iBGP Deceptions: More Sessions, Fewer Routes

Laurent Vanbever
University of Louvain, Louvain-la-Neuve, Belgium
laurent.vanbever@uclouvain.be

Joint work with:
Stefano Vissicchio (Roma Tre University),
Luca Cittadini (Roma Tre University)
Olivier Bonaventure (University of Louvain)

IEEE INFOCOMM'12
Thursday, March 29 2012

Breaking News

Adding a single iBGP session can disrupt iBGP ability of distributing routing information

iBGP Deceptions:
More Sessions, Fewer Routes

Introduction and Motivation

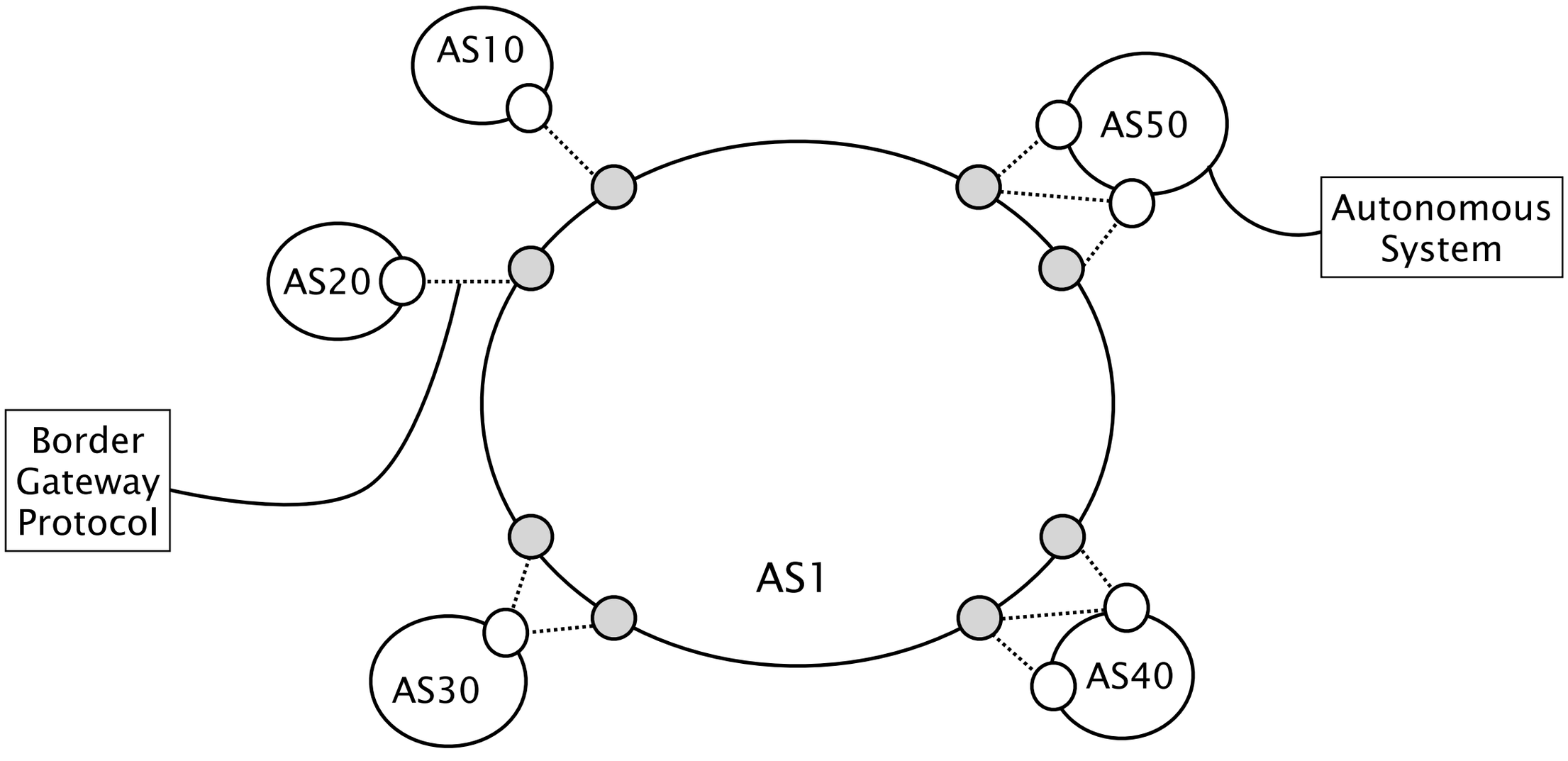
Dissemination correctness

Revisiting the state-of-the-art

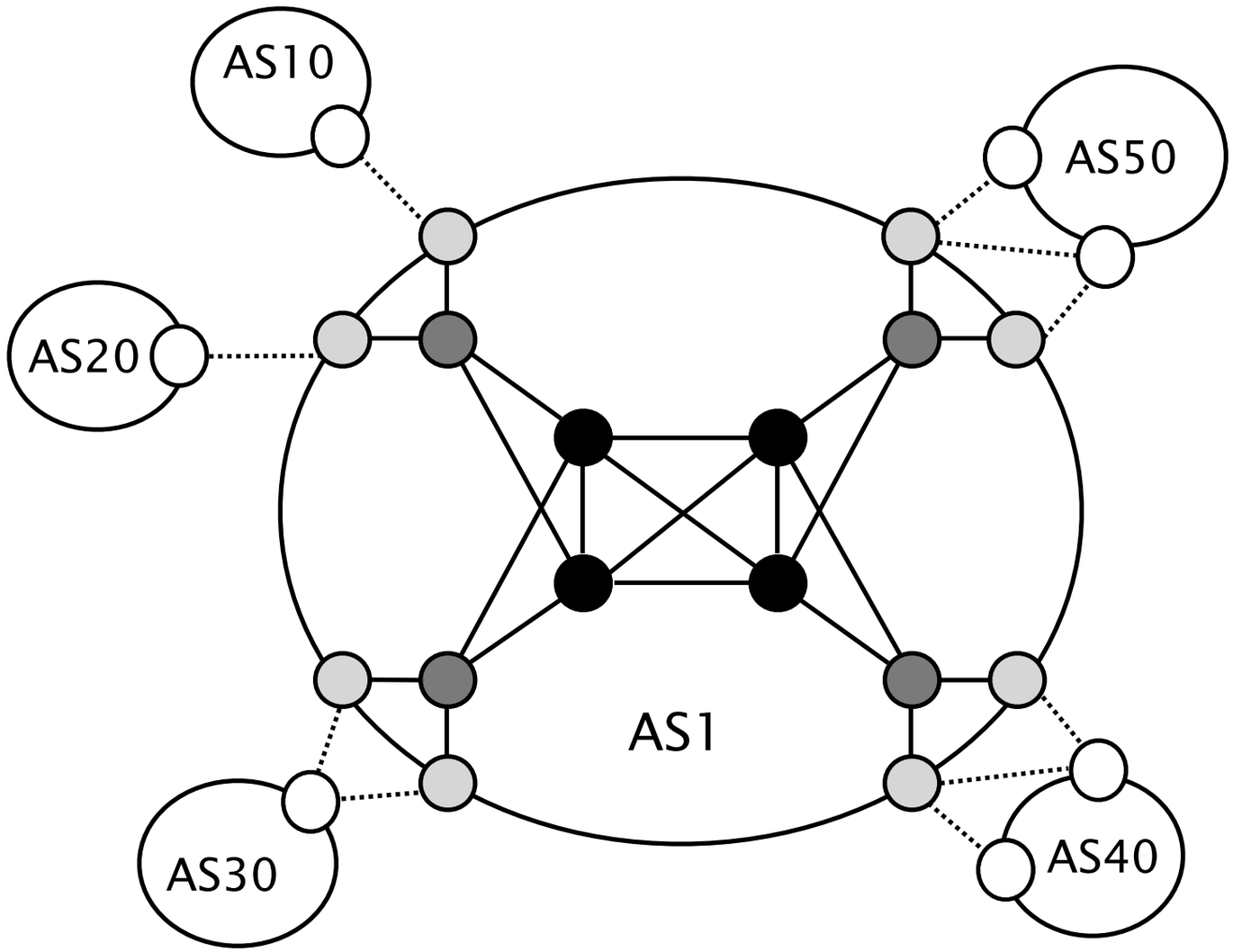
Conclusion

iBGP Deceptions:
More Sessions, Fewer Routes

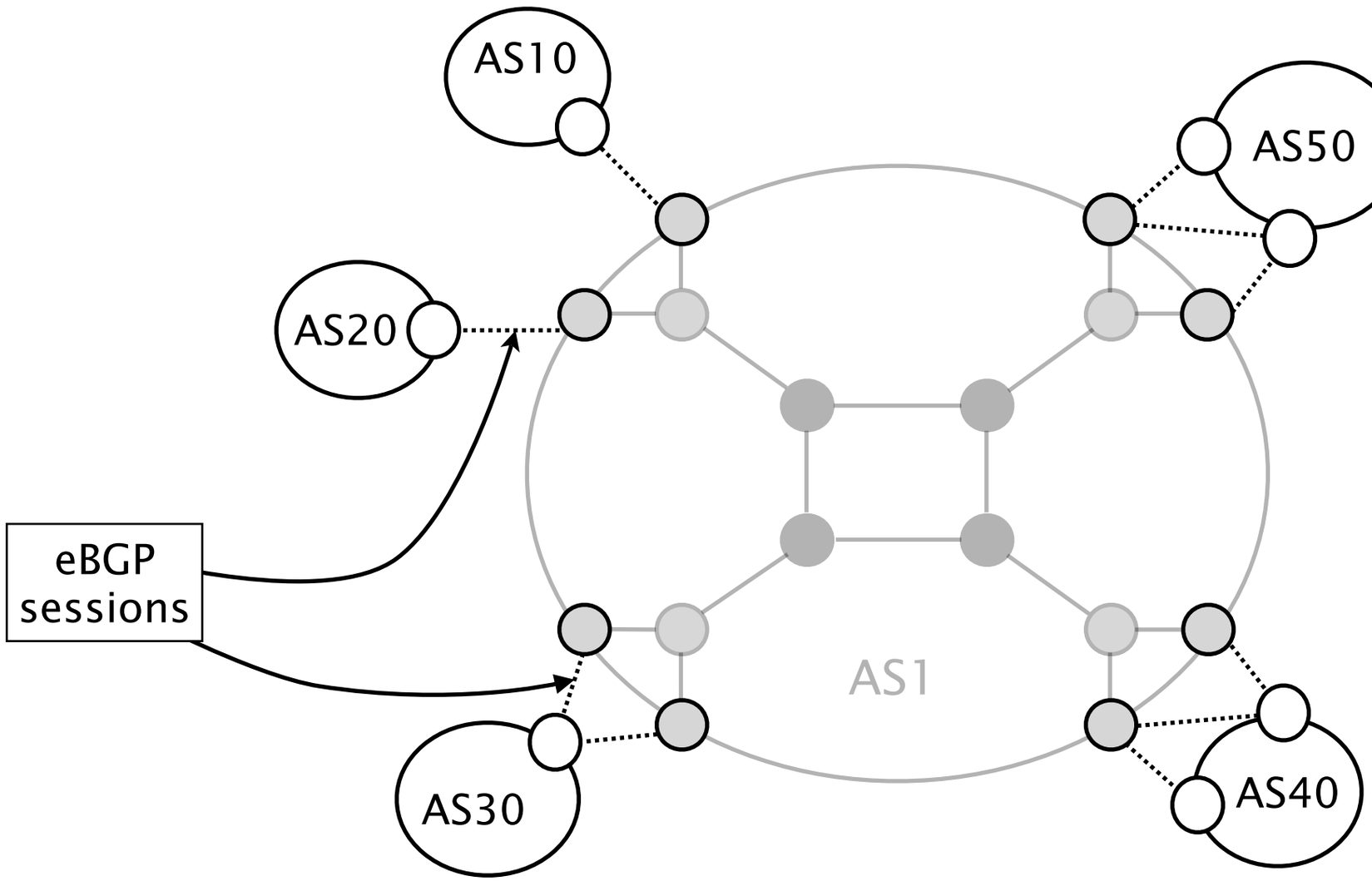
BGP is *the* inter-domain routing protocol used today



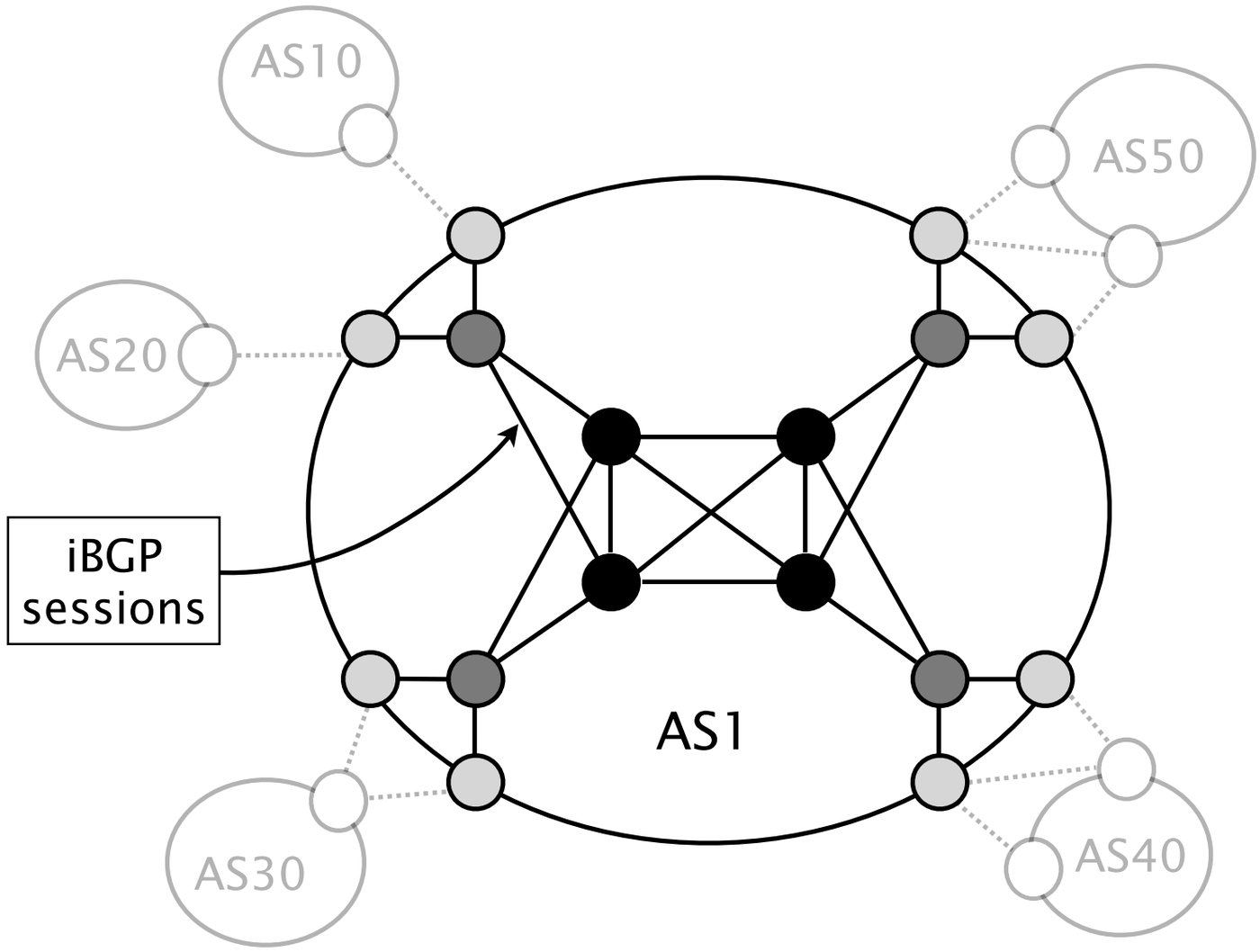
BGP comes in two flavors



external BGP (eBGP) exchanges reachability information between ASes

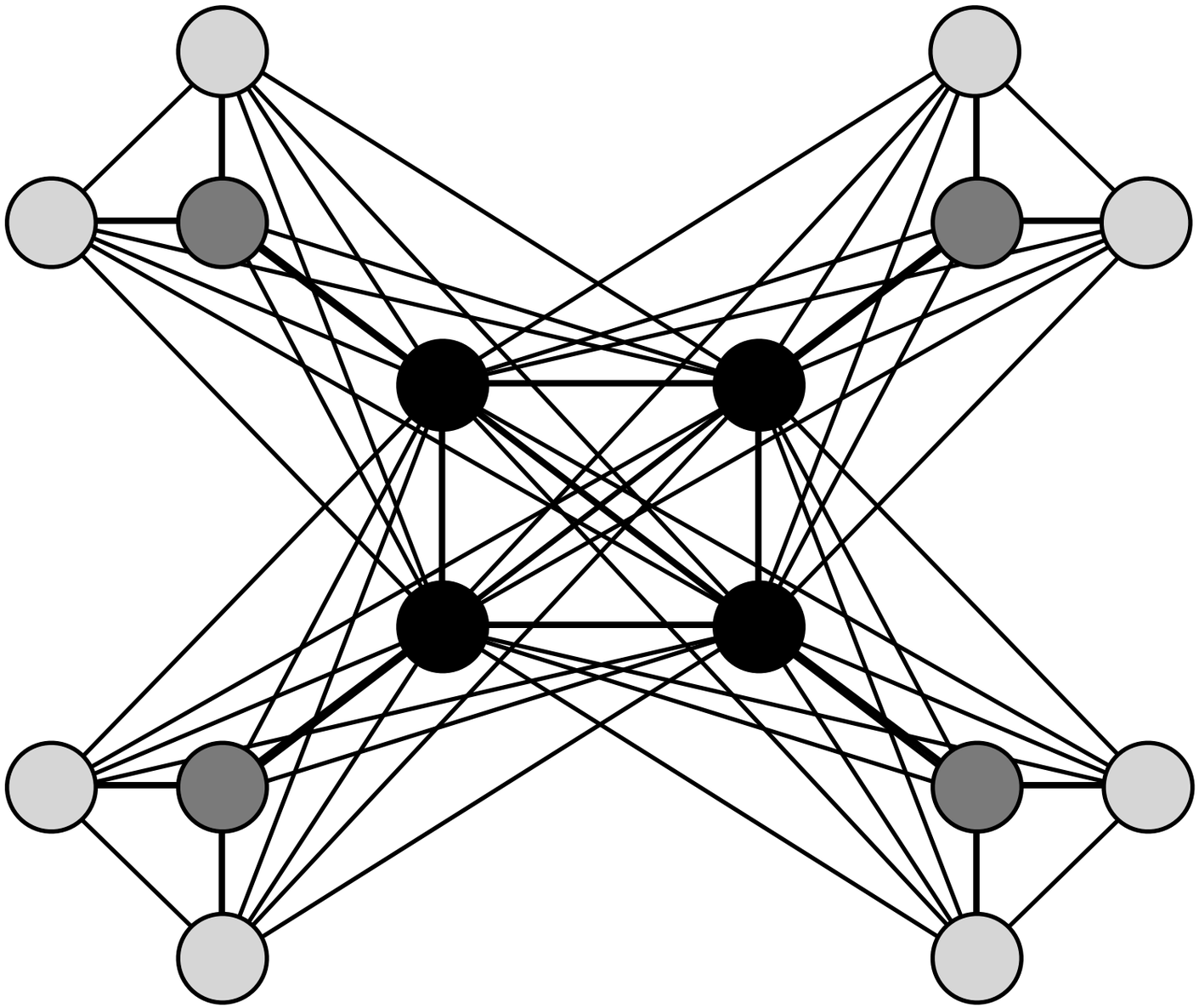


internal BGP (iBGP) distributes externally learned routes within the AS



In this talk, we take the perspective of a single AS and focus on iBGP

Originally, updates cannot be forwarded, mandating a full-mesh of iBGP sessions

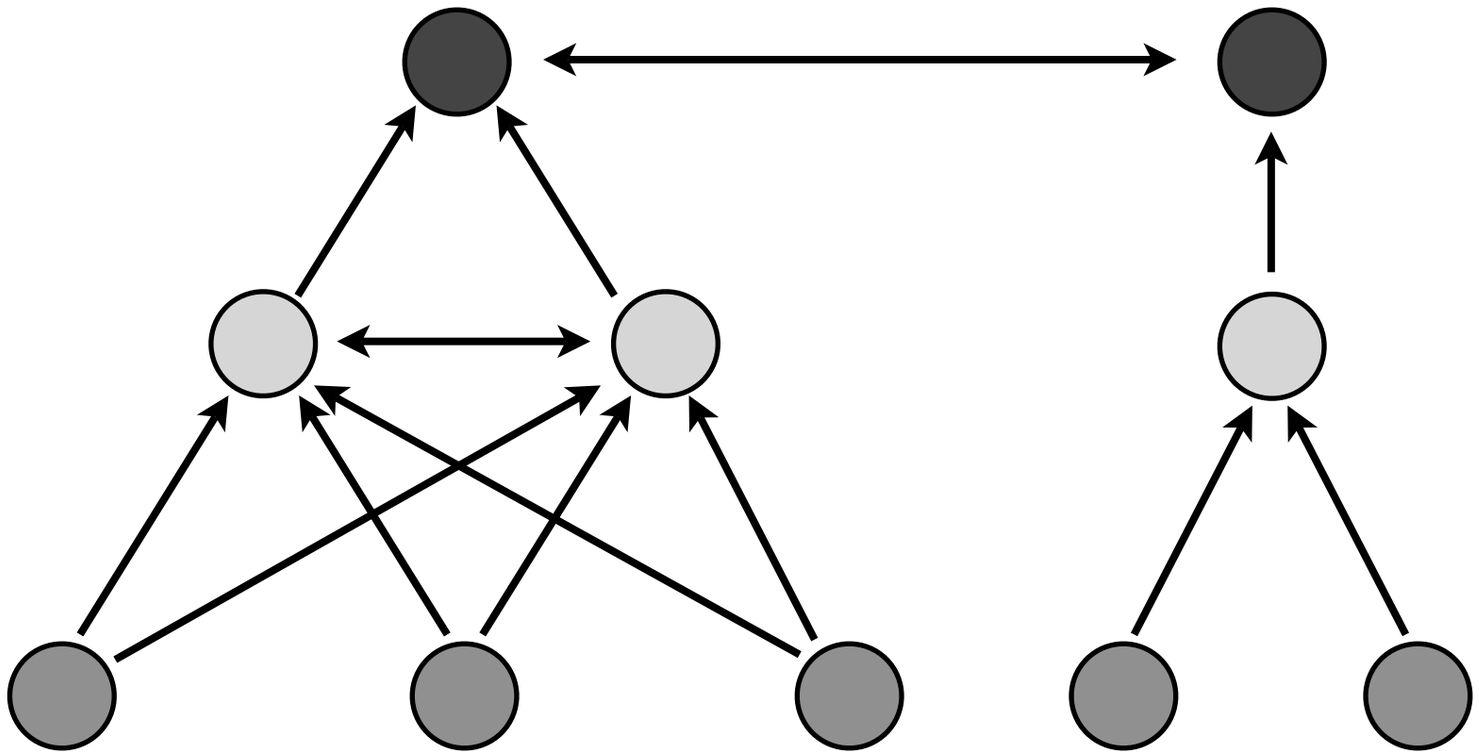


$O(n^2)$ iBGP sessions where *n* is the number of routers

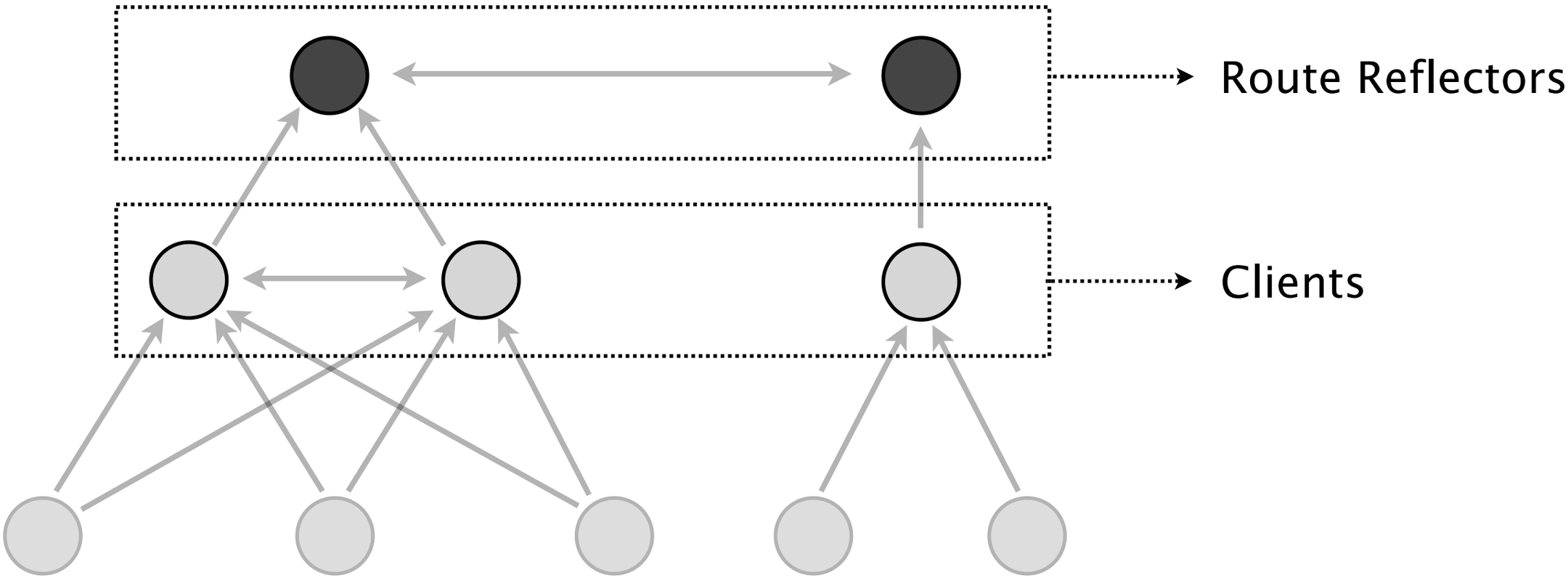
... quickly becomes totally *unmanageable*

Fair warning: some sessions are missing

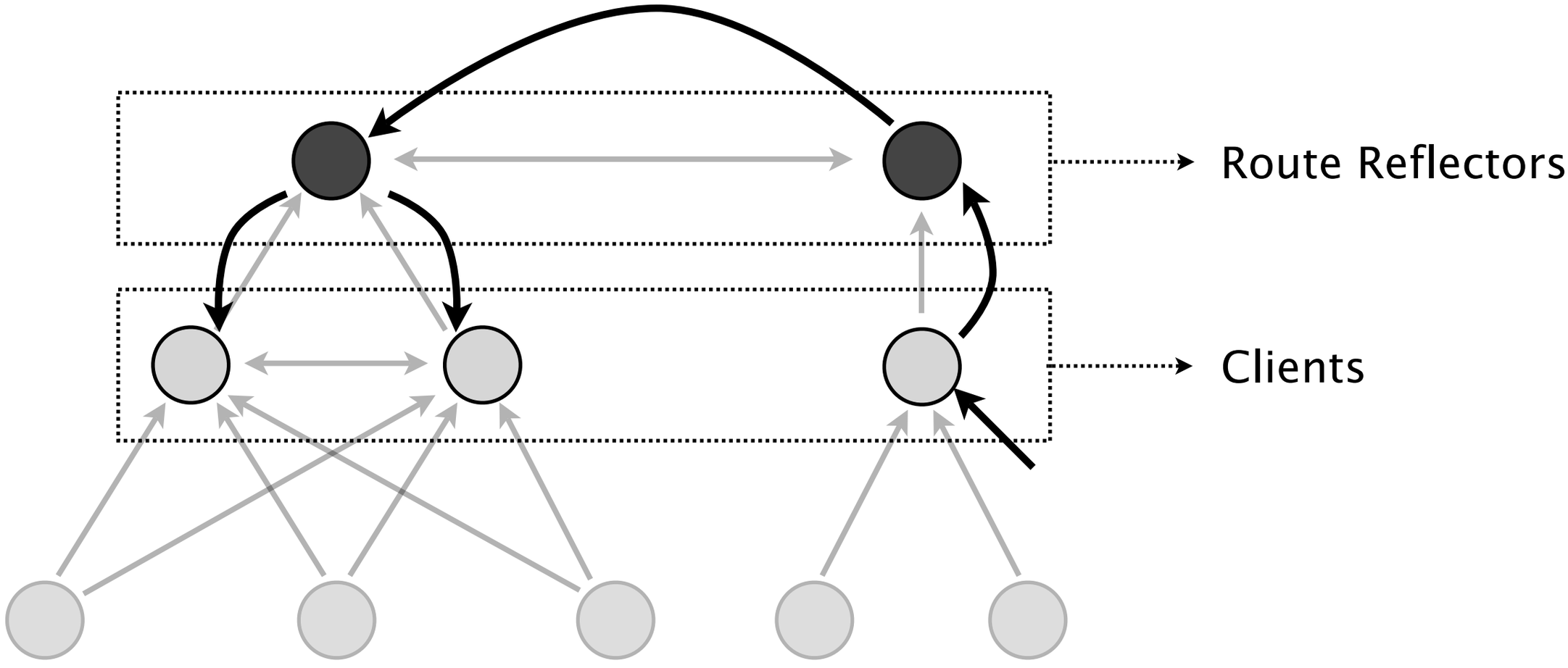
Thanks to Route Reflection, iBGP routers can be organized in a hierarchy



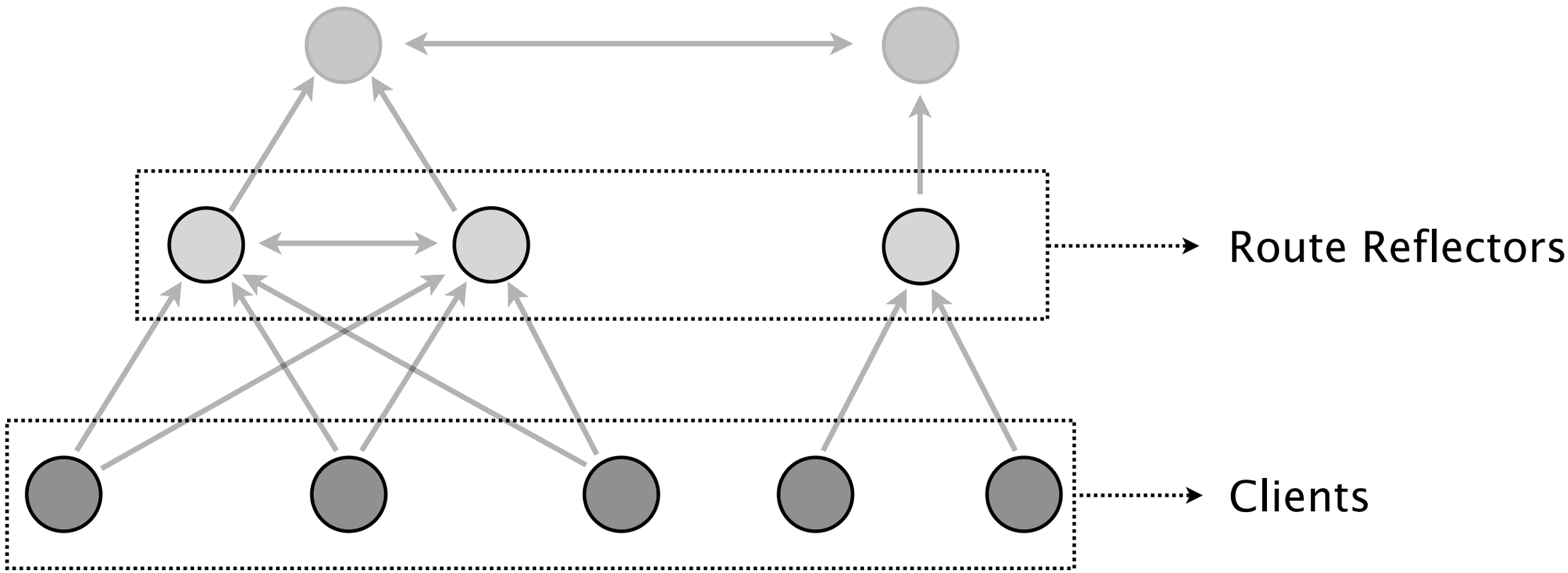
Route Reflector are allowed to forward updates between iBGP peers



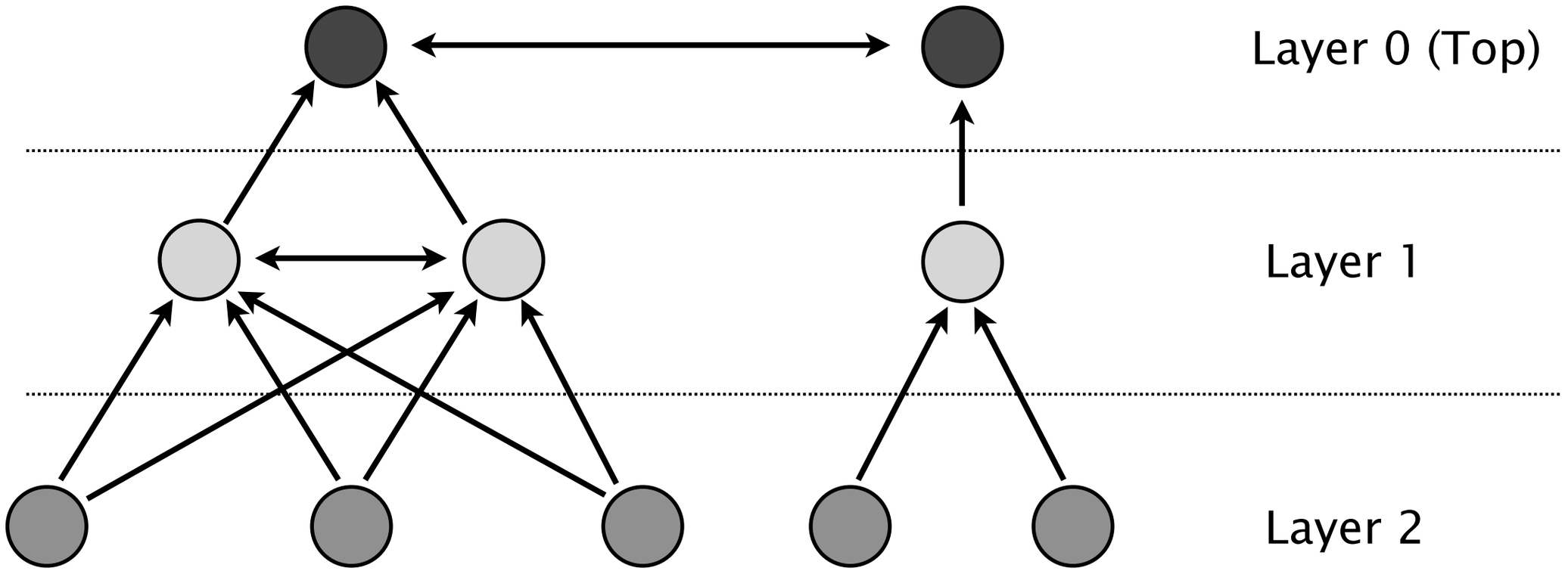
Route Reflector reflects updates between iBGP peers



Several layers of Route Reflection can be built

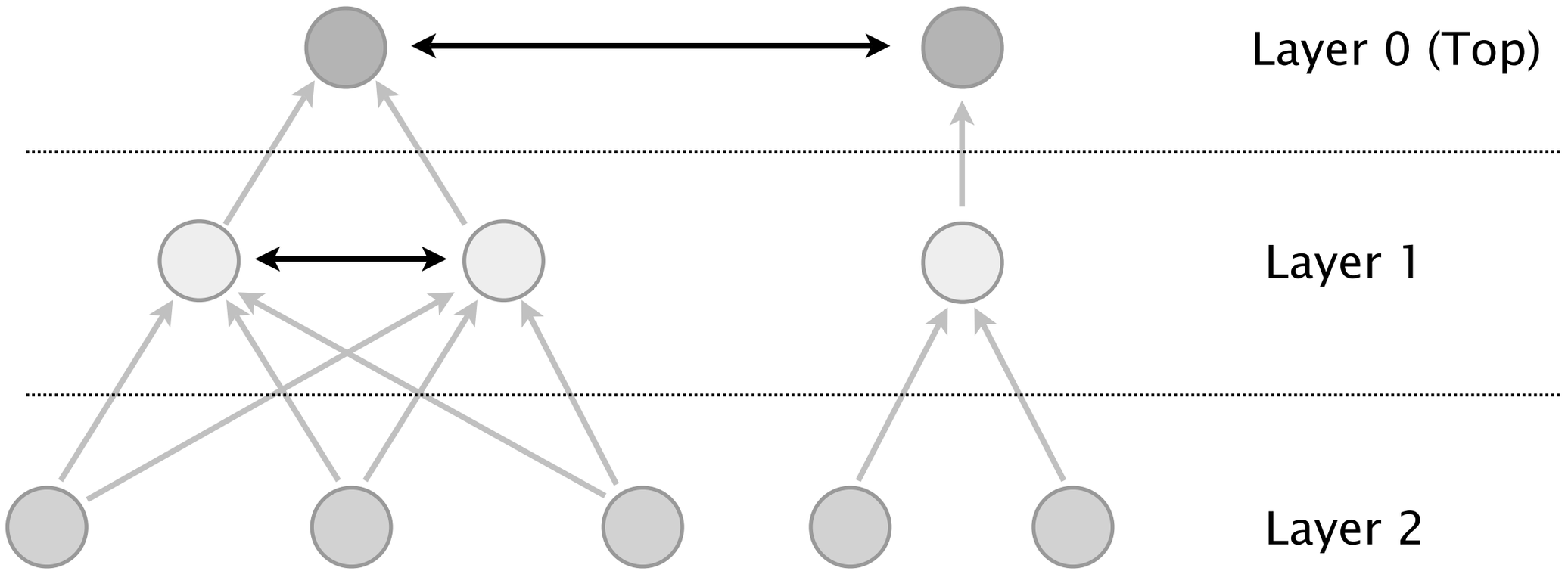


Several layers of Route Reflection can be built

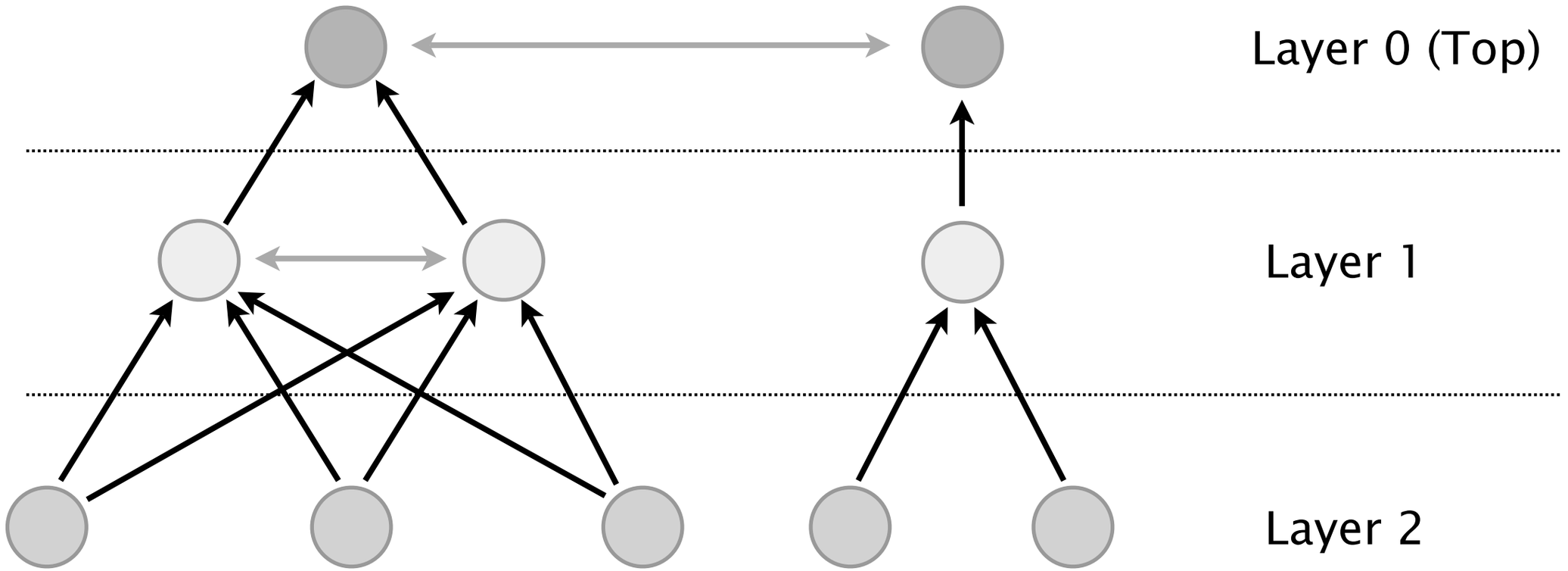


Lower layers rely on upper layers to propagate and receive routing information

OVER sessions connects iBGP peers

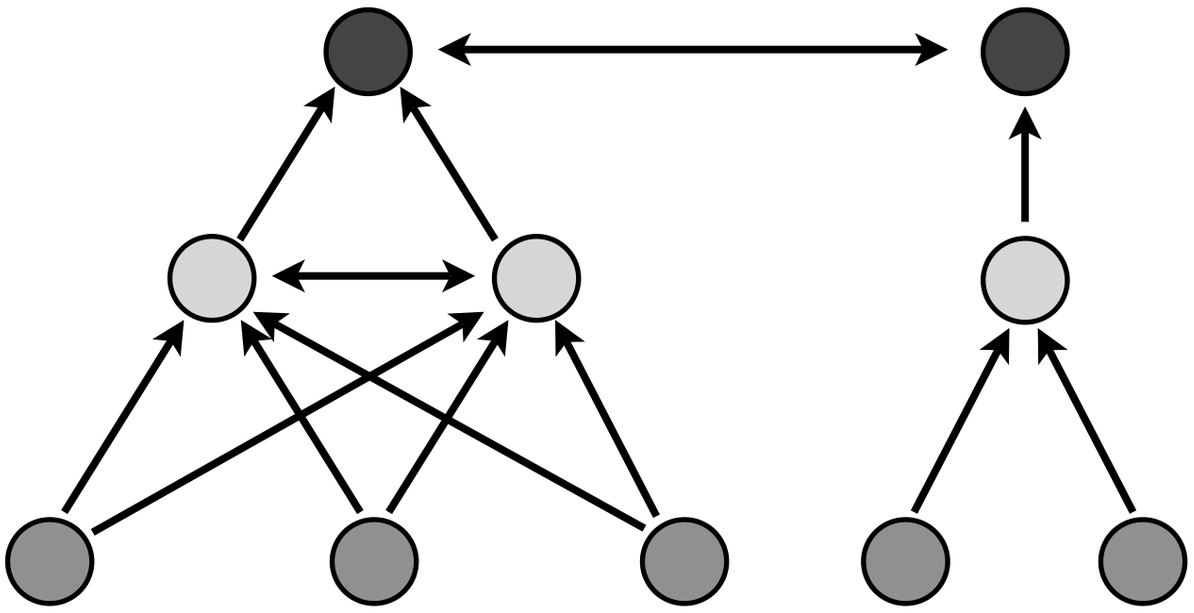


UP/DOWN sessions connect a Route Reflector to its client(s)



Best routes are allowed to flow on valid signaling paths only

Valid signaling path match the UP* OVER? DOWN* regular expression

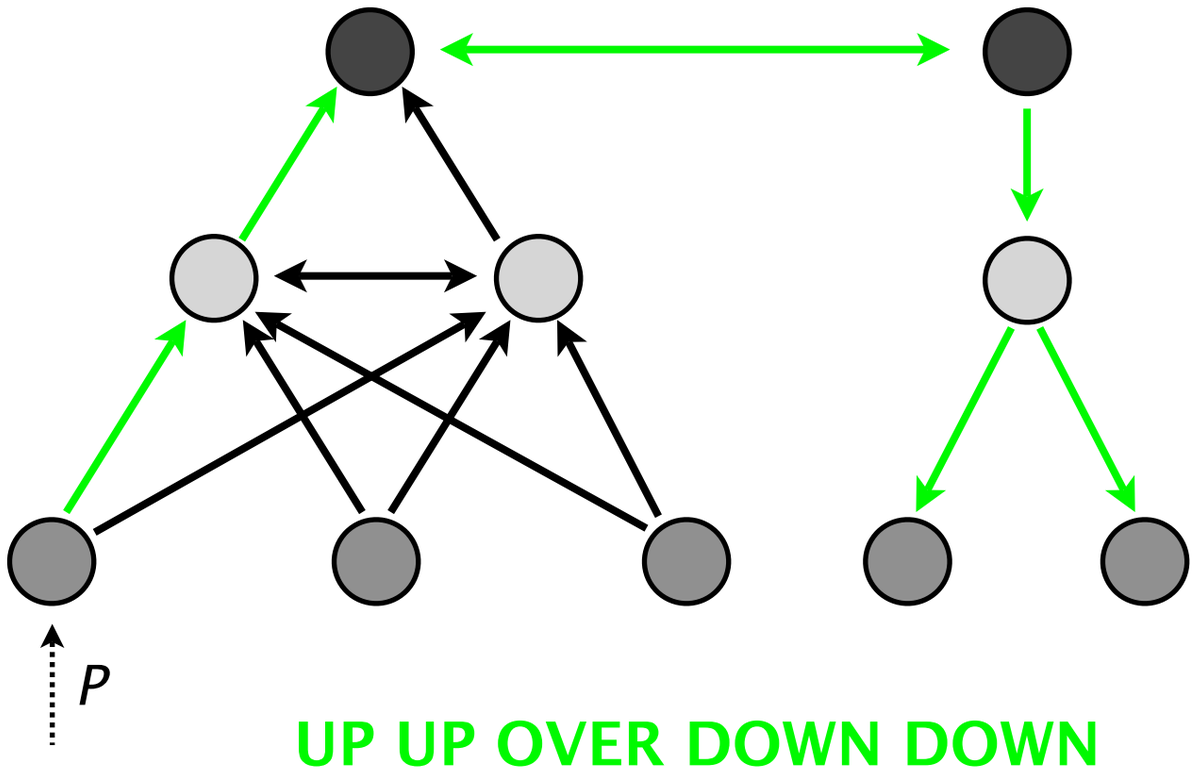


BGP Propagation rules

	To client	*To peer/RR*
From client	✓	✓
From peer/RR	✓	✗

Routes are allowed to flow on valid signaling paths only

Valid signaling path match the UP* OVER? DOWN* regular expression



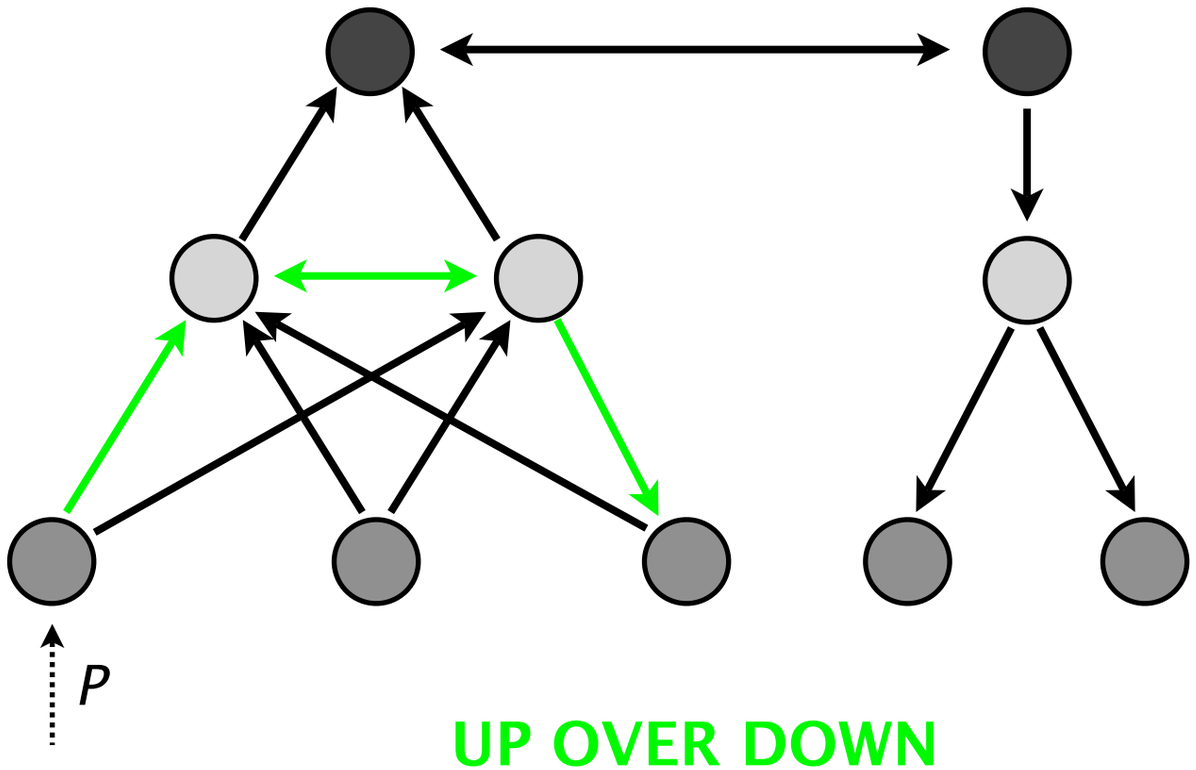
UP UP OVER DOWN DOWN

BGP Propagation rules

	To client	*To peer/RR*
From client	✓	✓
From peer/RR	✓	×

Routes are allowed to flow on valid signaling paths only

Valid signaling path match the UP* OVER? DOWN* regular expression



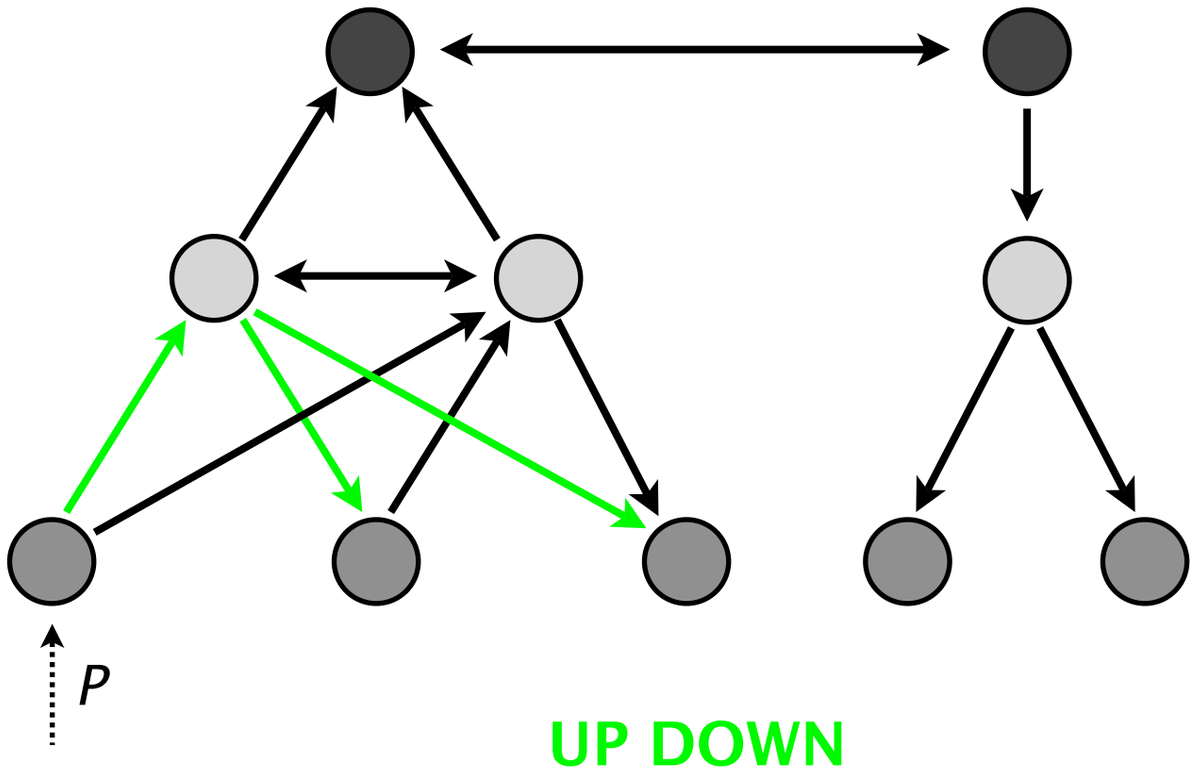
UP OVER DOWN

BGP Propagation rules

	To client	*To peer/RR*
From client	✓	✓
From peer/RR	✓	✗

Routes are allowed to flow on valid signaling paths only

Valid signaling path match the UP* OVER? DOWN* regular expression

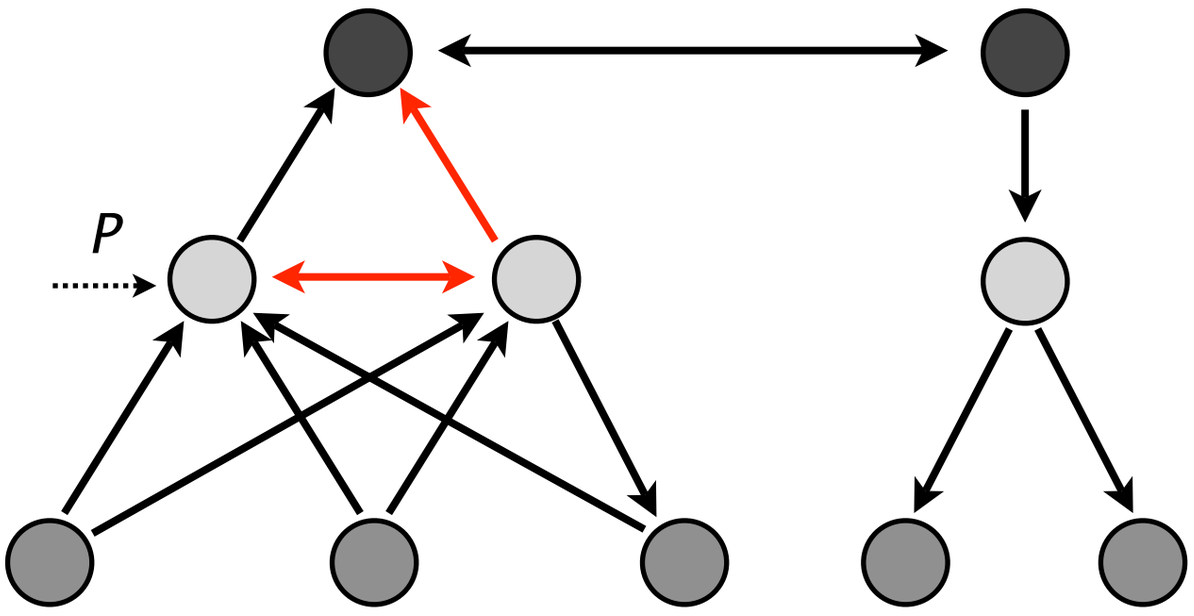


BGP Propagation rules

	To client	*To peer/RR*
From client	✓	✓
From peer/RR	✓	✗

Routes are allowed to flow on valid signaling paths only

Valid signaling path match the UP* OVER? DOWN* regular expression



OVER UP

BGP Propagation rules

	To client	*To peer/RR*
From client	✓	✓
From peer/RR	✓	✗

Breaking News

Adding a single iBGP session can disrupt iBGP ability of distributing routing information

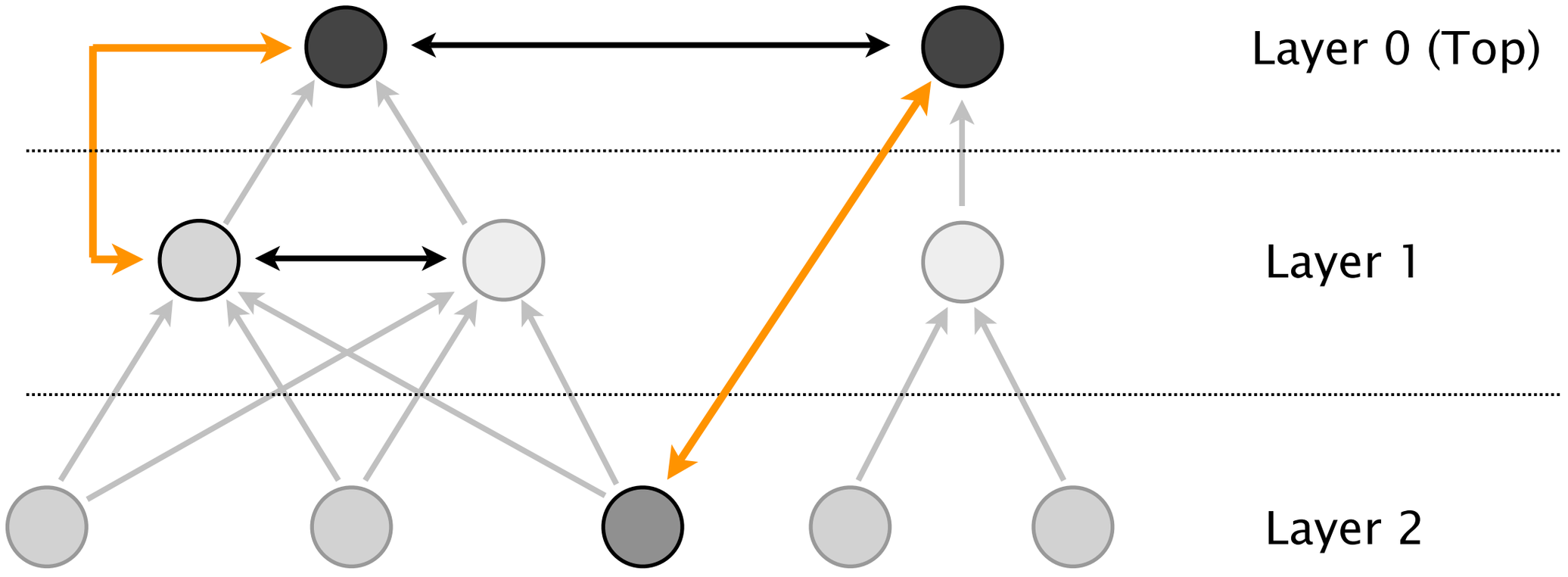
Breaking News

Adding *a single spurious OVER* can disrupt iBGP ability of distributing routing information

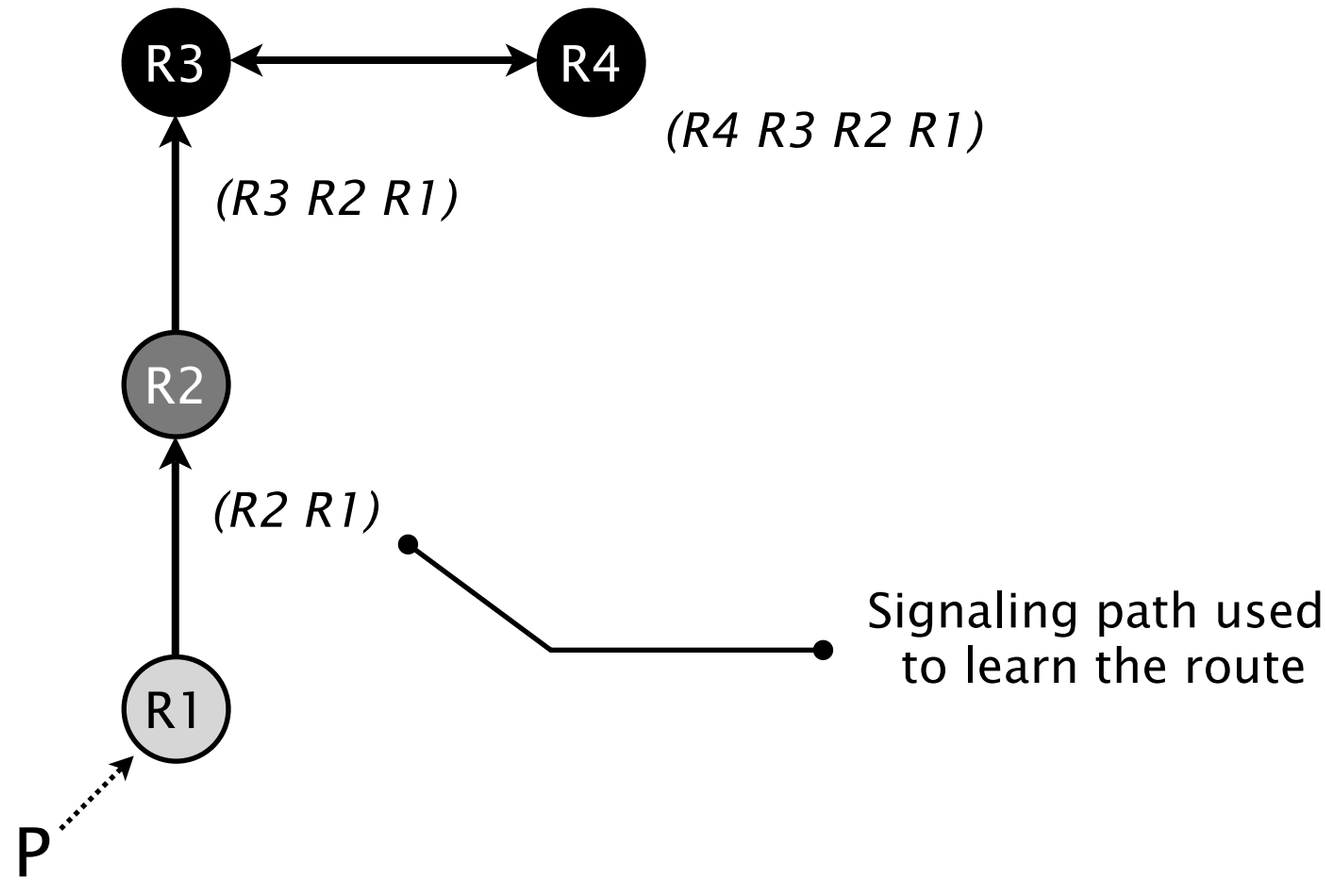
A spurious OVERs is a special type of OVER

Spurious OVER

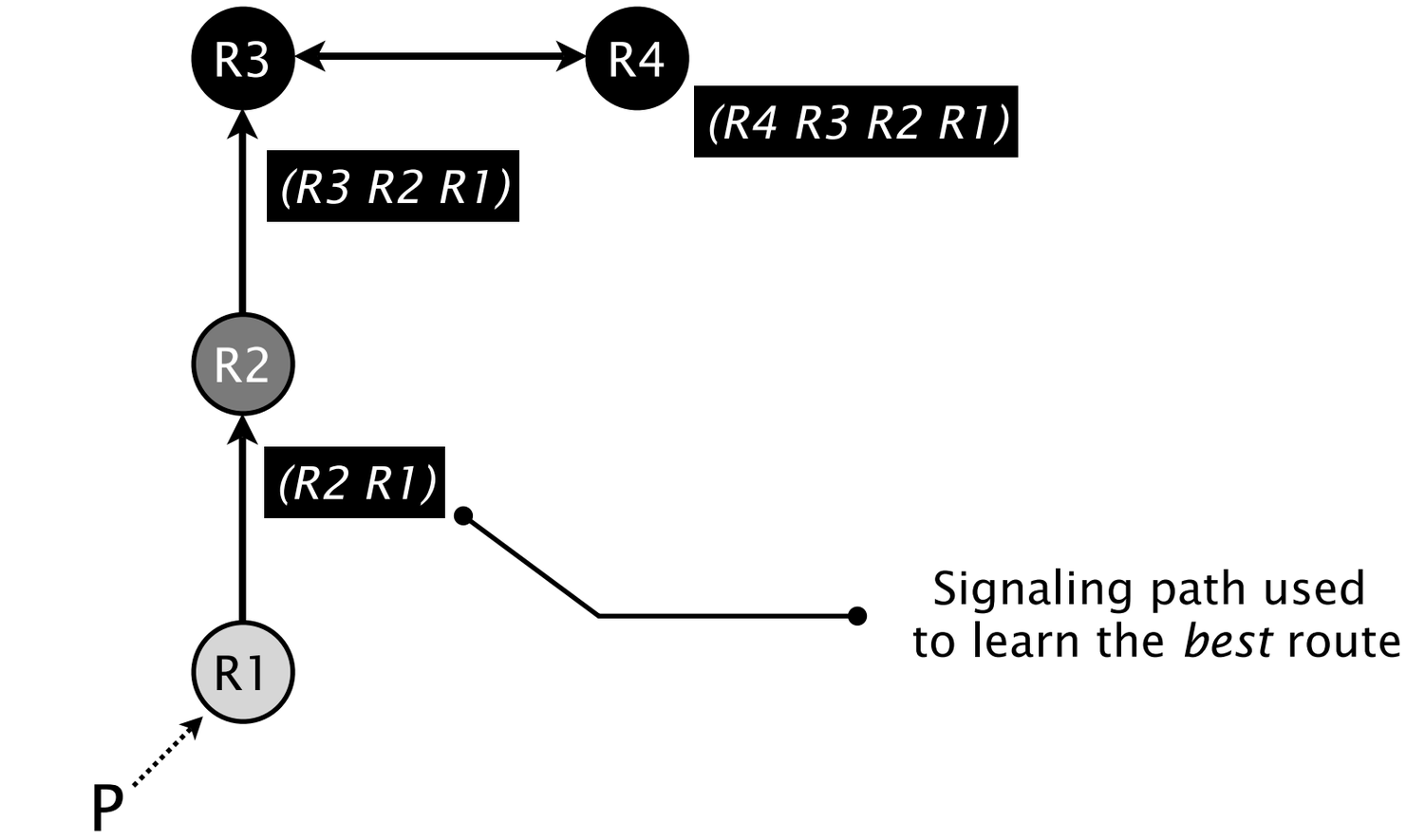
An OVER session between two routers *x* and *y* such that either *x* or *y* is not in the RR top layer



Let’s consider a simple example

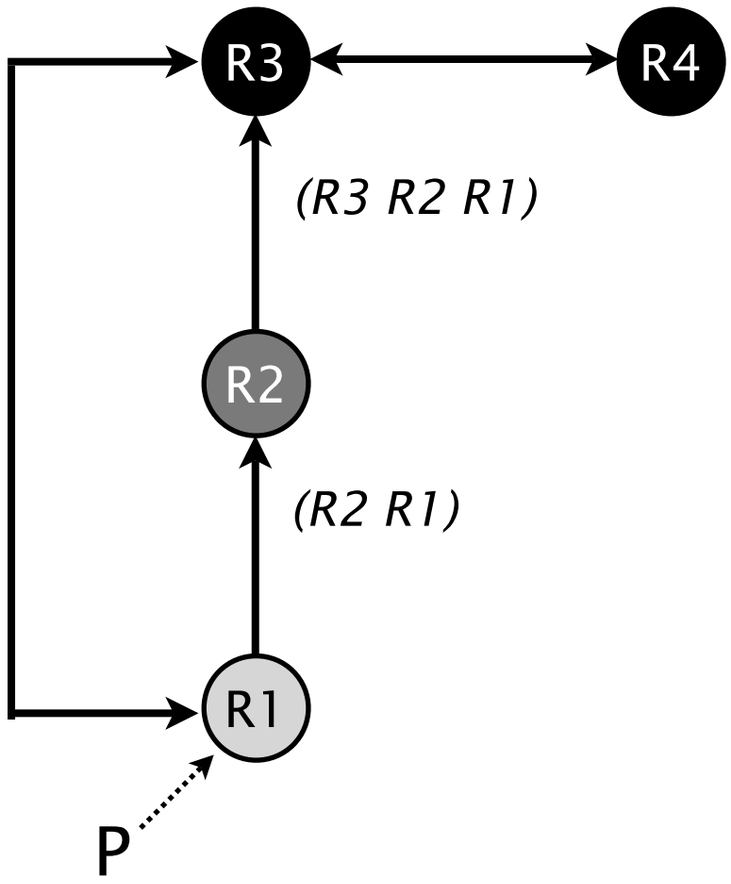


OVER-RIDE GADGET



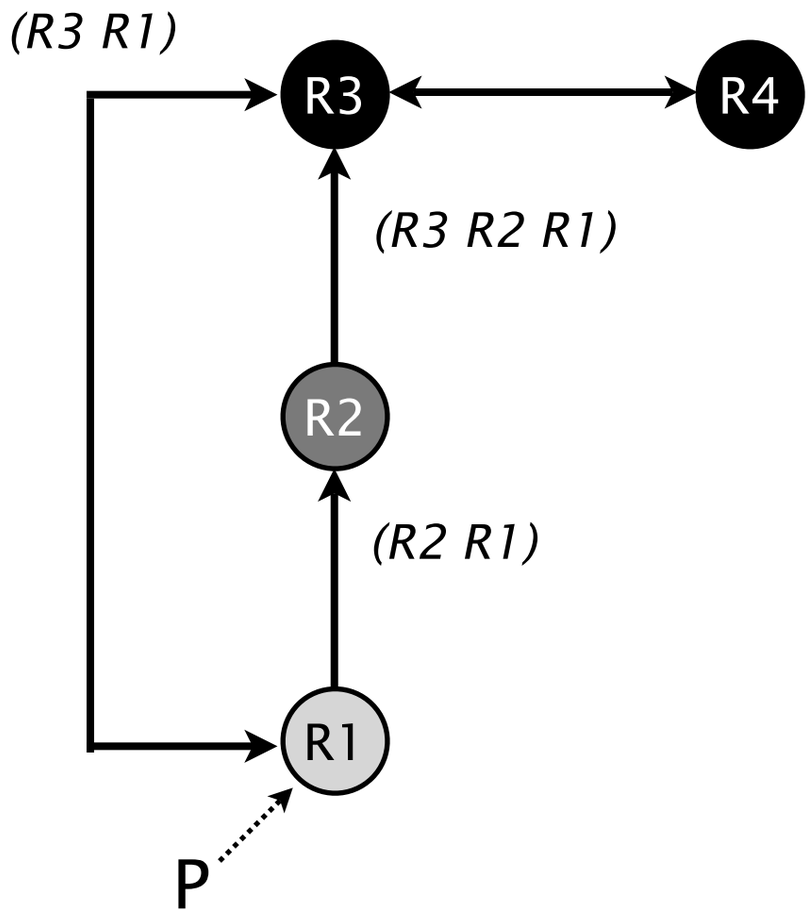
OVER-RIDE GADGET

Let’s add a spurious OVER session between R3 and R1



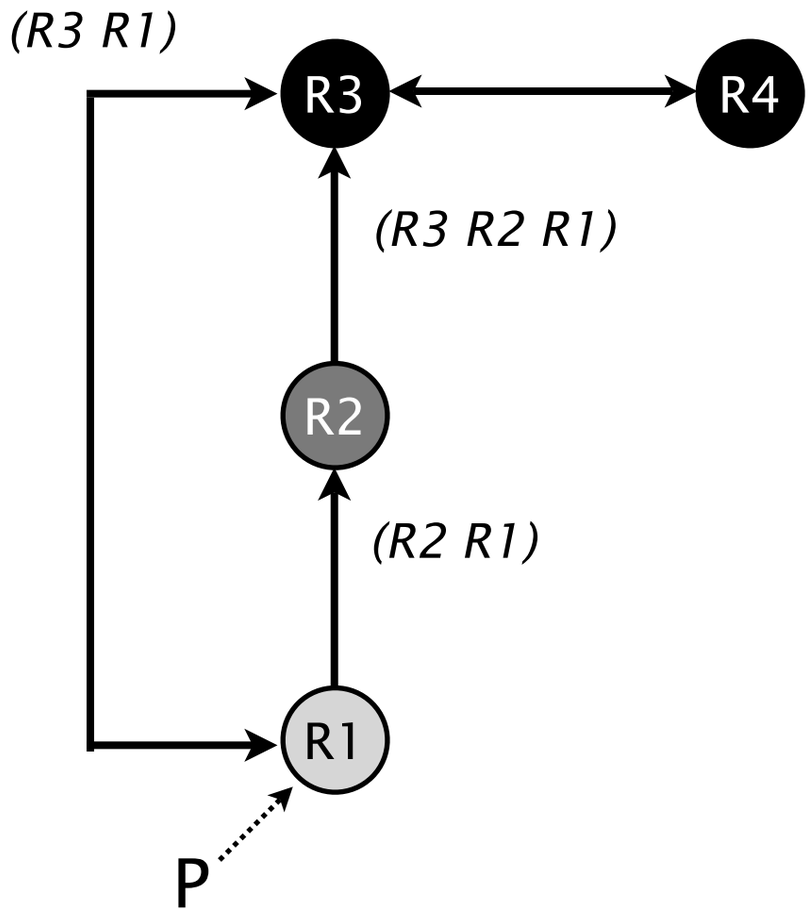
OVER-RIDE GADGET

Now, R3 learns P via two signaling paths



OVER-RIDE GADGET

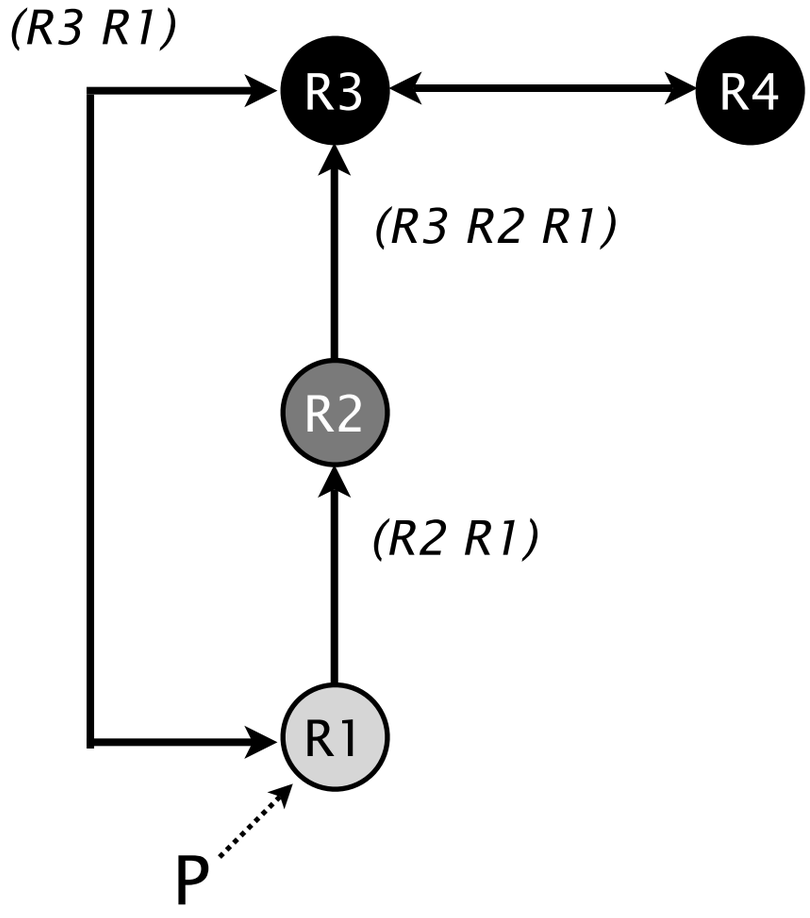
R3 BGP Decision Process is used to select one of them



OVER-RIDE GADGET

BGP Decision Process *(R3 R1)* *(R3 R2 R1)*

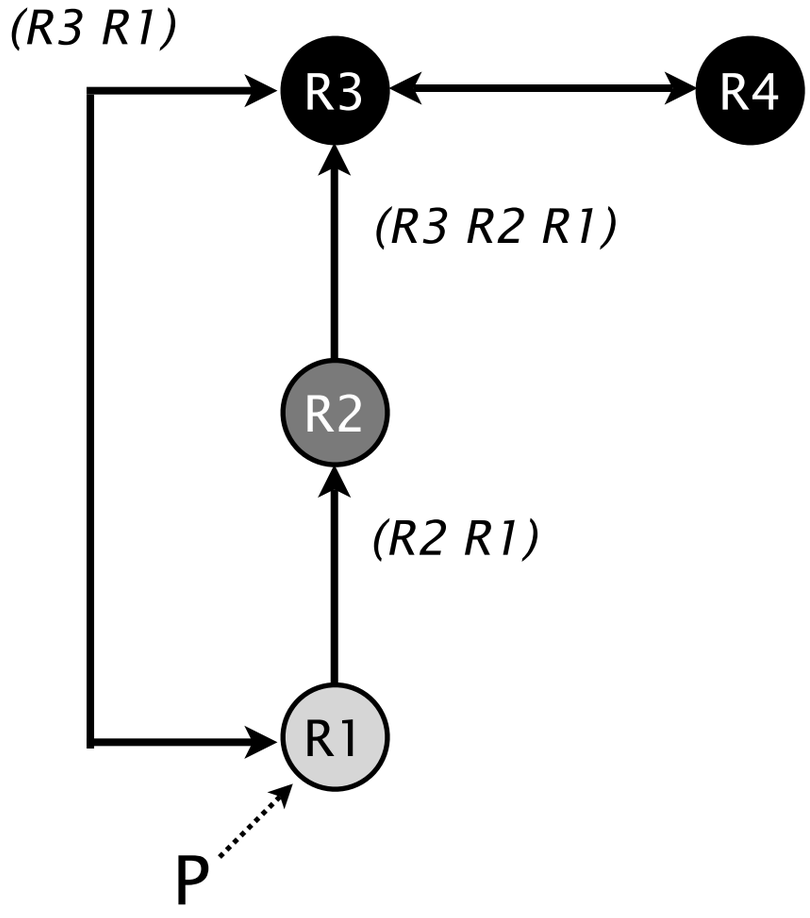
1. Higher Local-preference
2. Shorter AS-Path
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP



OVER-RIDE GADGET

BGP Decision Process

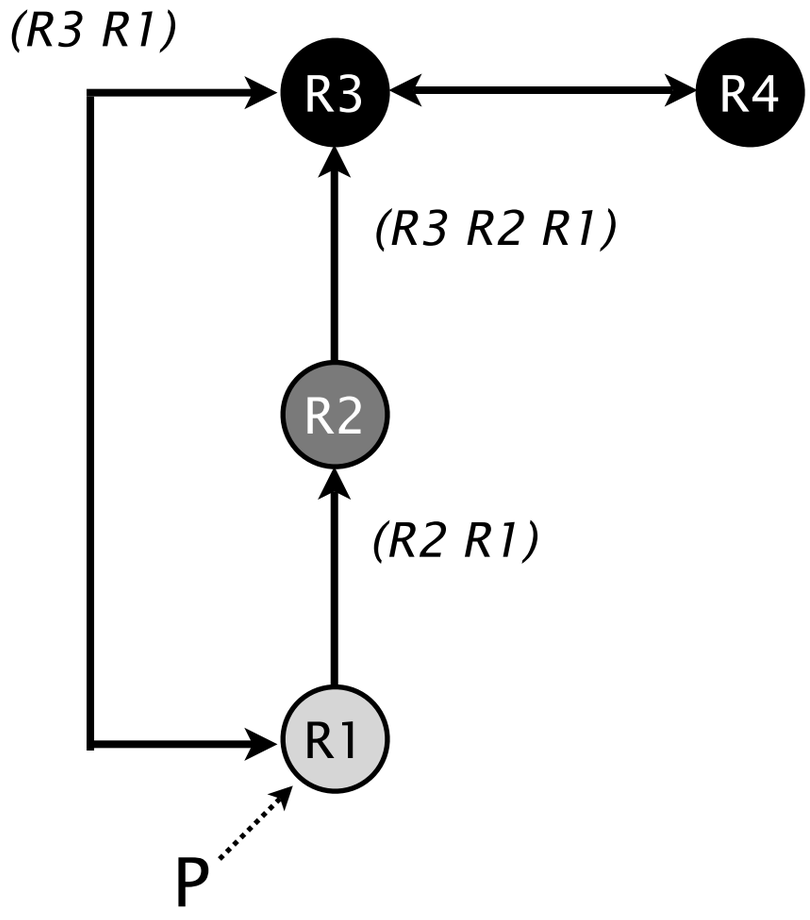
1. Higher Local-preference *(R3 R1)* *(R3 R2 R1)*
2. Shorter AS-Path
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP



OVER-RIDE GADGET

BGP Decision Process

1. Higher Local-preference
2. Shorter AS-Path *(R3 R1)* *(R3 R2 R1)*
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP



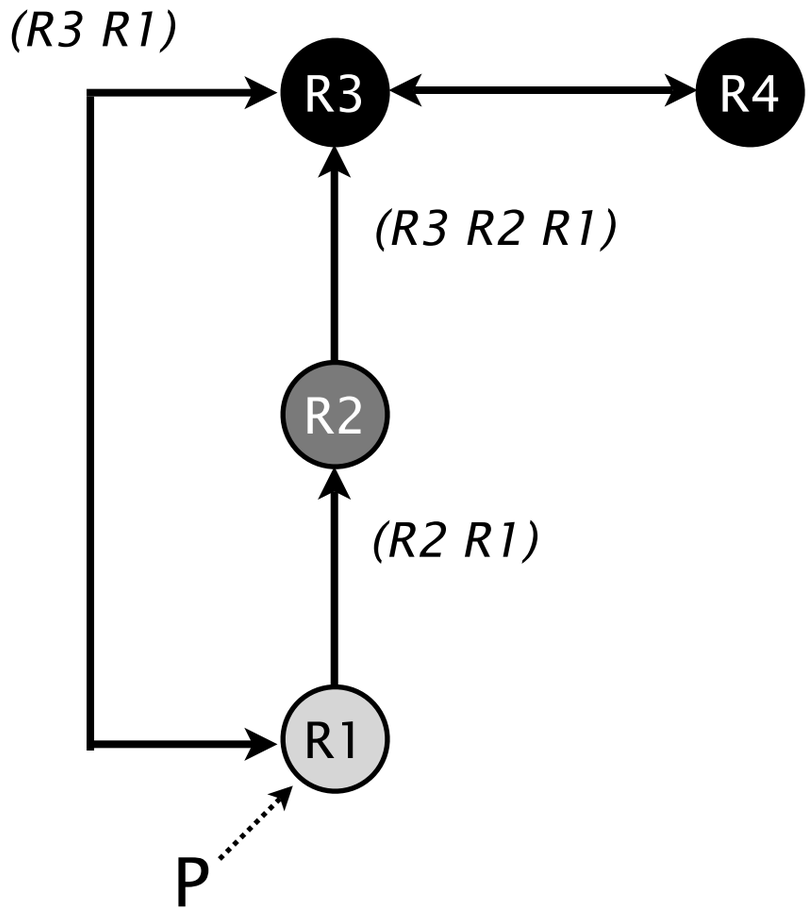
OVER-RIDE GADGET

BGP Decision Process

1. Higher Local-preference
2. Shorter AS-Path
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP

(R3 R1)

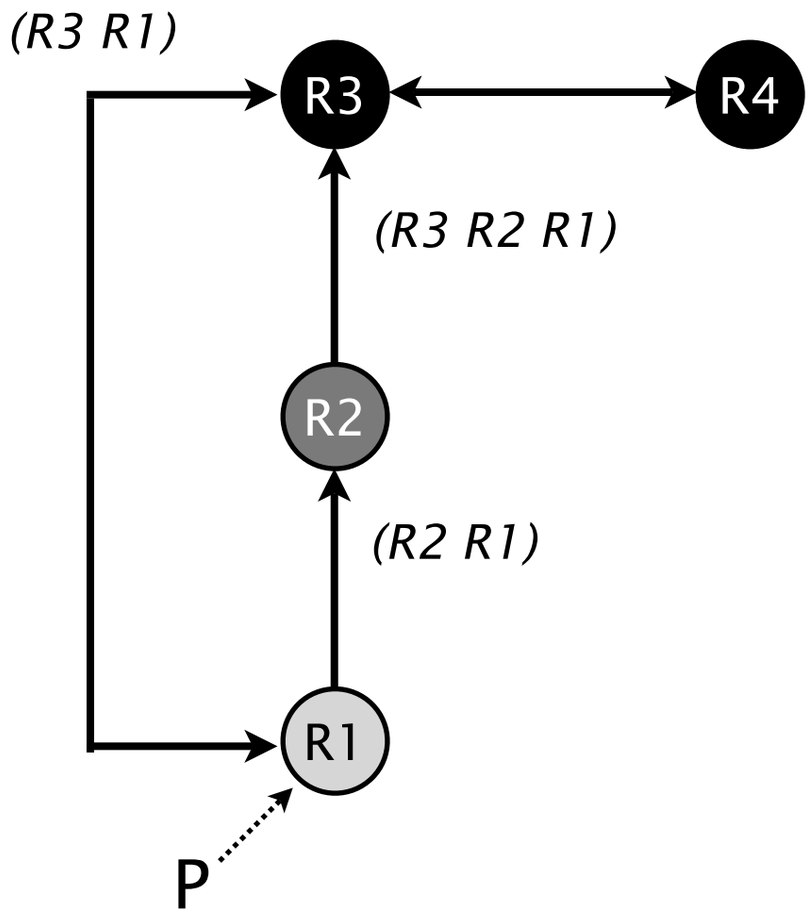
(R3 R2 R1)



OVER-RIDE GADGET

BGP Decision Process

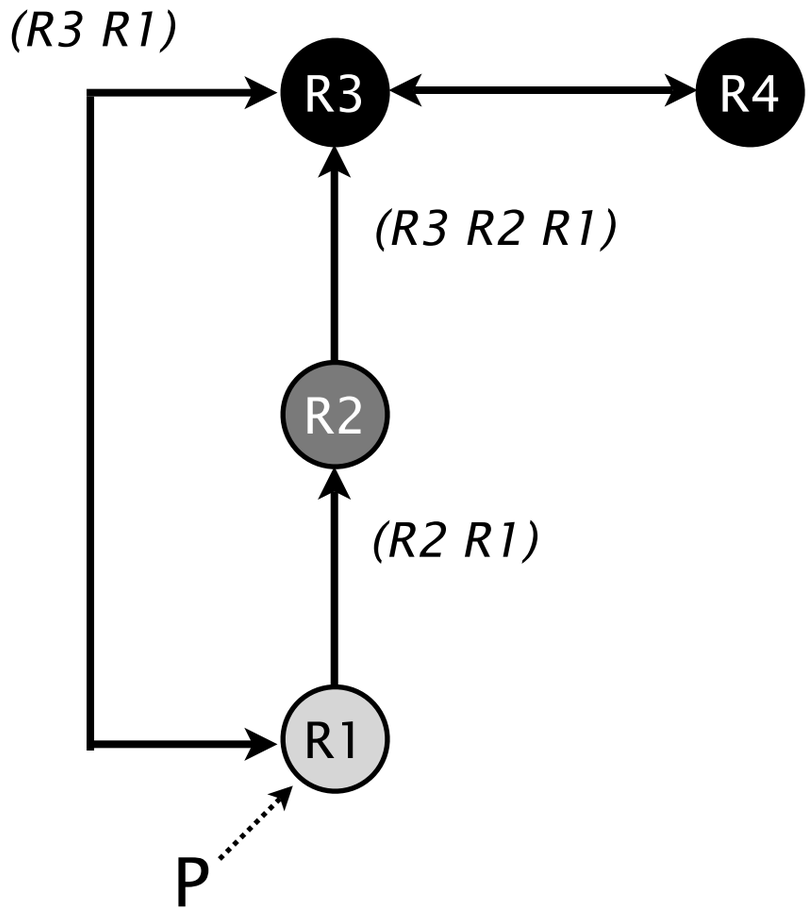
1. Higher Local-preference
2. Shorter AS-Path
3. Lower Origin
4. Lower MED *(R3 R1)* *(R3 R2 R1)*
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP



OVER-RIDE GADGET

BGP Decision Process

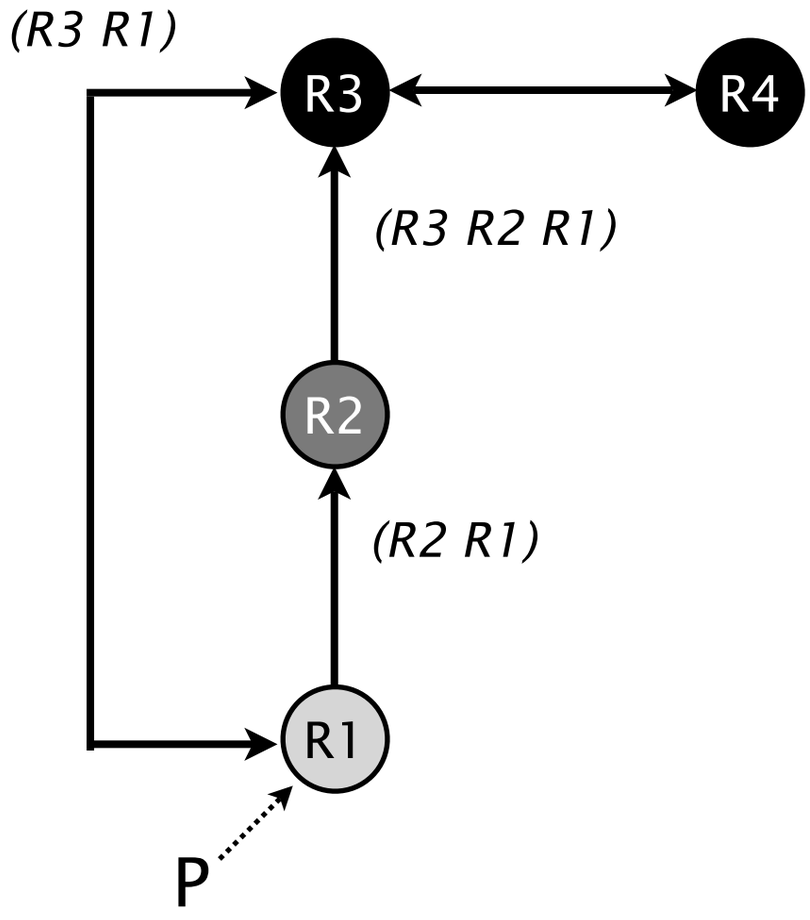
1. Higher Local-preference
2. Shorter AS-Path
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP *(R3 R1)* *(R3 R2 R1)*
6. Lower IGP metric to NH
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP



OVER-RIDE GADGET

BGP Decision Process

1. Higher Local-preference
2. Shorter AS-Path
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH *(R3 R1)* *(R3 R2 R1)*
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP

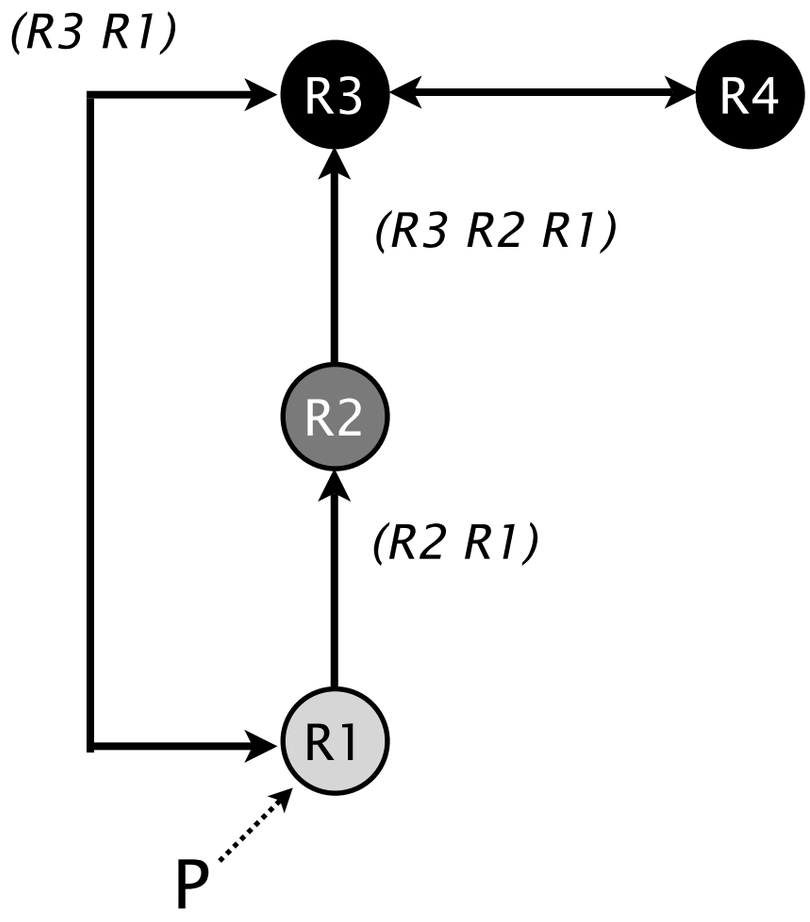


OVER-RIDE GADGET

BGP Decision Process

1. Higher Local-preference
2. Shorter AS-Path
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH
7. Lower Router ID *(R3 R1)* *(R3 R2 R1)*
8. Shorter cluster-list
9. Lower neighbor IP

(R3 R1) wins since it has no cluster-list



OVER-RIDE GADGET

BGP Decision Process

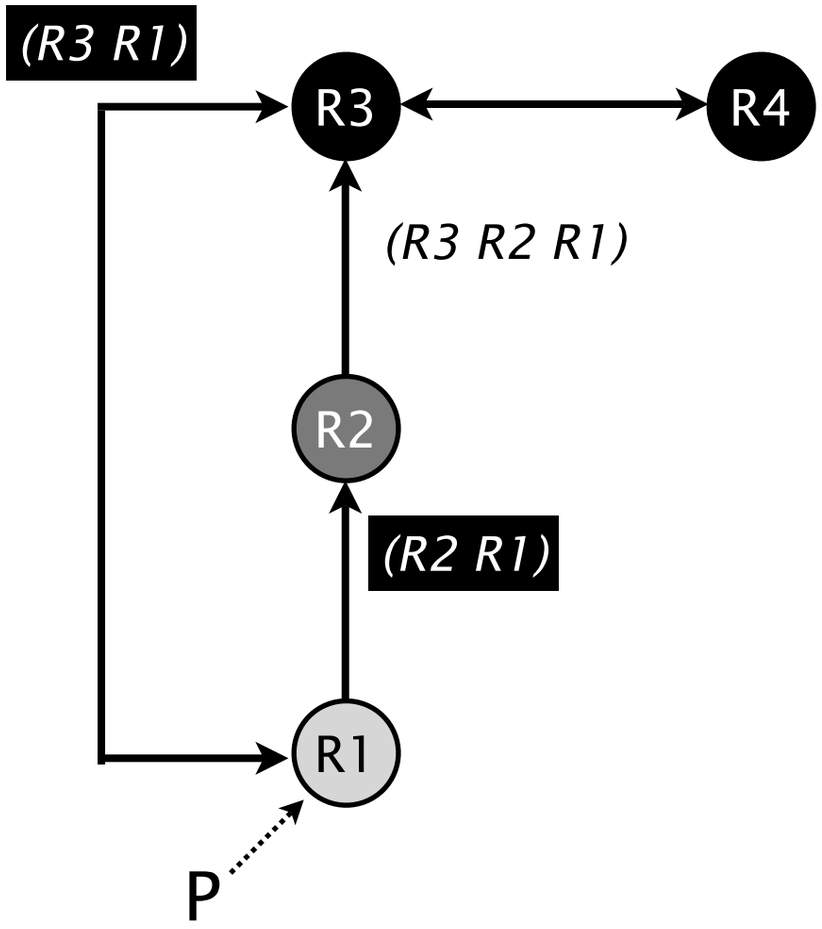
1. Higher Local-preference
2. Shorter AS-Path
3. Lower Origin
4. Lower MED
5. Prefer eBGP over iBGP
6. Lower IGP metric to NH
7. Lower Router ID
8. Shorter cluster-list
9. Lower neighbor IP

(R3 R1) *(R3 R2 R1)*

↓

wins

Due to BGP Propagation rules,
R3 does not announce the route to R4 anymore

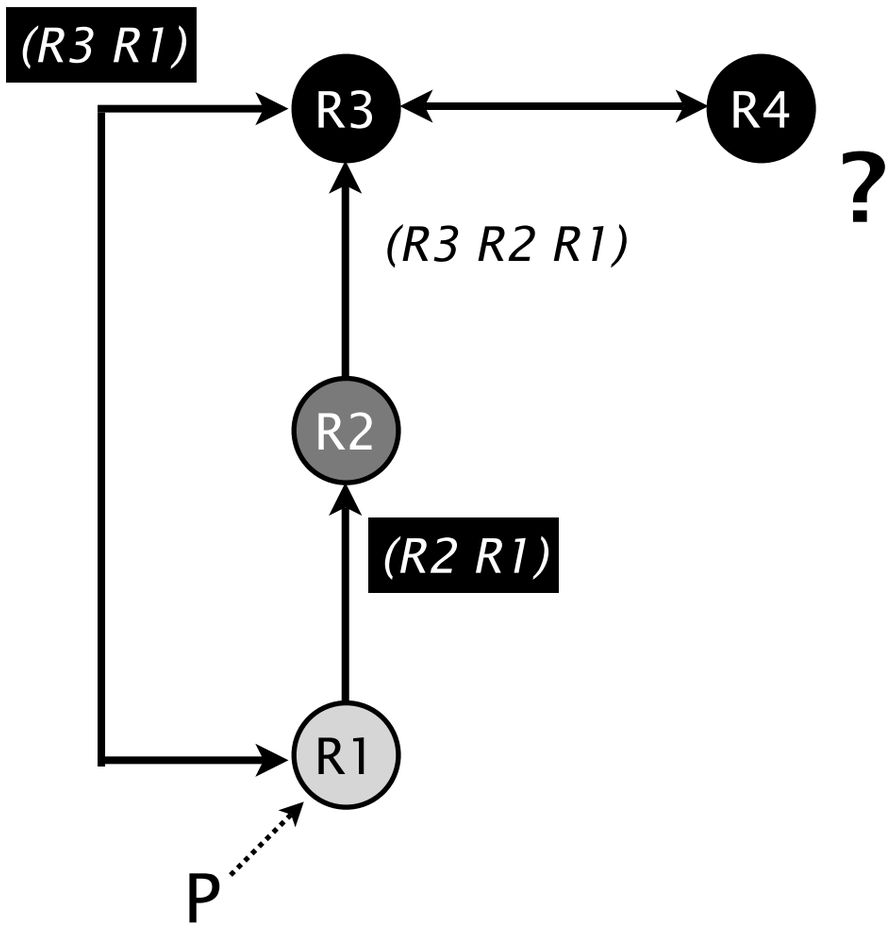


BGP Propagation rules

	To client	***To peer***
From client	✓	✓
From peer/RR	✓	×

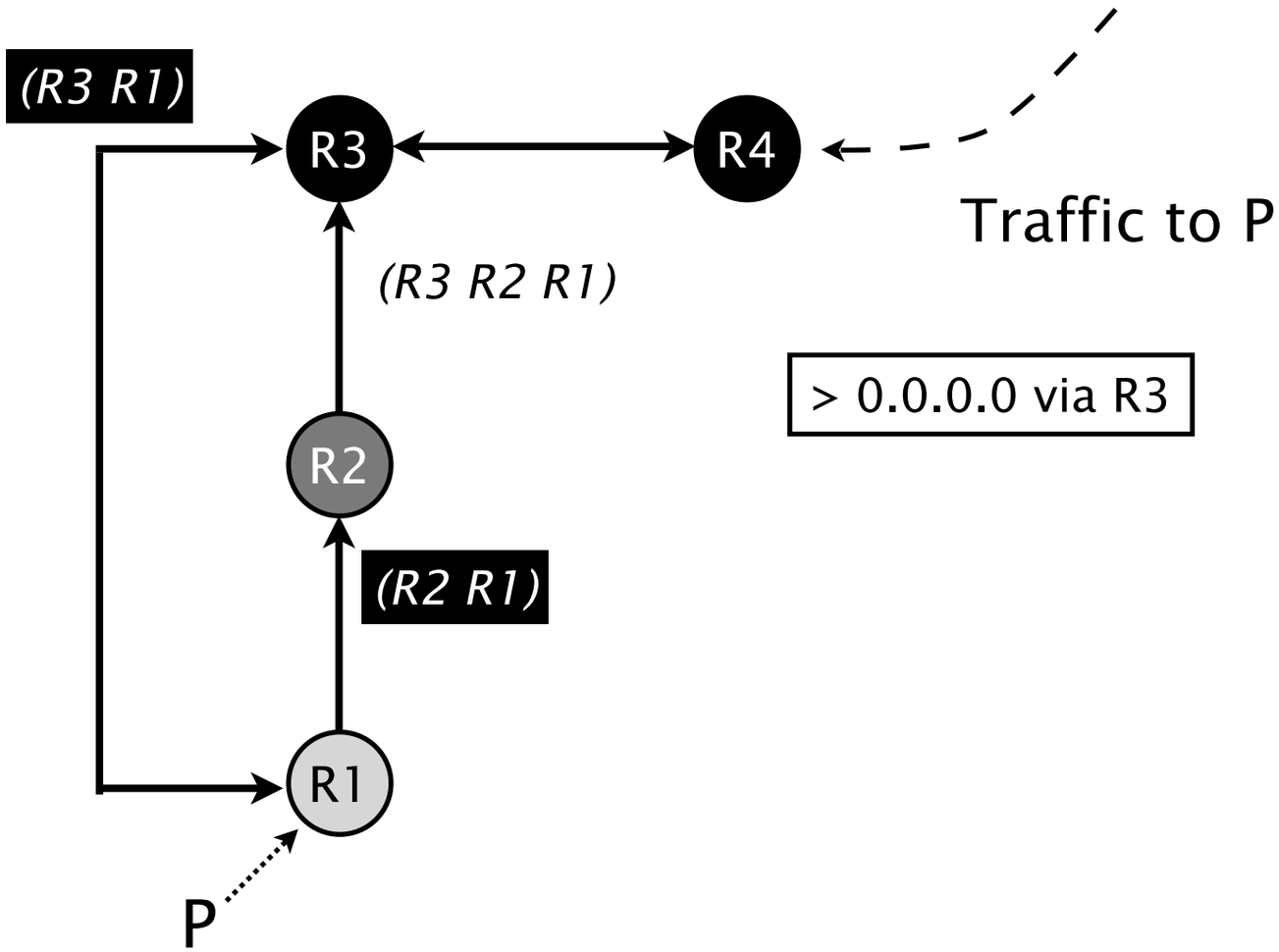
OVER-RIDE GADGET

R4 does not receive any route for P



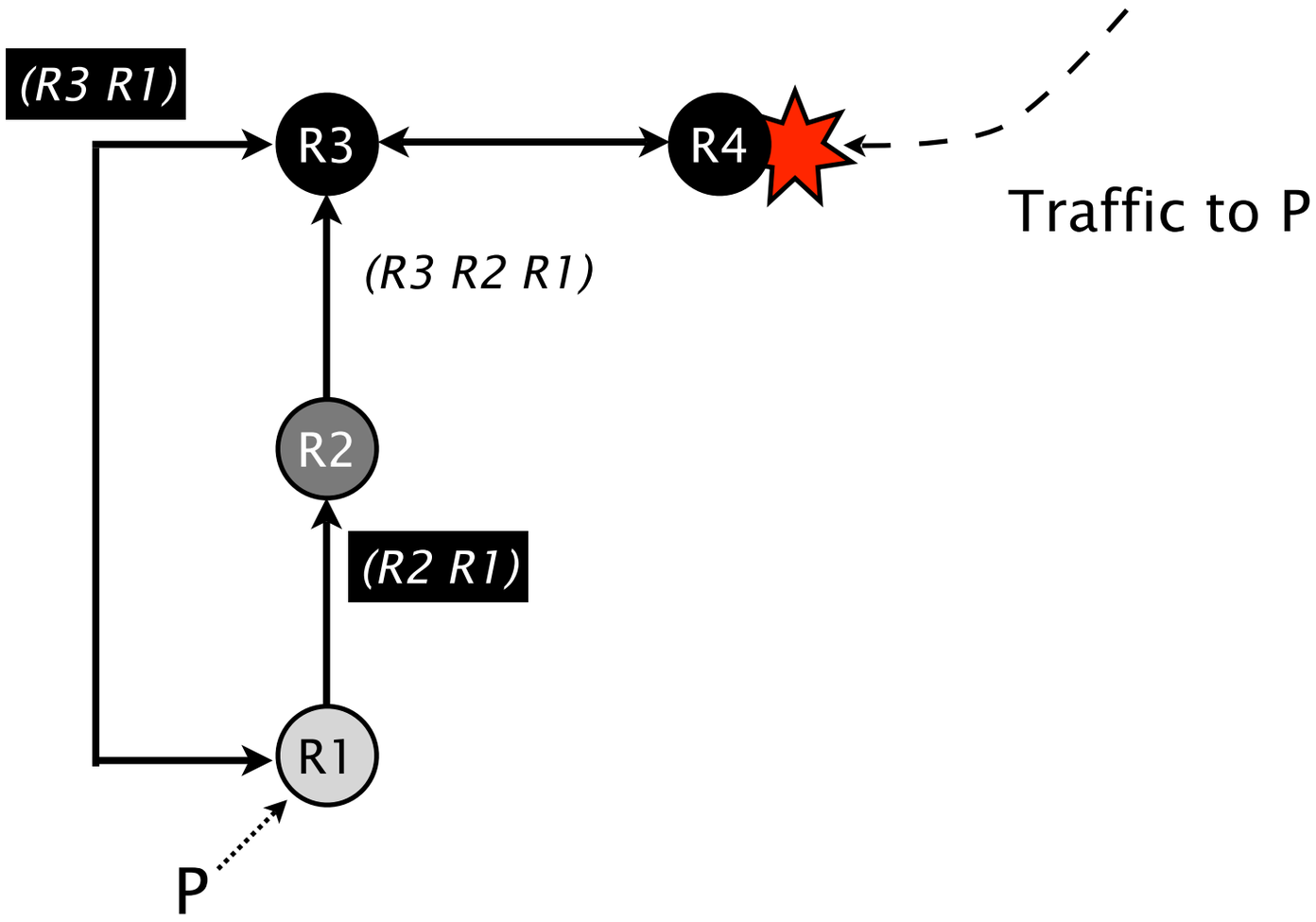
OVER-RIDE GADGET

R4 might then use a less specific route
which can create forwarding deflections and loops



OVER-RIDE GADGET

If R4 does not learn a less-specific route
a blackhole is effectively created



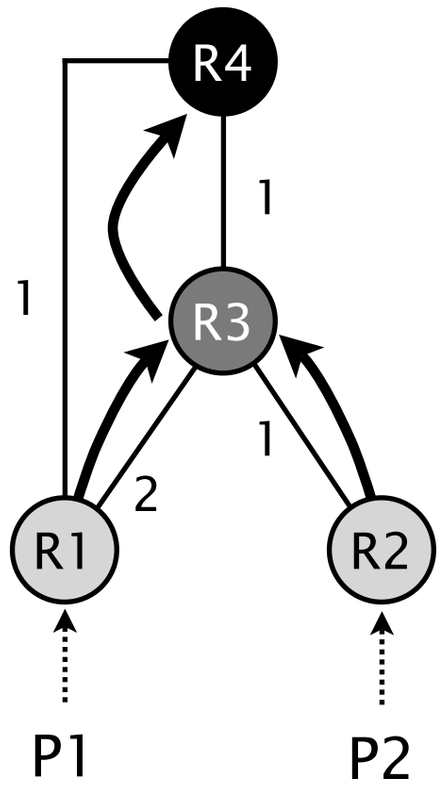
OVER-RIDE GADGET

Although uncommon, spurious OVER might appear in real-world network

Spurious OVERs

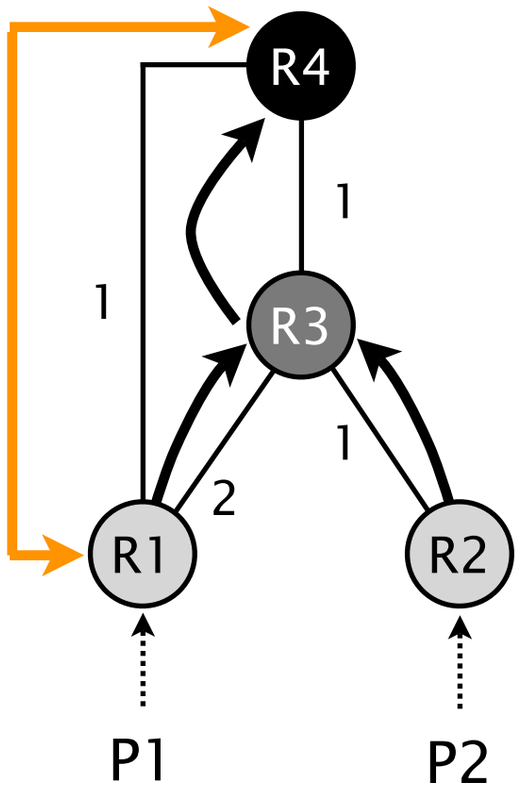
- have been found in real network [Feamster05, Park11]
- act as an easy-visibility fix [Pelsser08, Pelsser10]
- could appear during reconfiguration [Herrero10]

A spurious OVER is an easy and tempting solution to solve route visibility issue



Although preferred, R3 does not receive P1 since R2 prefers P2 (IGP cost)

A spurious OVER is an easy and tempting solution to solve route visibility issue



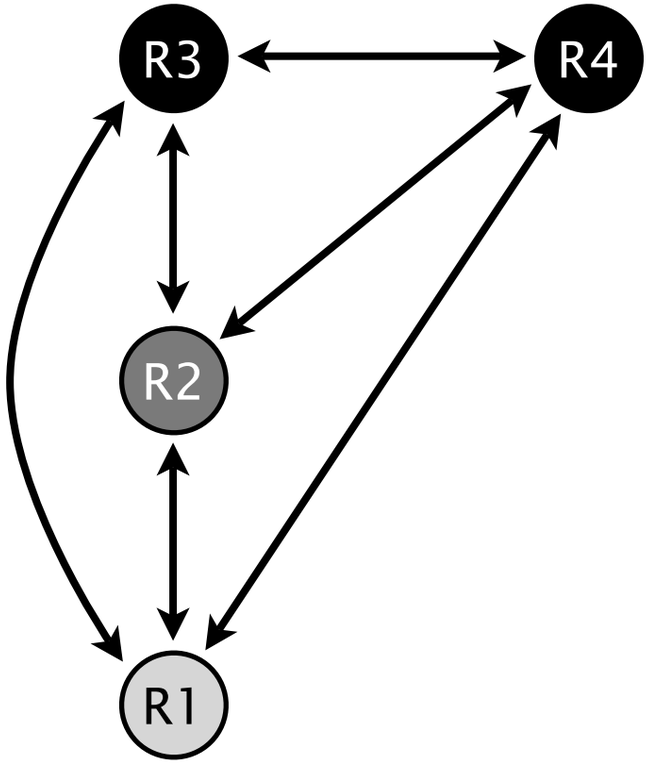
Adding a spurious OVER, improves R3's visibility

[Pelsser08, Pelsser10]

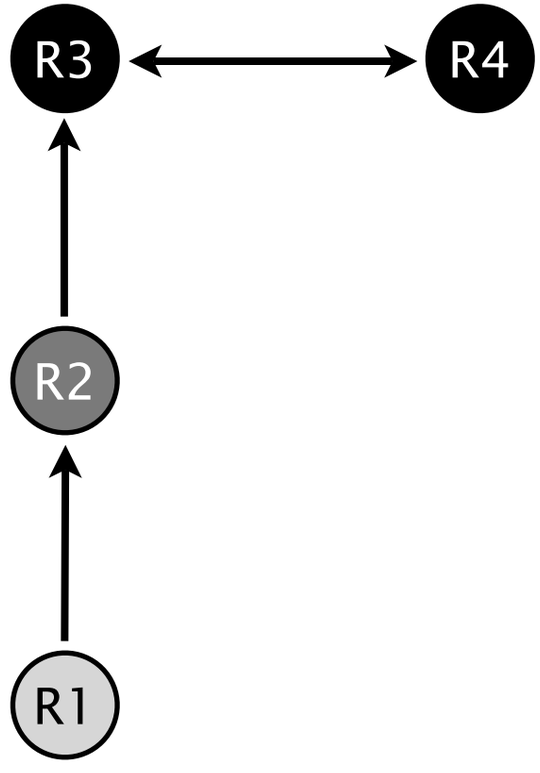
Spurious OVER are likely to appear during iBGP reconfiguration

Best practices: Introduce UP before tearing OVER down [Herrero10]

going from



to

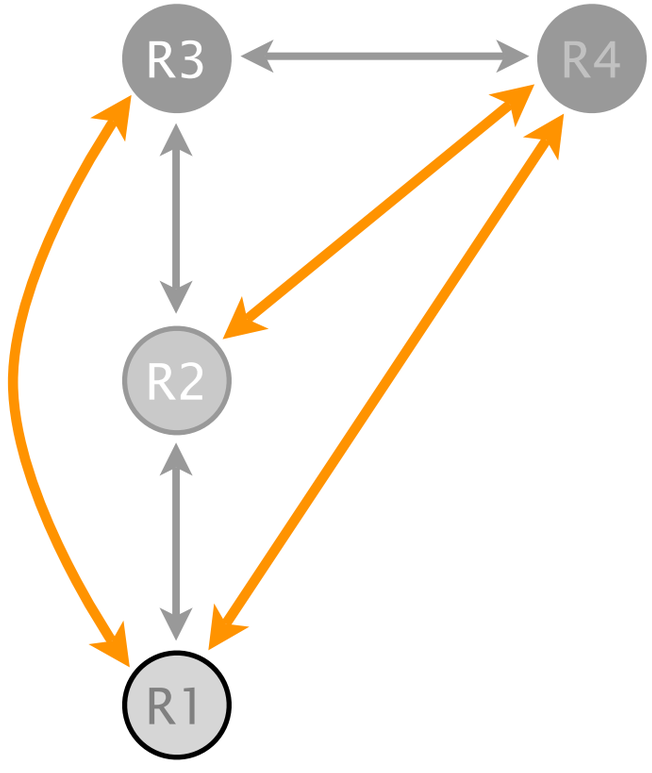


Spurious OVER are likely to appear during iBGP reconfiguration

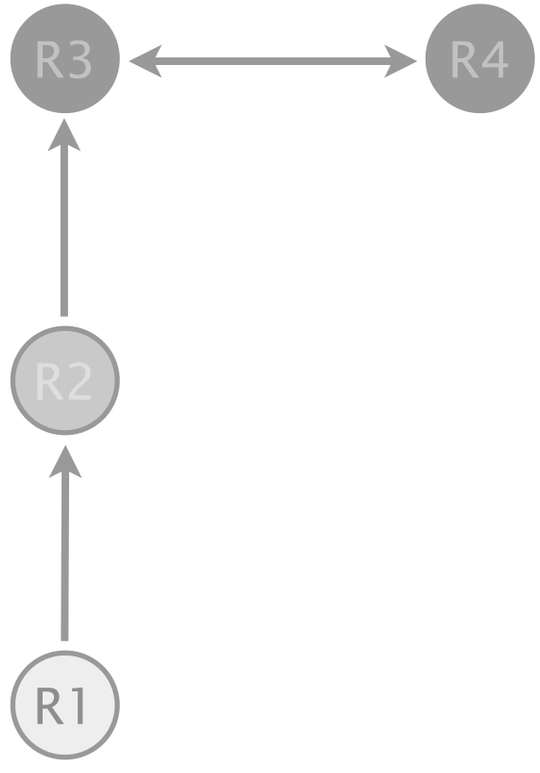
Best practices: Introduce UP before tearing OVER down [Herrero10]

▶ **OVERs potentially spurious during the process**

going from



to



Valid signaling path is not a good abstraction to study route propagation

- Spurious OVER improves visibility *locally*, but potentially worsen it *globally*

- Having a valid signaling path is *necessary*, not *sufficient*

- A connected iBGP topology does not guarantee correct route propagation

iBGP Deceptions:
More Sessions, Fewer Routes

Introduction and Motivation

Dissemination correctness

Revisiting the state-of-the-art

Conclusion

Route reflection is prone to both routing and forwarding anomalies

An iBGP configuration is correct if it respects the following two properties [Griffin02]:

- *signaling correctness*
  BGP will always converge to a stable, unique routing state

- *forwarding correctness*
  No forwarding deflection arises along any BGP forwarding path

One property is missing: *dissemination correctness*

An iBGP configuration is correct if it respects the following two properties [Griffin02]:

- *signaling correctness*
  BGP will always converge to a stable, unique routing state

- *forwarding correctness*
  Absence of deflection along any BGP forwarding path

Dissemination correctness deals with issues in the route propagation process

An iBGP configuration is correct if it respects the following **three** properties:

- *signaling correctness*
  BGP will always converge to a stable, unique routing state

- *forwarding correctness*
  Absence of deflection along any BGP forwarding path

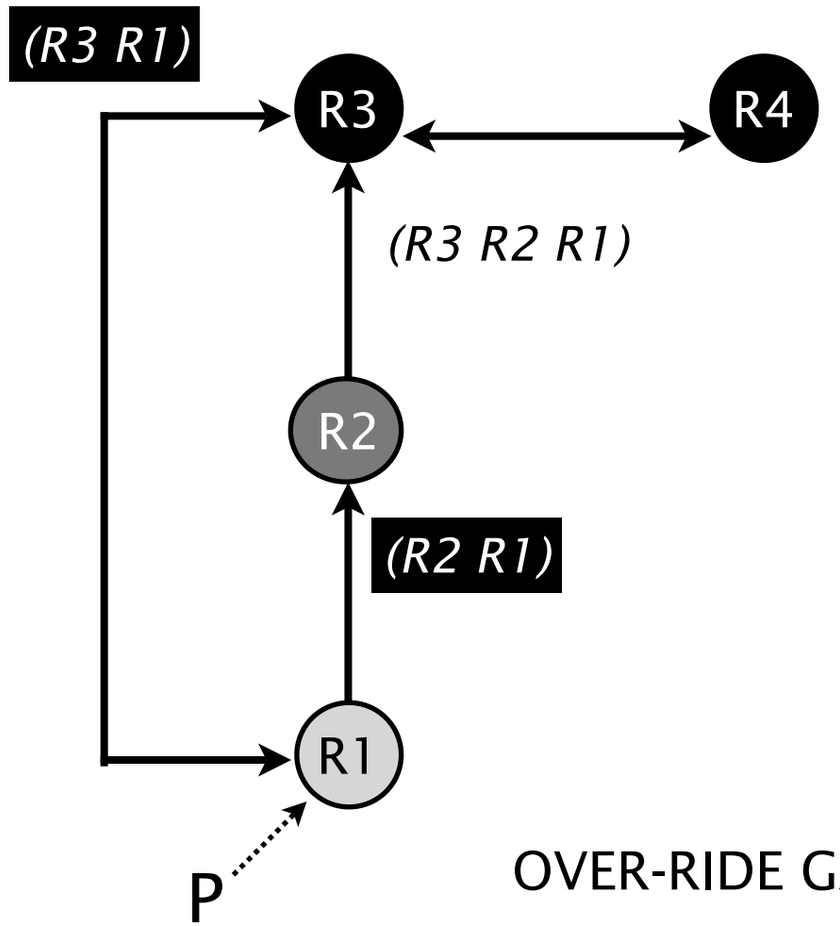
- *dissemination correctness*
  all BGP routers are guaranteed to receive a route to all prefixes

Signaling, dissemination and forwarding correctness complement each other

- Signaling correct does not imply dissemination correct

Signaling, dissemination and forwarding correctness complement each other

- Signaling correct does not imply dissemination correct

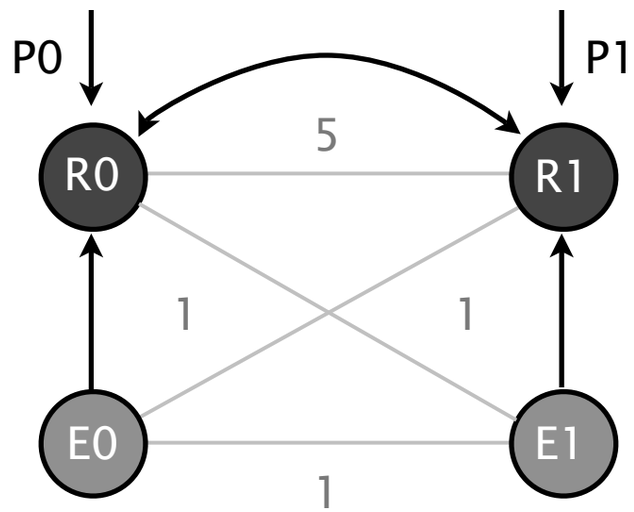


Example of iBGP topology which is *signaling* correct, but not *dissemination* correct

OVER-RIDE GADGET

Signaling, dissemination and forwarding correctness complement each other

- Signaling correct does not imply dissemination correct
- Dissemination correct does not imply forwarding correct



Example of iBGP topology which is *dissemination* correct, but not *forwarding* correct

Dealing with dissemination correctness is computationally hard

Dissemination Correctness Problem (DCP):

Given a signaling correct iBGP topology *B*
and the underlying IGP topology *I*,

Decide if *B* is dissemination correct

DCP is coNP-hard
P-time reduction from 3-SAT complement

Prior knowledge of correctness is useless

One More Session Problem (OMSP):

Given a dissemination correct iBGP topology *B*,
and the underlying IGP topology *I*,

Decide if adding a spurious OVER session to *B*
will result in a dissemination correct topology

OMSP is coNP-hard
P-time reduction from 3-SAT complement

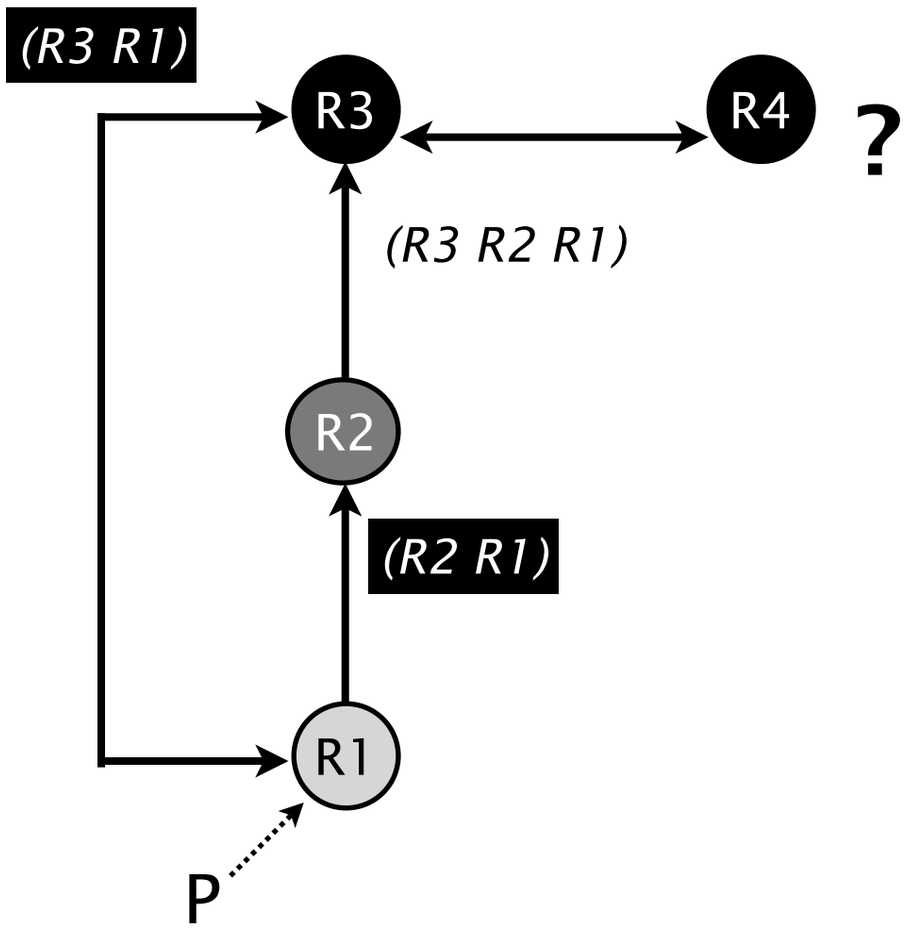
There exist sufficient conditions that guarantee *dissemination correctness*

Either of the following conditions guarantees a signaling correct iBGP topology to be dissemination correct

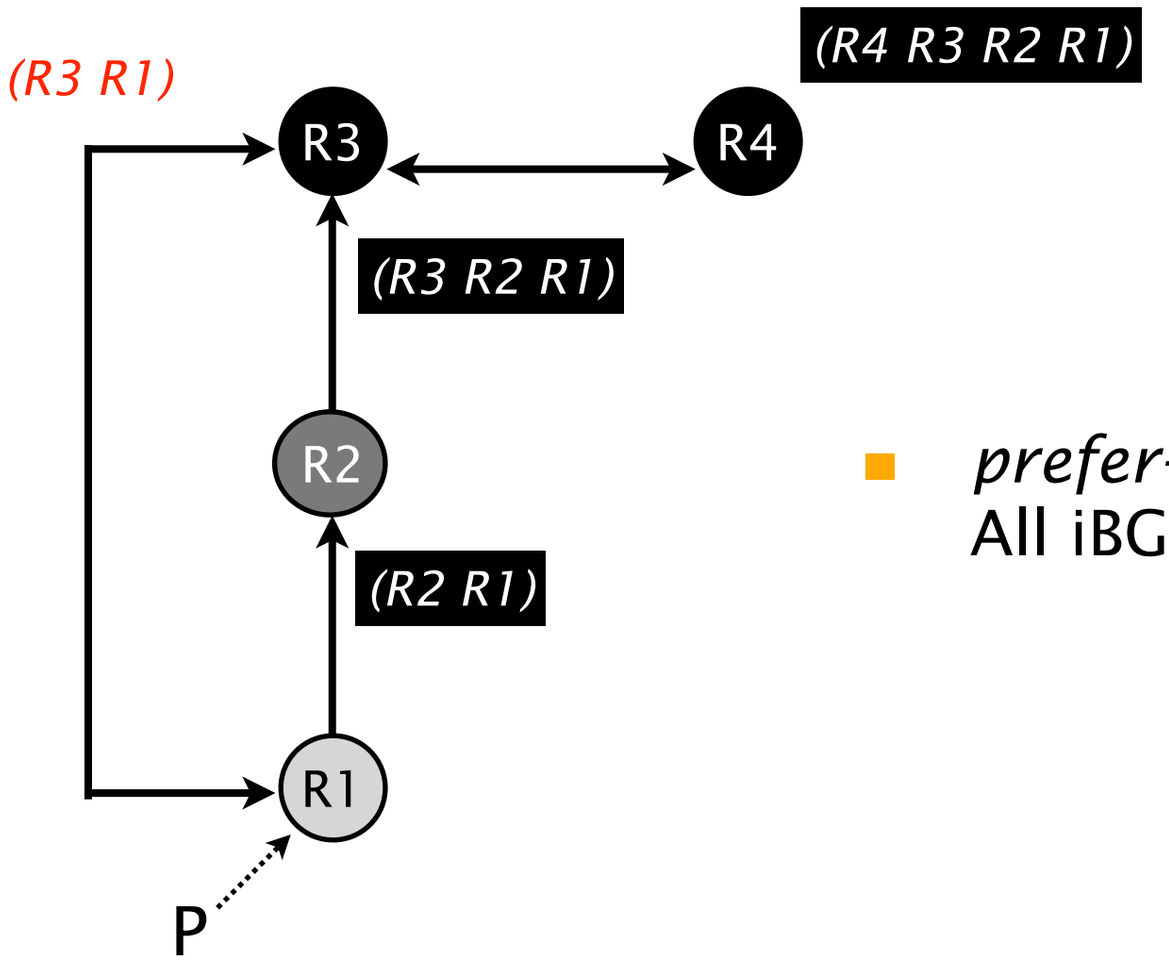
- *prefer-client*
  All iBGP routers strictly prefer client routes

- *no-spurious-OVER*
  The iBGP topology contains no spurious OVERs

There exist sufficient conditions that guarantee *dissemination correctness*

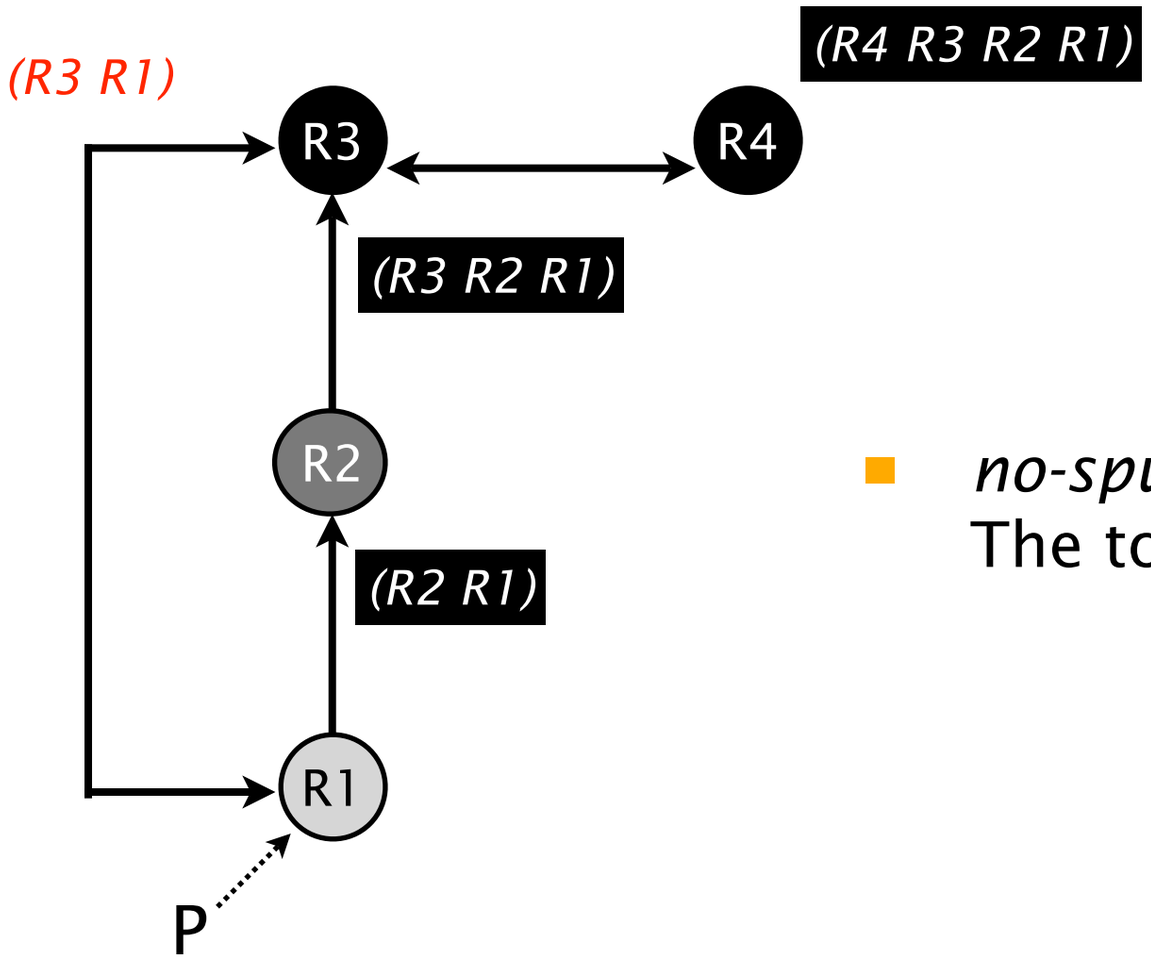


There exist sufficient conditions that guarantee *dissemination correctness*



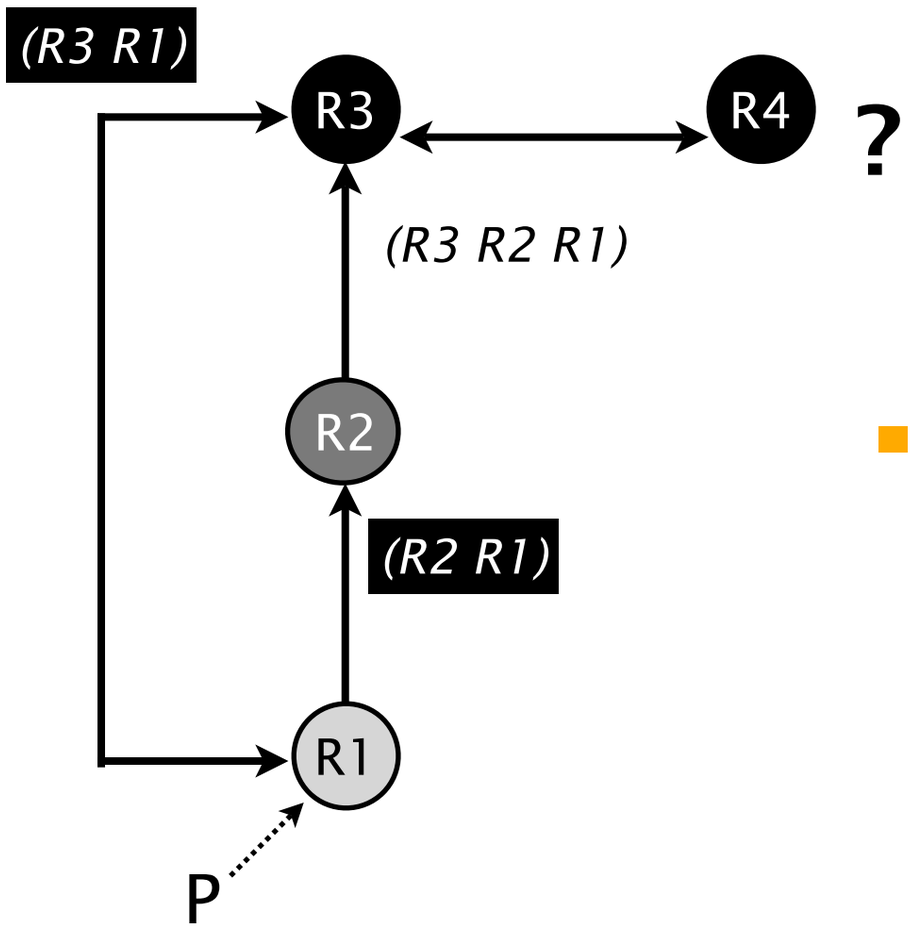
- *prefer-client*
  All iBGP routers strictly prefer client routes

There exist sufficient conditions that guarantee *dissemination correctness*



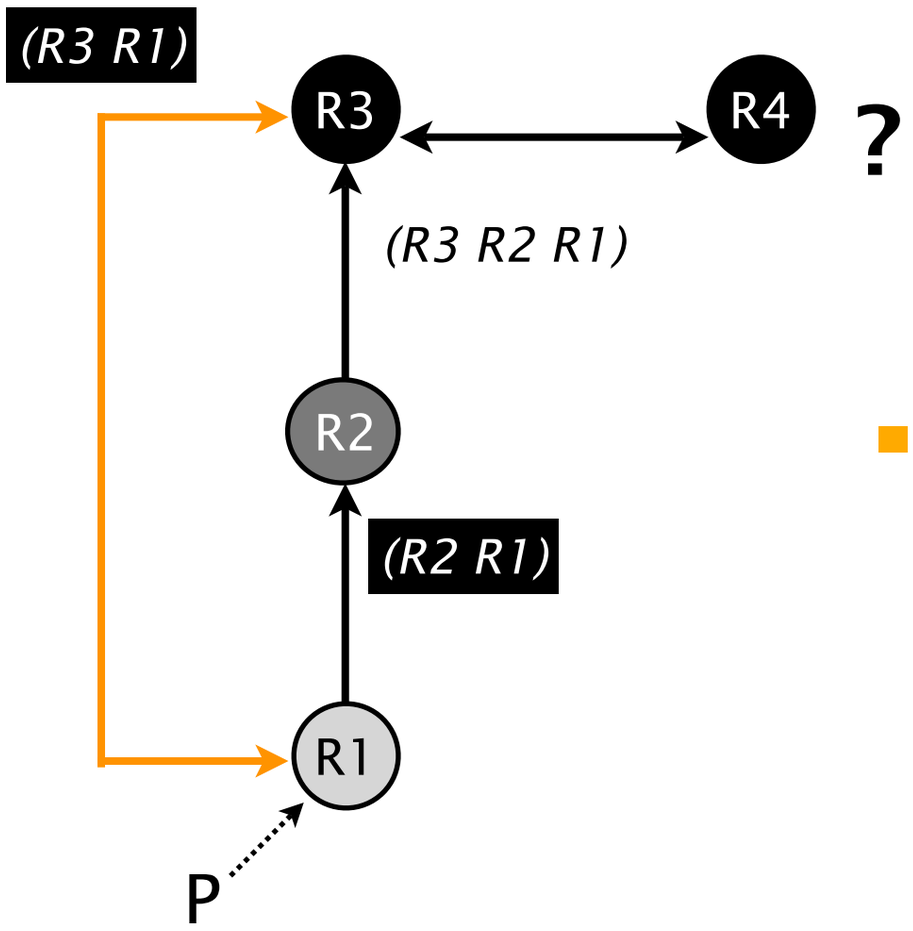
- *no-spurious-OVER*
  The topology contains no spurious OVERs

There exist sufficient conditions that guarantee *dissemination correctness*



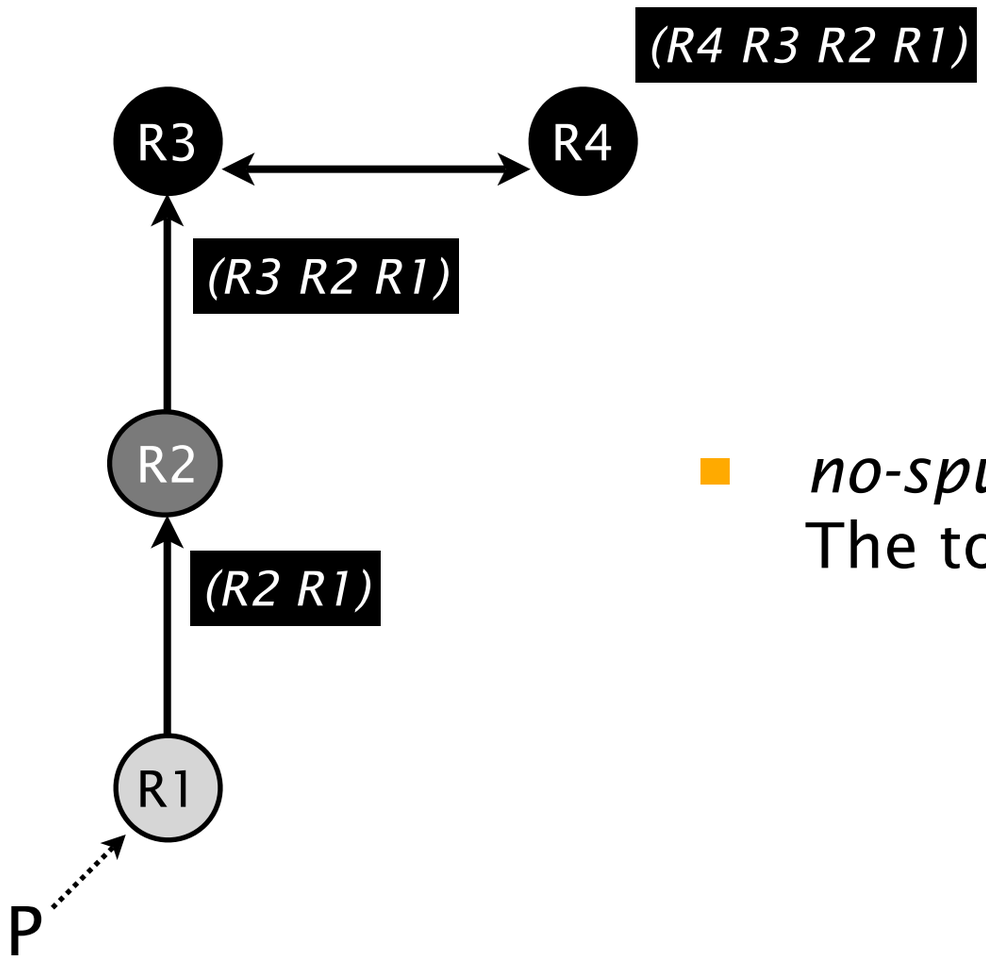
- *no-spurious-OVER*
  The topology contains no spurious OVERs

There exist sufficient conditions that guarantee *dissemination correctness*



- *no-spurious-OVER*
  The topology contains no spurious OVERs

There exist sufficient conditions that guarantee *dissemination correctness*



- *no-spurious-OVER*
  The topology contains no spurious OVERs

iBGP Deceptions:
More Sessions, Fewer Routes

Introduction and Motivation

Dissemination correctness

Revisiting the state-of-the-art

Conclusion

Some results already encompass dissemination correctness

Sufficient conditions guaranteeing signaling, forwarding correctness

On the correctness of IBGP configuration [Griffin, SIGCOMM02]

Some results already encompass dissemination correctness

Sufficient conditions guaranteeing signaling, forwarding correctness

On the correctness of IBGP configuration [Griffin, SIGCOMM02]

i) *B* has no cycles of UP sessions only
ii) **Route-reflector prefers paths propagated by clients**
iii) All-shortest-paths must also be valid signaling paths

implies dissemination correctness

Dissemination is often overlooked

Relaxed sufficient conditions for signaling or forwarding correctness

Preventing persistent oscillations and loops
in IBGP configuration with route reflection [Rawat, Comput.Netw.06]

Checking for optimal egress points in iBGP routing [Buob, DRCN07]

[Buob, Networking08]

Such conditions do not imply
dissemination correctness
(e.g. OVER-RIDE gadget)

Dissemination is often overlooked

Guarantee iBGP convergence by modifying the decision process

Stable and flexible iBGP [Flavel, SIGCOMM09]

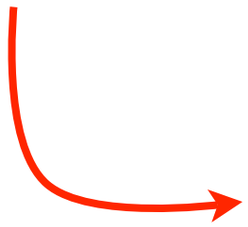
Modified iBGP does not guarantee dissemination (e.g., OVER-RIDE gadget)

Dissemination is often overlooked

Improve route diversity by adding spurious OVERs

Improving route diversity through the design of iBGP topologies [Pelsser, ICC08]

Providing scalable NH-diverse iBGP route redistribution to achieve sub-second switch-over time [Pelsser, Comput. Netw.10]



adding spurious OVERs increase the diversity only *locally*, but may worsen it *globally*

Dissemination is often overlooked

iBGP topology design guidelines

How to Construct a Correct and Scalable iBGP Configuration [Vutukuru, INFOCOM06]

> Lemma 3
>
> "If there exists a signaling chain between routers *A* and *B* [...] then A learns of the best route via B [...]"

Not true in presence of spurious OVERs

Having a valid signaling path is *necessary, not sufficient*

Summary of our contributions

In this work, we

- showed that iBGP Propagation rules plays a big role in iBGP
- introduced dissemination correctness
  - studied its complexity
  - provided sufficient conditions and guidelines to enforce it
- showed that dissemination is often overlooked

iBGP Deceptions:
More Sessions, Fewer Routes

Introduction and Motivation

Dissemination correctness

Revisiting the state-of-the-art

Conclusion

iBGP semantic is more complex than what is commonly assumed

Having a valid signaling path is *necessary*, not *sufficient*

Spurious OVER can invalidate simple assumptions that apparently hold in any iBGP topology

It provides new motivations to recent proposals for decoupling route *propagation* from route *selection*