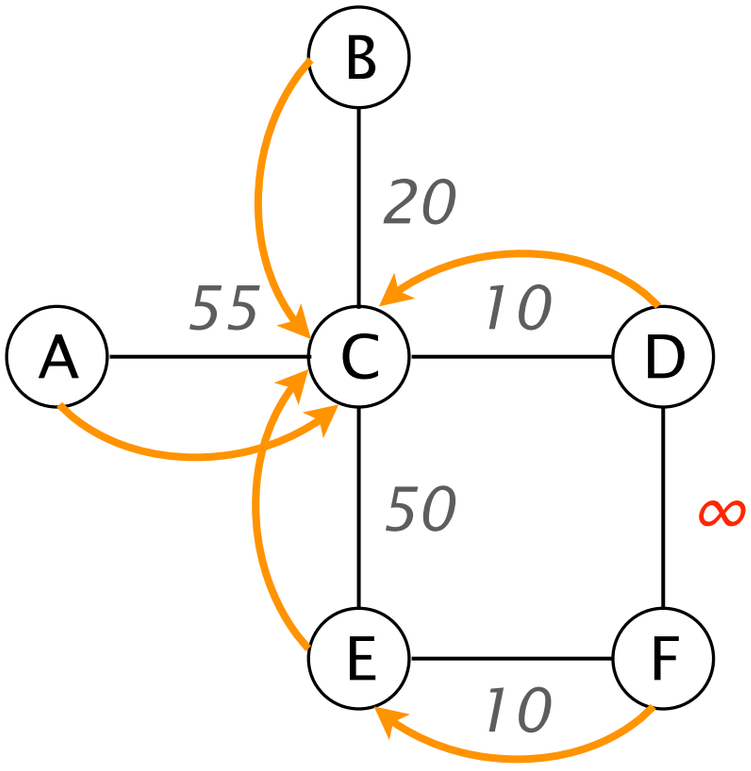


final IGP

IGP reconfiguration can impact the forwarding paths



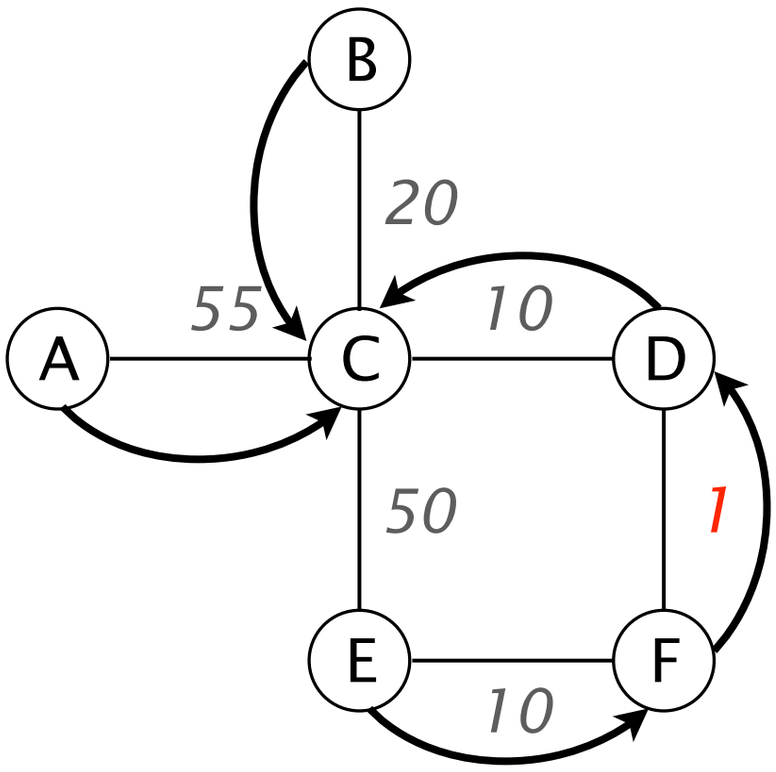
initial IGP



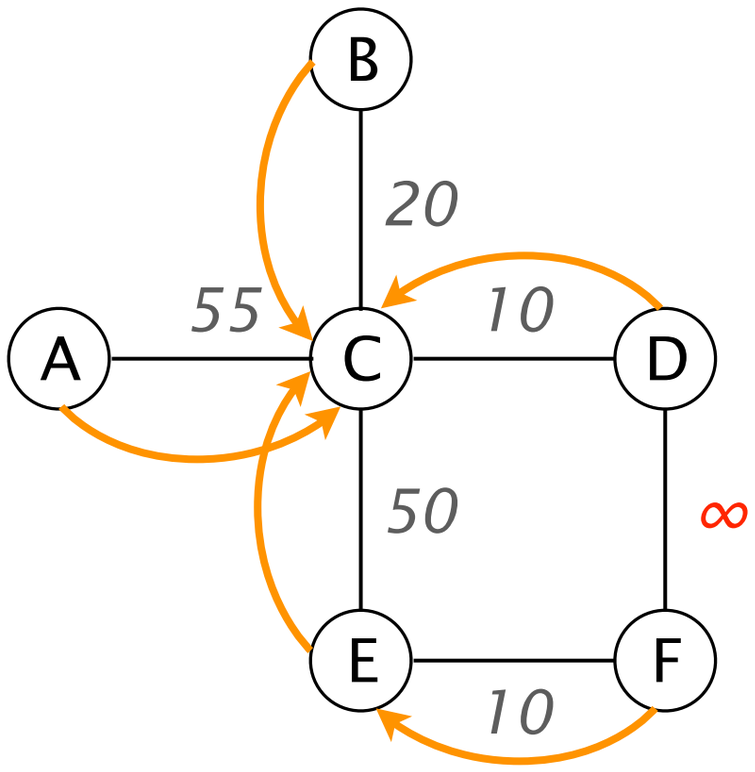
final IGP

If not handled correctly, paths change can lead to forwarding loops

Changing the metric of link (D,F) from 1 to ∞ can create a loop



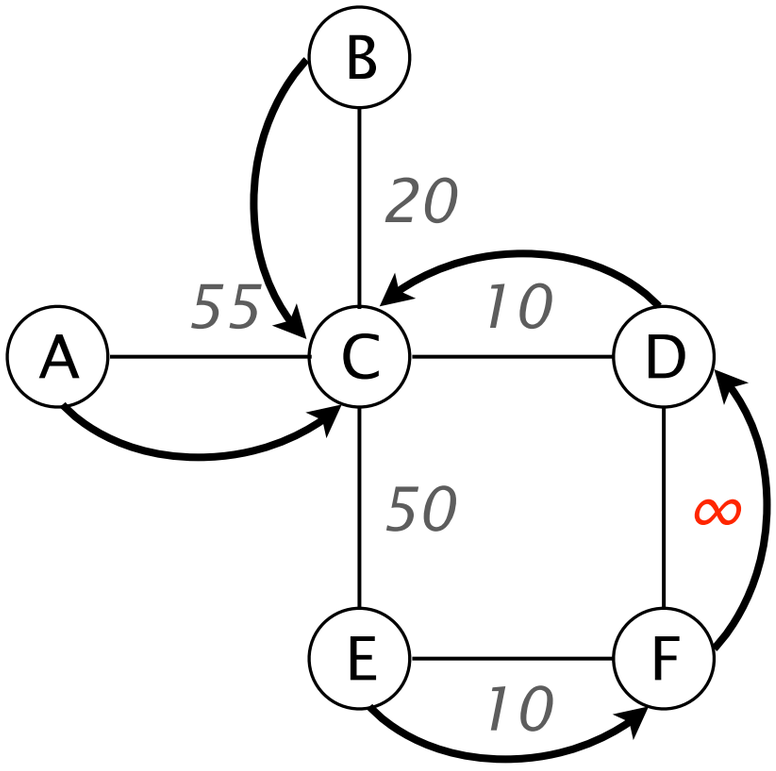
initial IGP

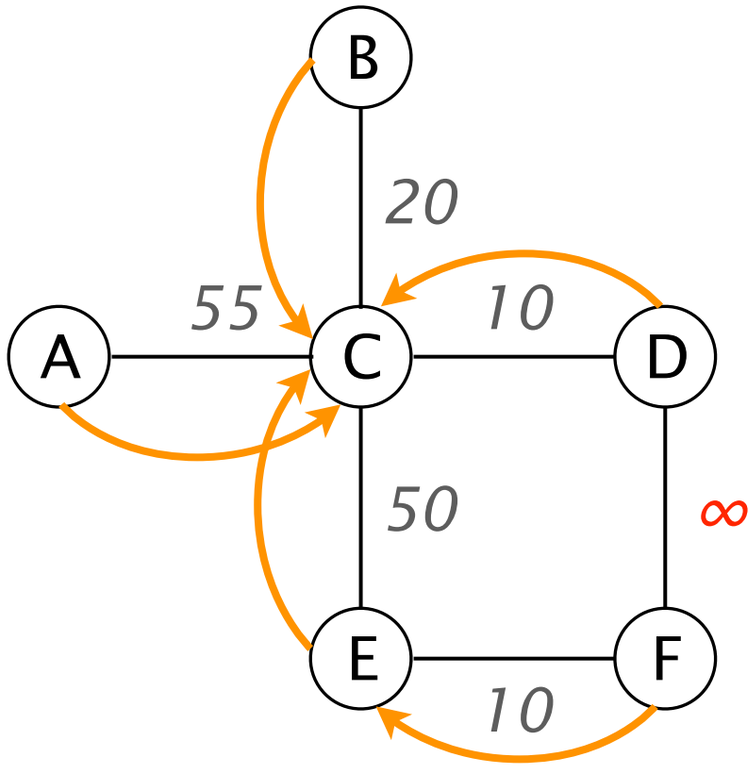


final IGP

If not handled correctly, paths change can lead to forwarding loops

Changing the metric of link (D,F) from 1 to ∞ can create a loop

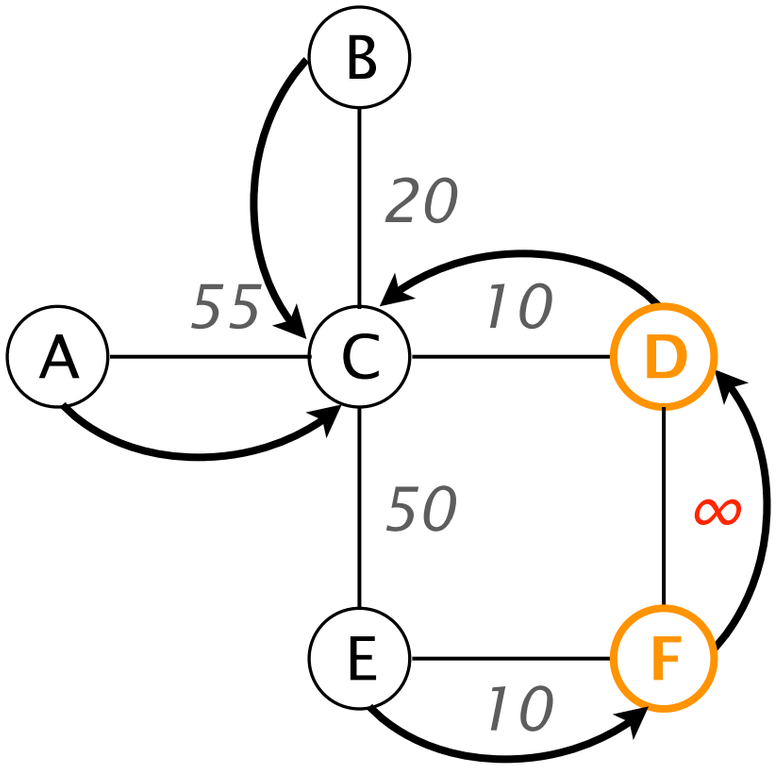


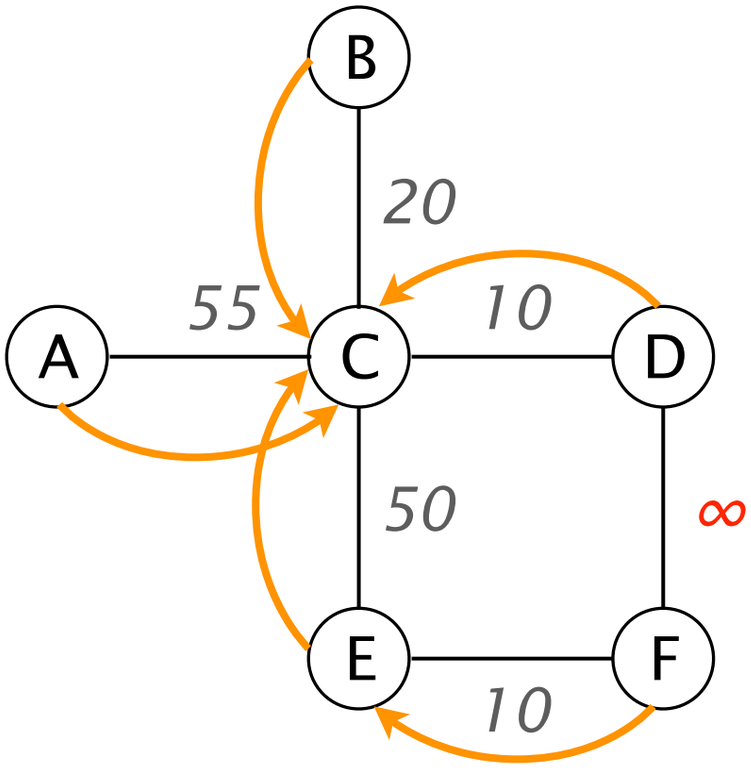


final IGP

If not handled correctly, paths change can lead to forwarding loops

F and D are the first to notice the change
and immediately update their forwarding table

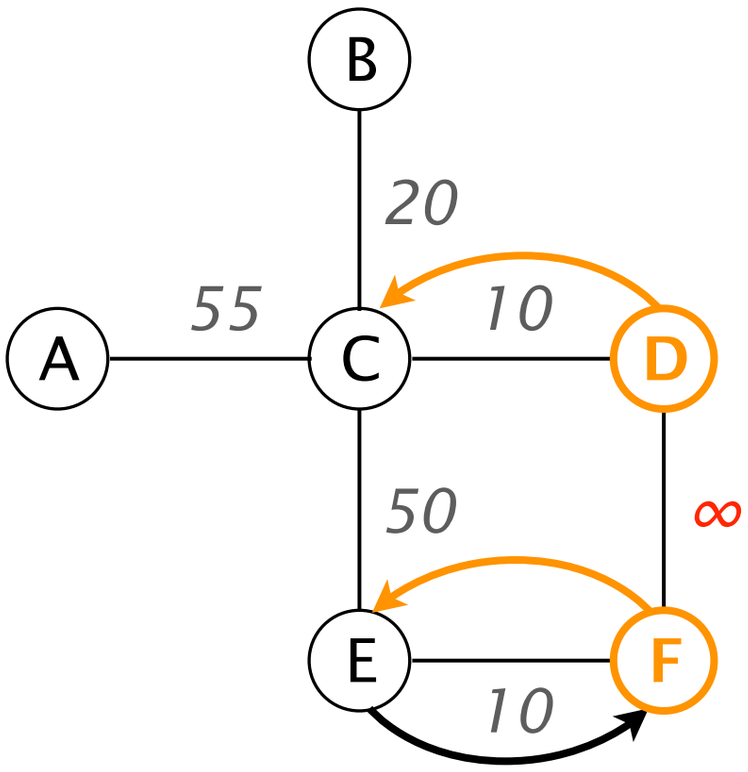


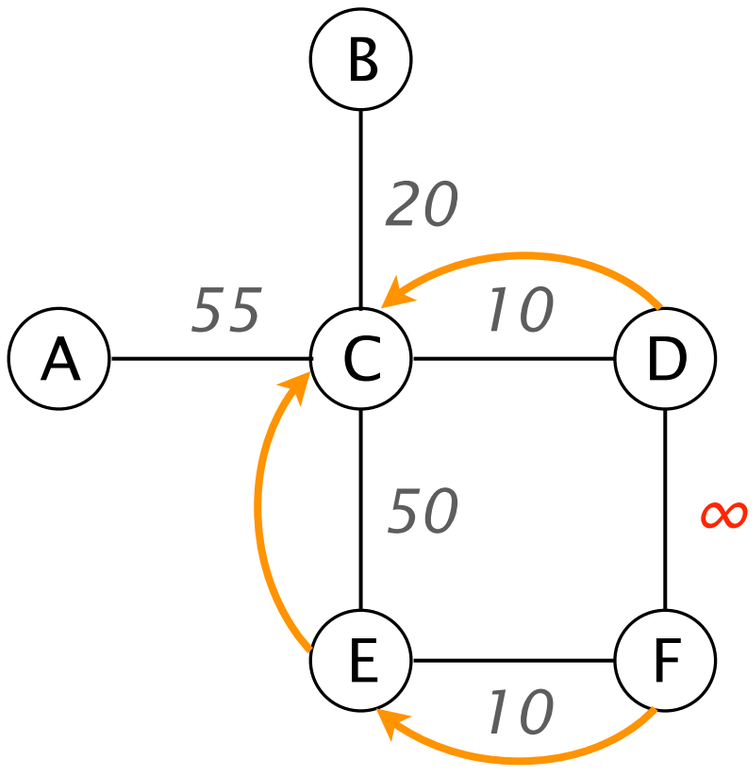


final IGP

If not handled correctly, paths change can lead to forwarding loops

F and D are the first to notice the change
and immediately update their forwarding table

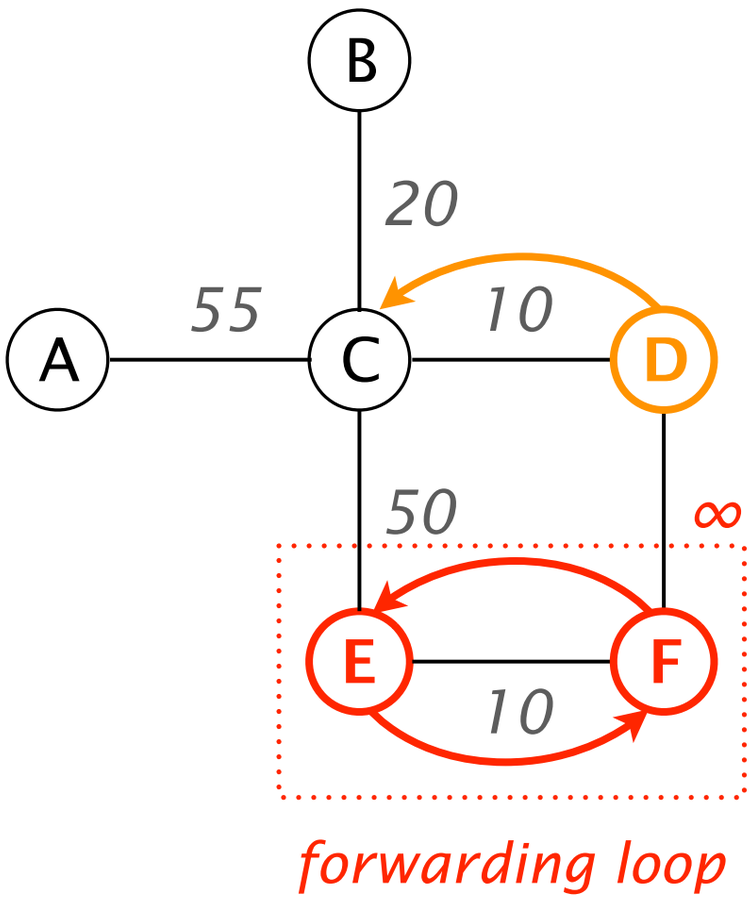




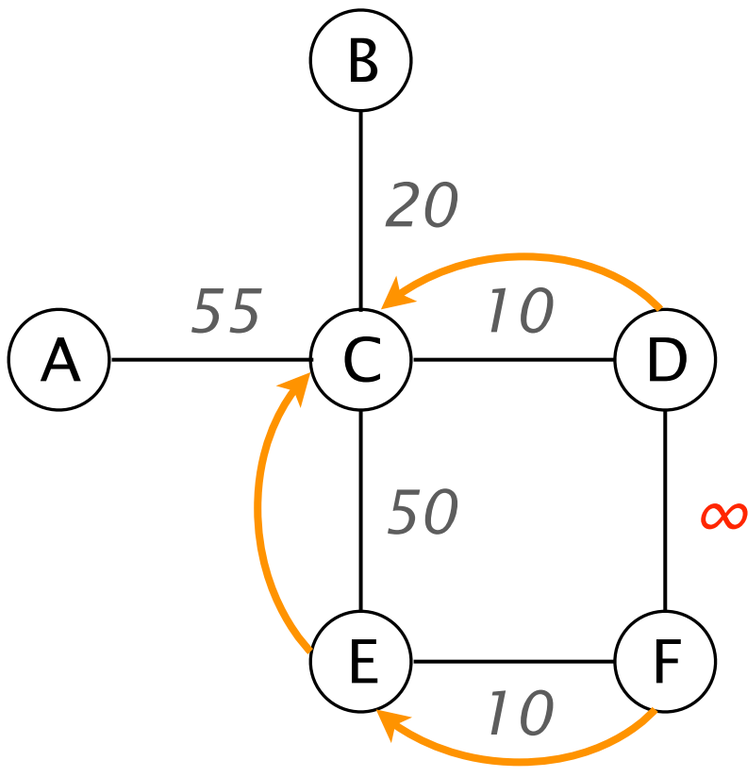
final IGP

If not handled correctly, paths change can lead to forwarding loops

A forwarding loop is created as long as E is not updated



forwarding loop



final IGP

Safe IGP reconfiguration techniques upgrade the forwarding entries in a precise order

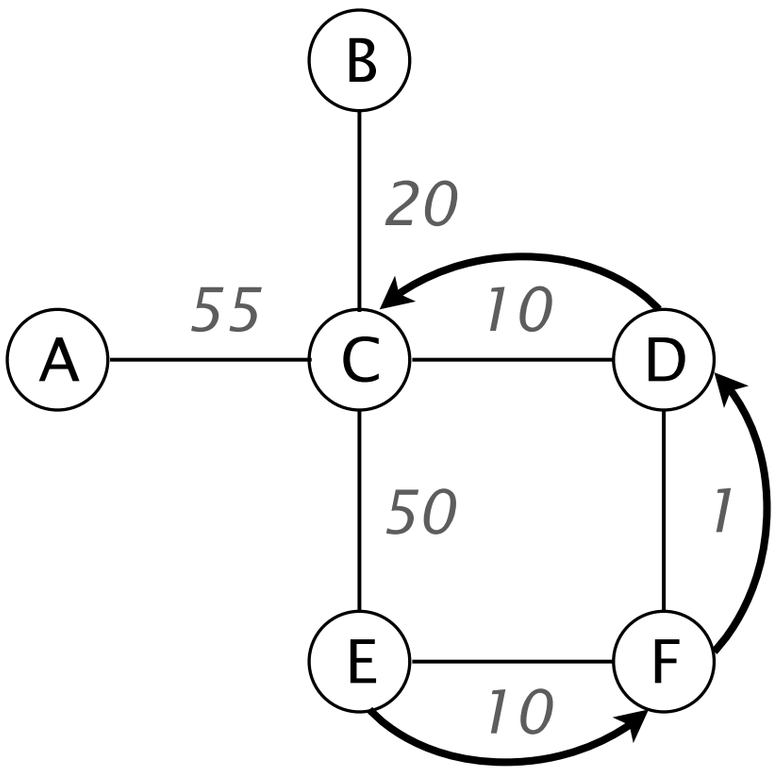
	Metric-Increment [Francois07]
Procedure	consecutive metric changes
Theoretical guarantees	YES, loop-freeness
Works Today	YES

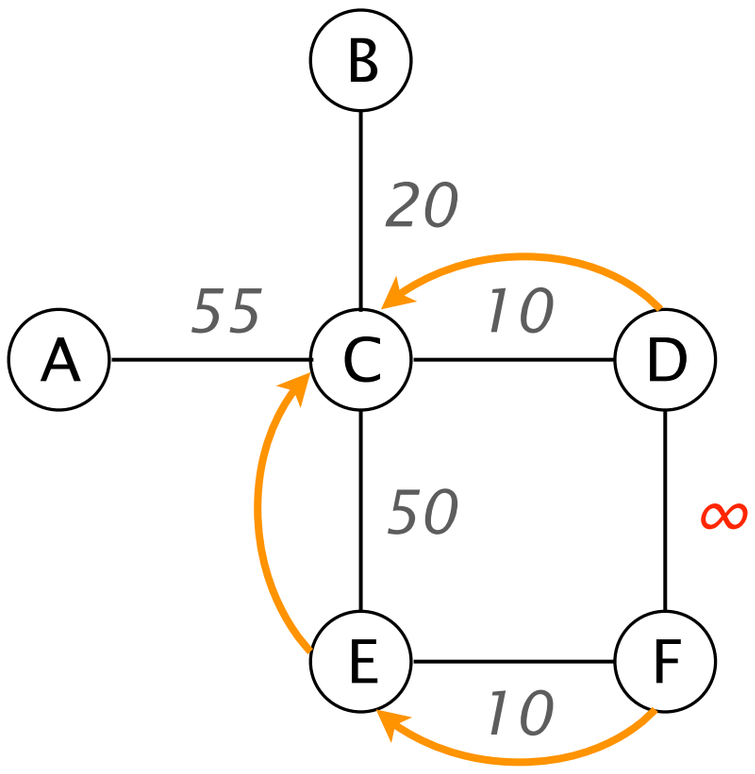
Safe IGP reconfiguration techniques upgrade the forwarding entries in a precise order

	Metric-Increment [Francois07]
Procedure	consecutive metric changes
Theoretical guarantees	YES, loop-freeness
Works Today	YES

Metric increment sequentially increases link metric to make remote routers transition first

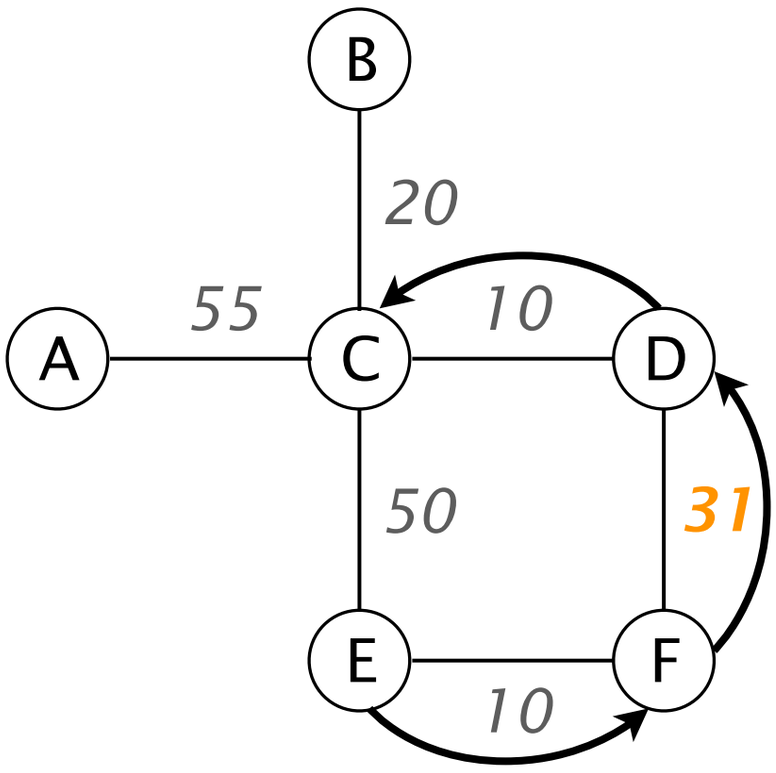
metric sequence [1,31,51,∞]

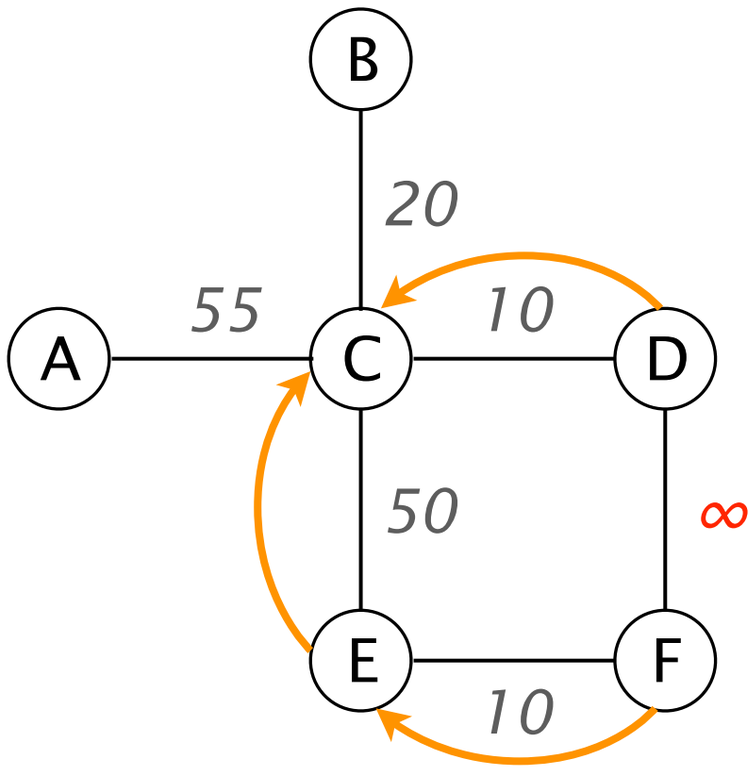




Metric increment sequentially increases link metric to make remote routers transition first

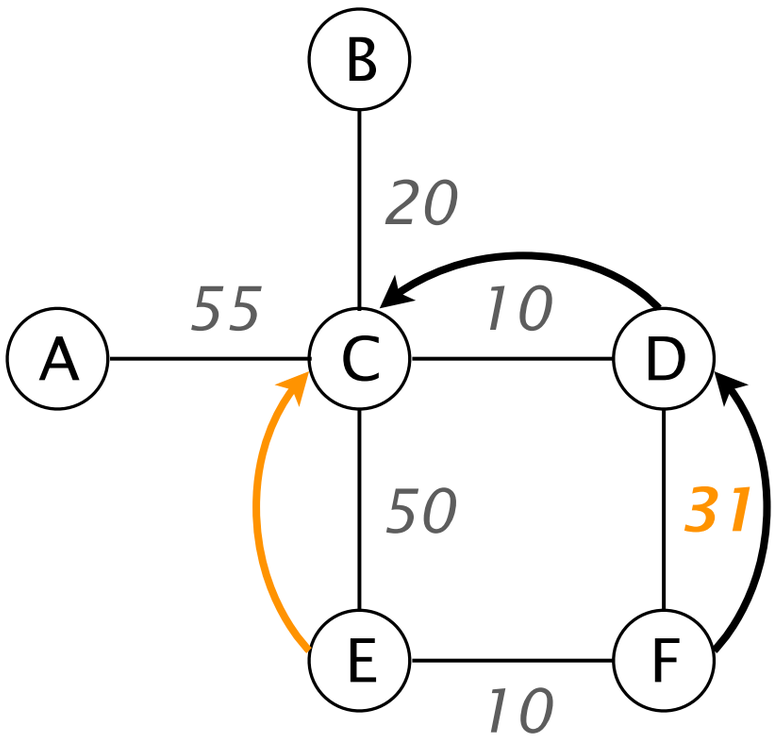
metric sequence [1,***31***,51,∞]

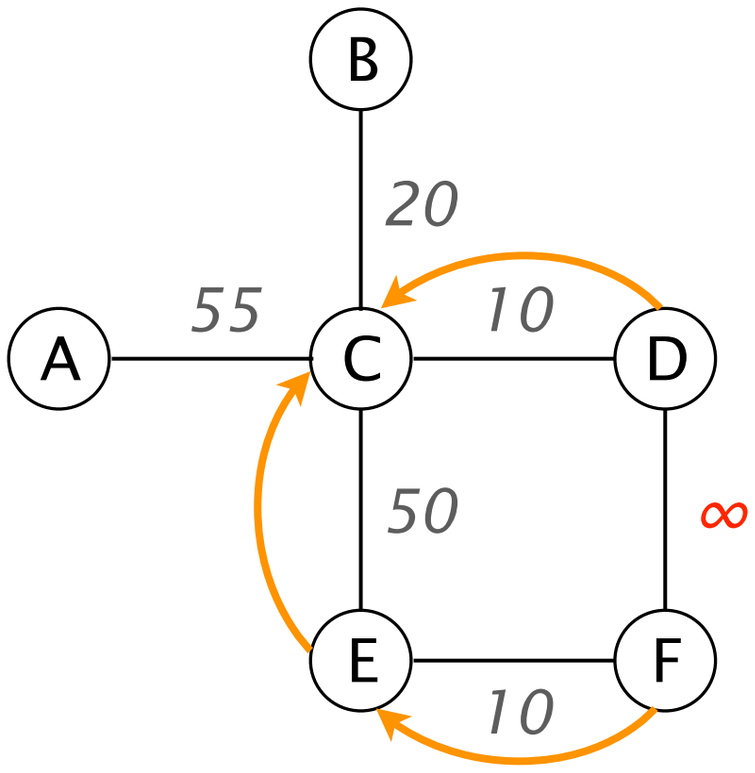




Metric increment sequentially increases link metric to make remote routers transition first

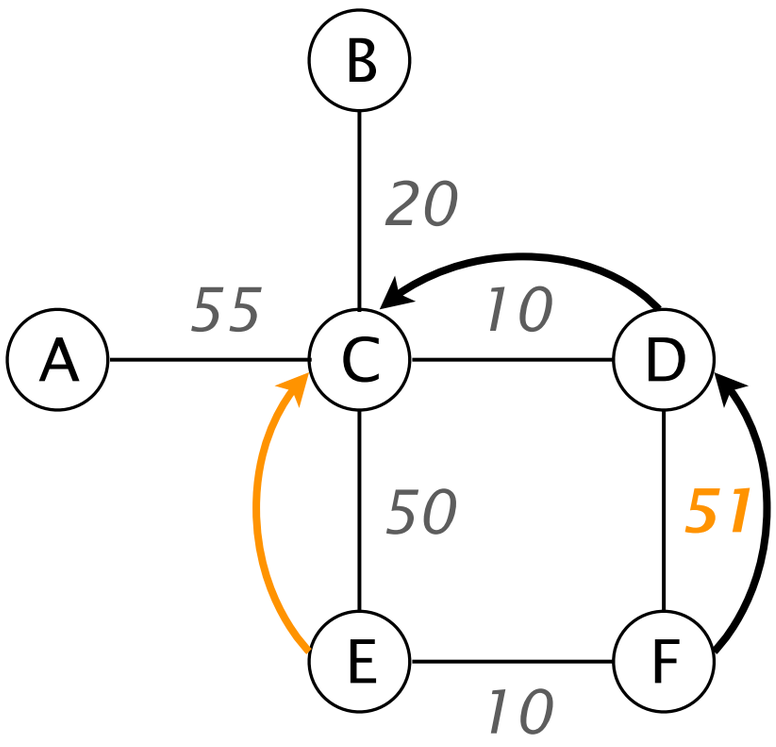
metric sequence [1,***31***,51,∞]

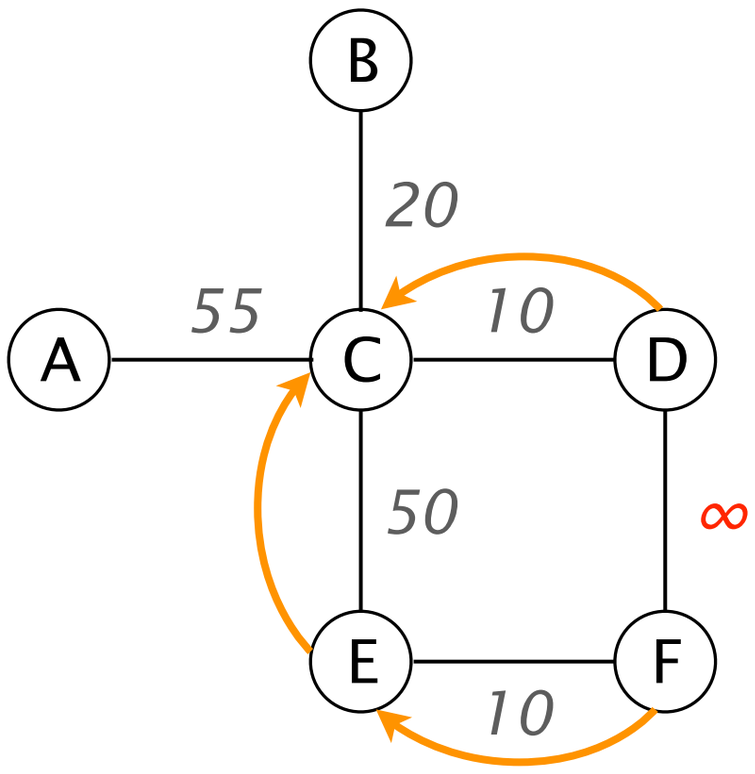




Metric increment sequentially increases link metric to make remote routers transition first

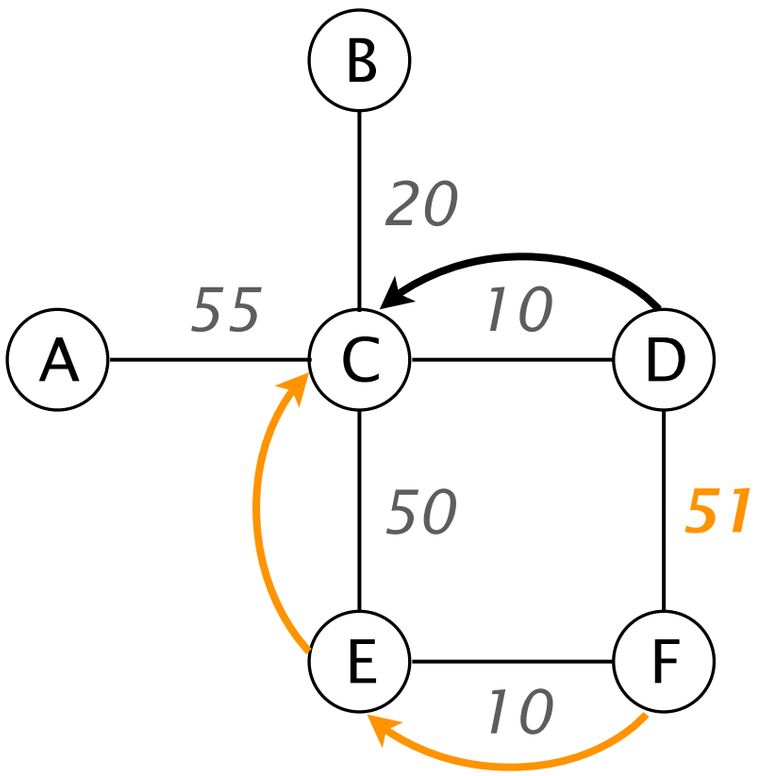
metric sequence [1,31,***51***,∞]

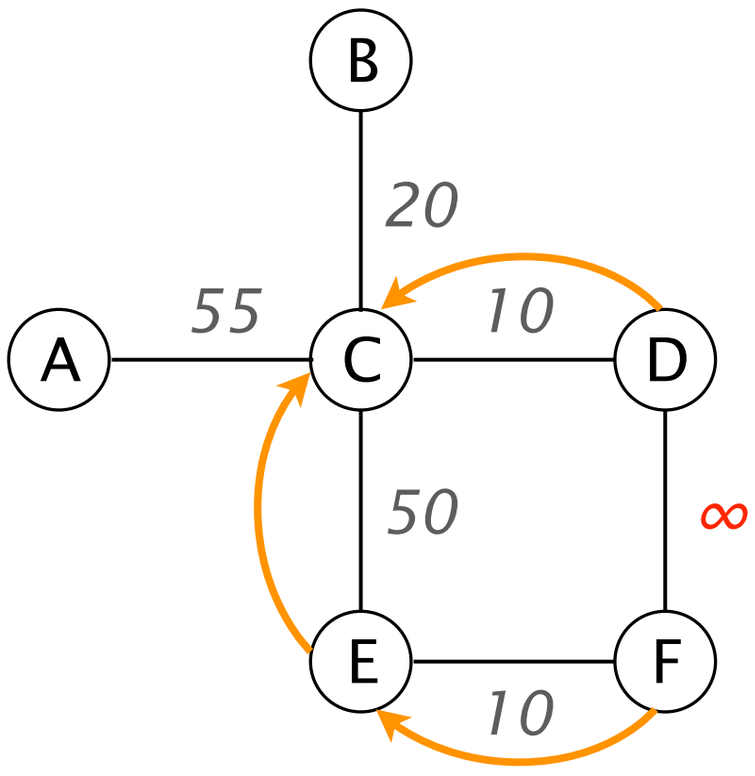




Metric increment sequentially increases link metric to make remote routers transition first

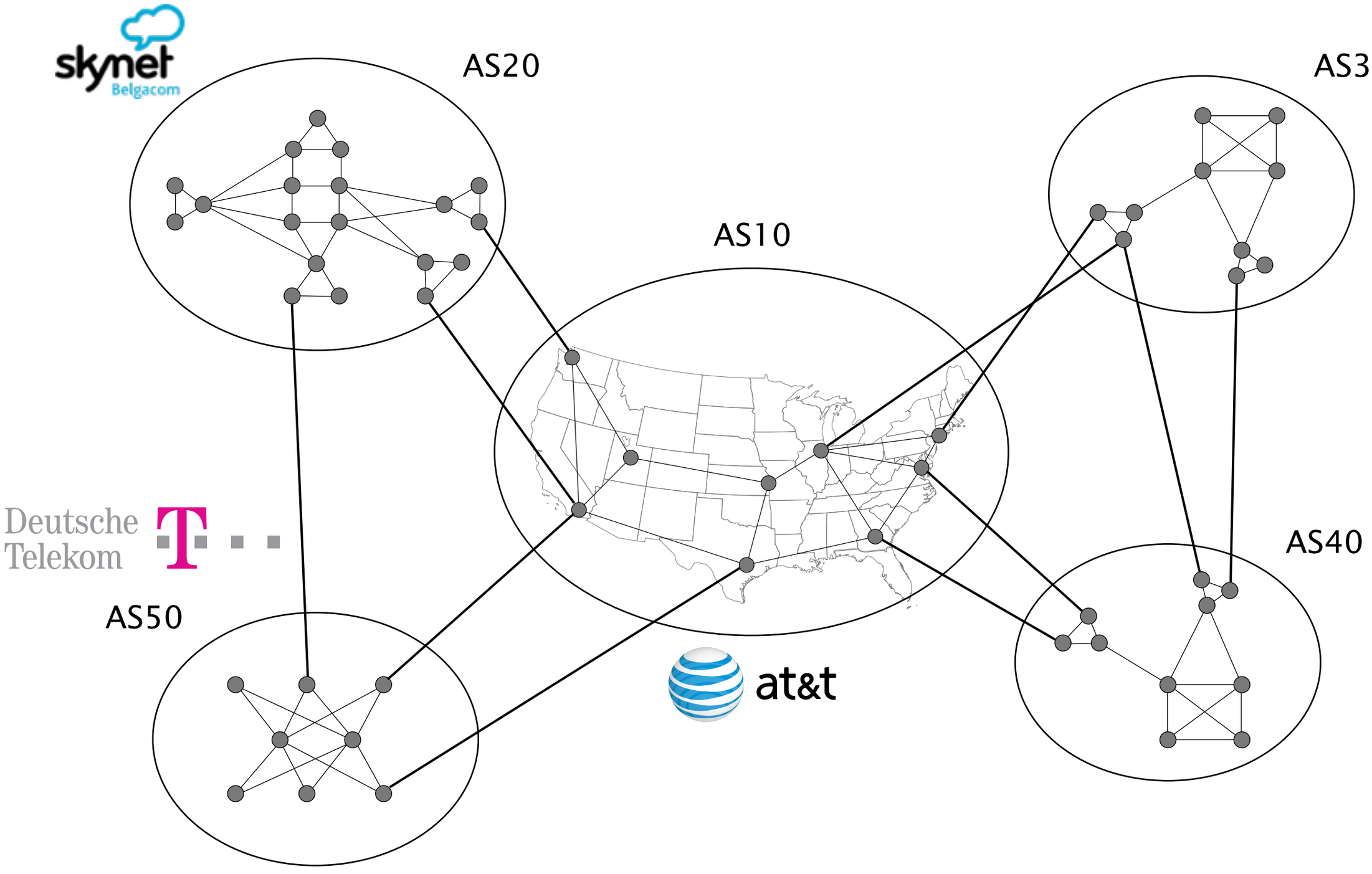
metric sequence [1,31,***51***,∞]



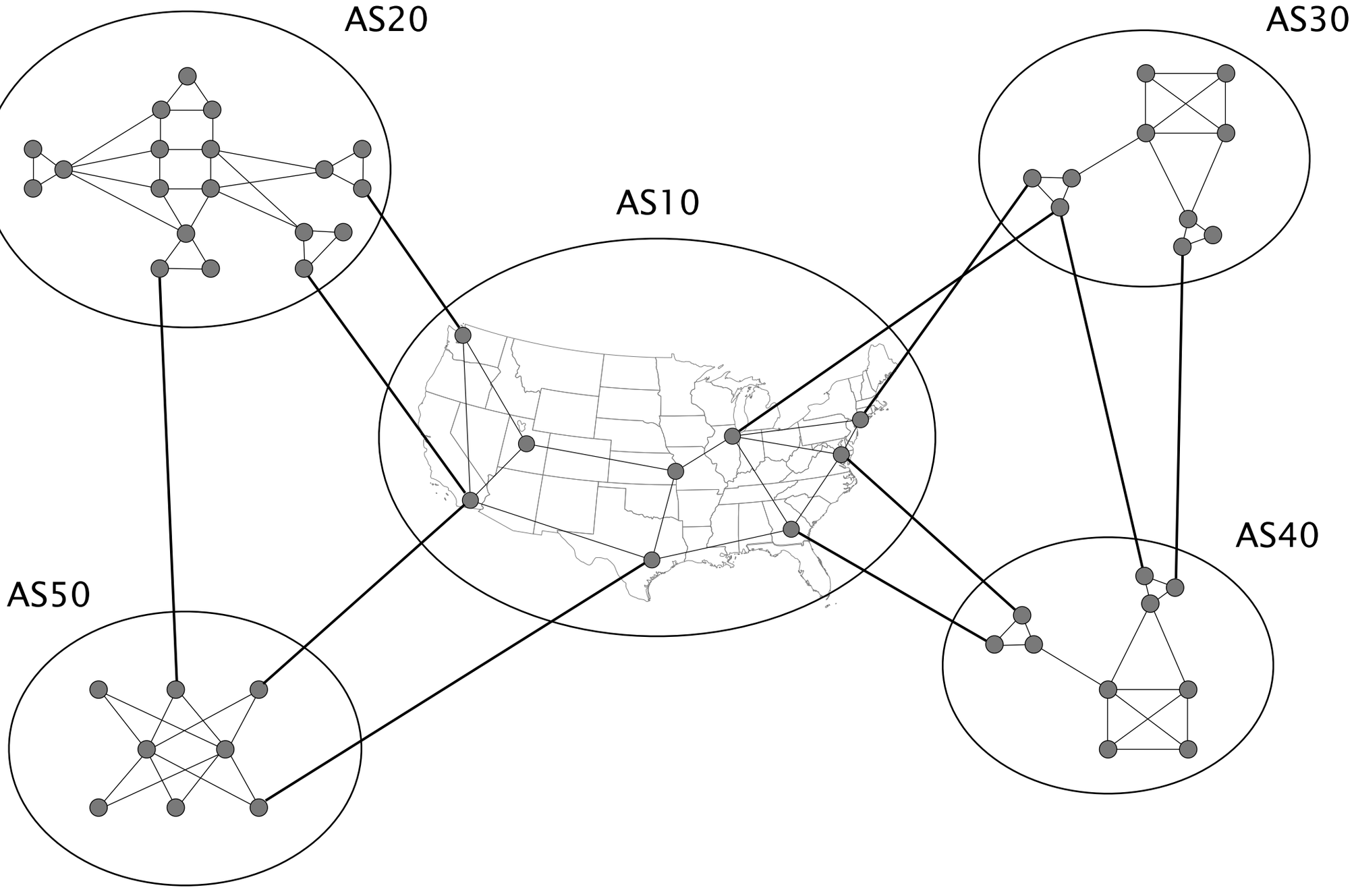


When the cure is worse than the disease: The impact of graceful IGP operations on BGP

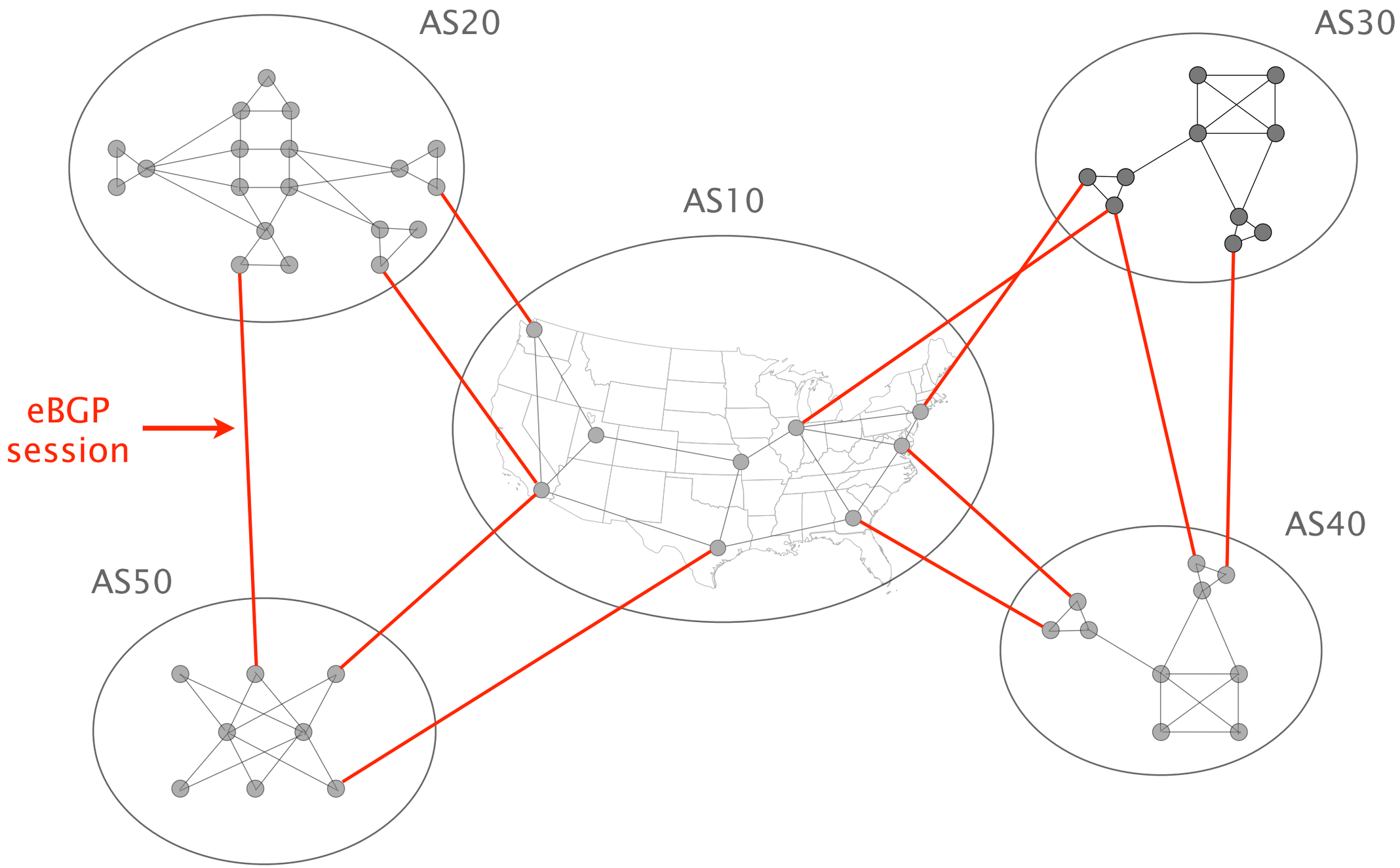
Interdomain routing protocols (BGP) rule traffic forwarding across routing domains



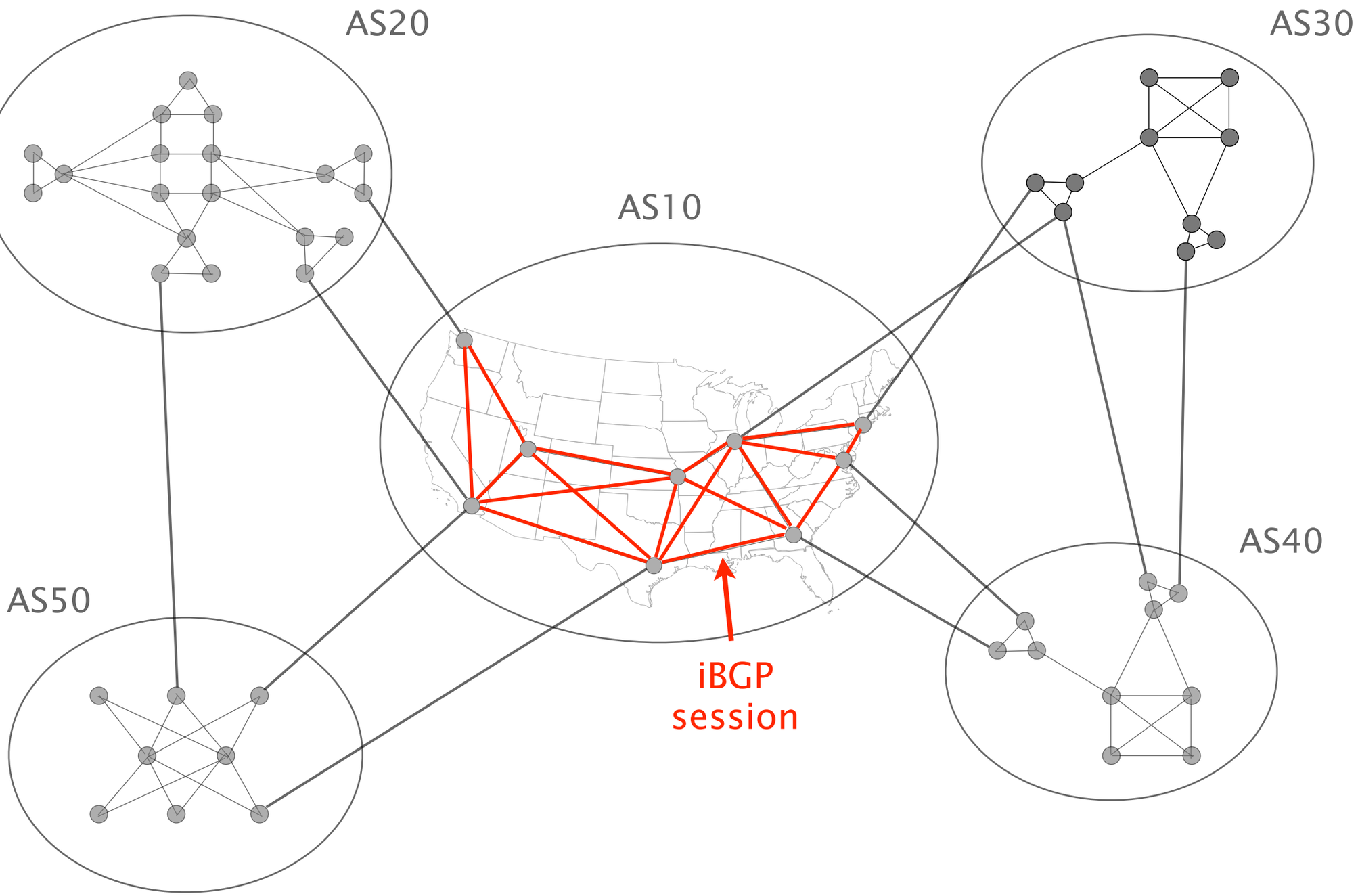
BGP comes in two flavors



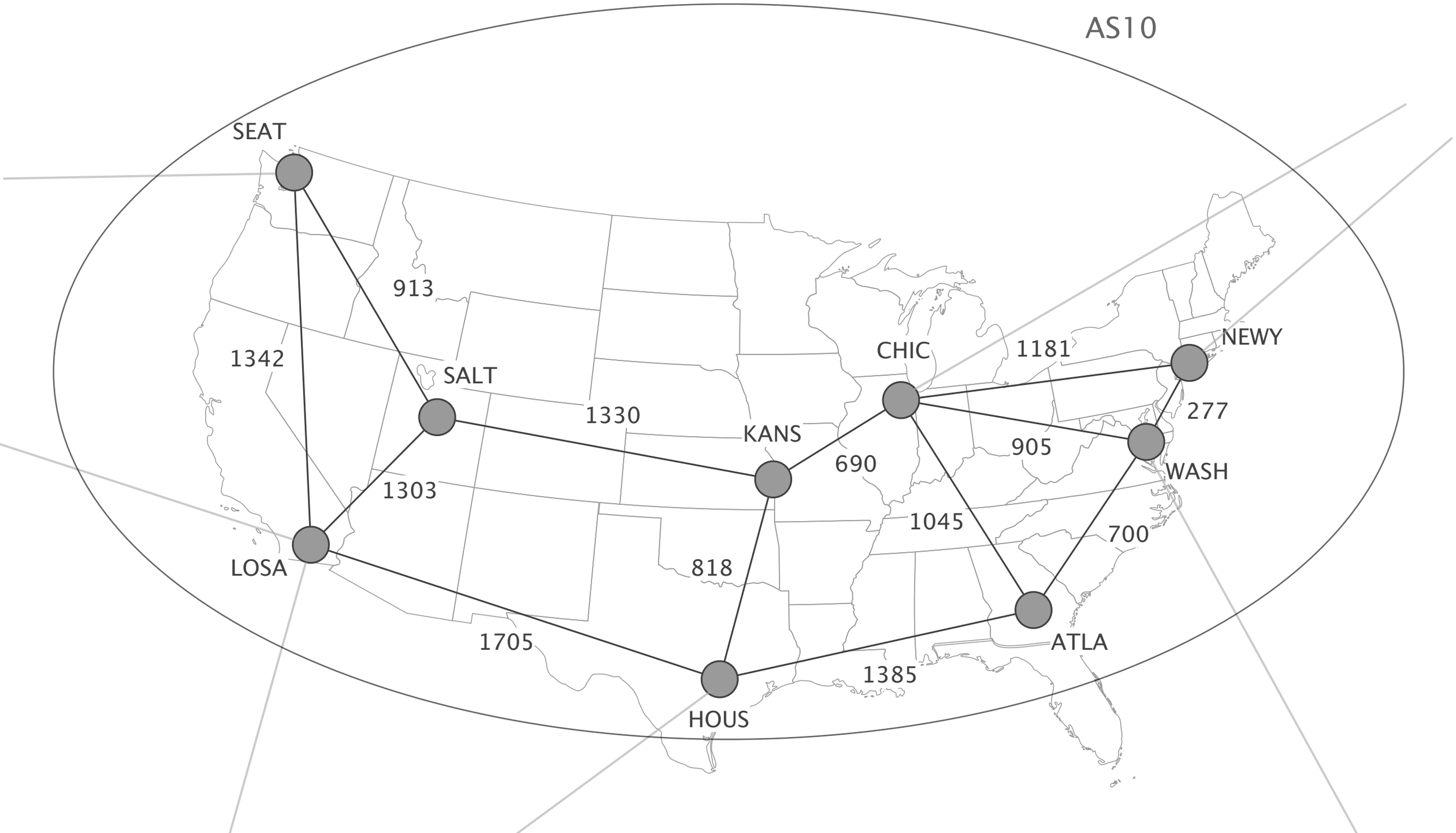
external BGP (eBGP) exchanges reachability information between ASes



internal BGP (iBGP) distributes externally learned routes internally

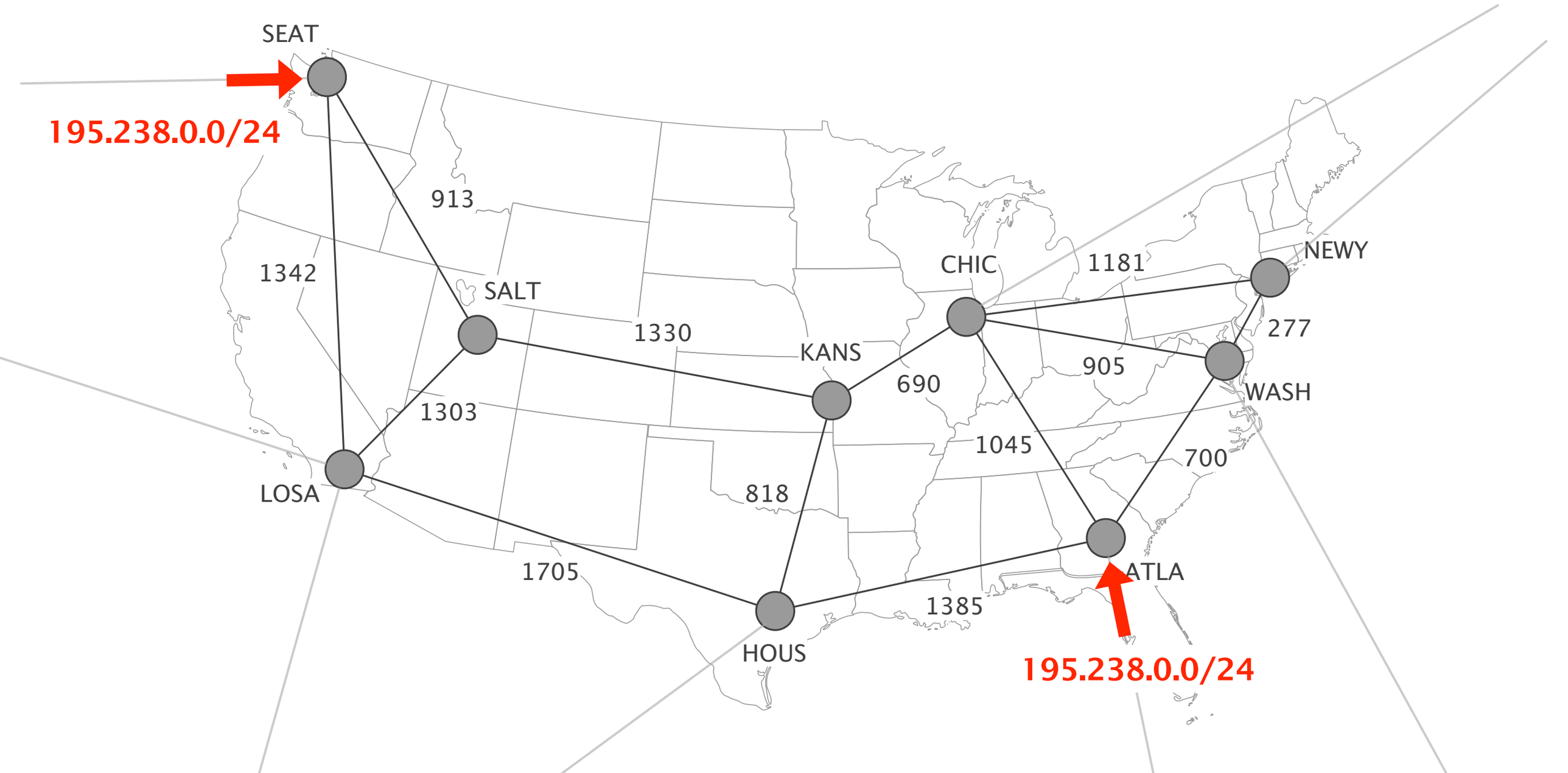


In this work, we focus on iBGP



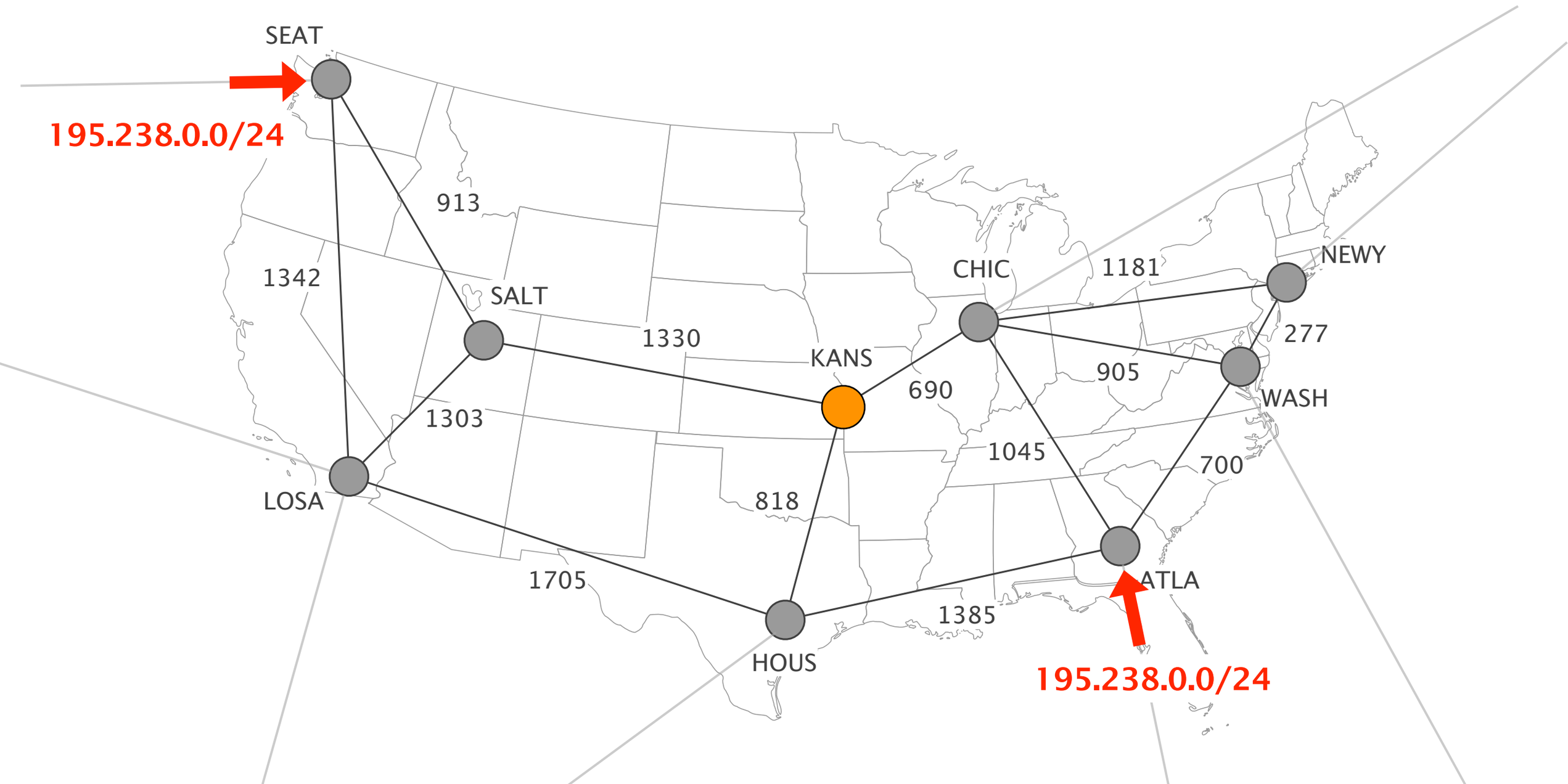
BGP is a single-route protocol.

Each router selects *one* route for each destination



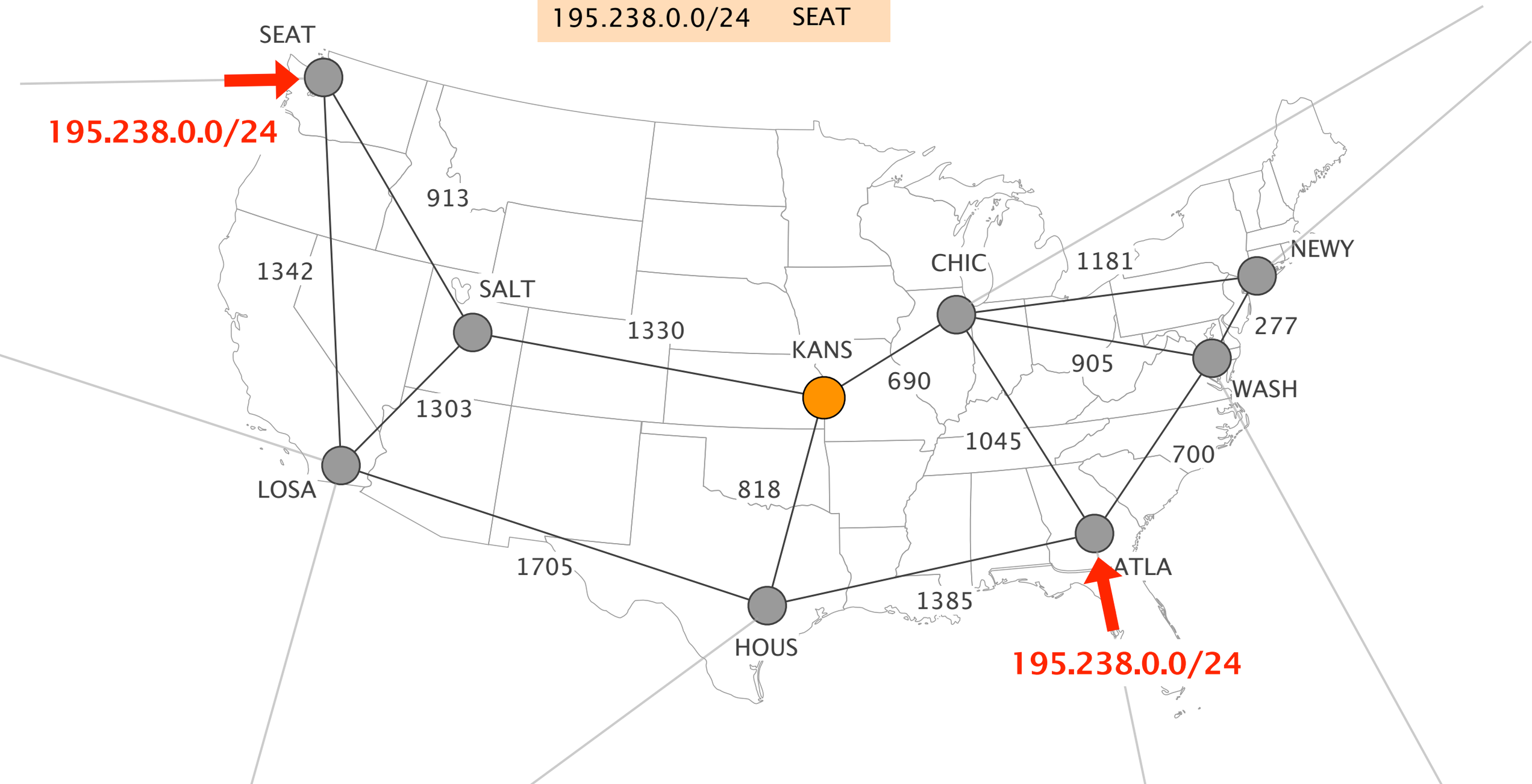
BGP is a single-route protocol.
Each router selects *one* route for each destination

When learning equivalent BGP routes, a router will prefer the closest one



KANS' BGP Routing table

dest	next-hop
195.238.0.0/24	ATLA
195.238.0.0/24	SEAT

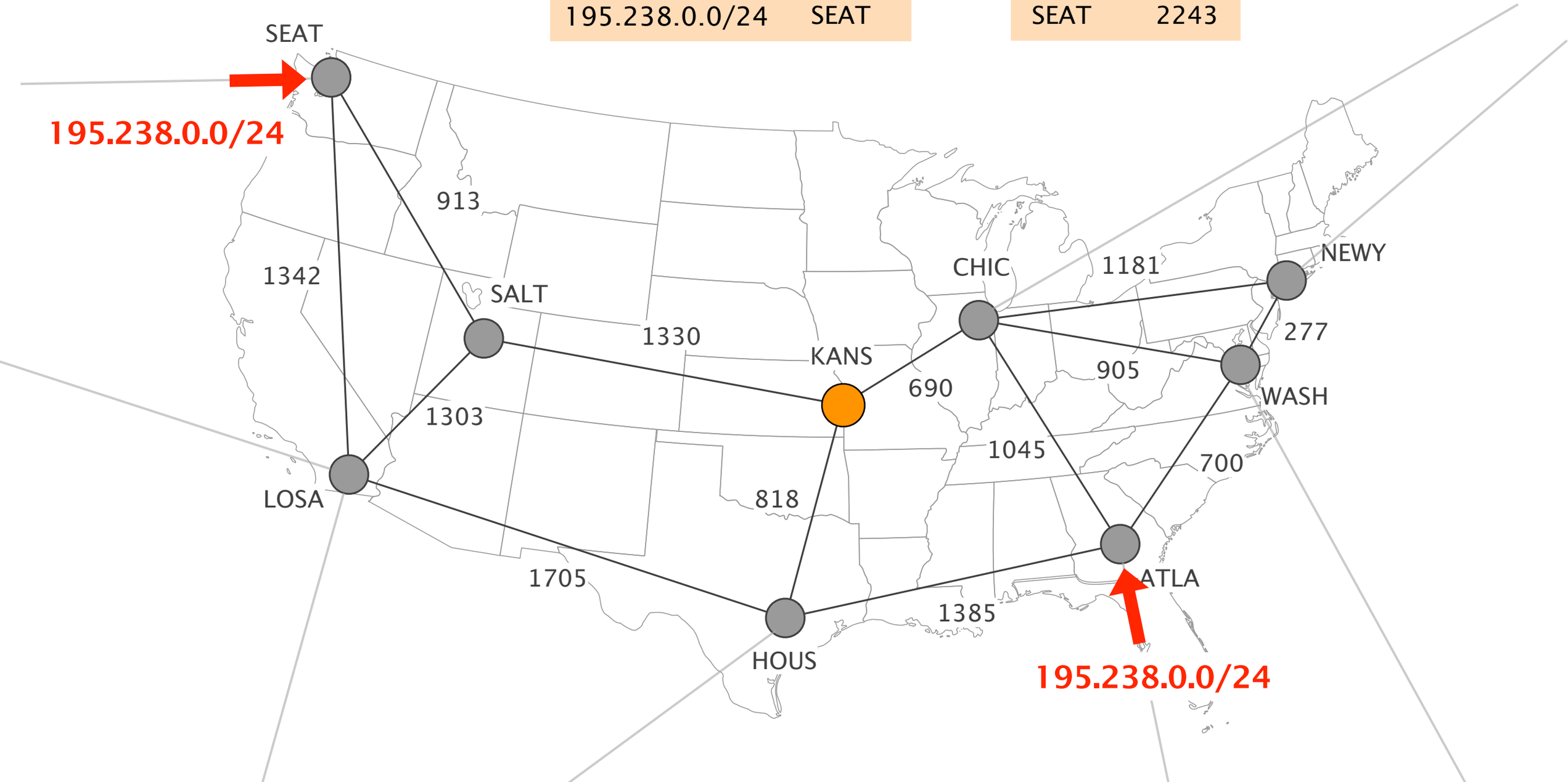


KANS' BGP Routing table

dest	next-hop
195.238.0.0/24	ATLA
195.238.0.0/24	SEAT

KANS' IGP Routing table

dest	weight
ATLA	1735
SEAT	2243

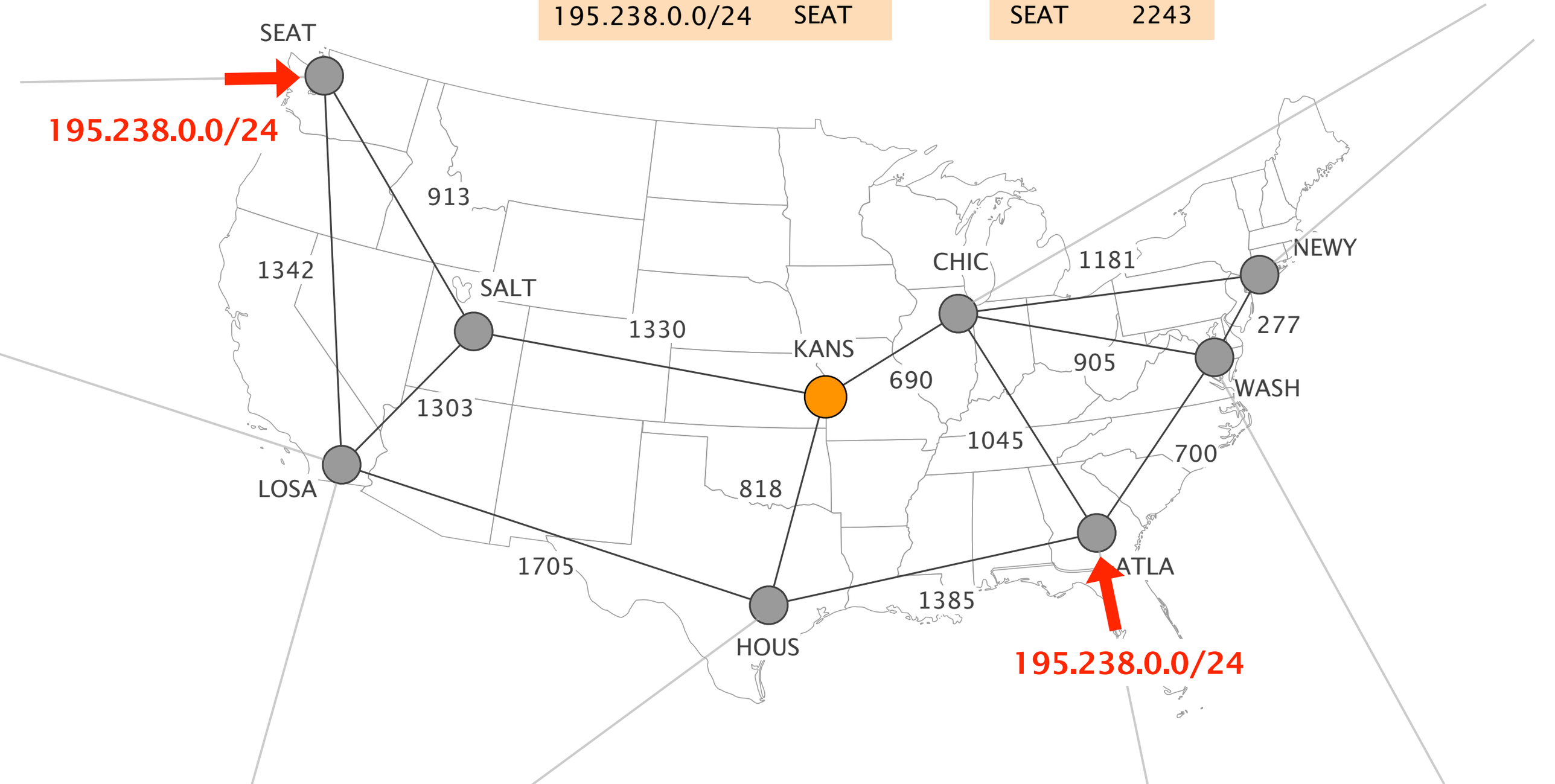


KANS' BGP Routing table

dest	next-hop
195.238.0.0/24	ATLA
195.238.0.0/24	SEAT

KANS' IGP Routing table

dest	weight
ATLA	1735
SEAT	2243



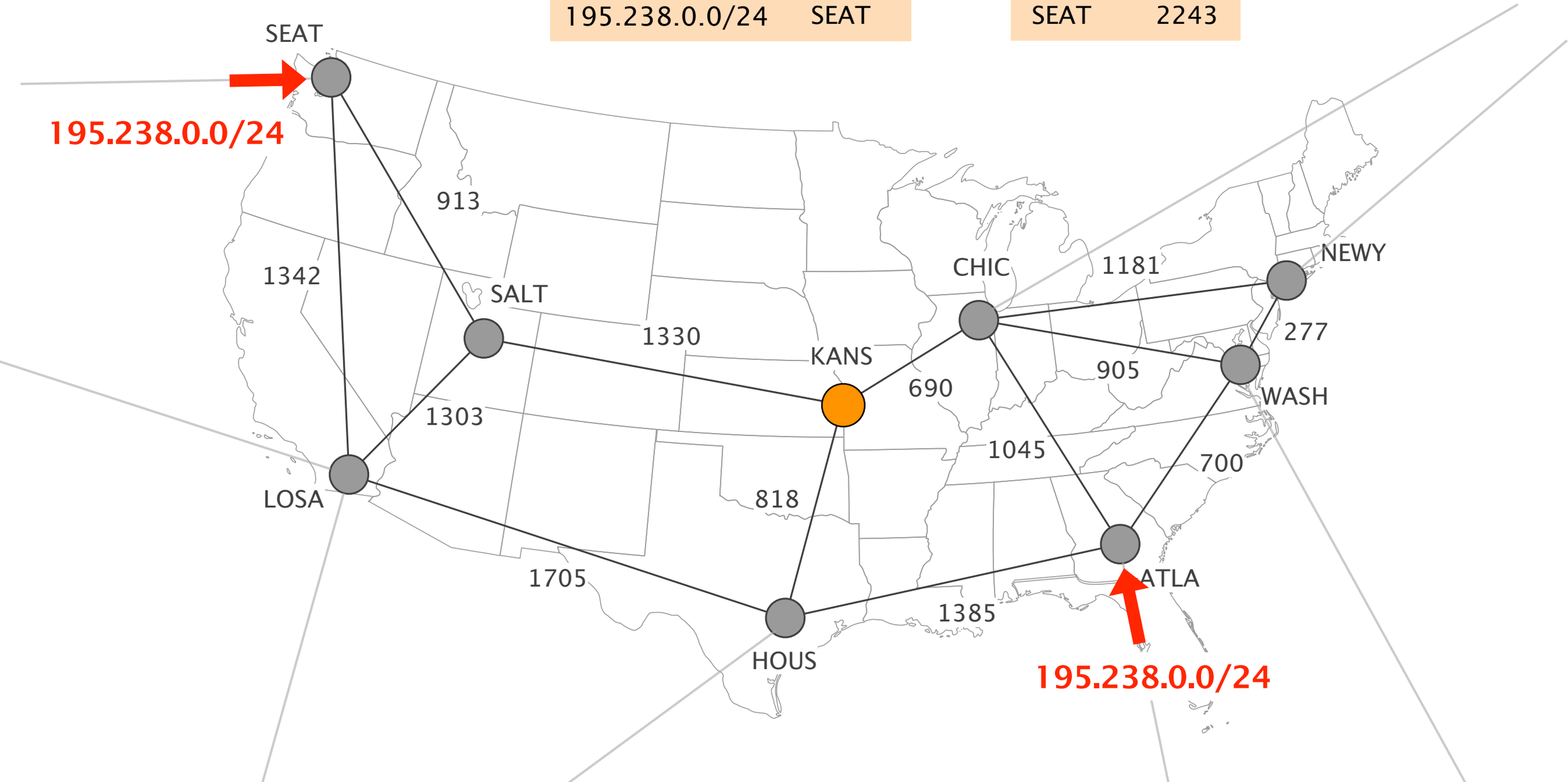
KANS' BGP Routing table

dest	next-hop
195.238.0.0/24	ATLA
195.238.0.0/24	SEAT

KANS' IGP Routing table

dest	weight
ATLA	1735
SEAT	2243

best BGP route



Reconfiguring the IGP can create any BGP anomaly

IGP reconfiguration can lead to *unavoidable* BGP-induced:

- forwarding loops
- routing oscillations
- network congestion
- blackholes

Reconfiguring the IGP can create any BGP anomaly

IGP reconfiguration can lead to *unavoidable* BGP-induced:

- forwarding loops
- routing oscillations
- network congestion
- blackholes

even if the initial and the final configurations are correct

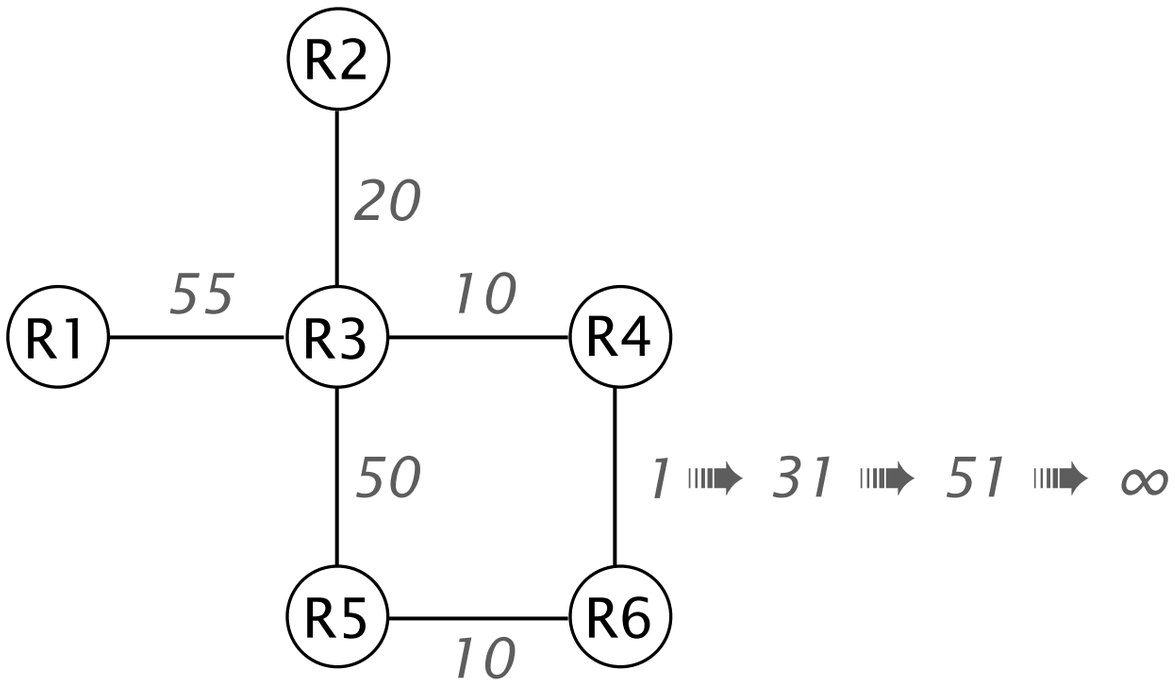
Reconfiguring the IGP can create any BGP anomaly

IGP reconfiguration can lead to *unavoidable* BGP-induced:

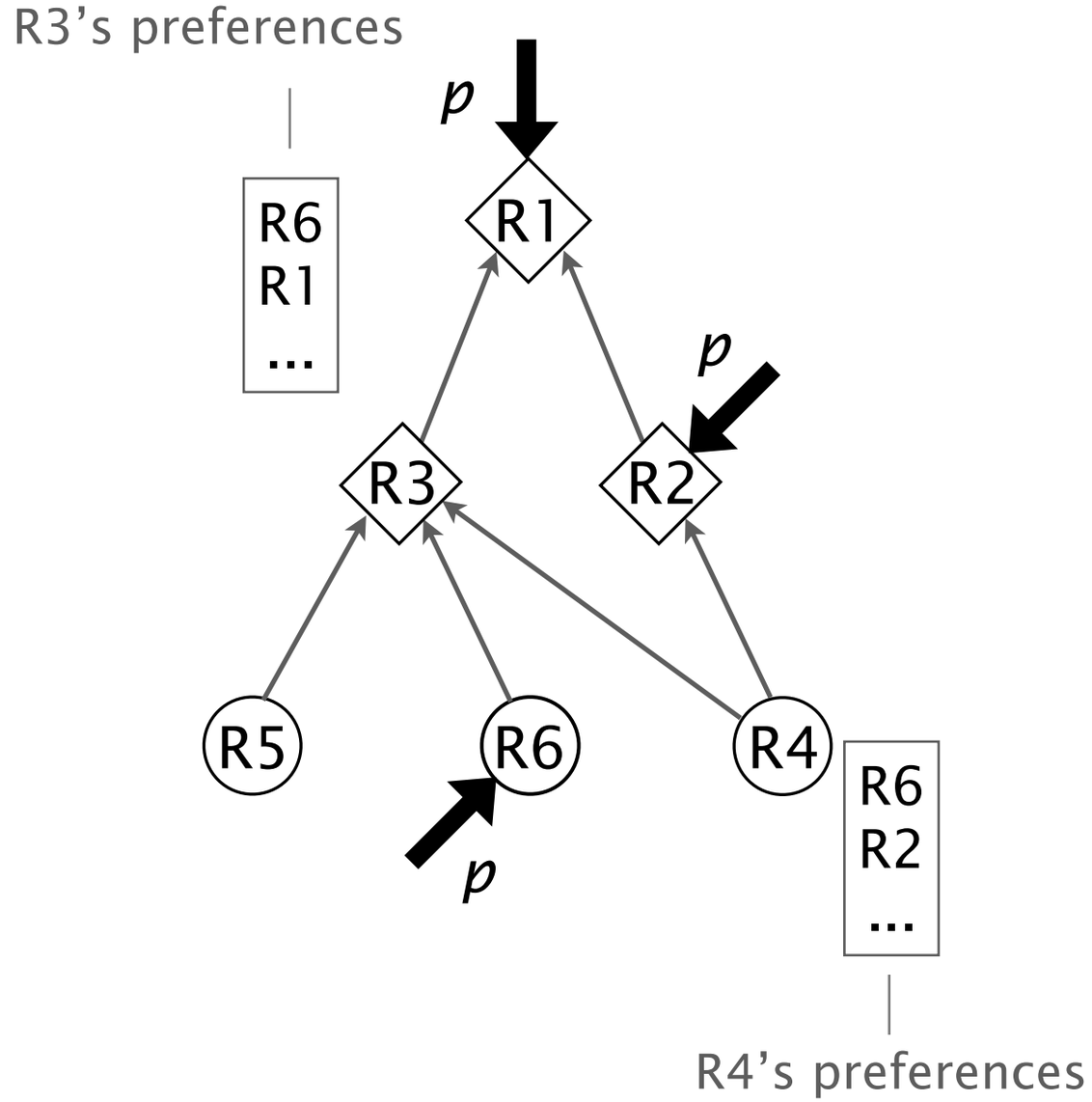
- **forwarding loops**
- routing oscillations
- network congestion
- blackholes

even if the initial and the final configurations are correct

Reconfiguring the IGP can create forwarding loops



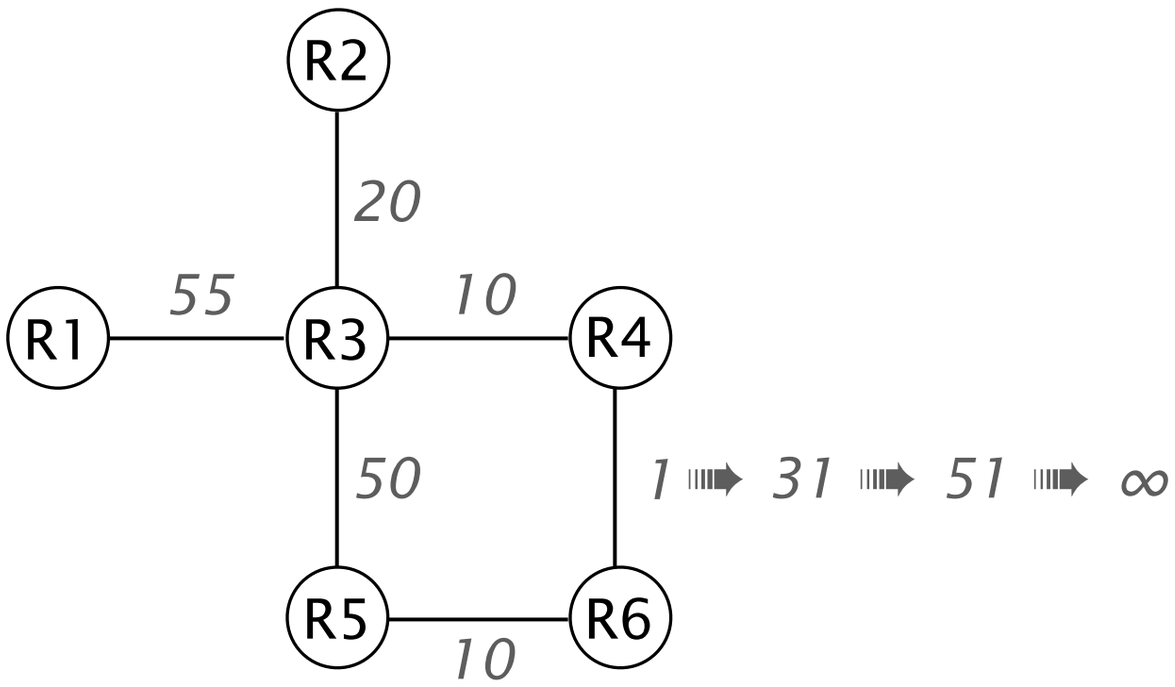
IGP topology



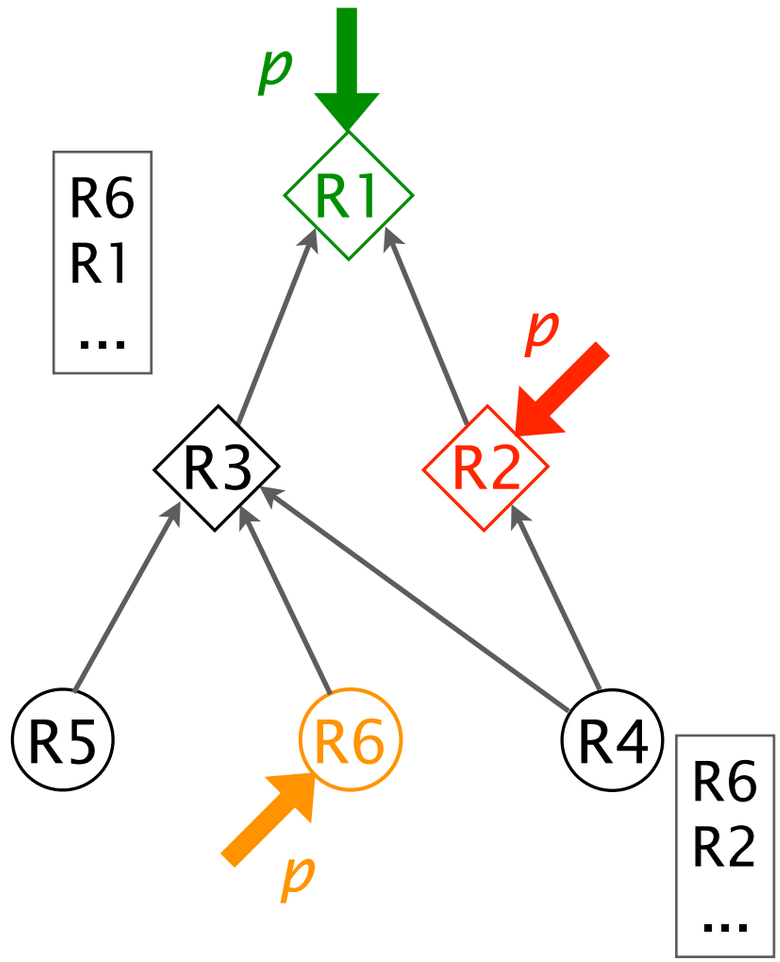
iBGP topology

Due to iBGP propagation rules,

R3 never learns the route propagated by R2

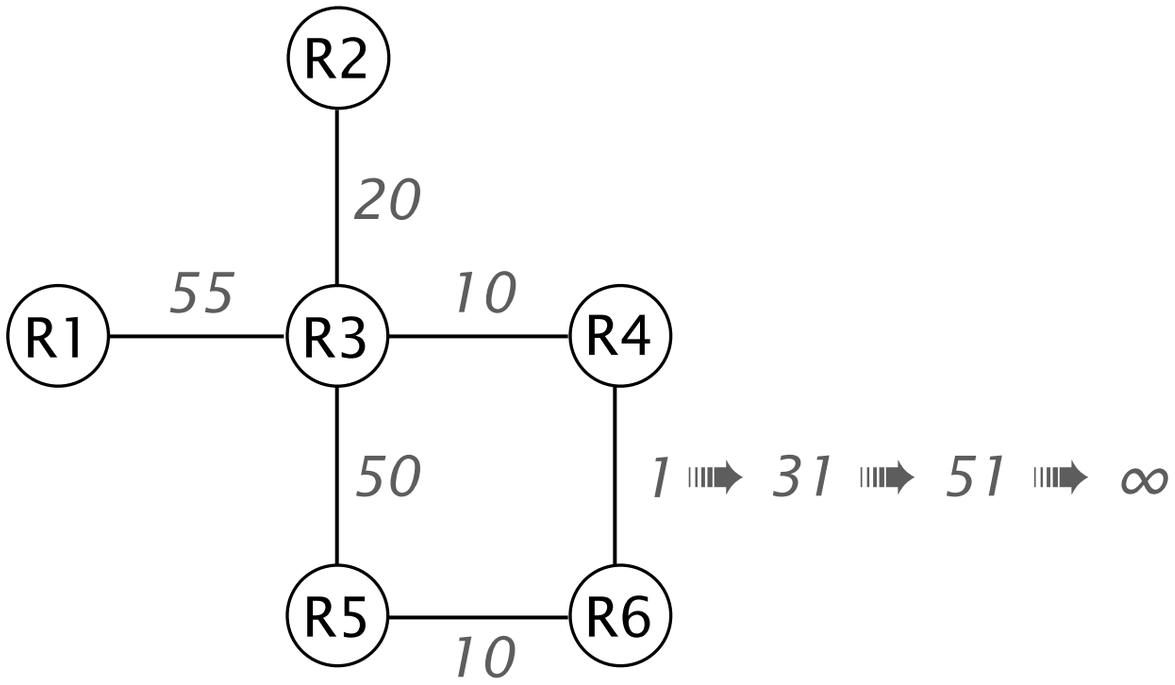


IGP topology

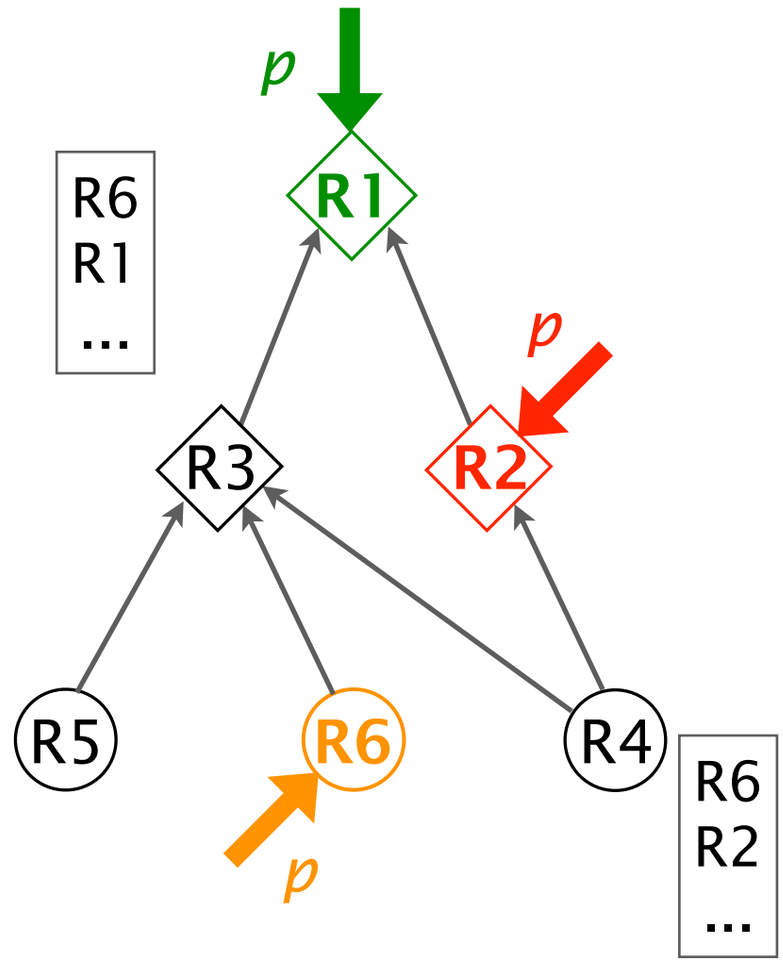


iBGP topology

By default, egress routers prefer their external routes

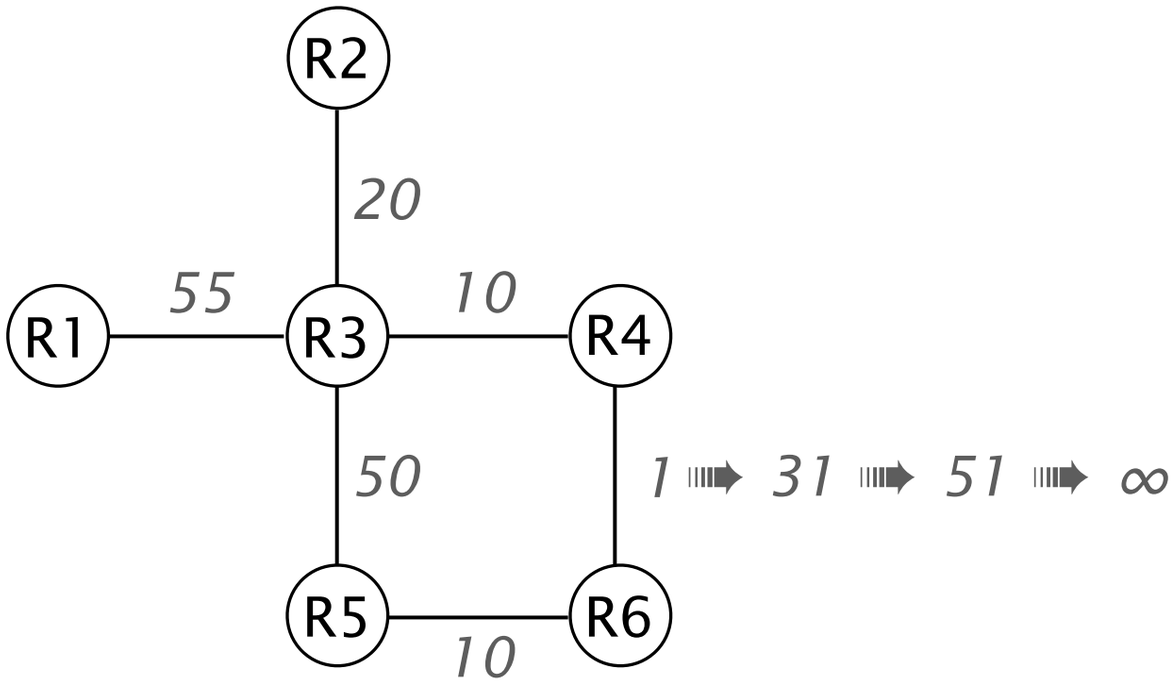


IGP topology

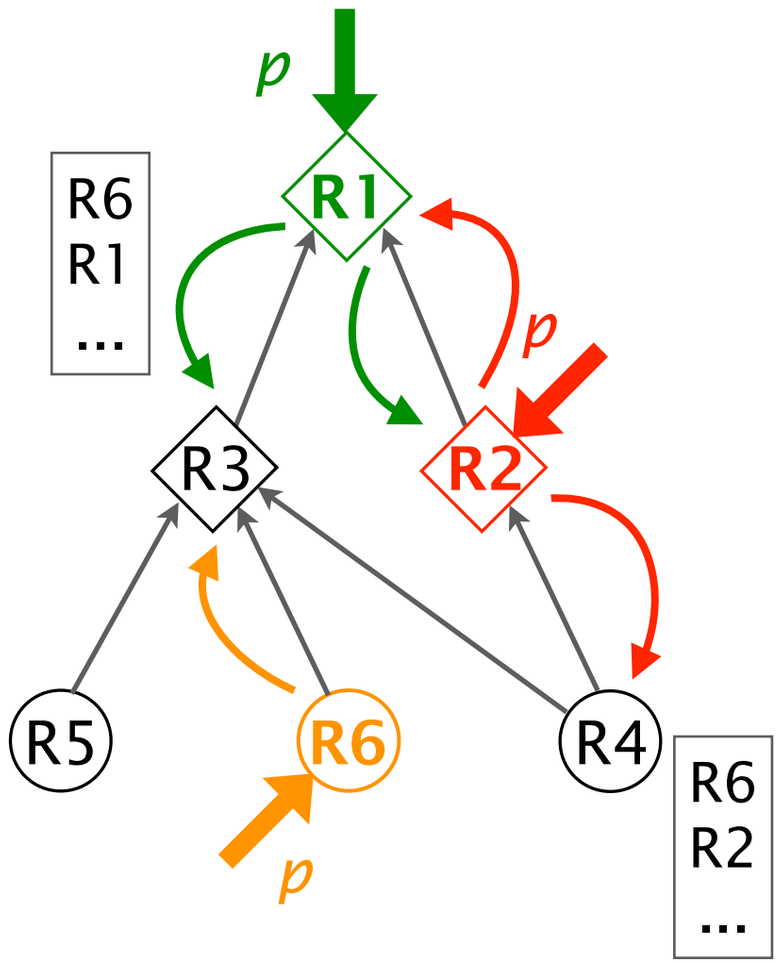


iBGP topology

R3 receives two routes, from R1 and R6,
and prefer R6 due to IGP distance

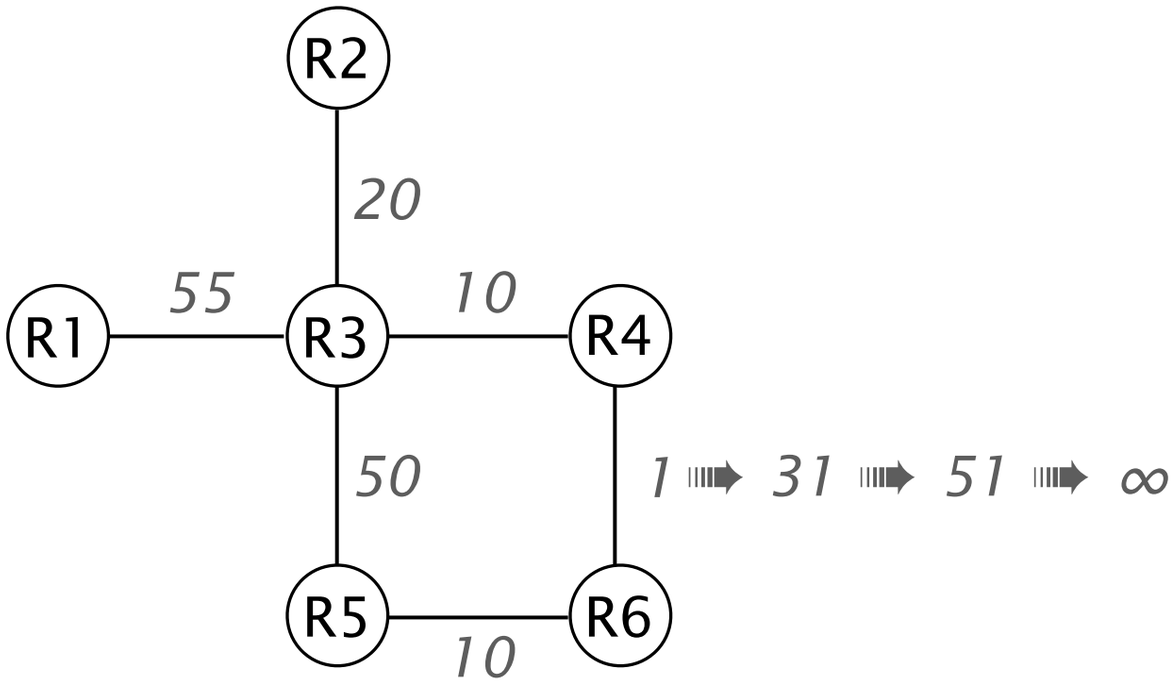


IGP topology

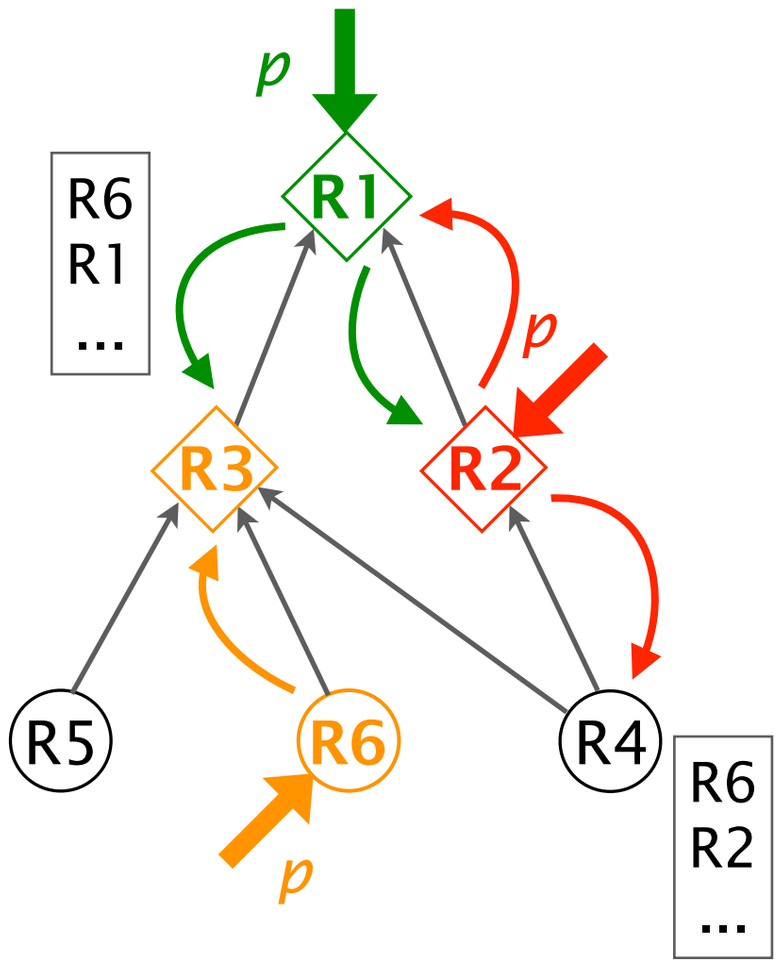


iBGP topology

R3 receives two routes, from R1 and R6,
and prefer R6 due to IGP distance

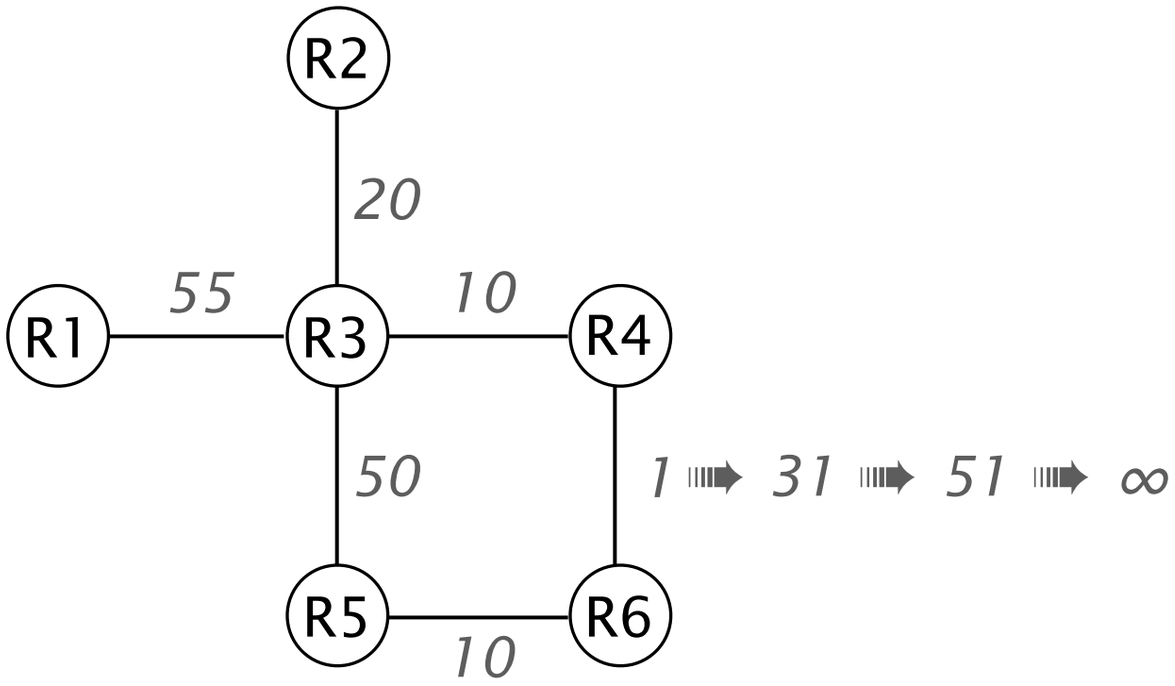


IGP topology

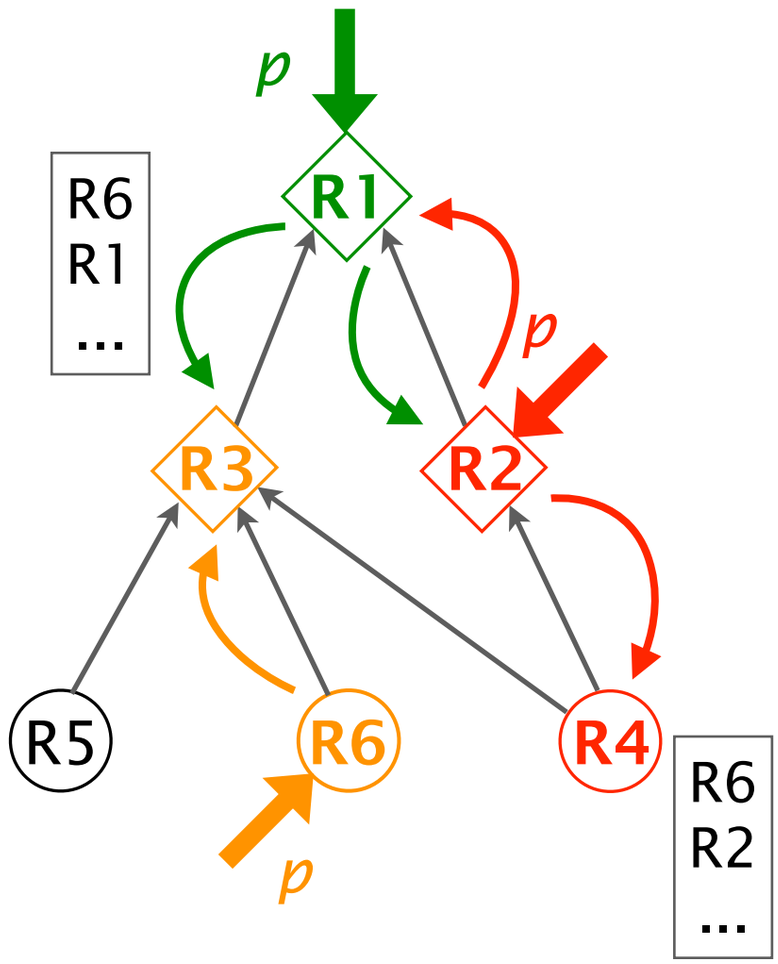


iBGP topology

R4 first receives the R2 route and prefers it



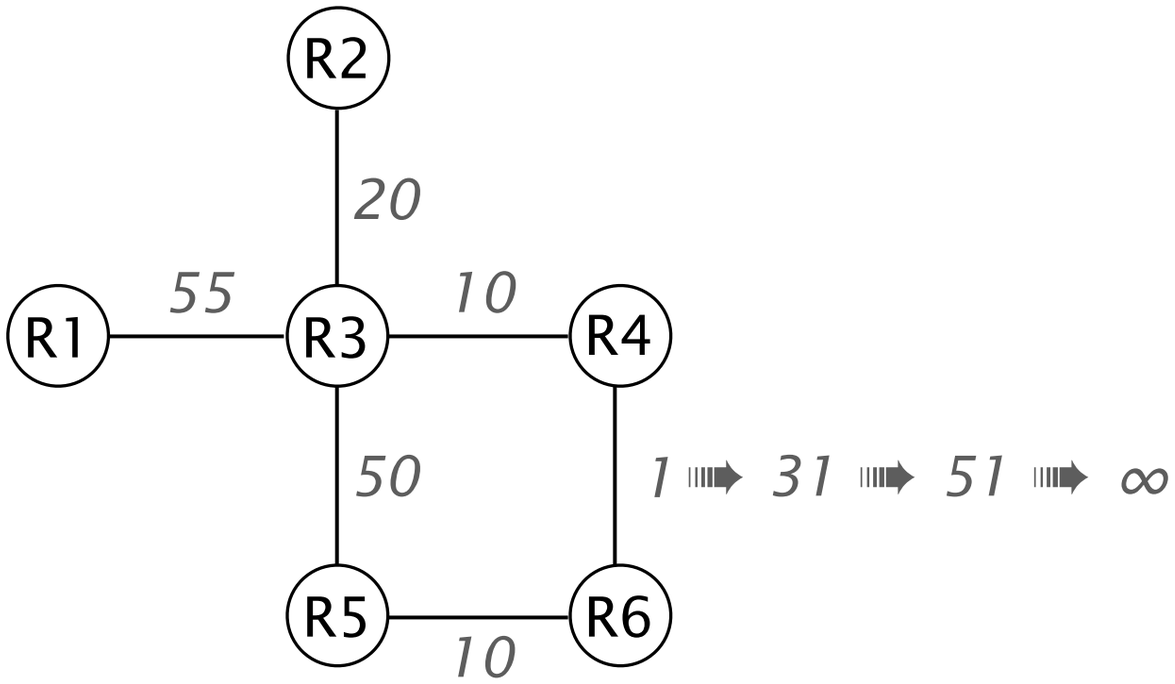
IGP topology



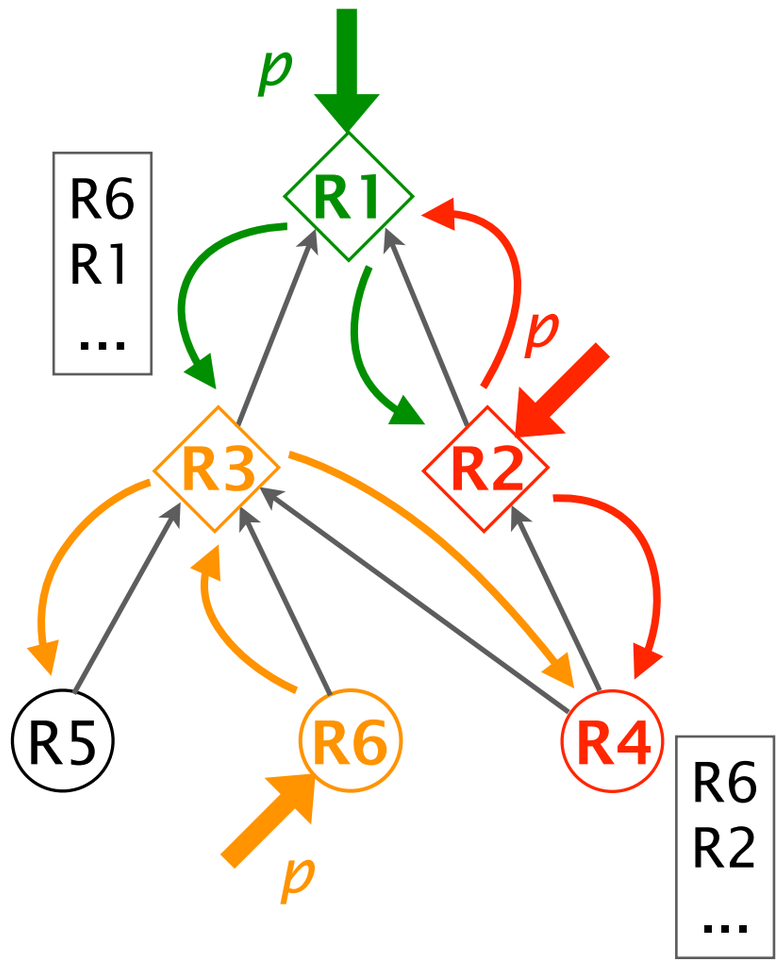
iBGP topology

R4 then learns the R6 route via R4 and prefers it

due to the IGP distance



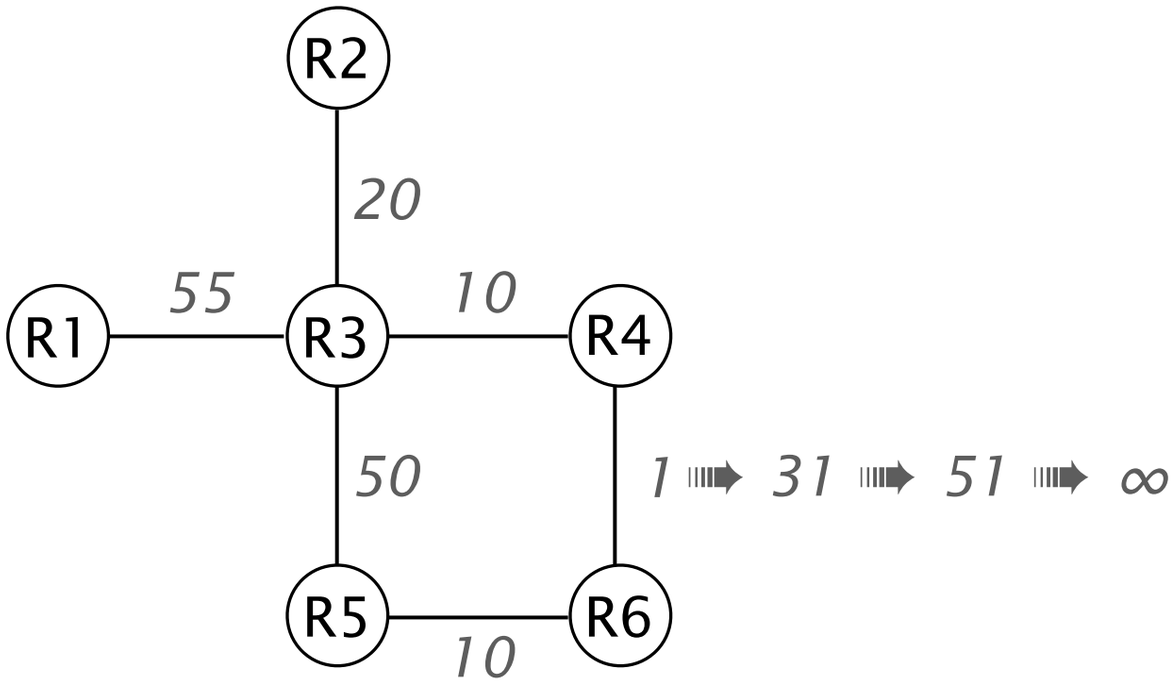
IGP topology



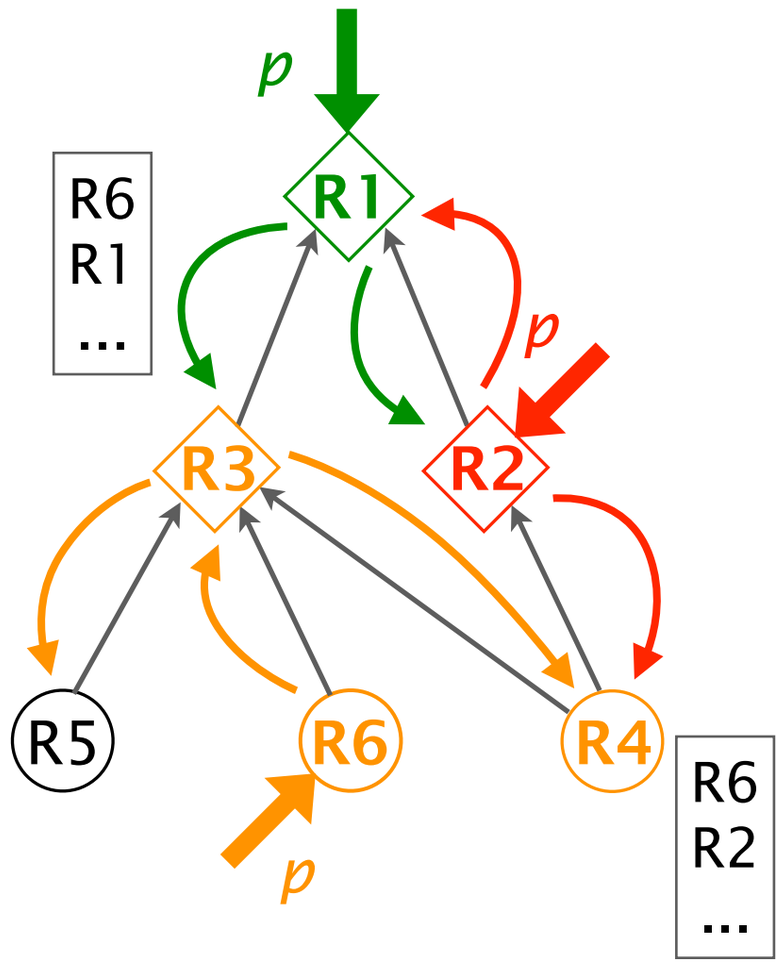
iBGP topology

R4 then learns the R6 route via R4 and prefers it

due to the IGP distance

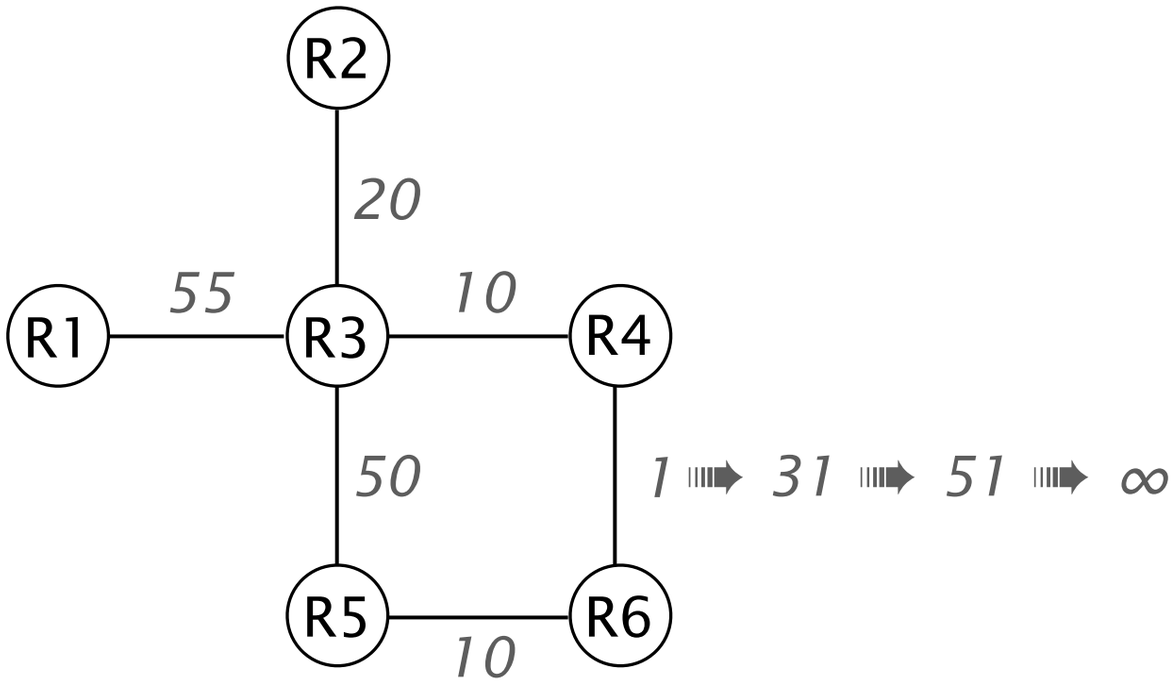


IGP topology

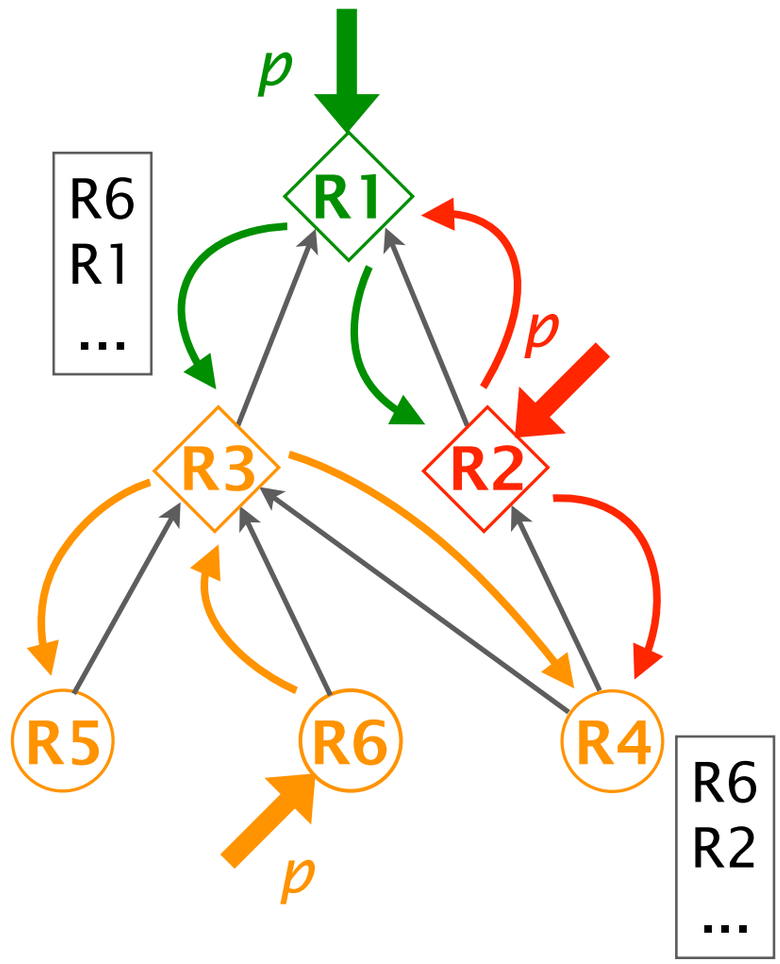


iBGP topology

R5 learns the R6 route via R3 and prefers it

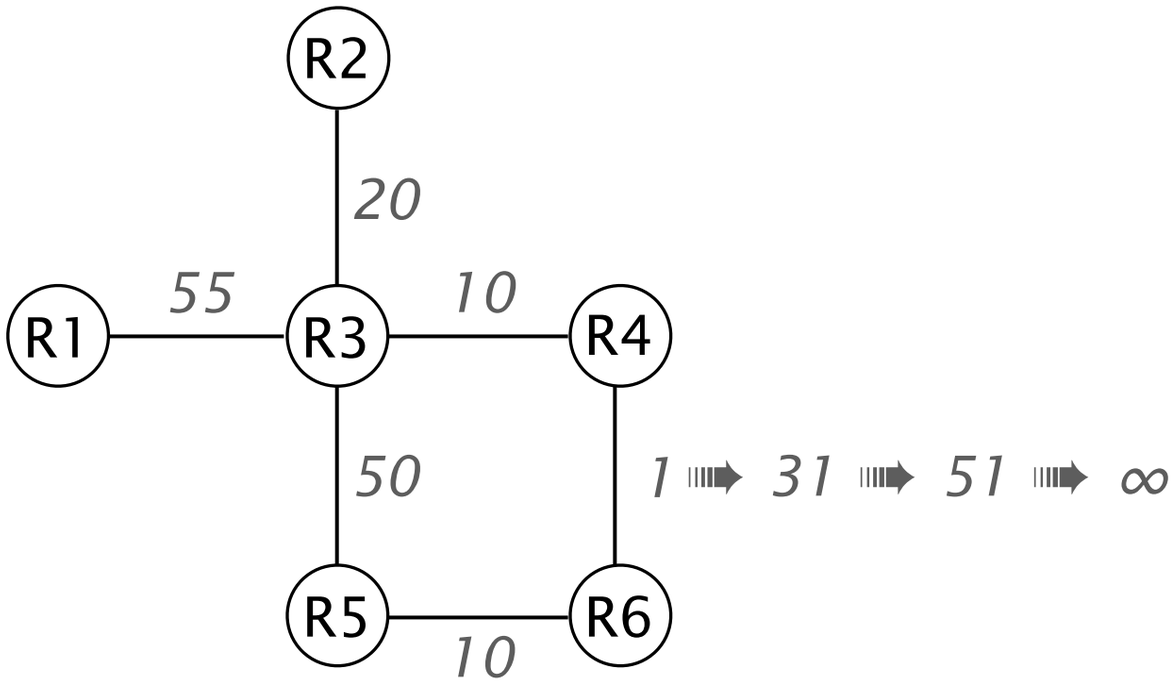


IGP topology

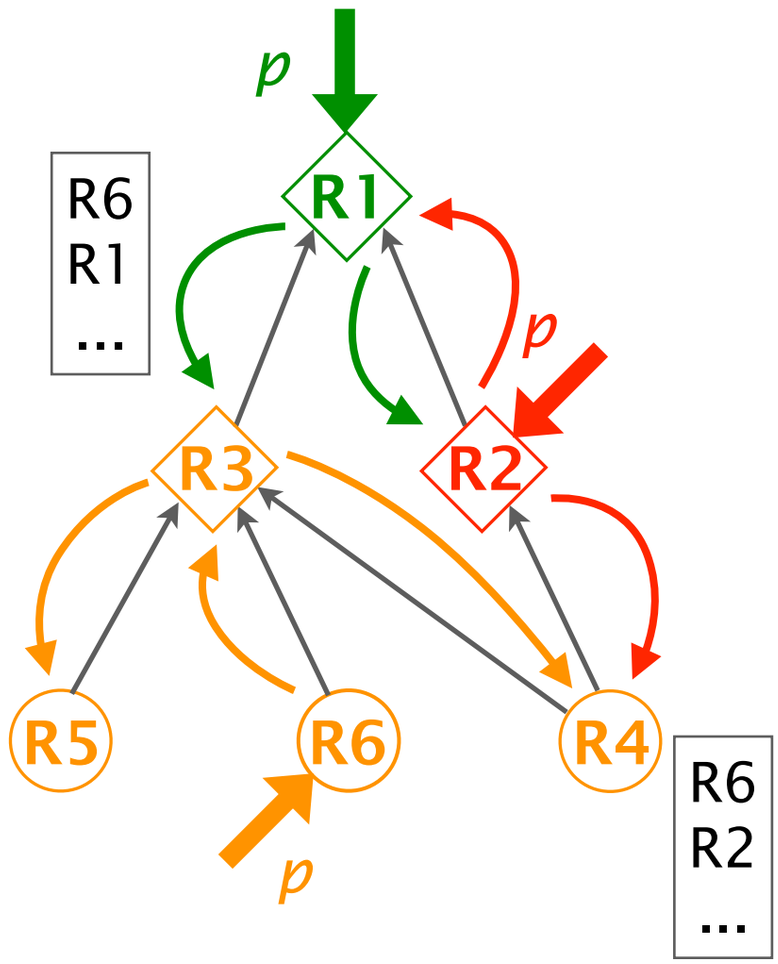


iBGP topology

The initial forwarding state is *loop-free*

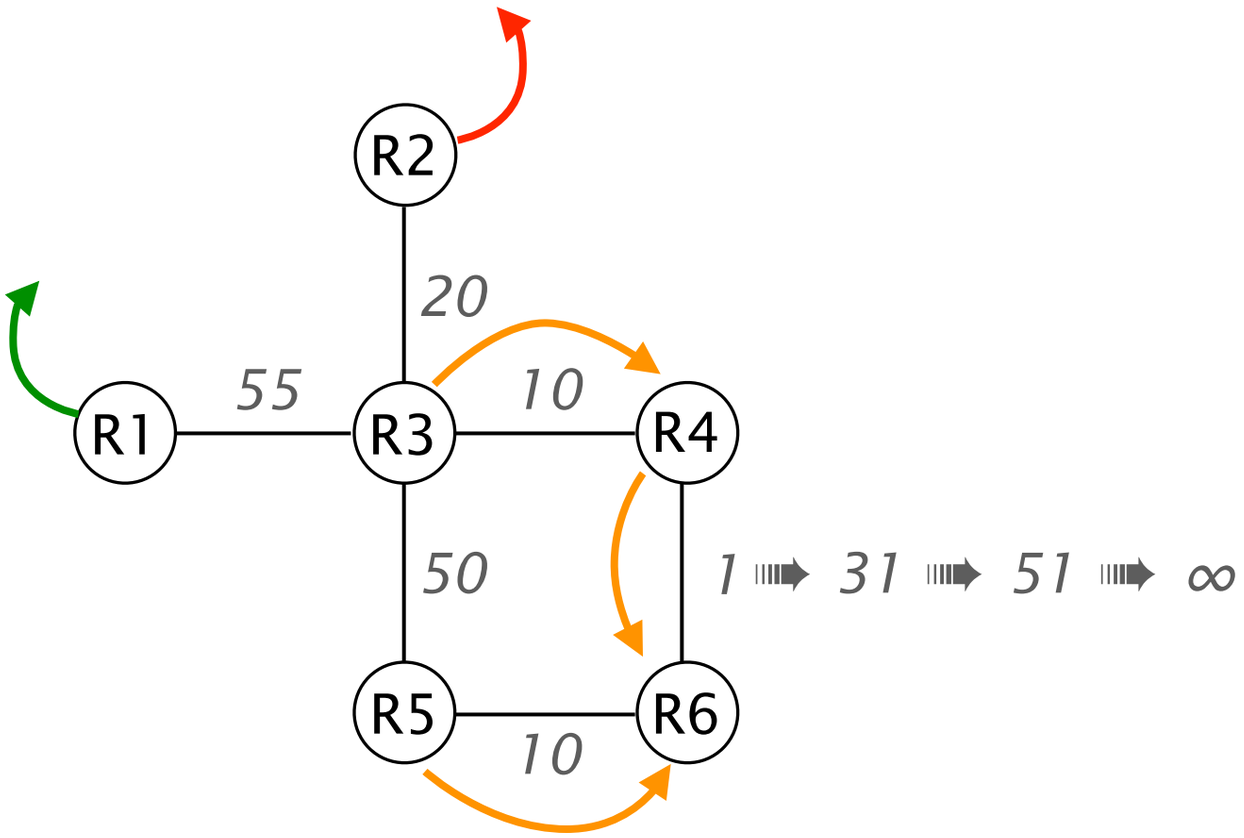


IGP topology

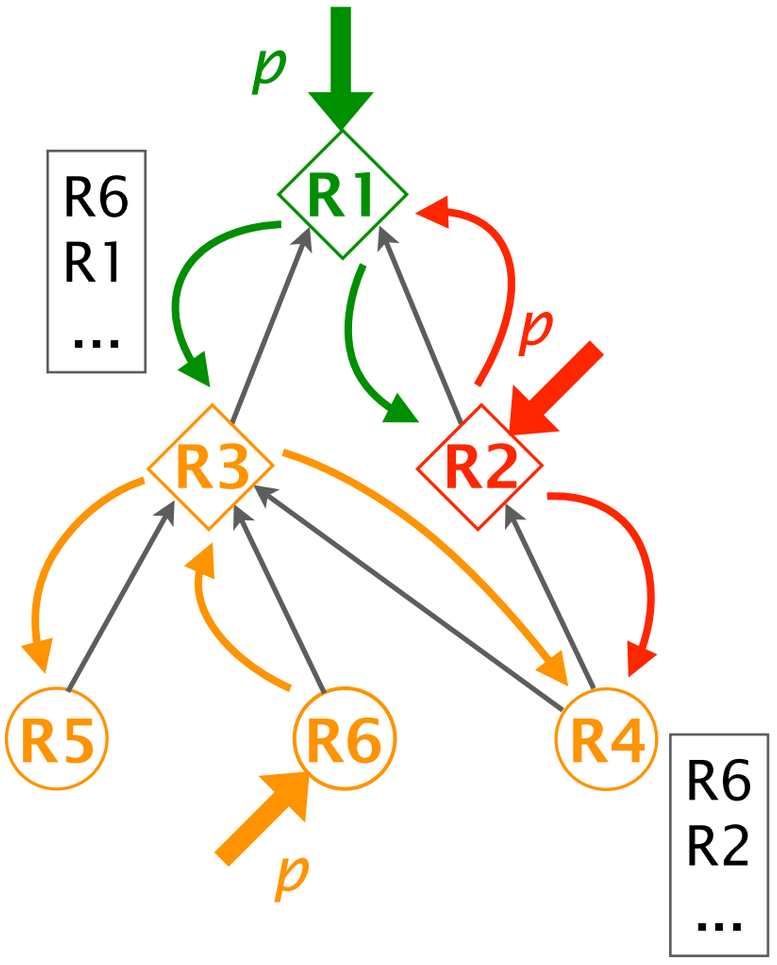


iBGP topology

The initial forwarding state is *loop-free*

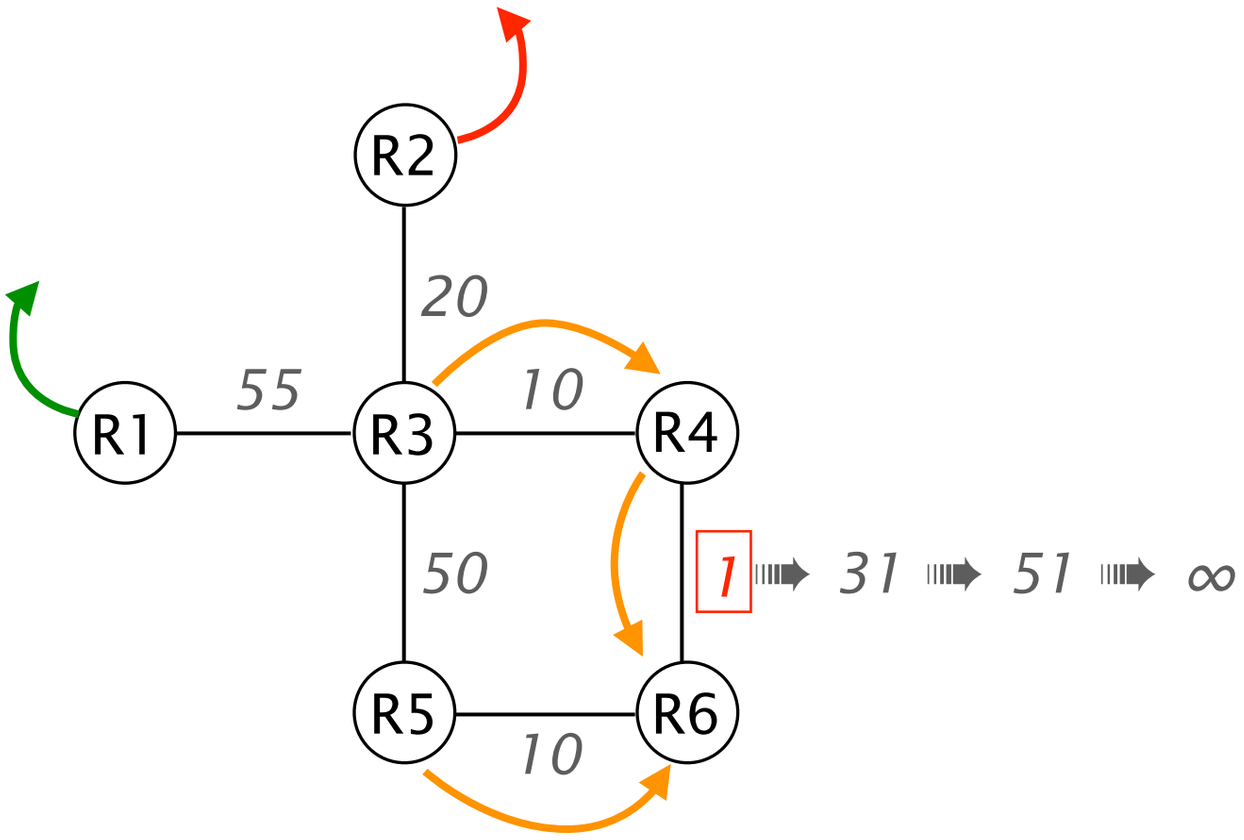


IGP topology

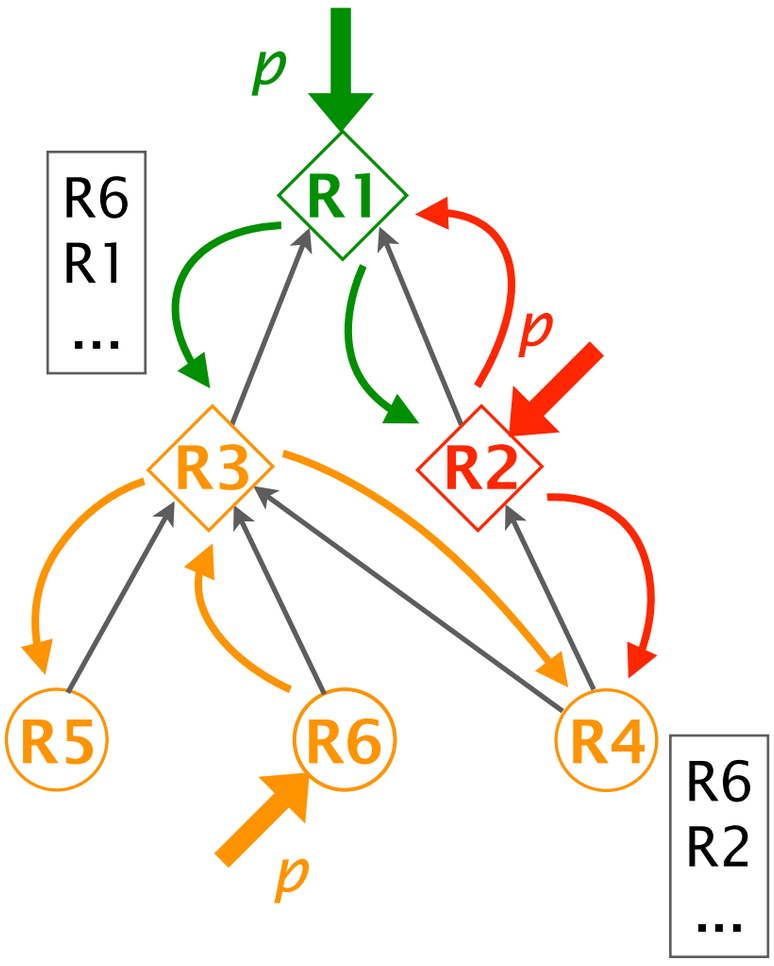


iBGP topology

Let’s proceed to the first metric-increment

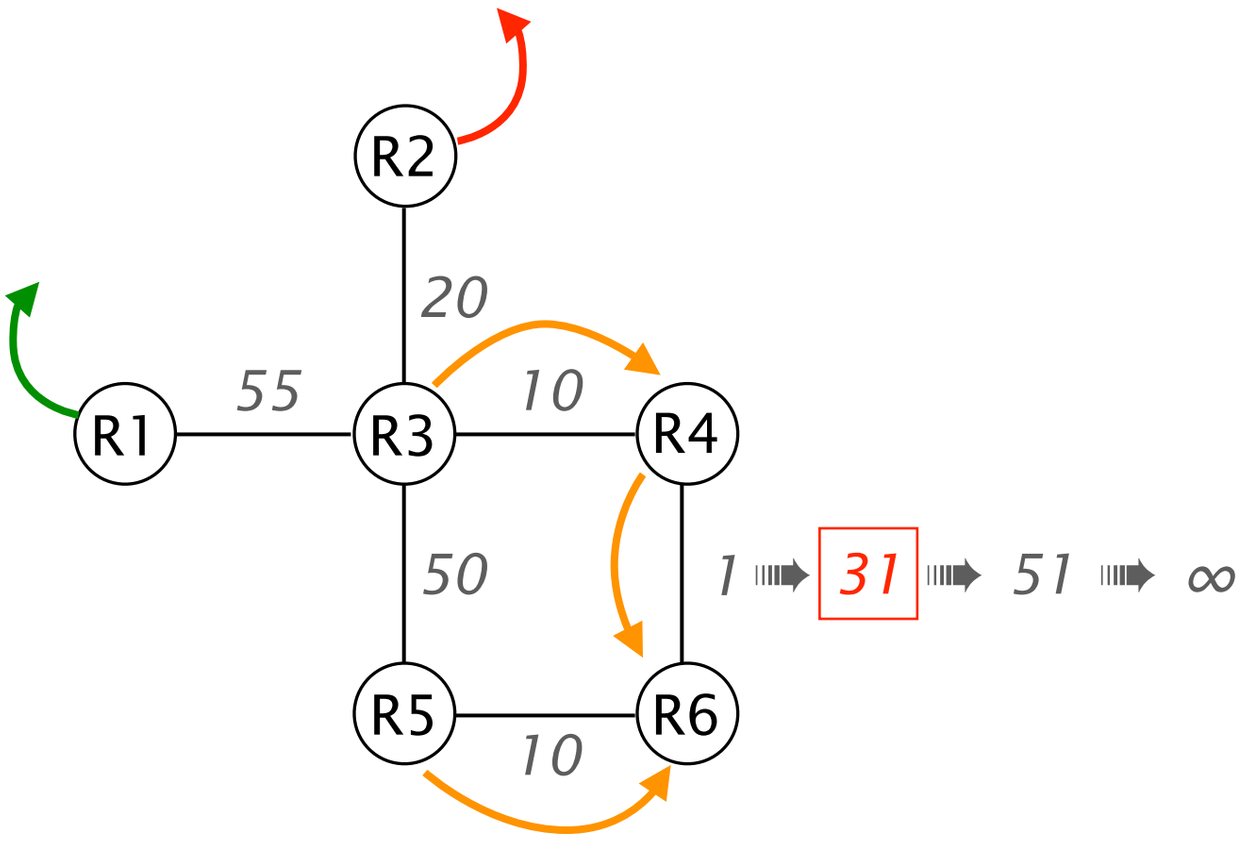


IGP topology

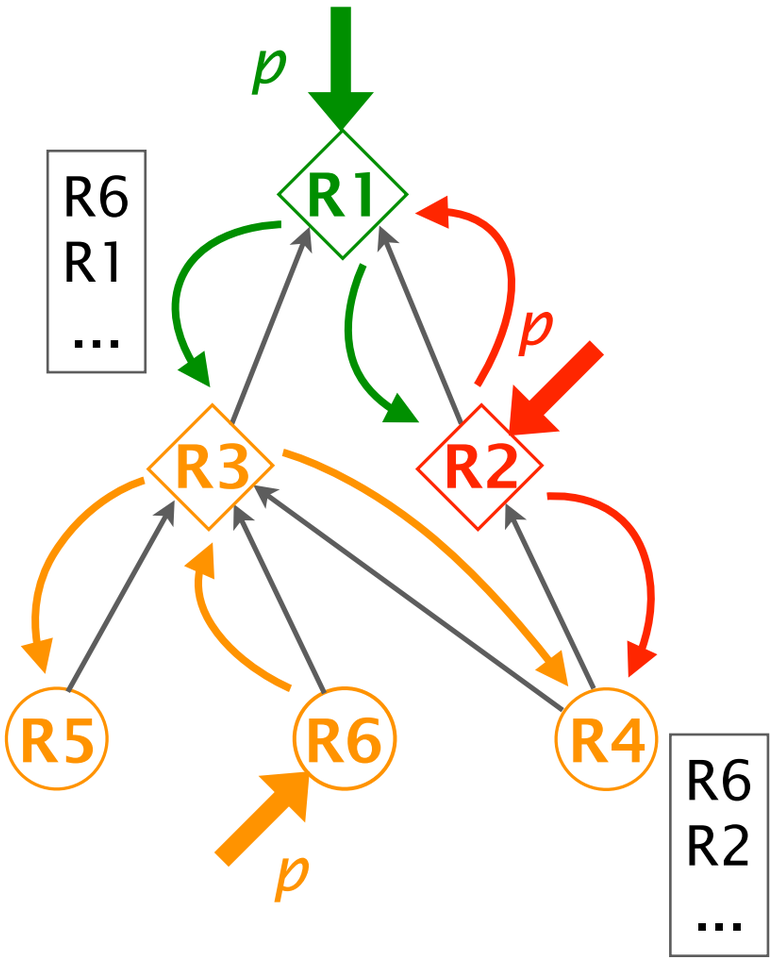


iBGP topology

Let’s proceed to the first metric-increment

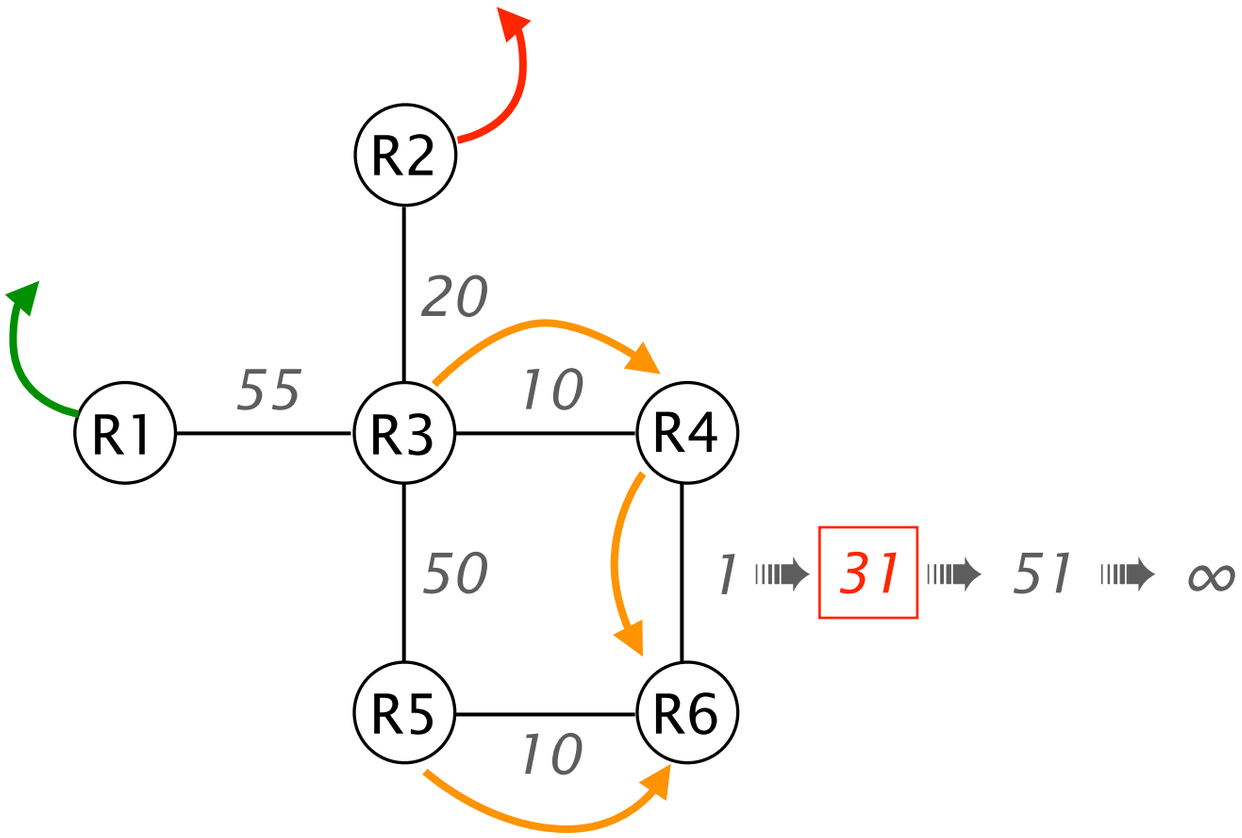


IGP topology

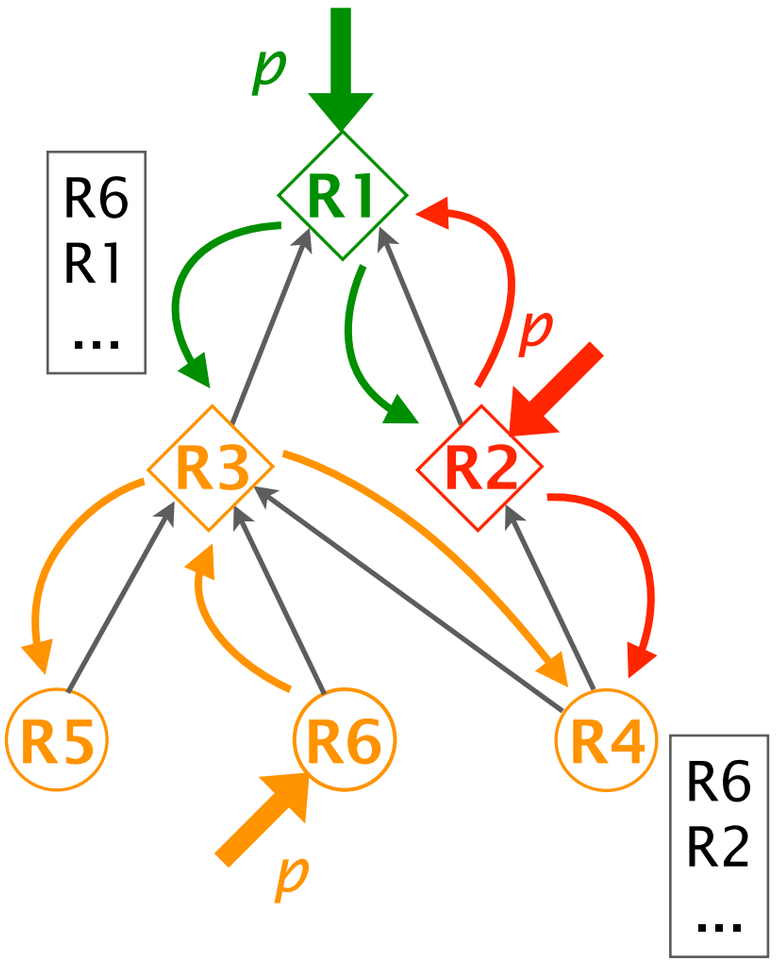


iBGP topology

R4 is now closer to R2 (distance 30) than R6 (distance 31)

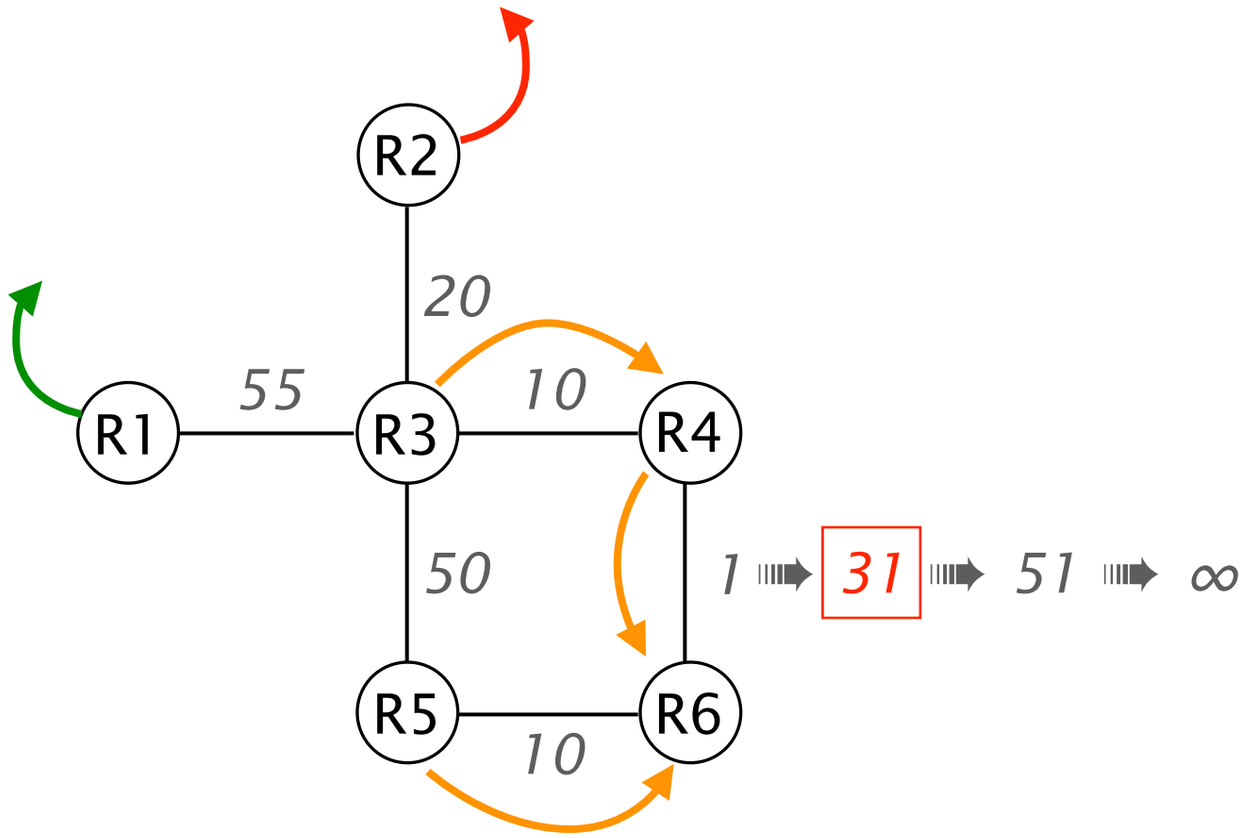


IGP topology

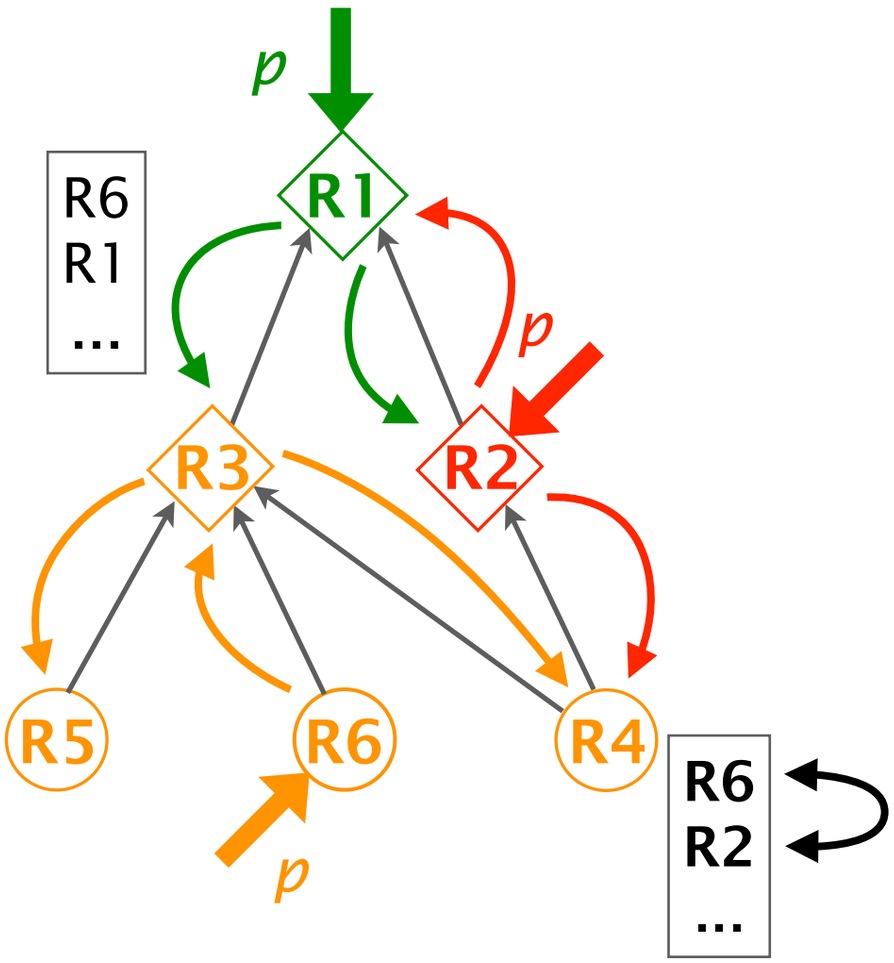


iBGP topology

R4 is now closer to R2 (distance 30) than R6 (distance 31)

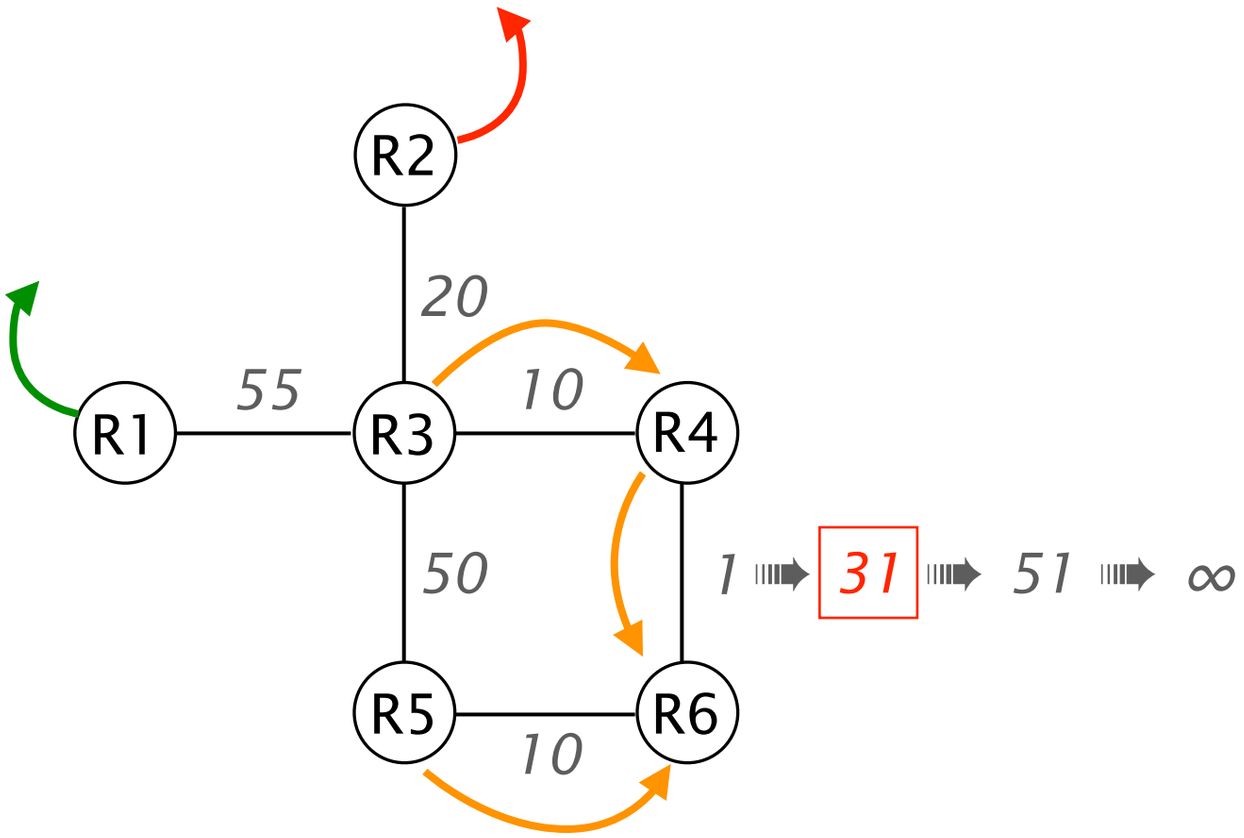


IGP topology

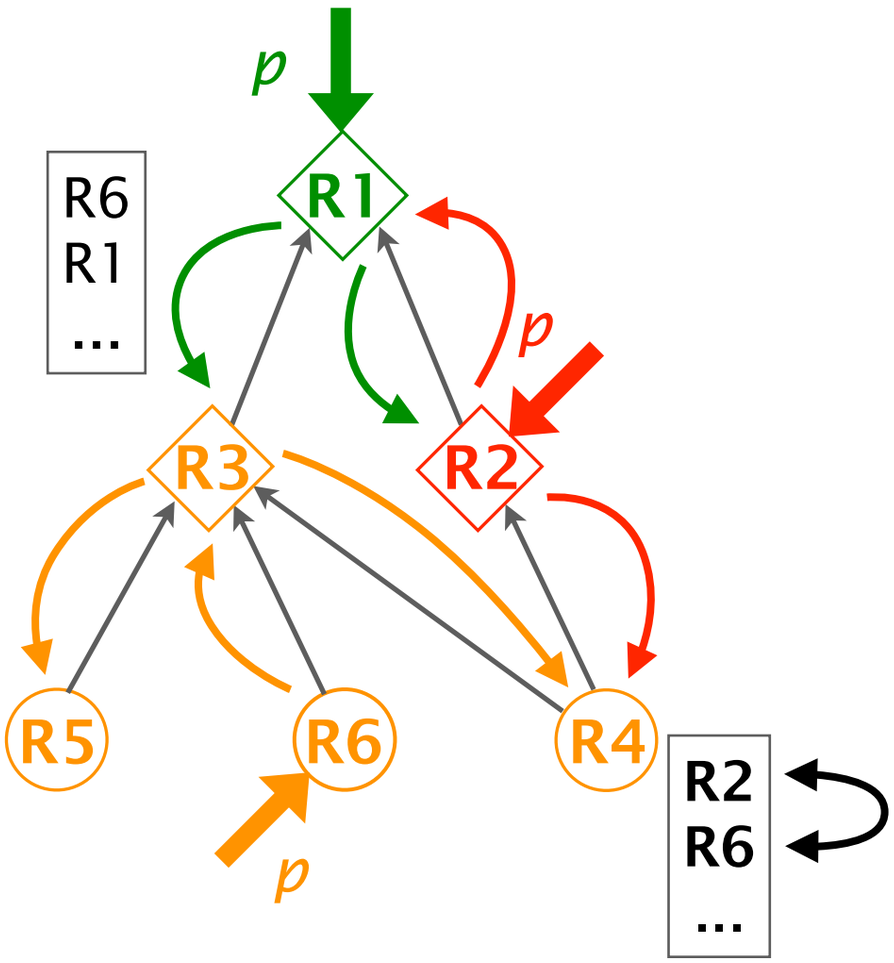


iBGP topology

R4 is now closer to R2 (distance 30) than R6 (distance 31)

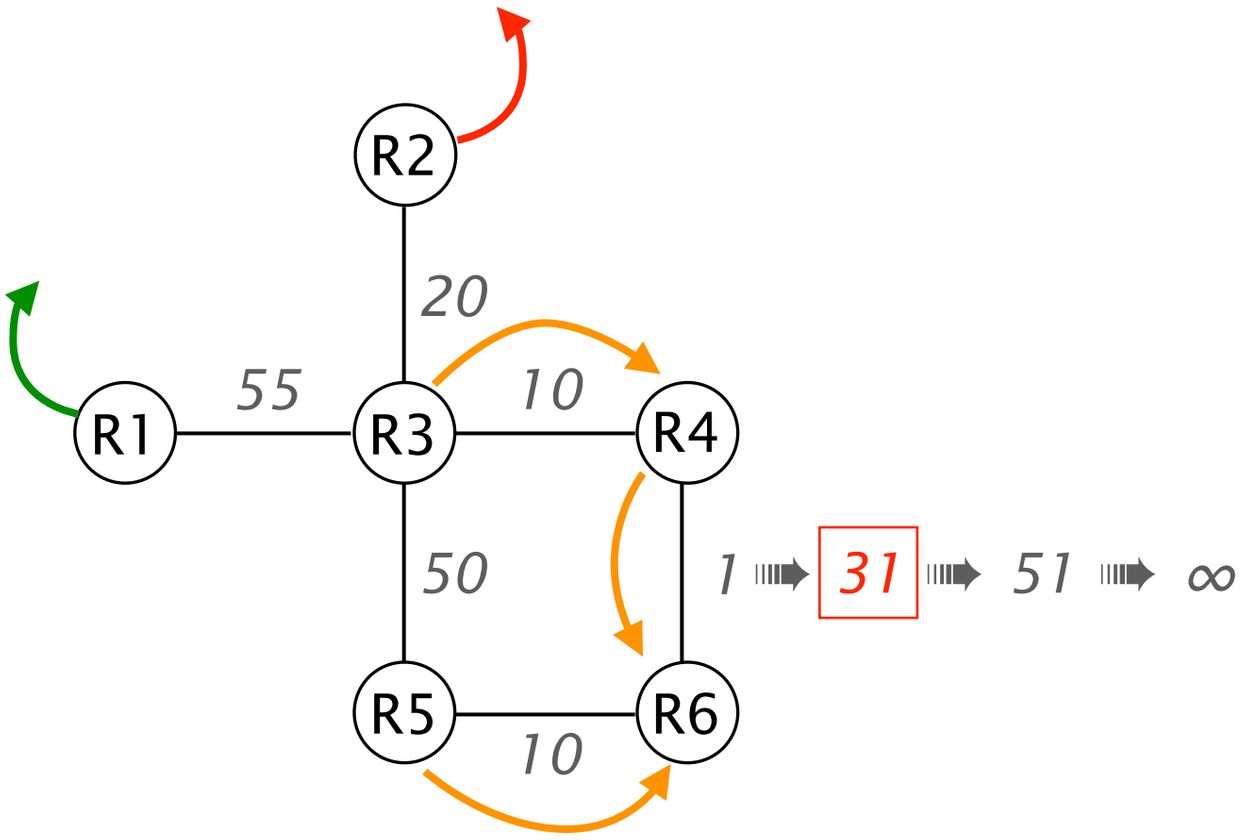


IGP topology

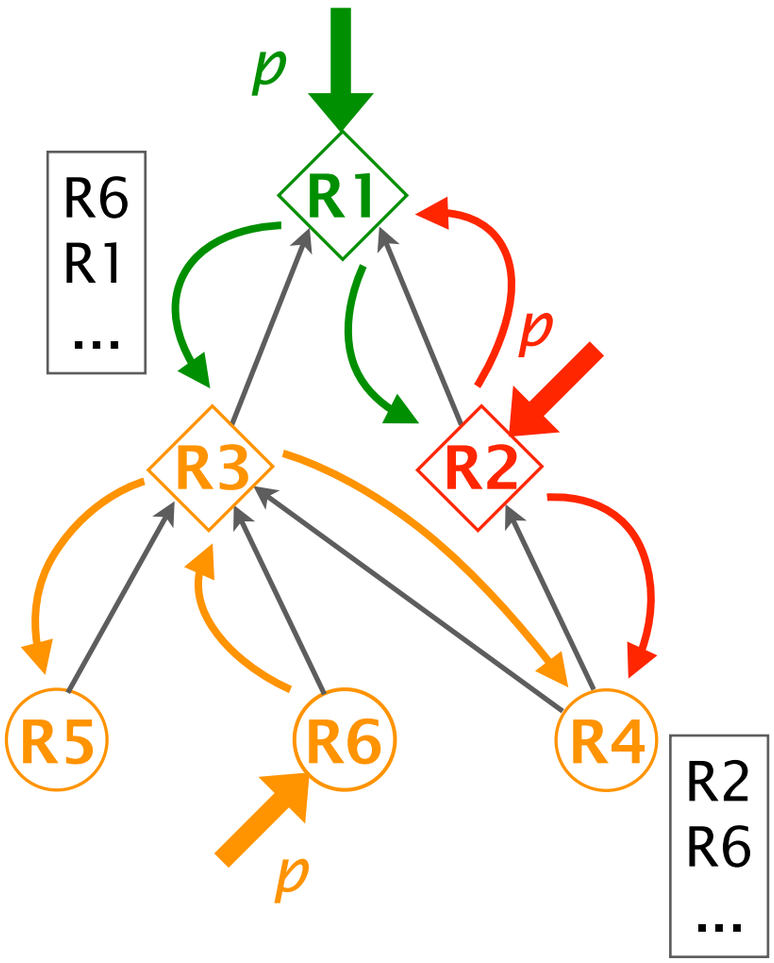


iBGP topology

Since R4 also receives the R2 route directly, it starts using it

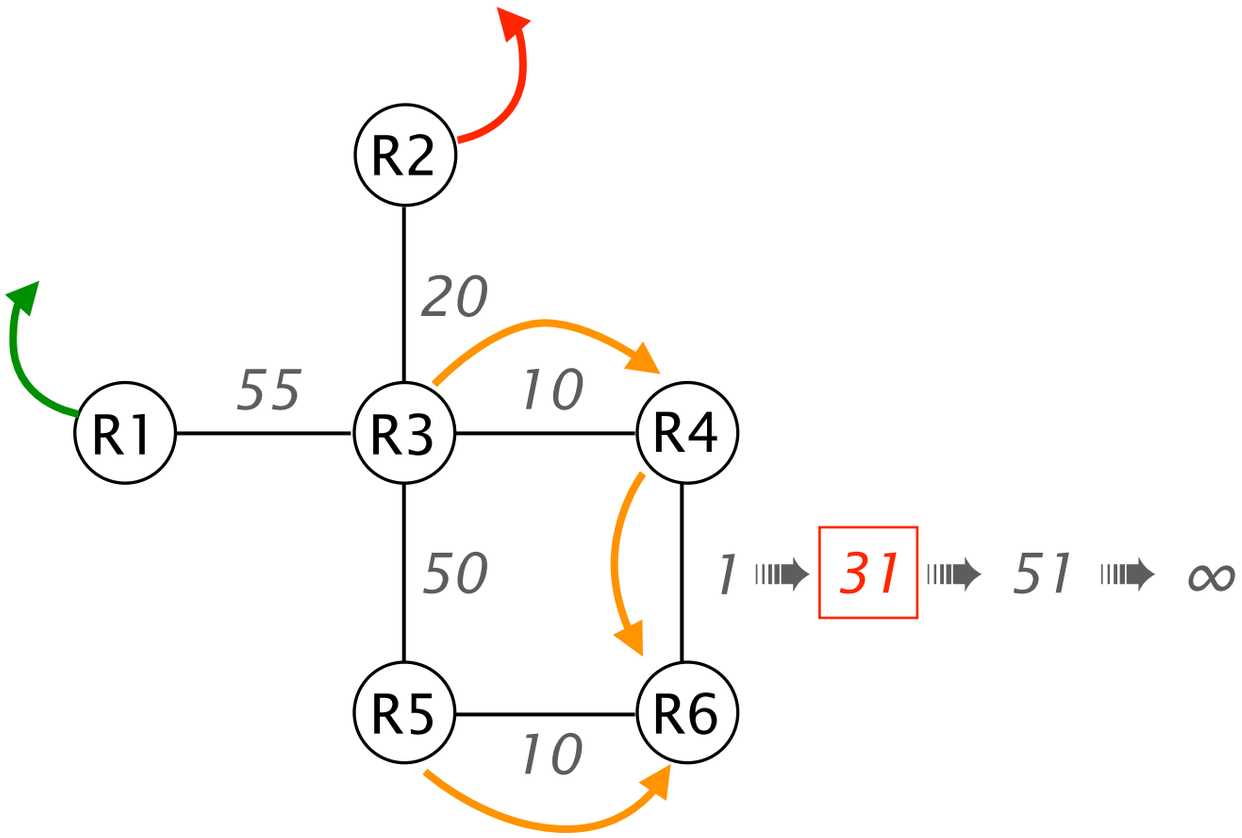


IGP topology

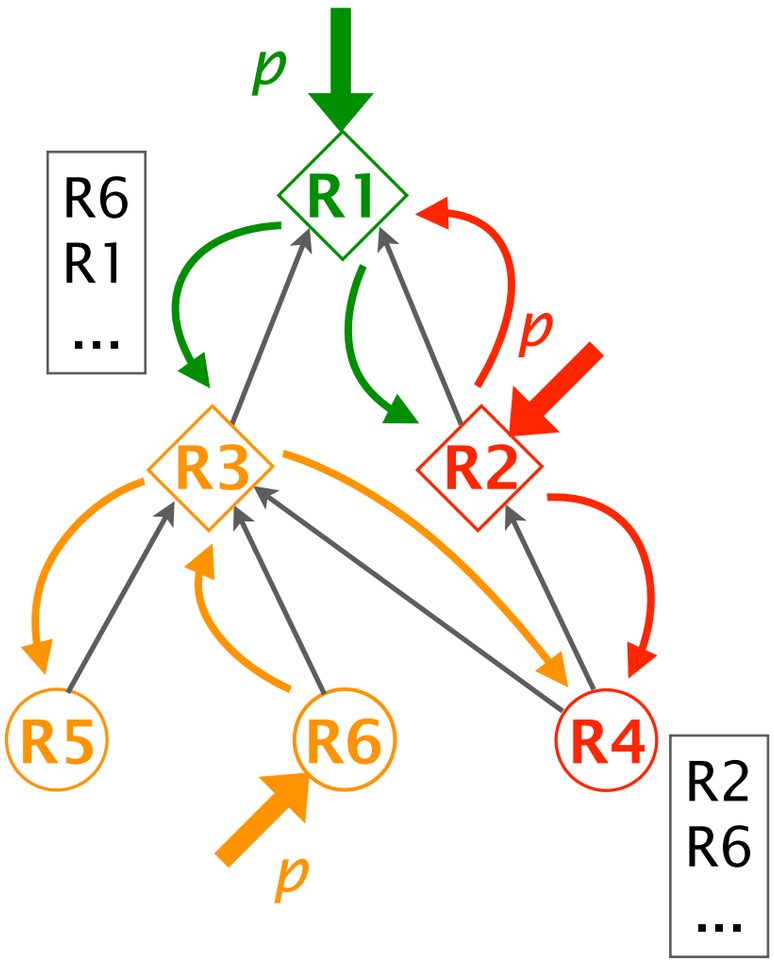


iBGP topology

Since R4 also receives the R2 route directly, it starts using it

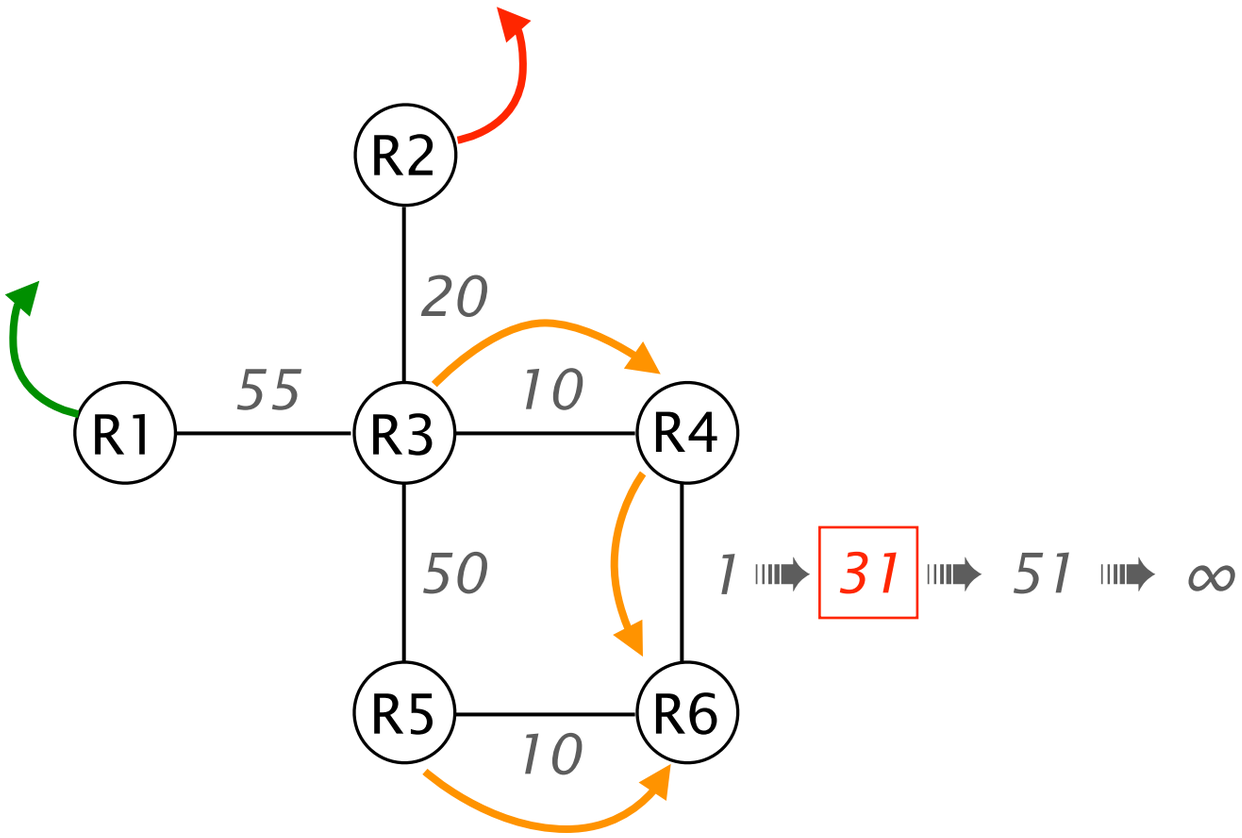


IGP topology

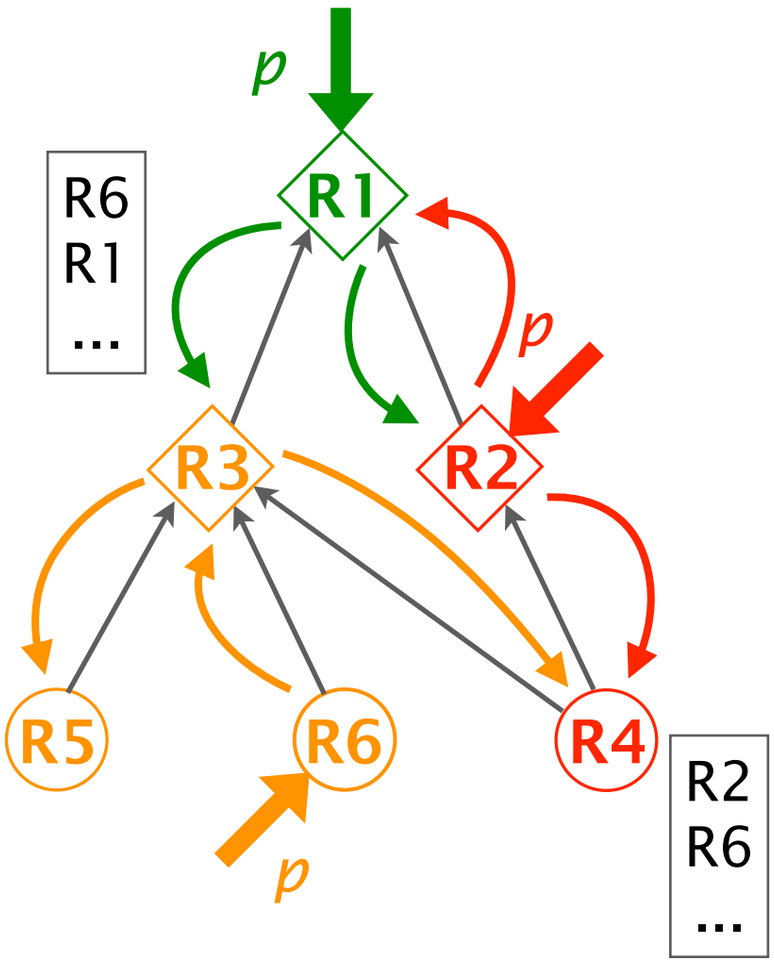


iBGP topology

R3 still prefers R6 (distance 41) to R1 (distance 55)

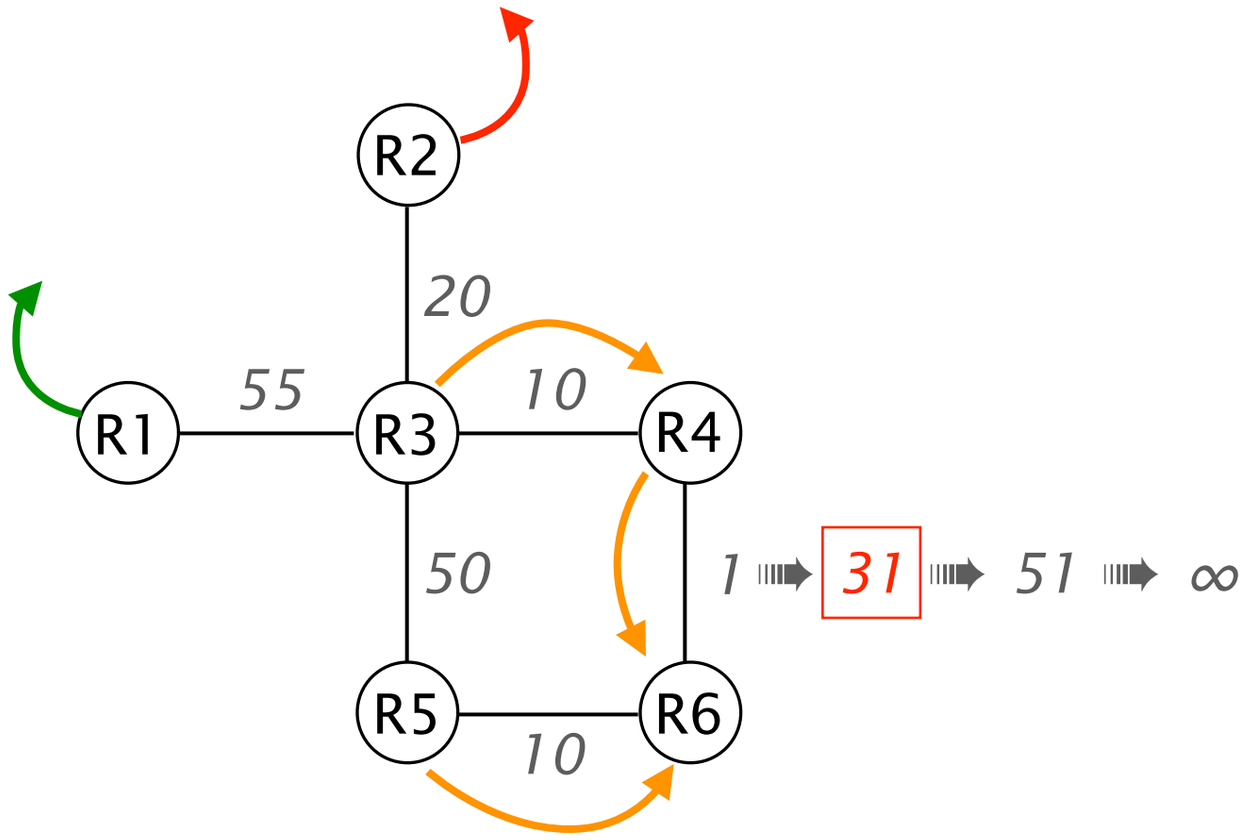


IGP topology

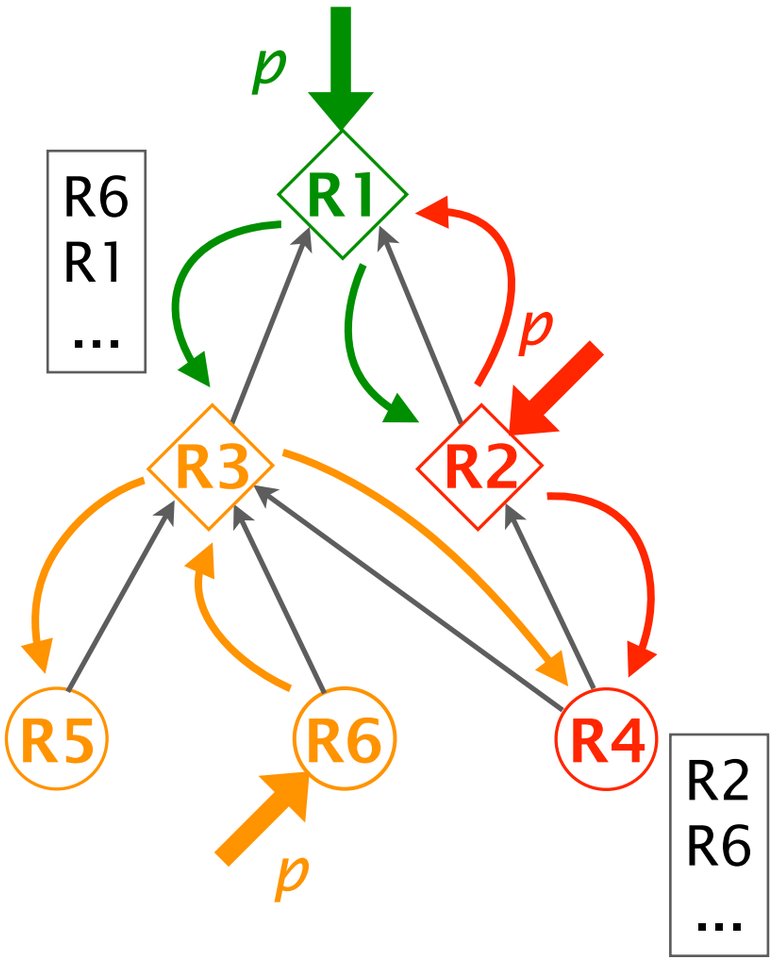


iBGP topology

A forwarding loop is created between R3 and R4
as R4 uses R3 to reach R2

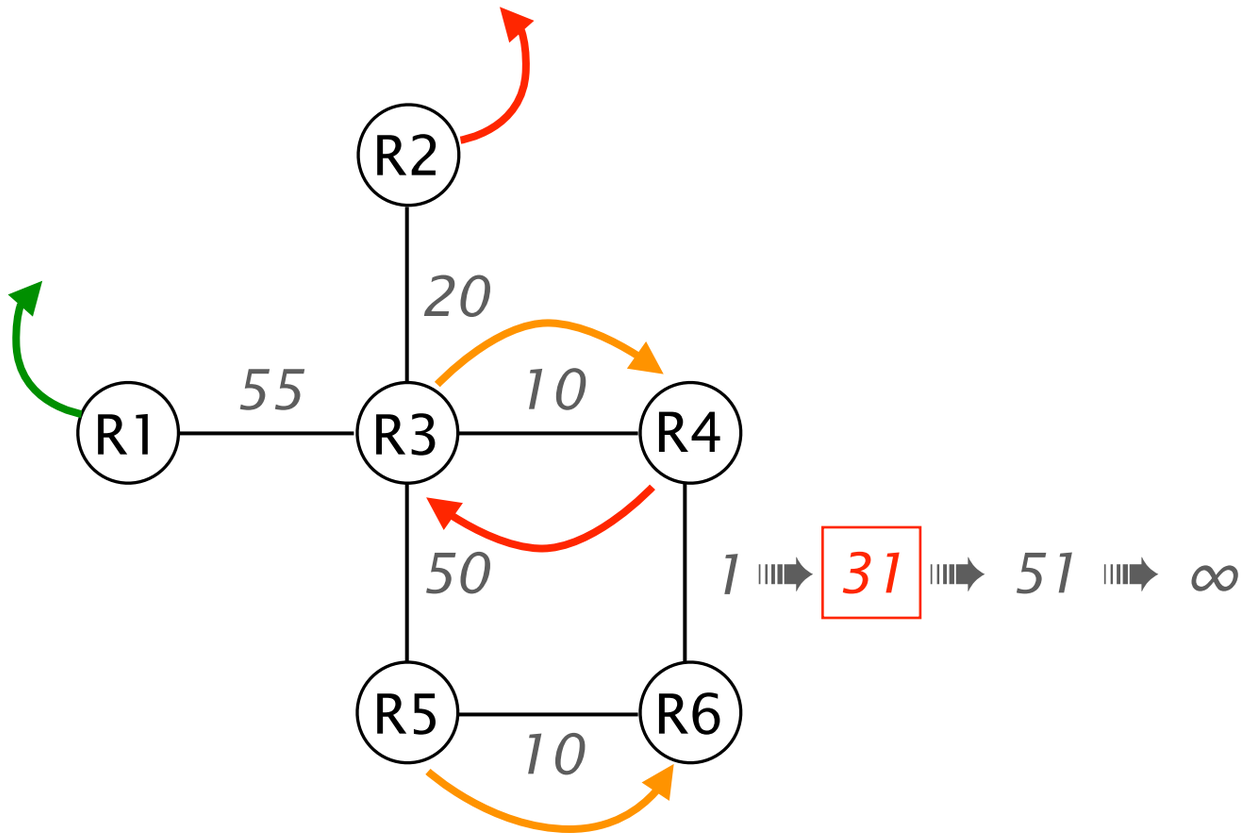


IGP topology

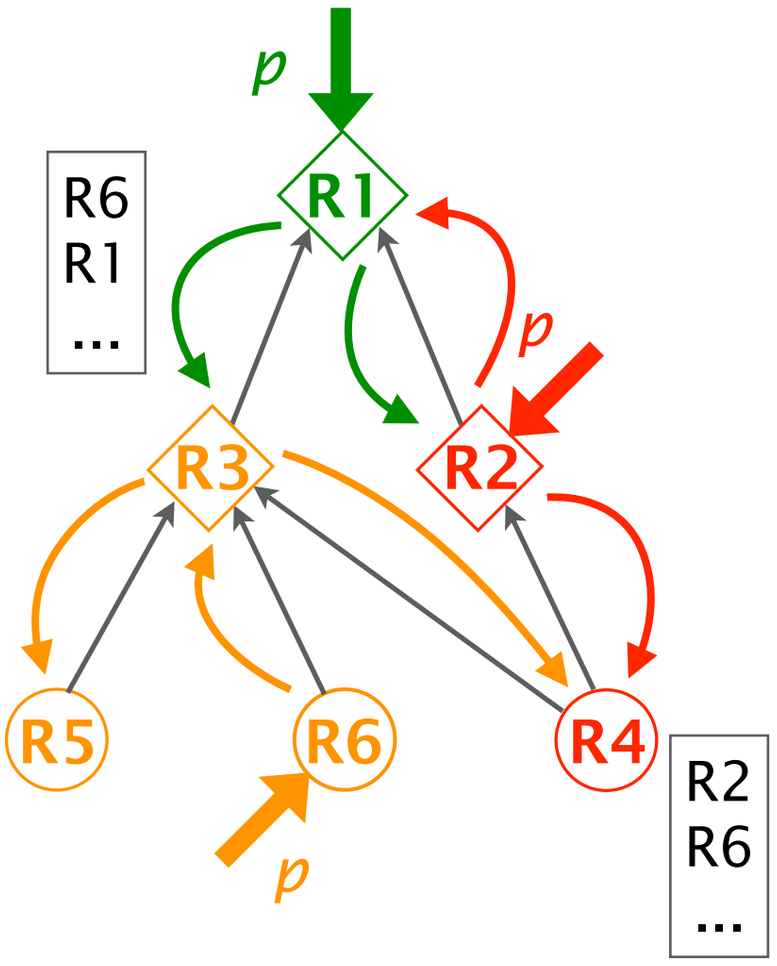


iBGP topology

A forwarding loop is created between R3 and R4
as R4 uses R3 to reach R2

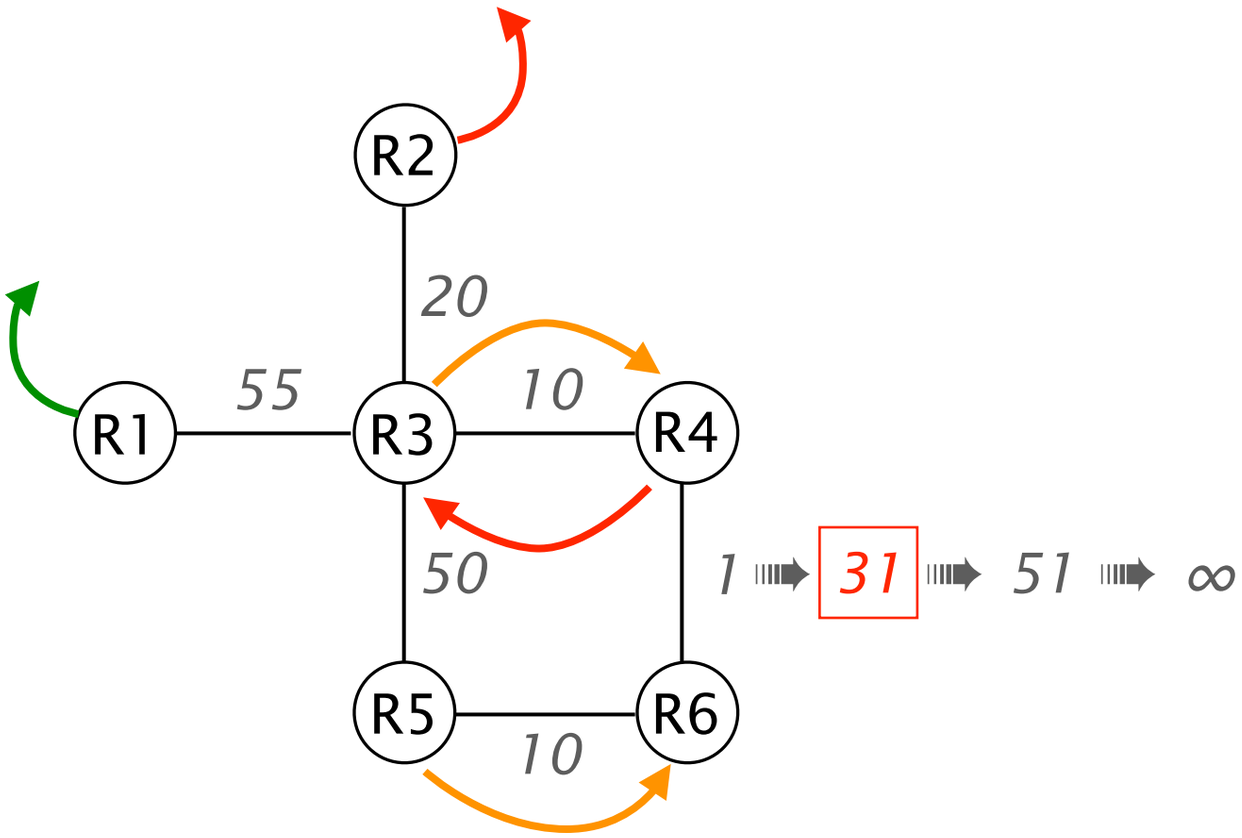


IGP topology

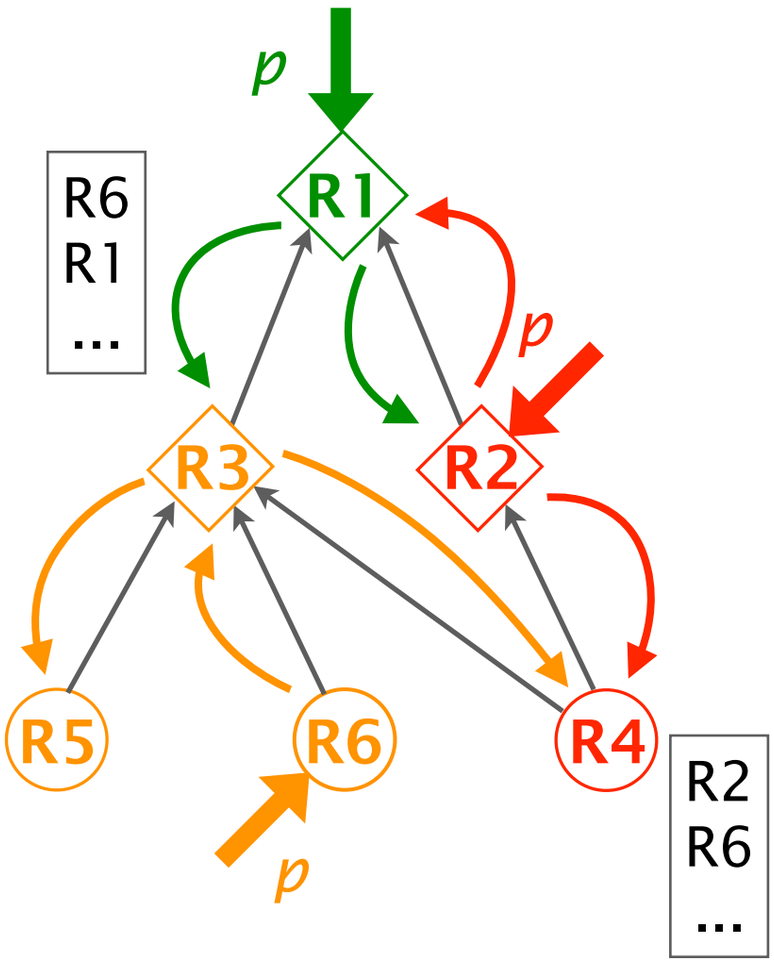


iBGP topology

The loop disappears when we proceed to the second increment

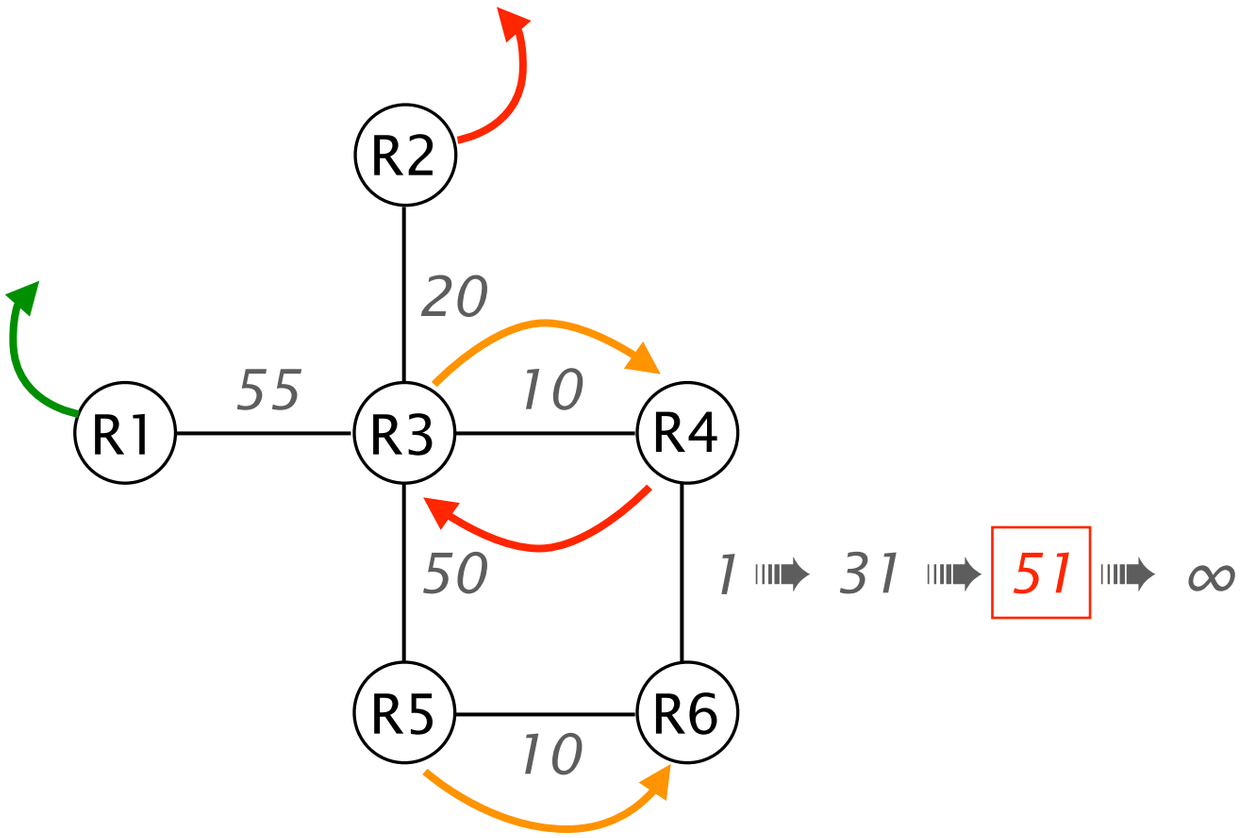


IGP topology

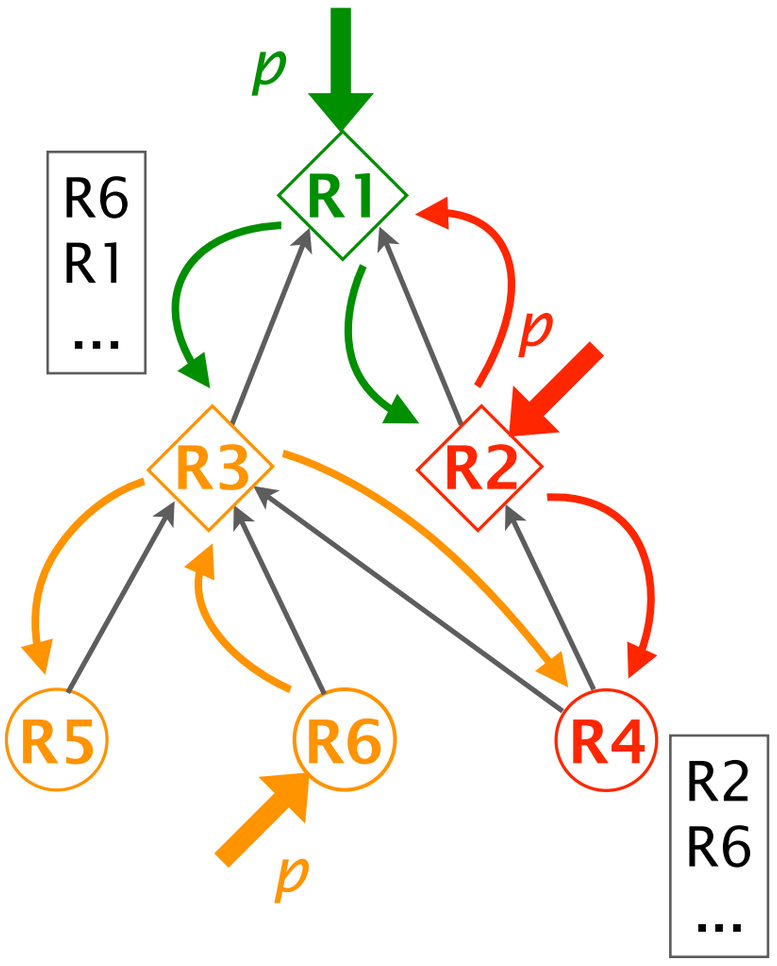


iBGP topology

Let’s now proceed to the second increment

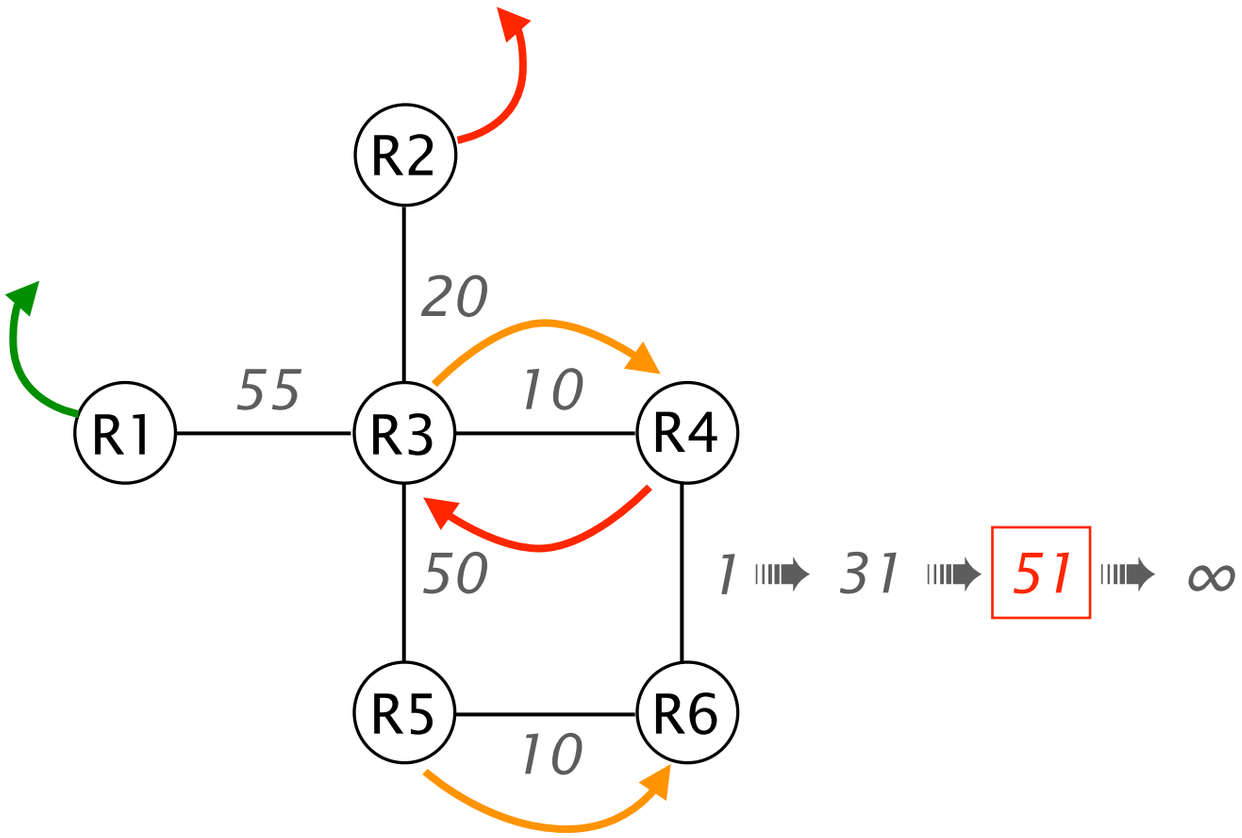


IGP topology

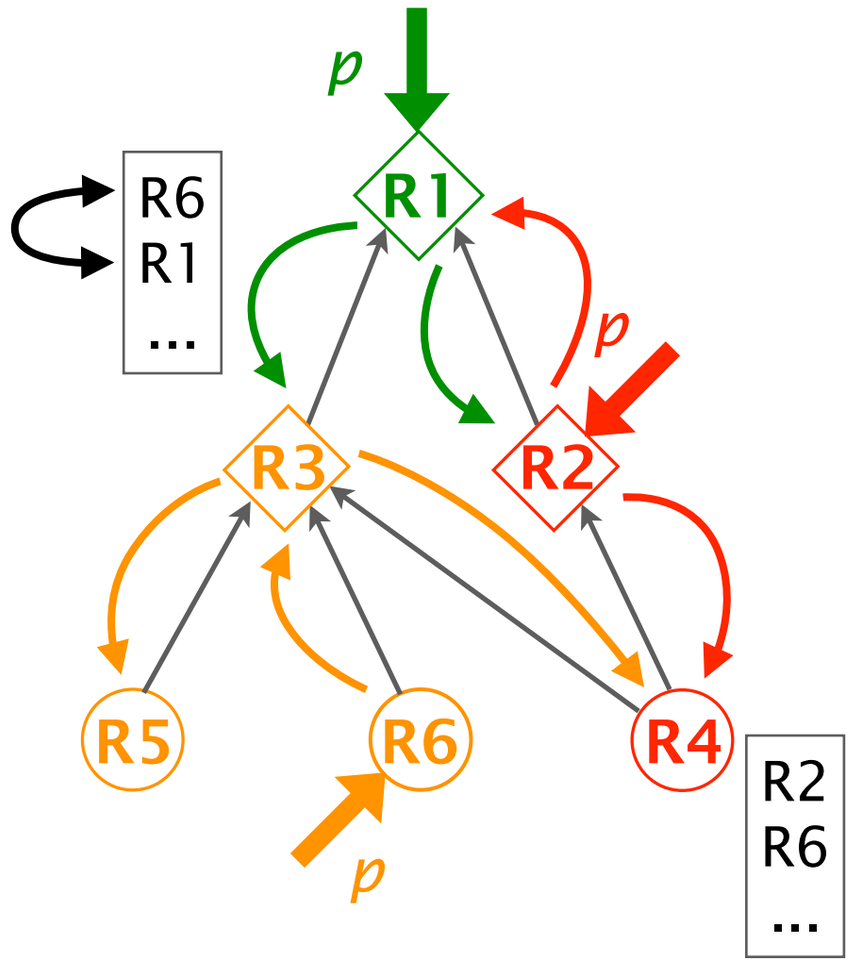


iBGP topology

R3 is now closer to R1 (distance 55) than R6 (distance 60)

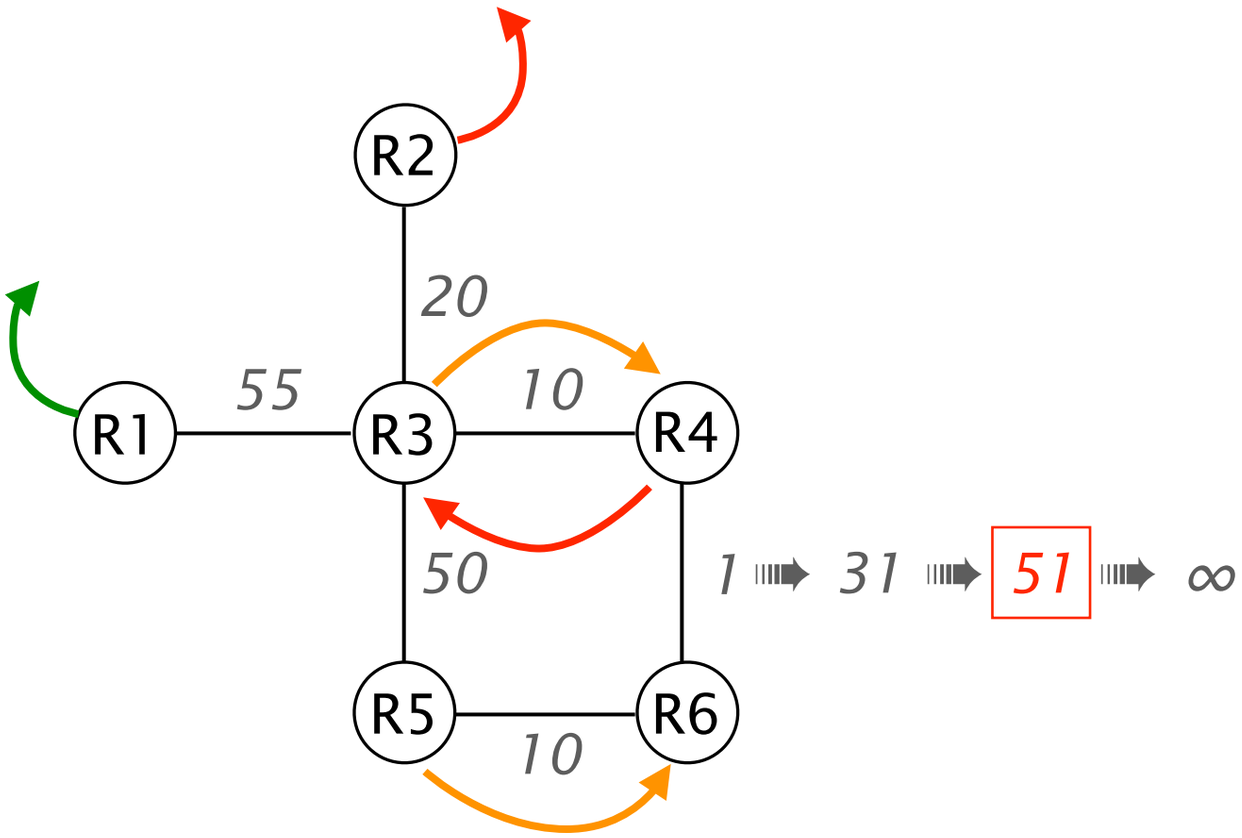


IGP topology

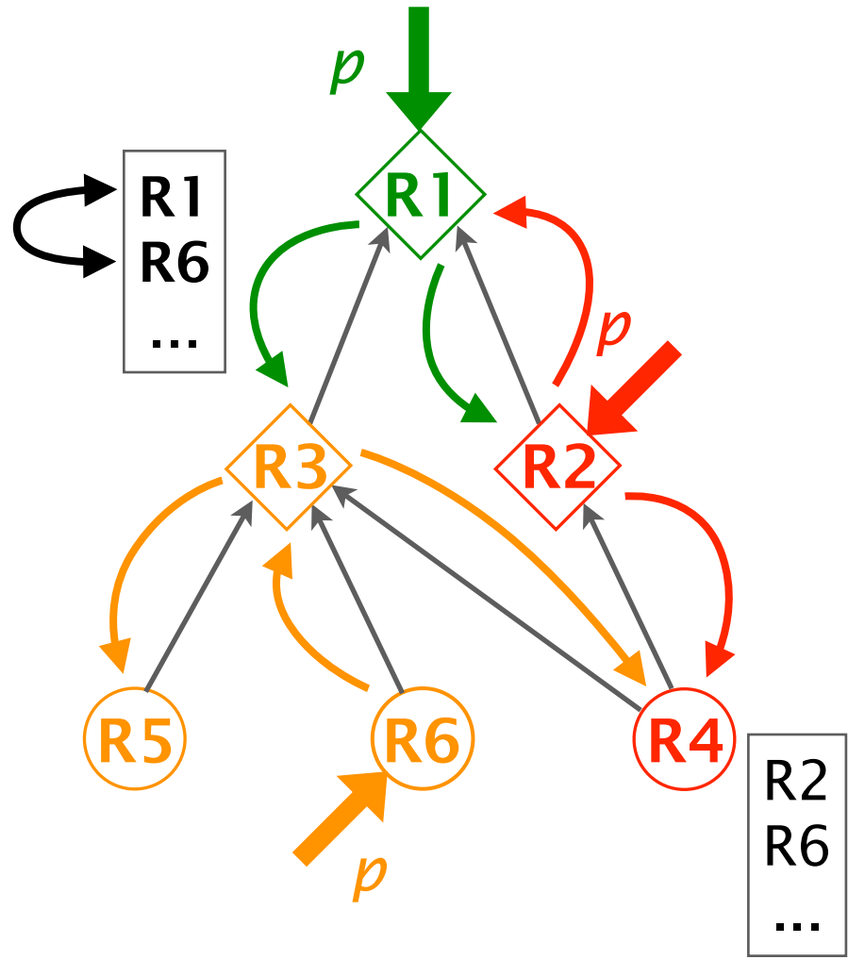


iBGP topology

R3 is now closer to R1 (distance 55) than R6 (distance 60)

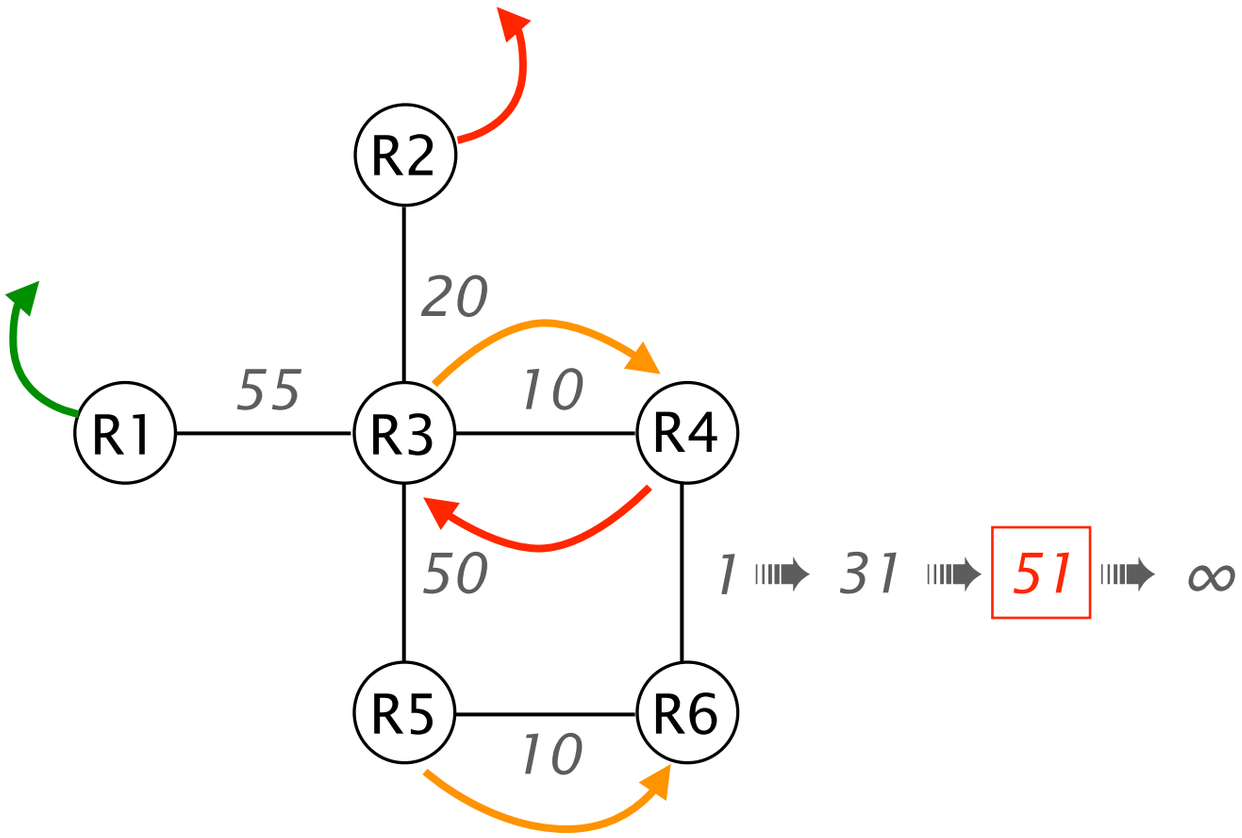


IGP topology

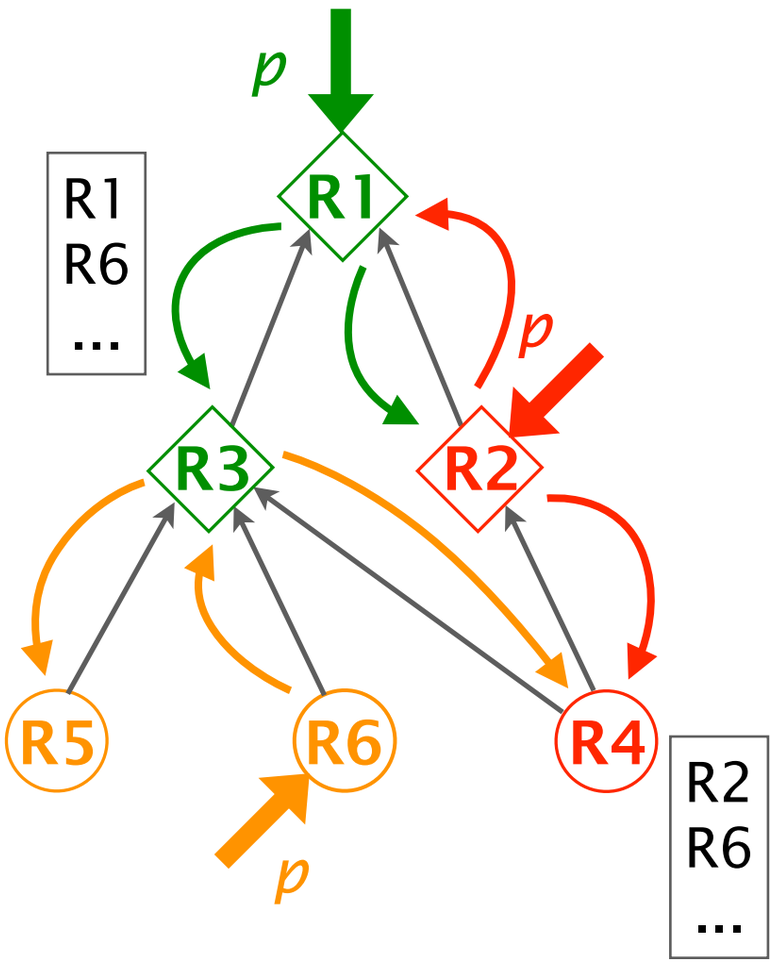


iBGP topology

Since R3 also receives the R1 route directly, it starts using it

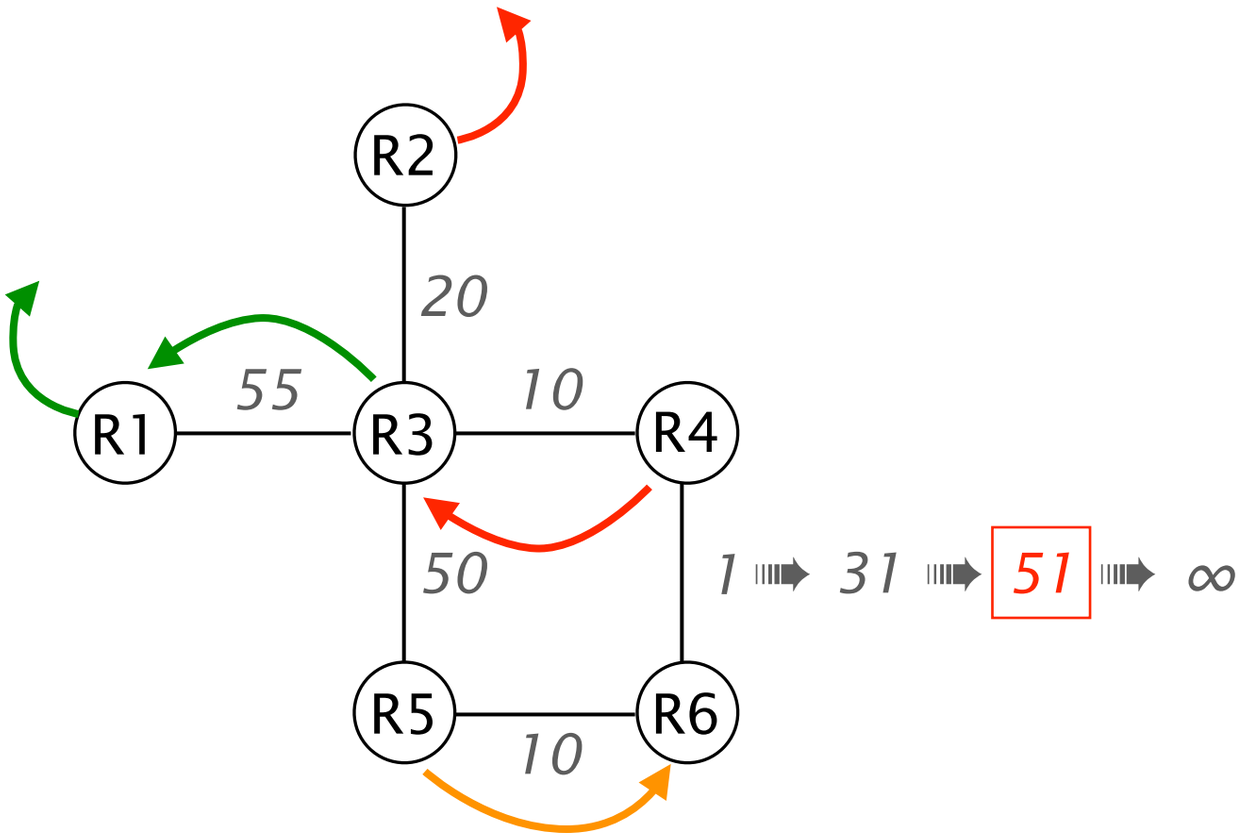


IGP topology

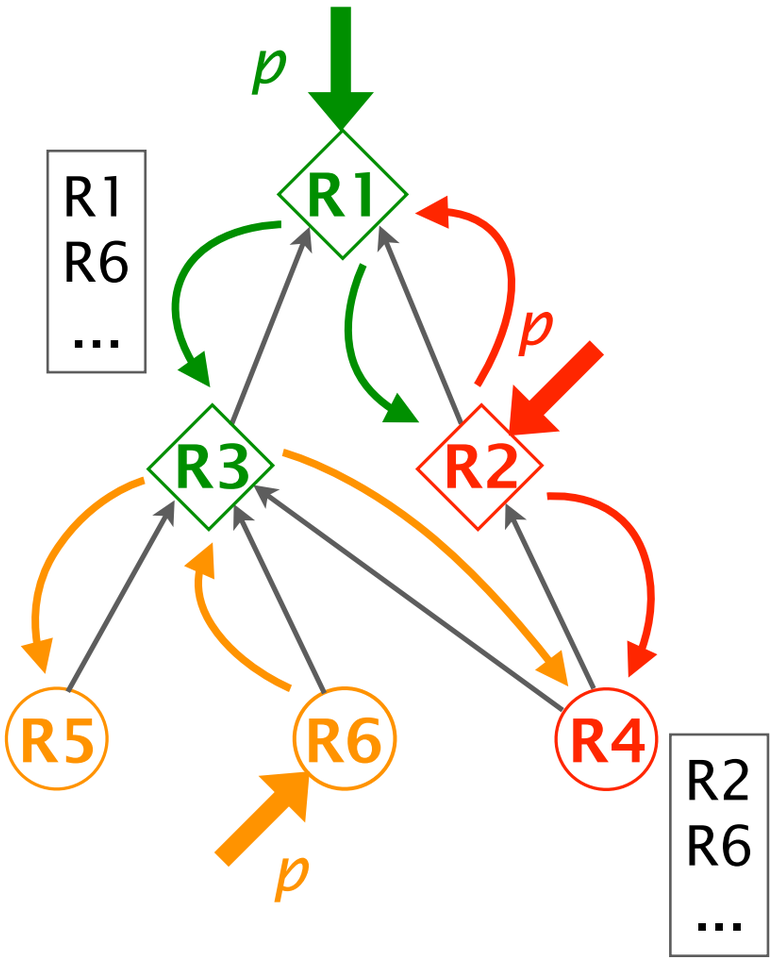


iBGP topology

... which solves the loop

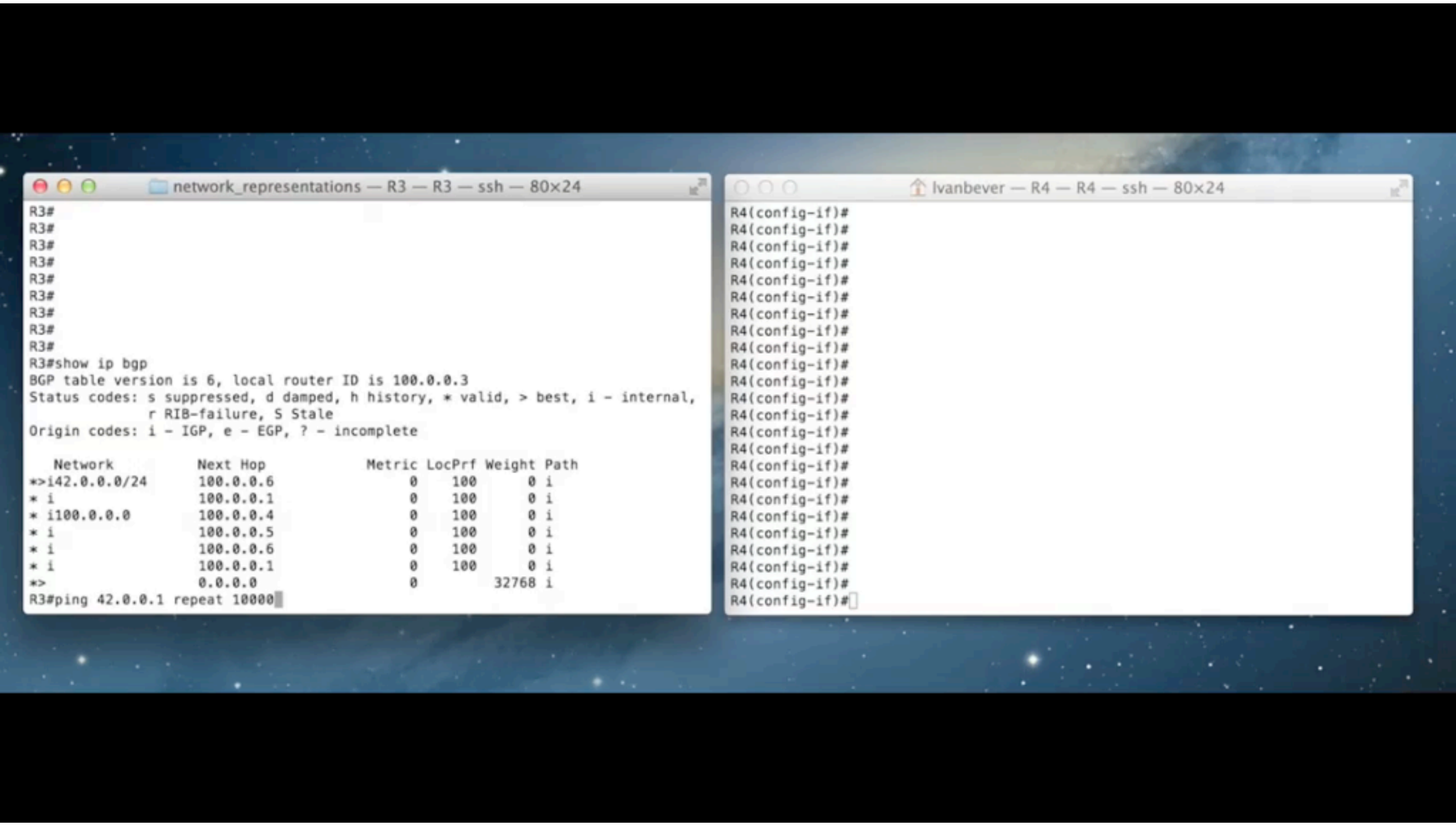


IGP topology



iBGP topology

A BGP-induced loop in the wild



Deciding if reconfiguring the IGP will create BGP anomaly is hard

Problem

Given one iBGP topology and two IGP topologies: *a* and b,

Decide if any IGP reconfiguration from *a* to *b* is free of any BGP anomaly

Deciding if reconfiguring the IGP will create BGP anomaly is hard

Problem

Given one iBGP topology and two IGP topologies: *a* and b,

Decide if any IGP reconfiguration from *a* to *b* is free of any BGP anomaly

This problem is NP-hard

When the cure is worse than the disease:
The impact of graceful IGP operations on BGP

The cure

IGP reconfiguration

The side effects

BGP induced anomalies

3 The solutions

sufficient conditions

Both IGP and BGP safety can be ensured

An IGP reconfiguration will not trigger BGP anomaly *if*

#1 the relative BGP preferences do not change

since no BGP router will change its decision

Both IGP and BGP safety can be ensured

An IGP reconfiguration will not trigger BGP anomaly *if*

#1 the relative BGP preferences do not change

since no BGP router will change its decision

#2 the BGP configuration complies with the two known sufficient conditions for ensuring routing correctness

the “prefer-client” and the “no-spurious OVER” conditions

Both IGP and BGP safety can be ensured

An IGP reconfiguration will not trigger BGP anomaly *if*

#1 the relative BGP preferences do not change

since no BGP router will change its decision

#2 the BGP configuration complies with the two known sufficient conditions for ensuring routing correctness

the “prefer-client” and the “no-spurious OVER” conditions

#3 an encapsulation mechanism is used for forwarding

as only one IP lookup is performed within the network

When the cure is worse than the disease:
The impact of graceful IGP operations on BGP

The cure

IGP reconfiguration

The side effects

BGP induced anomalies

The solutions

sufficient conditions

For truly safe network reconfiguration,
the entire protocol stack must be considered

IGP reconfiguration techniques can create BGP anomalies
leading to more disruption than the one they aim to avoid

Guaranteeing BGP safety is hard, in the general case
sufficient conditions exist, for particular cases

Decoupling BGP from the IGP solves the problem
but require protocol changes

When the cure is worse than the disease: The impact of graceful IGP operations on BGP

Laurent Vanbever
www.vanbever.eu

IEEE INFOCOM

April 18, 2013