Methods and Techniques for Disruption-free Network Reconfiguration

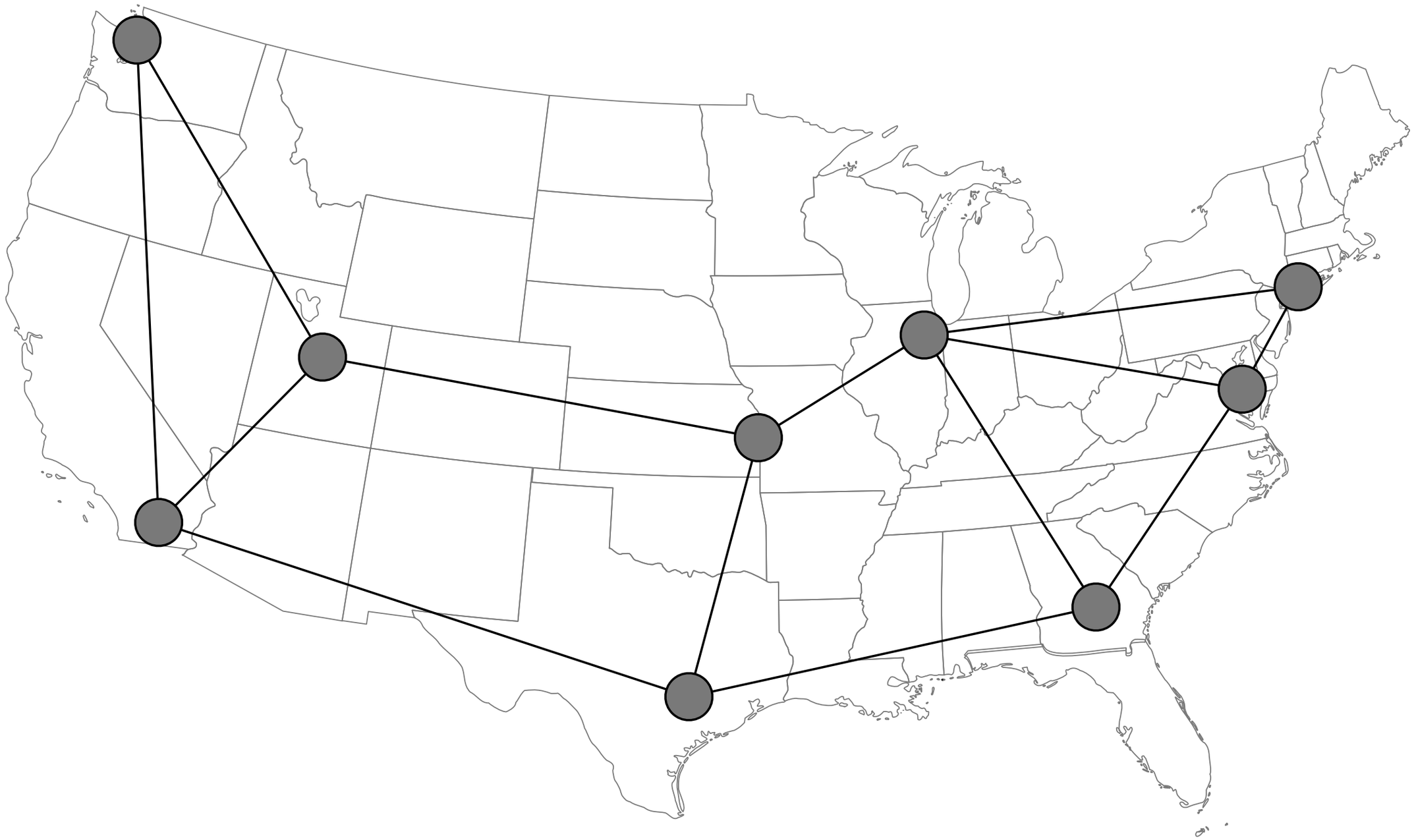
Laurent Vanbever

laurent.vanbever@uclouvain.be

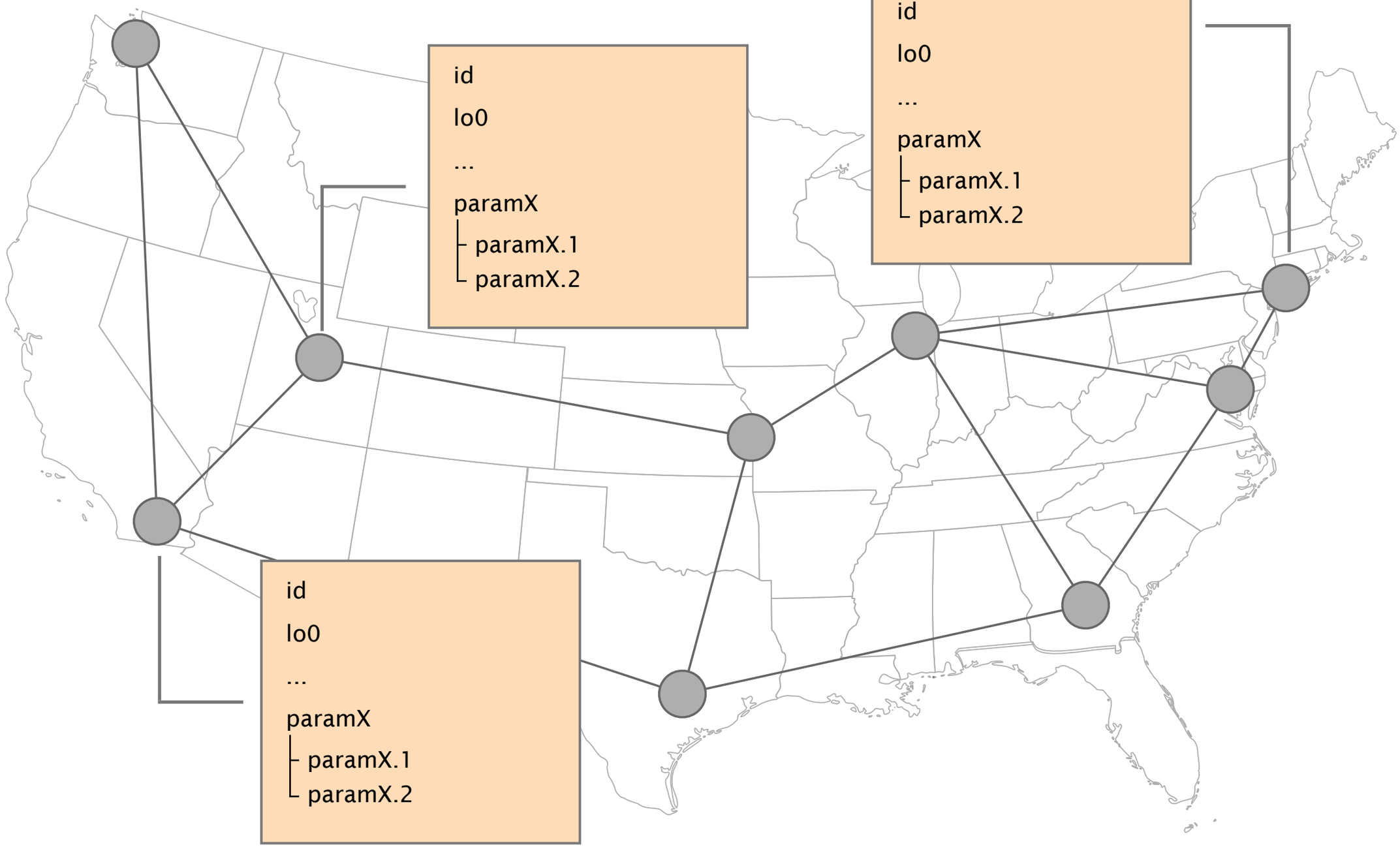
FIArch Workshop

September 27, 2012

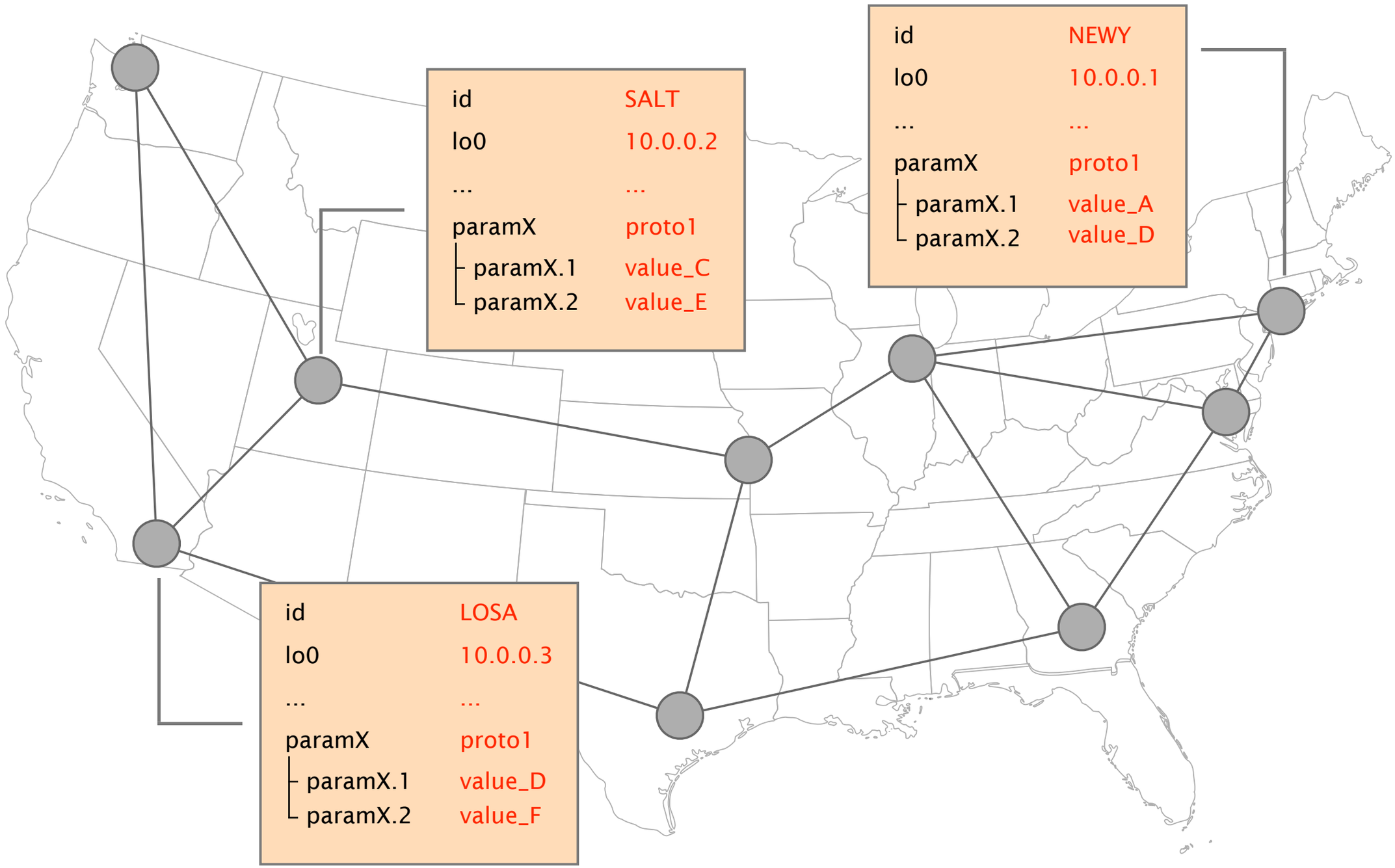
A network is a distributed system



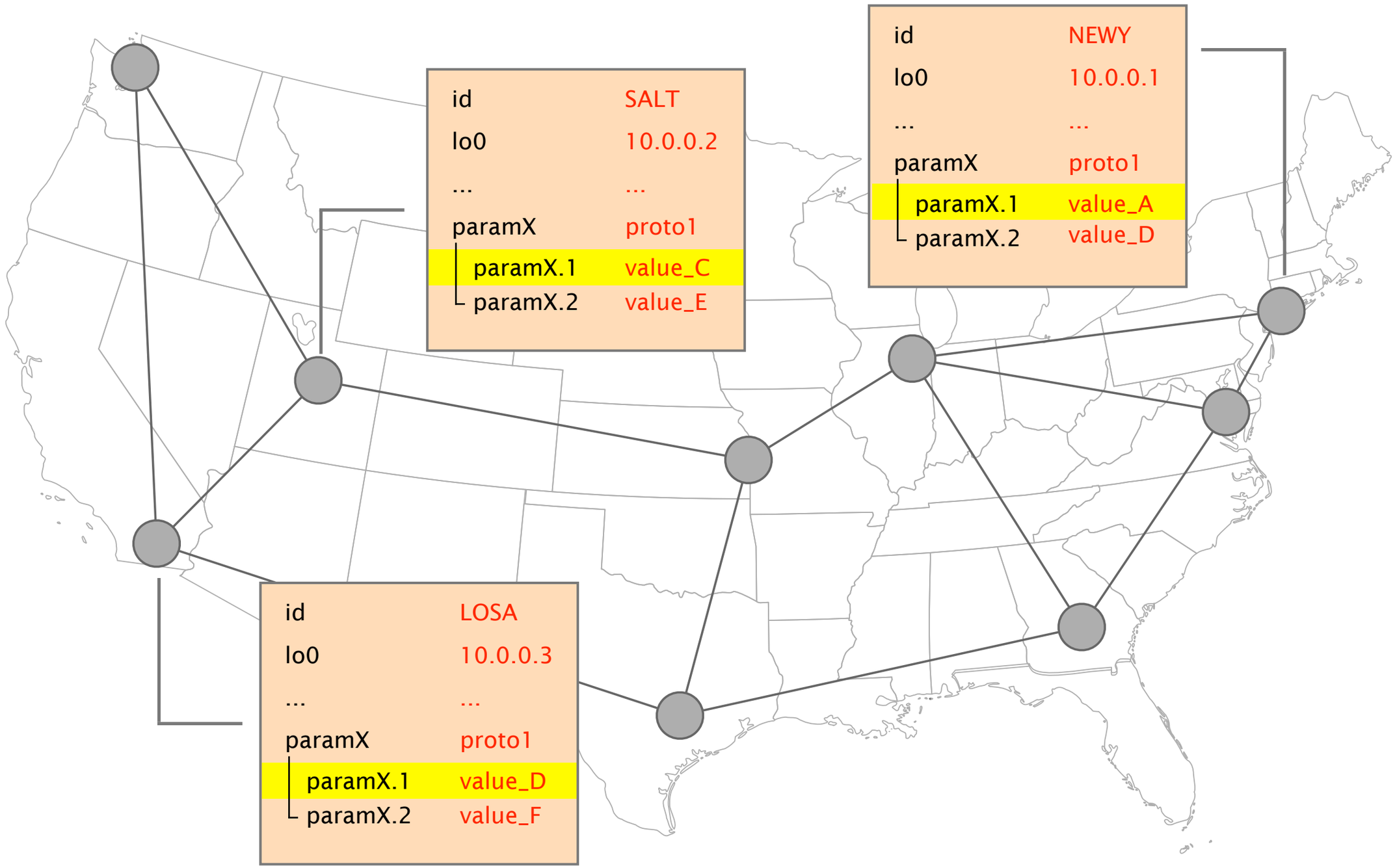
A network is a distributed system
with a distributed configuration



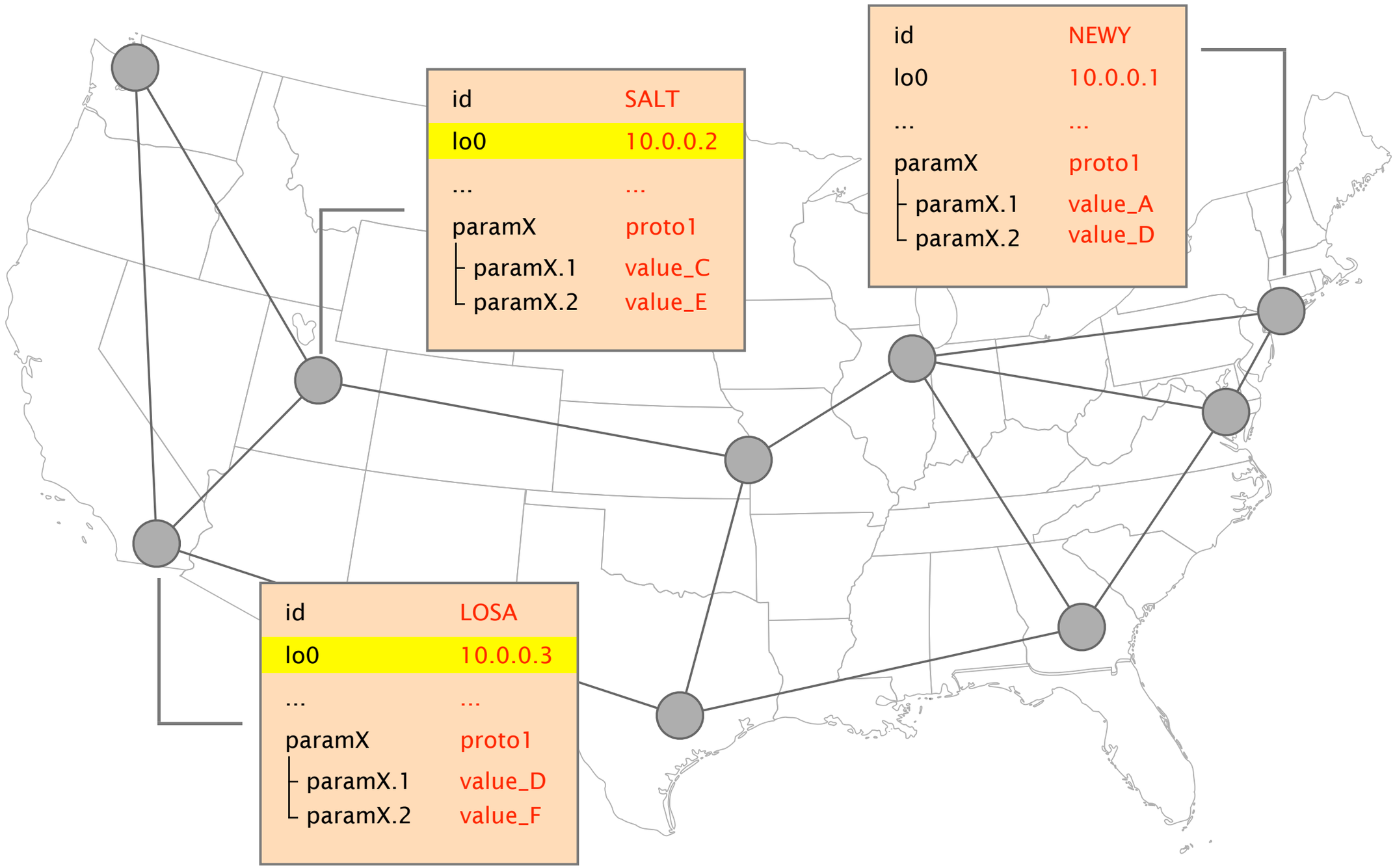
Configuring the network consists in defining the value of each parameter



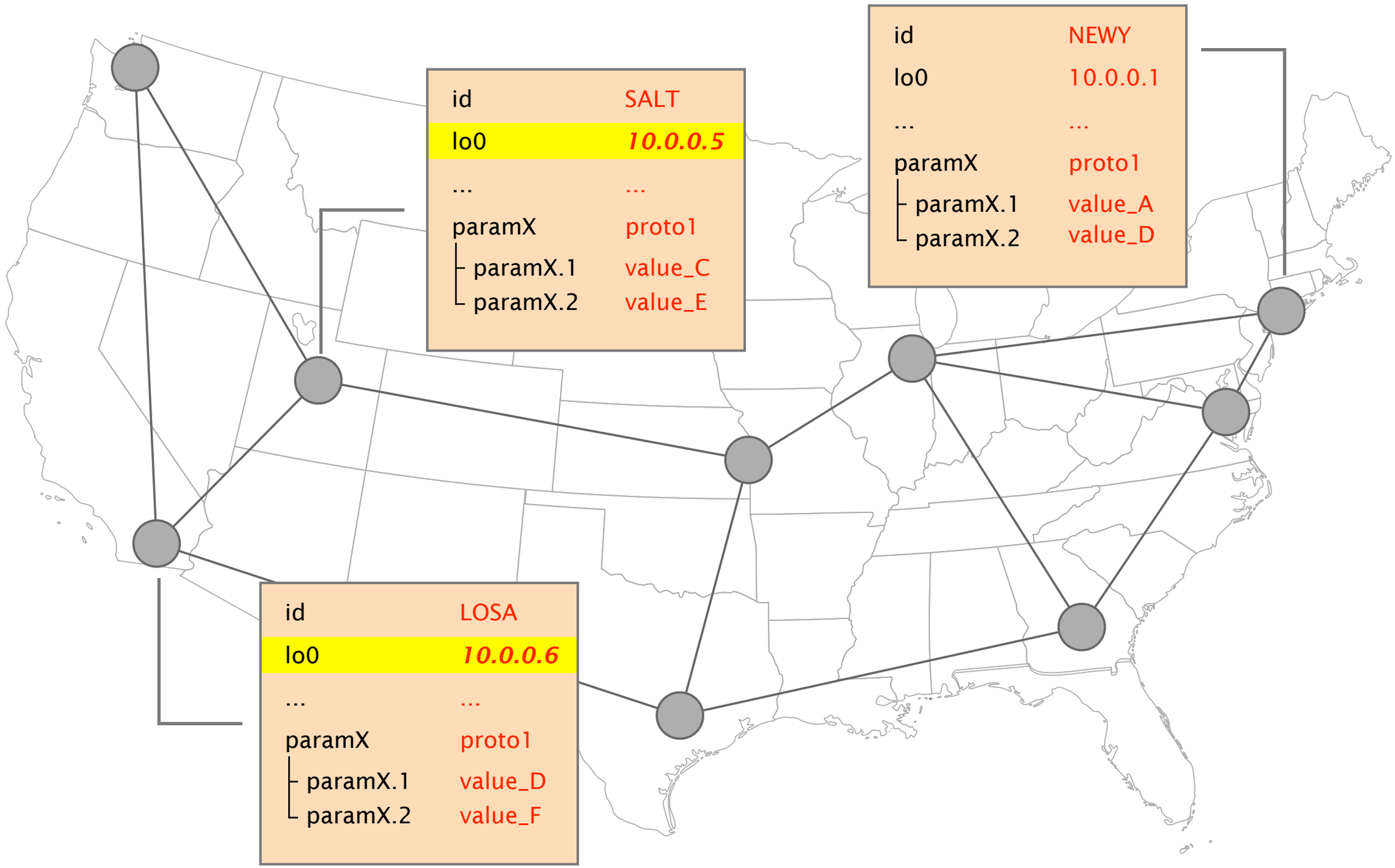
For the network to work properly, each parameter must be consistent network-wide



Reconfiguring the network consists in modifying some configuration parameters



Reconfiguring the network consists in modifying some configuration parameters



Network reconfiguration is a day-to-day task

Configuring the network from scratch is done only once

Everything change after is a reconfiguration

Typical reconfigurations scenario include

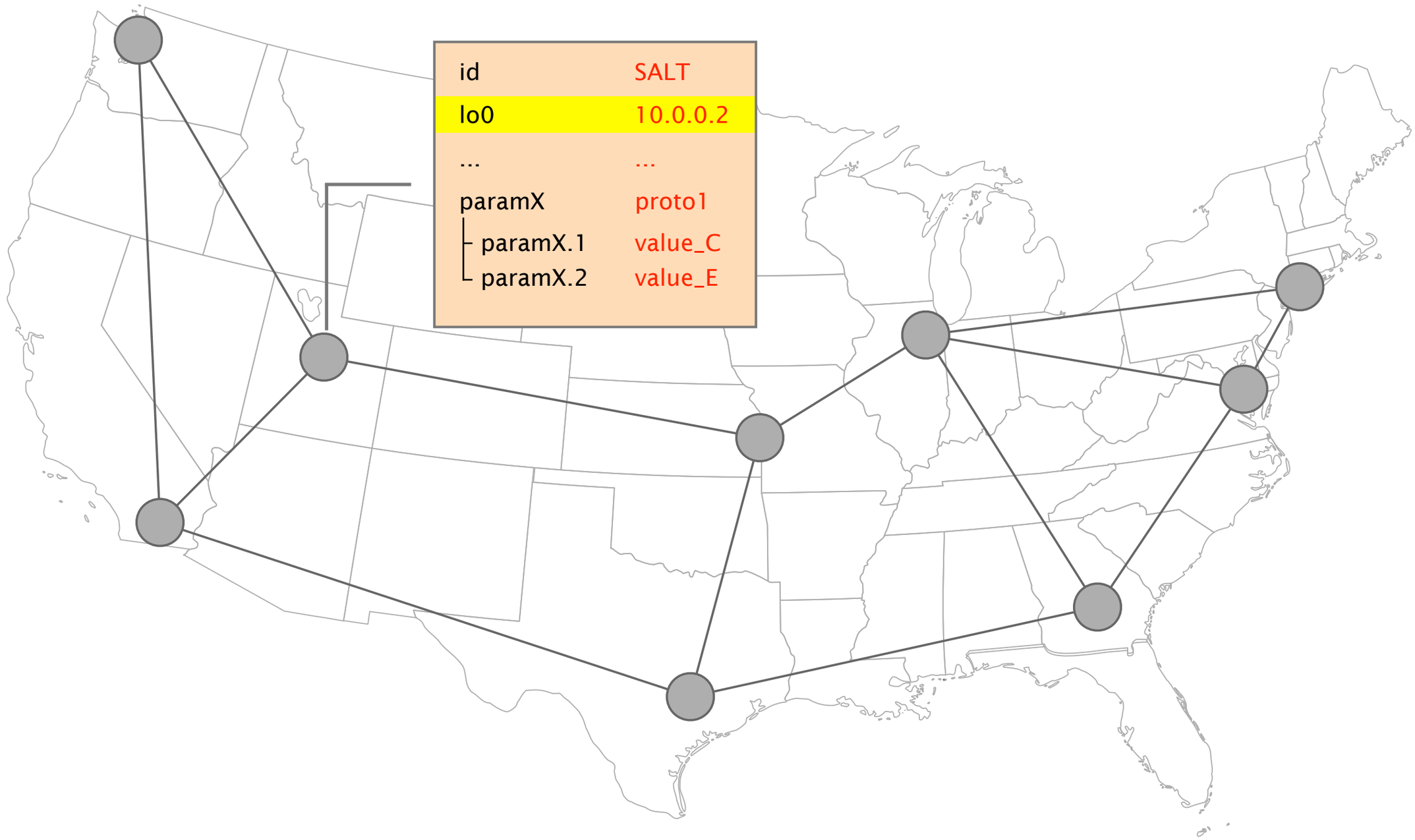
- Updating the physical or logical infrastructure
- Managing resources (e.g., bandwidth, CPU, memory)
- Deploying new services

Network reconfiguration is hardly done right

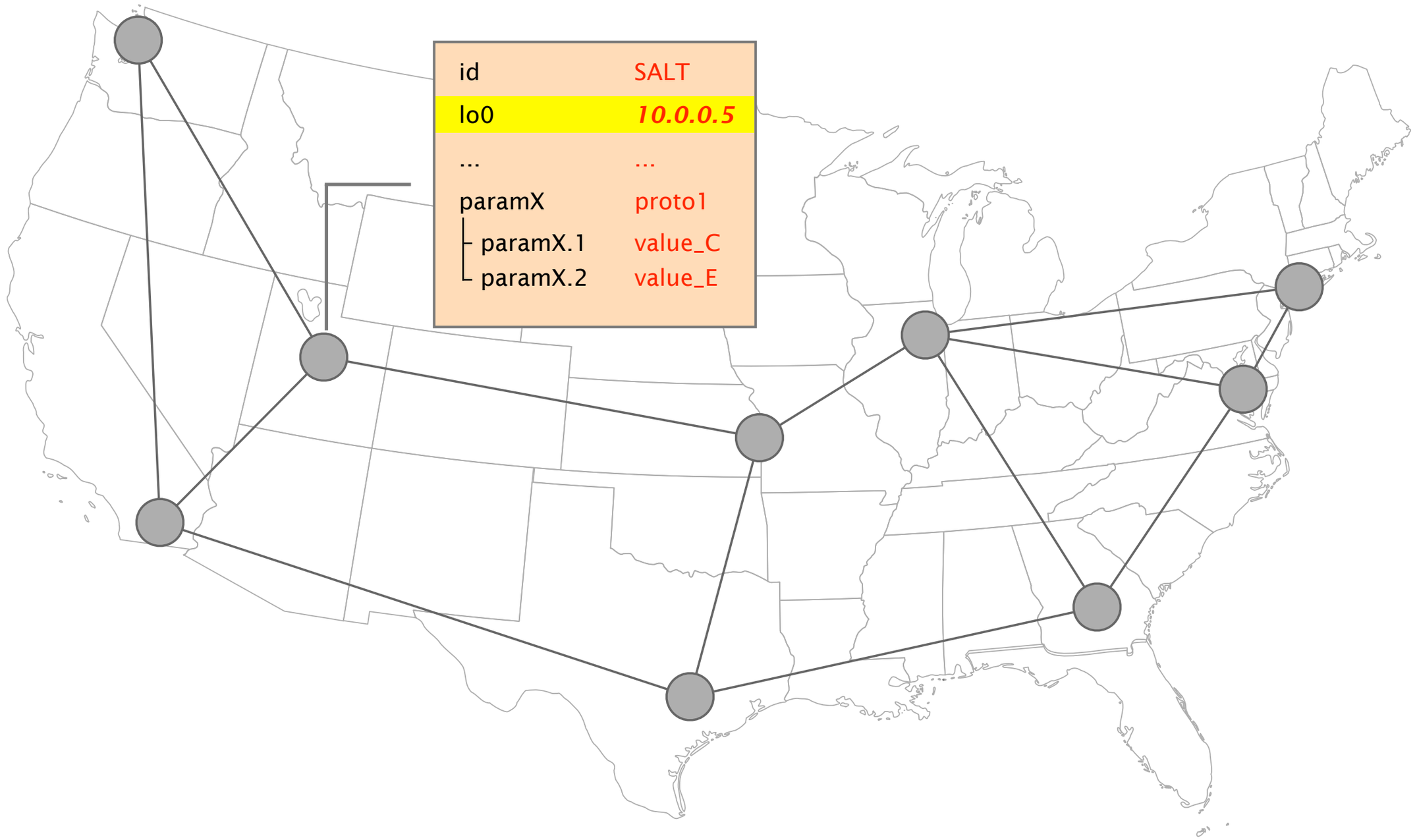
Manually change a running network

device-by-device, using proprietary, low-level CLI interfaces

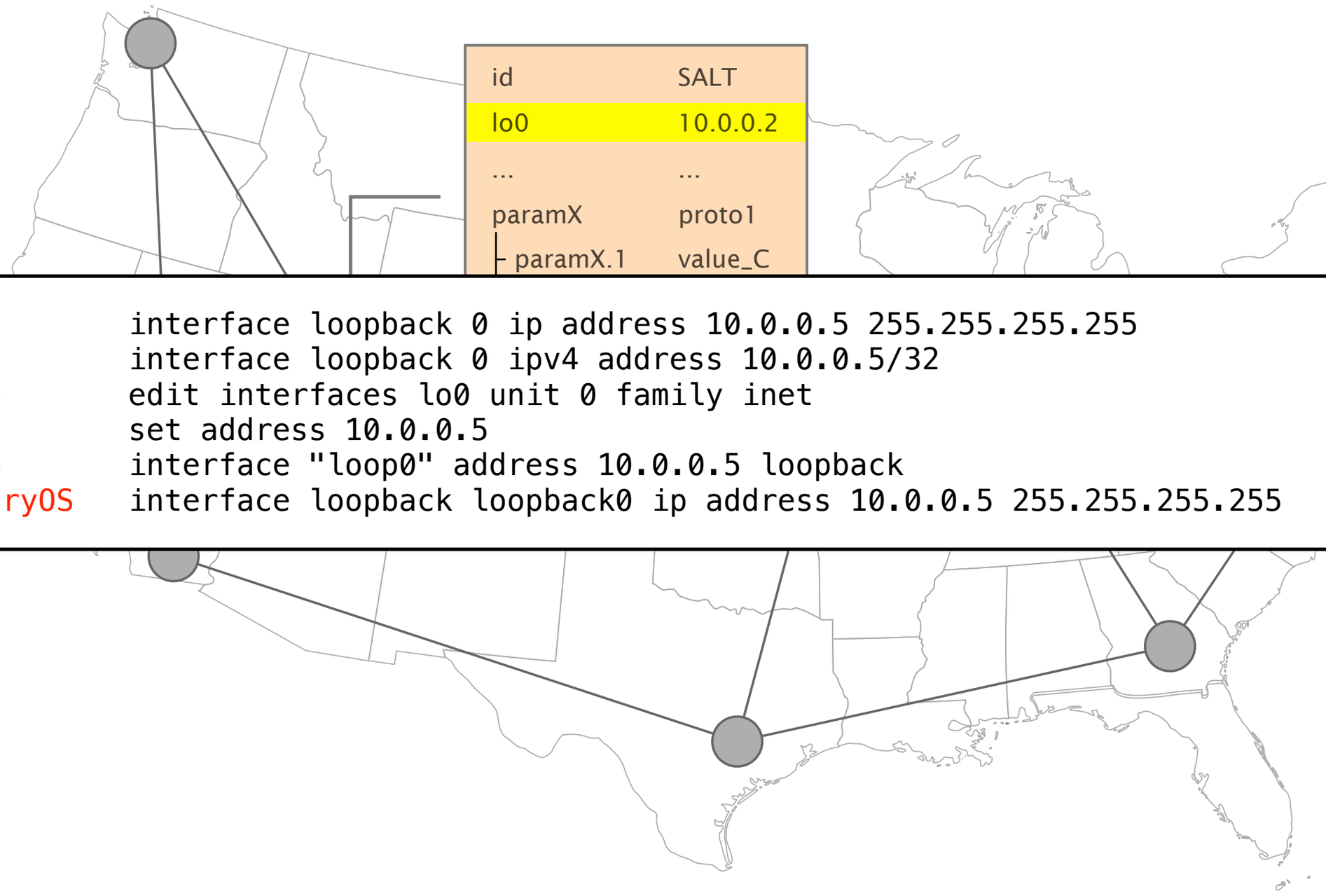
Most networks are still configured manually using heterogenous low-level interfaces



Most networks are still configured manually using heterogenous low-level interfaces



Most networks are still configured manually using heterogenous low-level interfaces




```
IOS          interface loopback 0 ip address 10.0.0.5 255.255.255.255
IOX          interface loopback 0 ipv4 address 10.0.0.5/32
JunOS        edit interfaces lo0 unit 0 family inet
             set address 10.0.0.5
TimOS        interface "loop0" address 10.0.0.5 loopback
FoundryOS    interface loopback loopback0 ip address 10.0.0.5 255.255.255.255
```

Network reconfiguration is hardly done right

Manually change a running network

device-by-device, using proprietary, low-level CLI interfaces

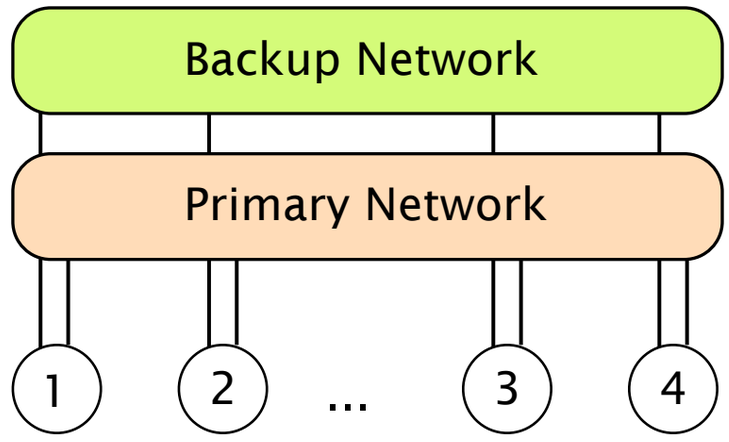
Ensuring consistency in every intermediate step

coordinating the changes across the entire network

Face routing and forwarding anomalies

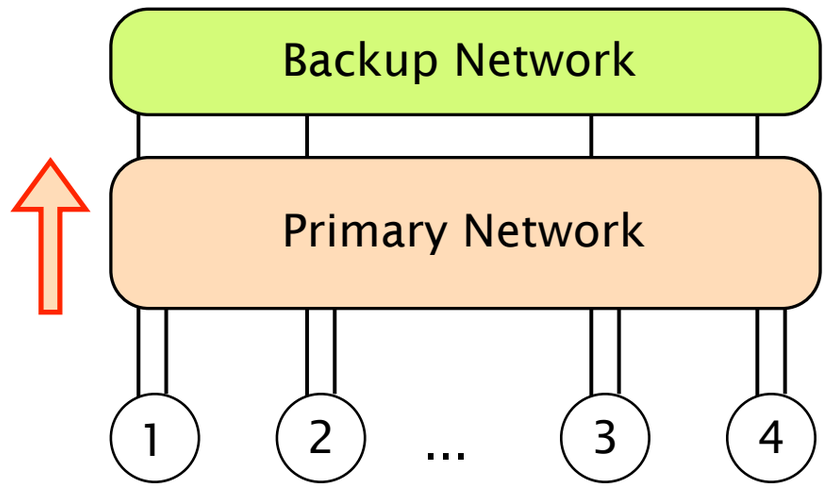
as non-reconfigured routers interact with reconfigured ones

At 12:47 AM PDT on April 21st, a network change was performed as part of our normal AWS scaling activities [...] The configuration change was to upgrade the capacity of the primary network.



Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

At 12:47 AM PDT on April 21st, a network change was performed as part of our normal AWS scaling activities [...] The configuration change was to upgrade the capacity of the primary network.

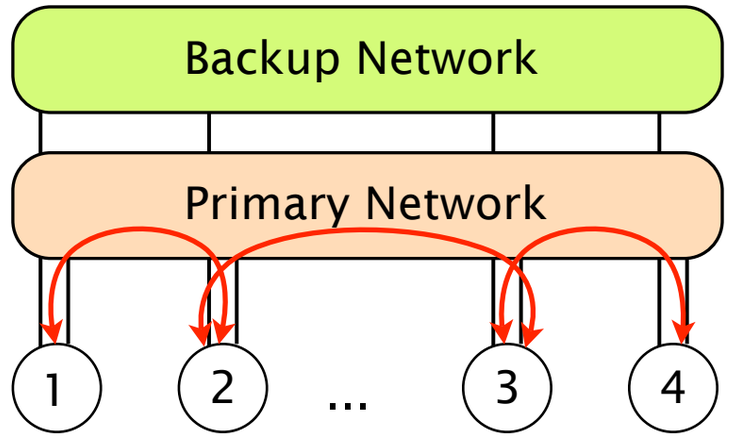


Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

amazon web services™

At 12:47 AM PDT on April 21st, a network change was performed as part of our normal AWS scaling activities [...]. The configuration change was to upgrade the capacity of the primary network.

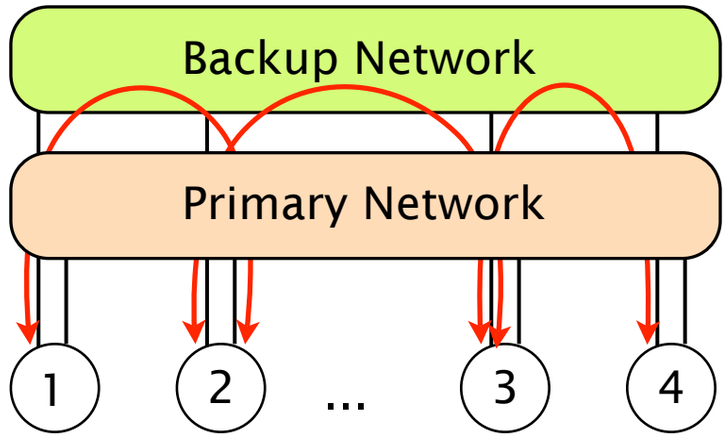
During the change, one of the standard steps is to shift traffic off of one of the redundant routers in the primary EBS network to allow the upgrade to happen.



At 12:47 AM PDT on April 21st, a network change was performed as part of our normal AWS scaling activities [...]. The configuration change was to upgrade the capacity of the primary network.

During the change, one of the standard steps is to shift traffic off of one of the redundant routers in the primary EBS network to allow the upgrade to happen.

The traffic shift was executed incorrectly and rather than routing the traffic to the other router on the primary network, the traffic was routed onto the lower capacity redundant EBS network [...]



Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

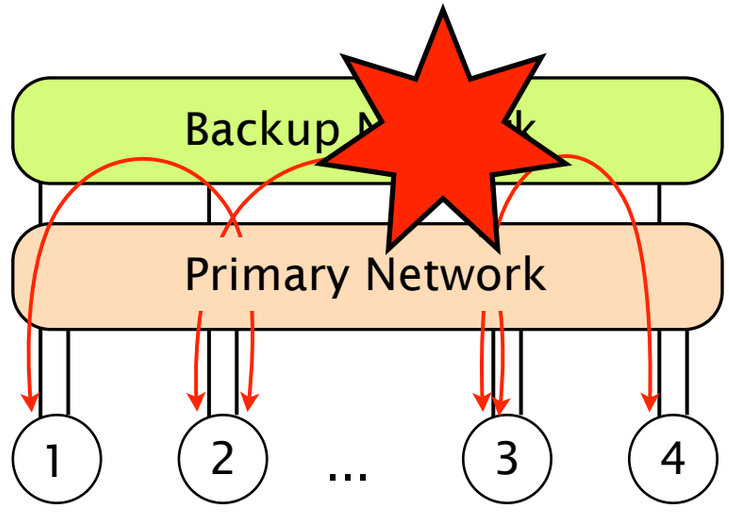
amazon web services™

At 12:47 AM PDT on April 21st, a network change was performed as part of our normal AWS scaling activities [...]. The configuration change was to upgrade the capacity of the primary network.

During the change, one of the standard steps is to shift traffic off of one of the redundant routers in the primary EBS network to allow the upgrade to happen.

The traffic shift was executed incorrectly and rather than routing the traffic to the other router on the primary network, the traffic was routed onto the lower capacity redundant EBS network [...]

Unlike a normal network interruption, this change disconnected both the primary and secondary network simultaneously, leaving the affected nodes completely isolated from one another.

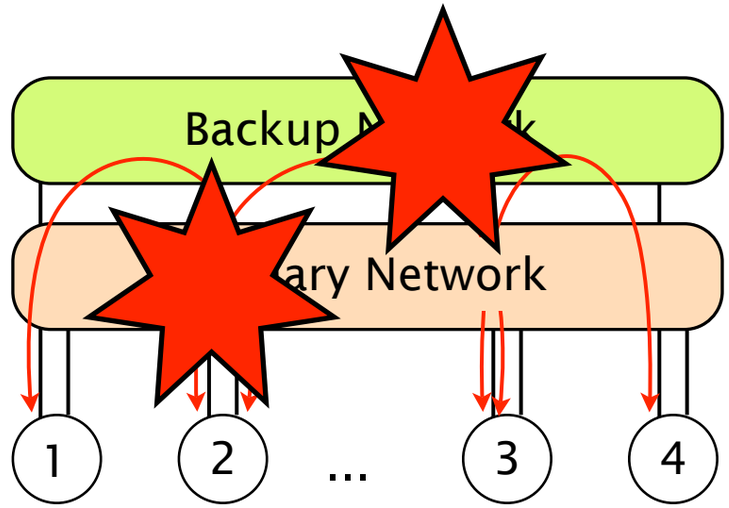


amazon web services™

At 12:47 AM PDT on April 21st, a network change was performed as part of our normal AWS scaling activities [...]. The configuration change was to upgrade the capacity of the primary network.

During the change, one of the standard steps is to shift traffic off of one of the redundant routers in the primary EBS network to allow the upgrade to happen.

The traffic shift was executed incorrectly and rather than routing the traffic to the other router on the primary network, the traffic was routed onto the lower capacity redundant EBS network [...]



Unlike a normal network interruption, this change disconnected both the primary and secondary network simultaneously, leaving the affected nodes completely isolated from one another.

Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

Amazon is currently experiencing a degradation. They are working on it. We are still waiting on them to get to our volumes. Sorry.

amazon web services™

change was
ling activities [...].
the capacity of the

teps is to shift traffic
primary EBS network

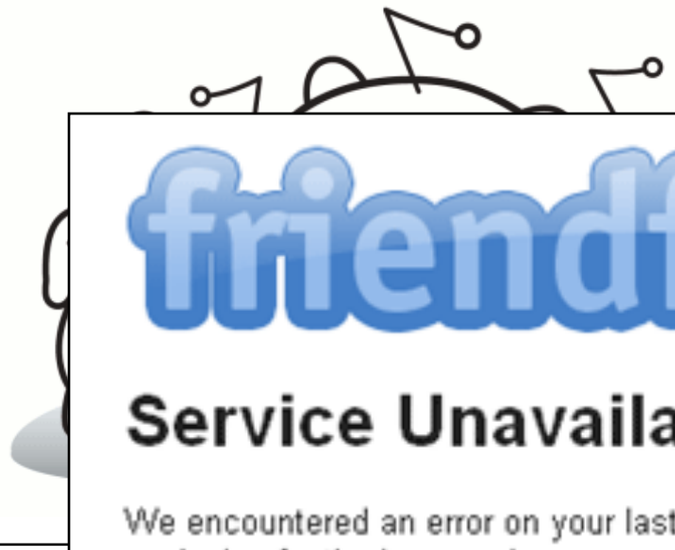
The traffic shift was executed incorrectly and rather than routing the traffic to the other router on the primary network, the traffic was routed onto the lower capacity redundant EBS network [...]

Unlike a normal network interruption, this change disconnected both the primary and secondary network simultaneously, leaving the affected nodes completely isolated from one another.

Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

Amazon is currently experiencing a degradation. They are working on it. We are still waiting on them to get to our volumes. Sorry.

reddit is down.



friendfeed

Service Unavailable

We encountered an error on your last request. Our service is new, and we are just working out the kinks. We apologize for the inconvenience.

amazon web services™

change was
ling activities [...].
the capacity of the

The
routing the traffic to the other router on the primary network, the traffic was routed onto the lower capacity redundant EBS network [...]

Unlike a normal network interruption, this change disconnected both the primary and secondary network simultaneously, leaving the affected nodes completely isolated from one another.

Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

Amazon is currently experiencing a degradation. They are working on it. We are still waiting on them to get to our volumes. Sorry.

reddit is down.

friendfeed

Serv

We enco
apologize

Quora

A continually improving collection of questions and answers created, edited, and organized by everyone who uses it.

We're currently having an unexpected outage, and are working to get the site back up as soon as possible. Thanks for your patience.

amazon web services™

change was
ling activities [...].
the capacity of the

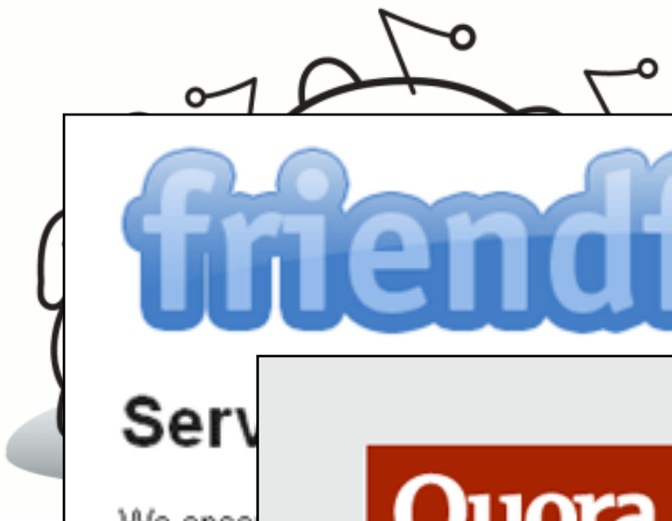
The
routing the tr
the traffic wa
network [...]

Unlike a normal network interruption, this change disconnected both the primary and secondary network simultaneously, leaving the affected nodes completely isolated from one another.

Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

Amazon is currently experiencing a degradation. They are working on it. We are still waiting on them to get to our volumes. Sorry.

reddit is down.



friendfeed

Serv

We enco
apologize

Quora

We're curre
site back up

A continually improving collection of questions and answers

foursquare

Sorry! We're having technical difficulties

Latest post from status.foursquare.com:

Thu Apr 21 2011

This morning's downtime and slowness

Hi all,

Our usually-amazing datacenter hosts, Amazon EC2, are having a few hiccups this morning, which affected us and a bunch of other services that use them. Everything looks to be getting back to normal now. We'll update this when we have the all clear. Thanks for your patience.

amazon web services™

change was
ling activities [...].
the capacity of the

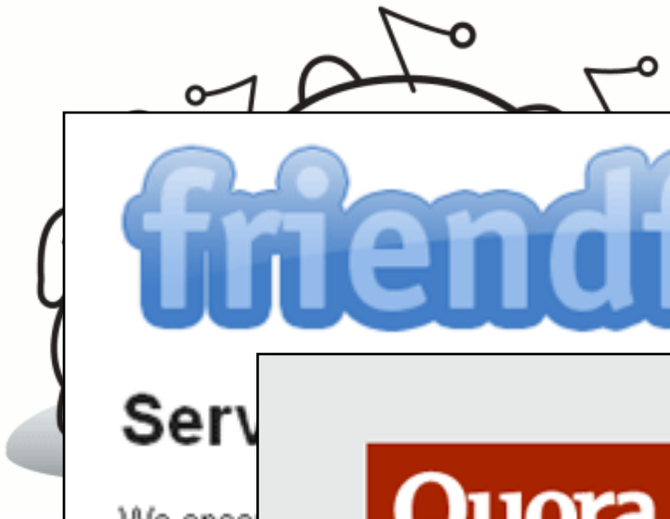
The
routing the tr
the traffic wa
network [...]

Unlike a normal network ir
both the primary and seco
the affected nodes comple

Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

Amazon is currently experiencing a degradation. They are working on it. We are still waiting on them to get to our volumes. Sorry.

reddit is down.



friendfeed

Serv

We enco
apologize

Quora

A continually improving collection of questions and answers

We're curre
site back up

foursquare

Sorry! We'
Latest post from s

Thu Apr 21 2011
This morning's do

Hi all,

Our usually-amazing datacent
which affected us and a bunch
back to normal now. We'll upd

Owls need a break sometimes too

We'll be back in action shortly -- in the meantime go outside and flap your arms around, you may find that flying ain't very easy.

In the meantime, if you can't wait to send a Tweet, head over to Twitter web to share your 140 character musings.

フクロウはときどき休まないといけないのです。

復旧するまでそれほど長い時間はかからないと思います。その間、ちょっと外に出掛けてみて、腕をぐっと伸ばし、そして空高く羽ばたくことは実際には結構難しいのではないかなどと考察してみるのはいかがでしょうか。

ツイートするのが待ちきれない方は、直接Twitterを開き、あなたの思惑を140字で投稿してみましょう。

hootsuite

amazon web services™

change was
ling activities [...].
the capacity of the

The
routing the tr
the traffic wa
network [...]

Unlike a normal network in
both the primary and seco
the affected nodes comple

Summary of the Amazon EC2 and Amaz

Amazon is currently experiencing a degradation. They are working on it. We are still waiting on them to get to our volumes. Sorry.

reddit is down.

change was
ling activities [...].
the capacity of the

friendfeed

The trigger for this event was a poorly executed network reconfiguration

the traffic wa
network [...]

site back up

Latest post from s

Thu Apr 21 2011

This morning's do

Hi all,

Our usually-amazing datacent
which affected us and a bunch
back to normal now. We'll upd

Unlike a normal network in
both the primary and seco
the affected nodes comple

Summary of the Amazon EC2 and Amaz

We'll be back in action shortly -- in the meantime go outside and flap your arms around, you may find that flying ain't very easy.

In the meantime, if you can't wait to send a Tweet, head over to Twitter web to share your 140 character musings.

フクロウはときどき休まないといけないのです。

復旧するまでそれほど長い時間はかからないと思います。その間、ちょっと外に出掛けてみて、腕をぐっと伸ばし、そして空高く羽ばたくことは実際には結構難しいのではないかなどと考察してみるのはいかがでしょうか。

ツイートするのが待ちきれない方は、直接Twitterを開き、あなたの思惑を140字で投稿してみましょう。

hootsuite

Our goal is to enable anomaly-free routing reconfiguration

Progressively modify the configuration of a running network without creating any anomaly

Our approach

Develop reconfiguration techniques which are provably correct, efficient, automatic, backward compatible

Methods and Techniques for Disruption-free Network Reconfiguration

IGP reconfiguration

BGP reconfiguration

Principles

Methods and Techniques for Disruption-free Network Reconfiguration

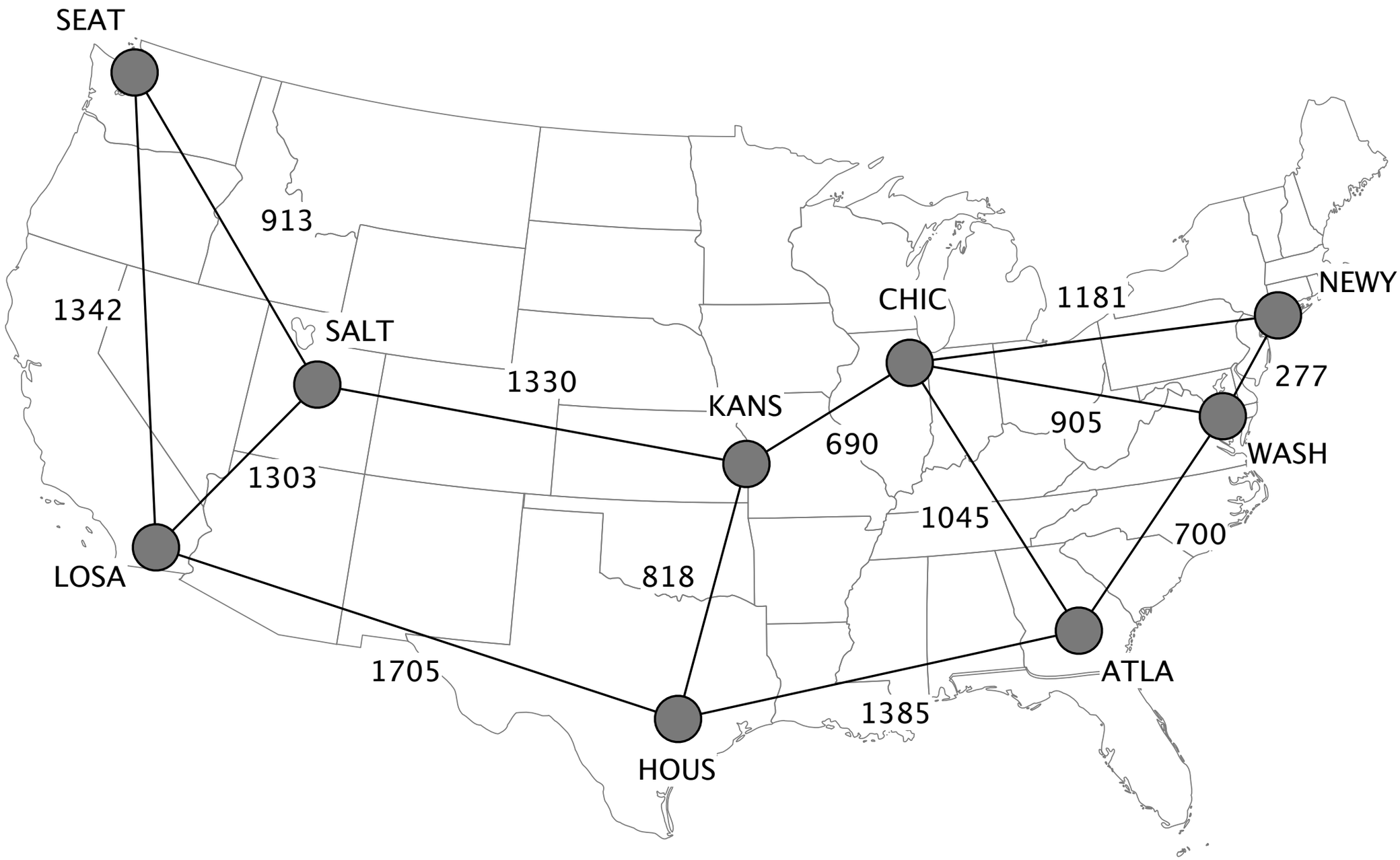
IGP reconfiguration

BGP reconfiguration

Principles

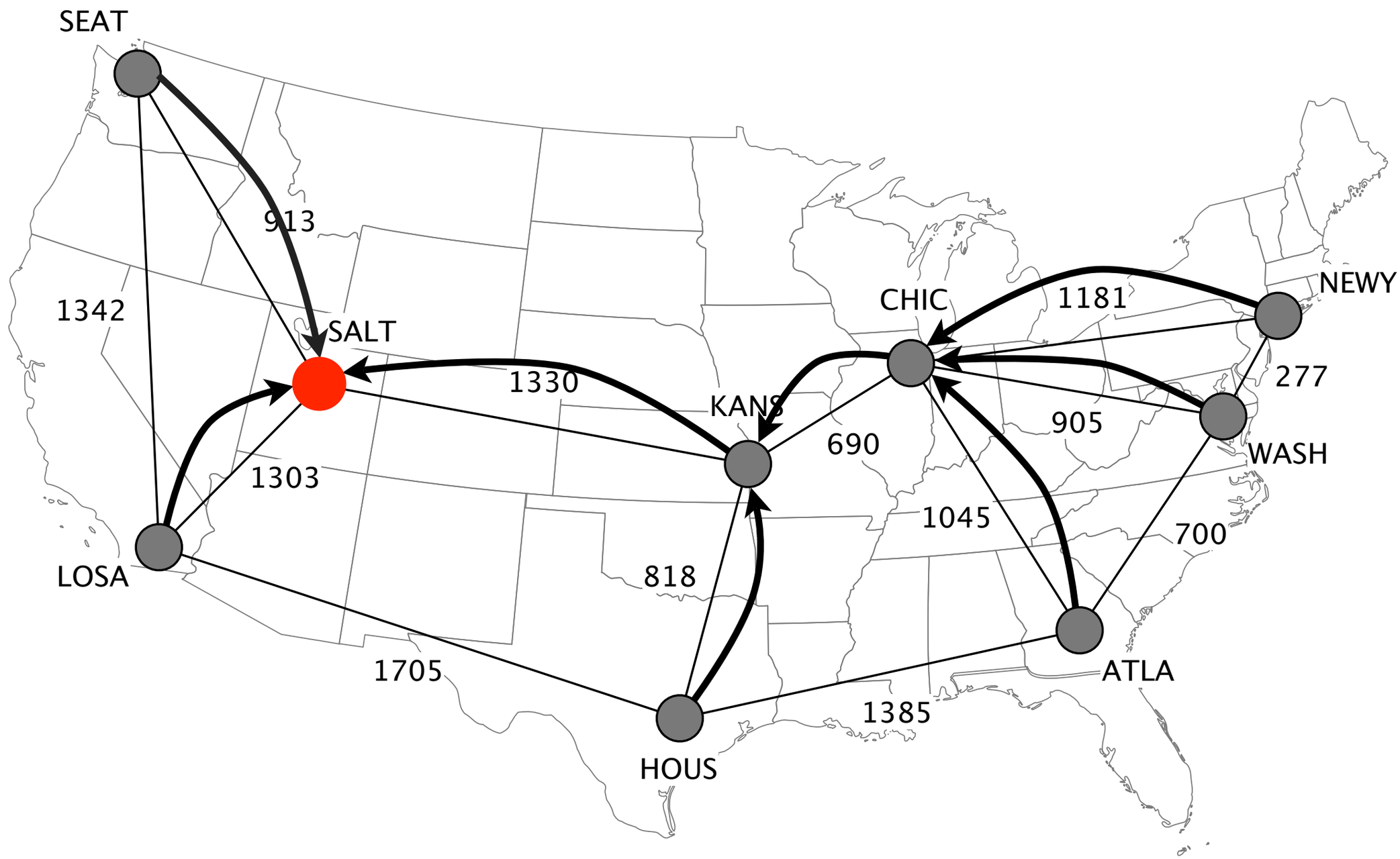
Intradomain routing protocols (IGP) rule traffic forwarding within a routing domain

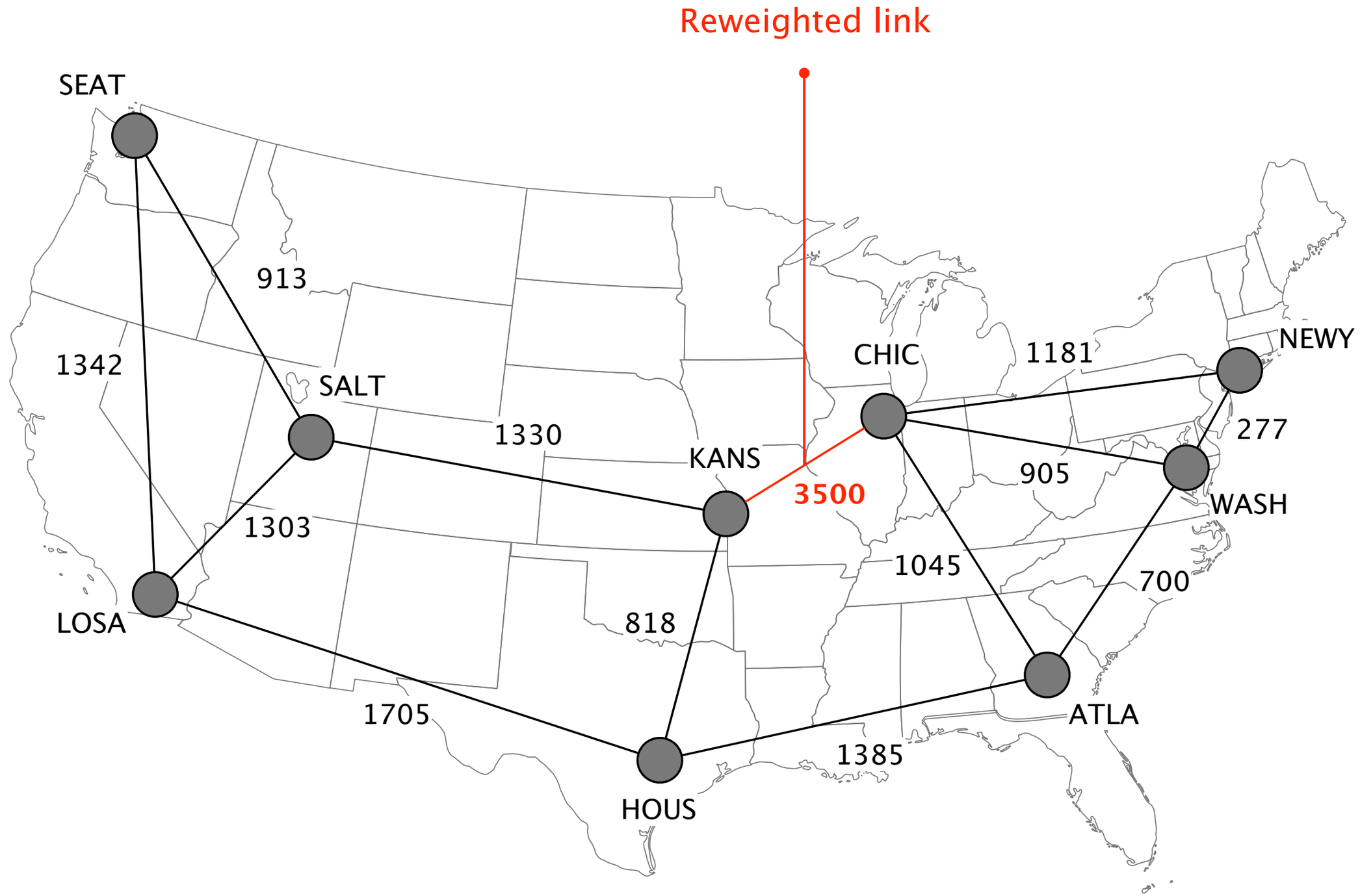
The US research network (Abilene/Internet2)



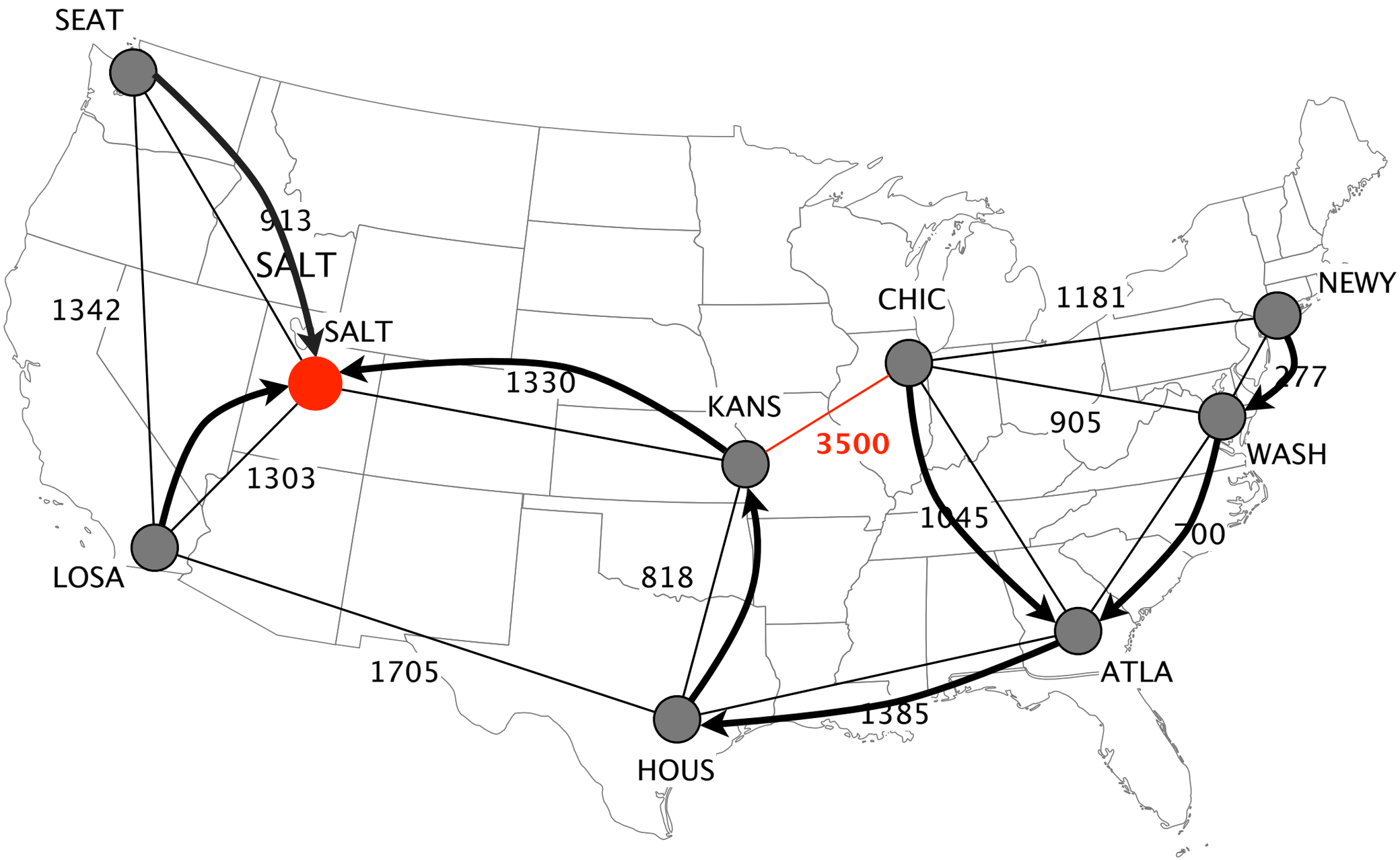
IGP enables each router to compute the shortest path to reach every other router

Forwarding paths towards SALT



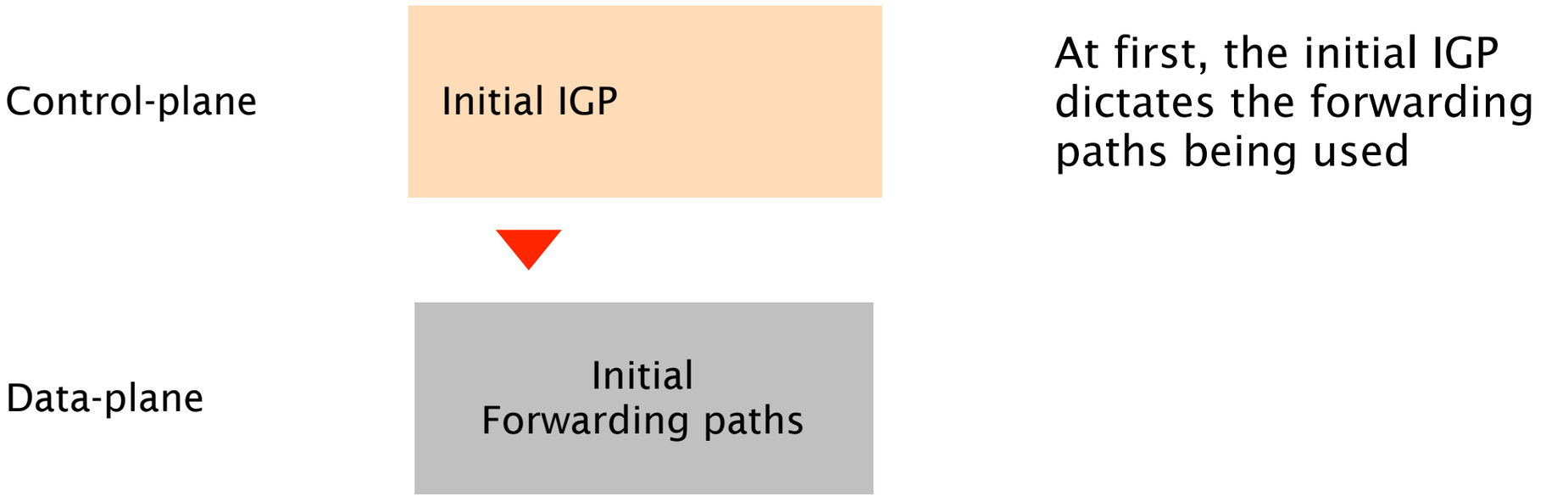

Reweighted link
SEAT
913
1342
SALT
1330
KANS
3500
CHIC
1181
NEWY
277
905
WASH
1303
1045
700
LOSA
818
1705
ATLA
1385
HOUS

Final forwarding paths towards SALT



Reconfiguring the IGP usually requires running two routing planes (*)

Abstract model of a router

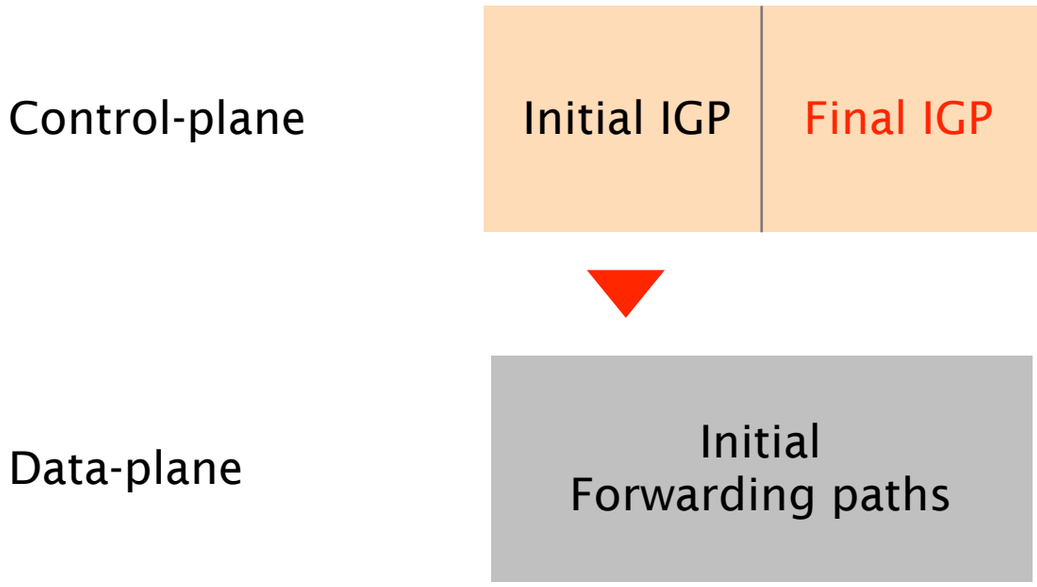


At first, the initial IGP dictates the forwarding paths being used

(*) [Gill03, Pepelnjak07, Herrero10, Smith12]

Reconfiguring the IGP usually requires running two routing planes (*)

Abstract model of a router

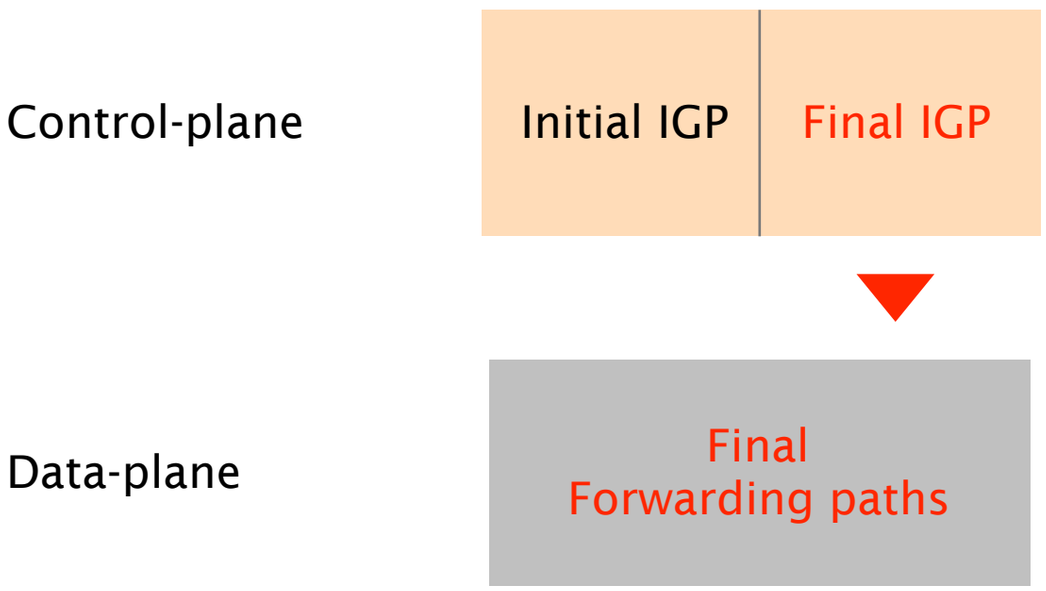


Then, the final IGP is introduced without changing the forwarding

(*) [Gill03, Pepelnjak07, Herrero10, Smith12]

Reconfiguring the IGP usually requires running two routing planes (*)

Abstract model of a router

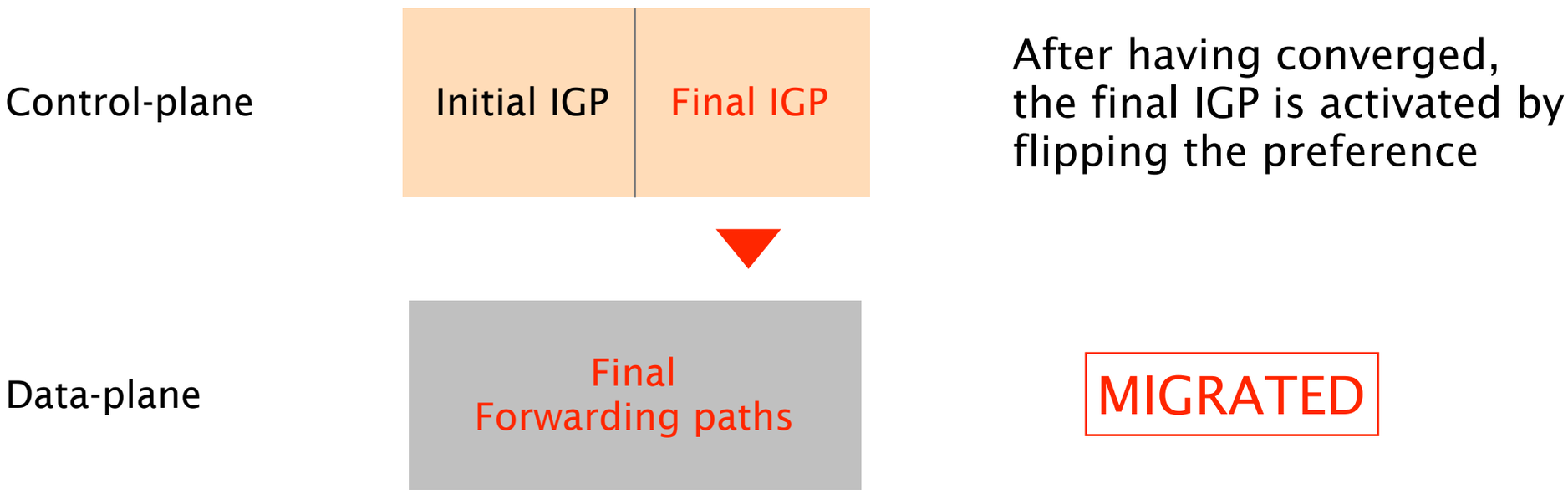


After having converged, the final IGP is activated by flipping the preference

(*) [Gill03, Pepelnjak07, Herrero10, Smith12]

Reconfiguring the IGP usually requires running two routing planes (*)

Abstract model of a router



After having converged, the final IGP is activated by flipping the preference

MIGRATED

(*) [Gill03, Pepelnjak07, Herrero10, Smith12]

problem

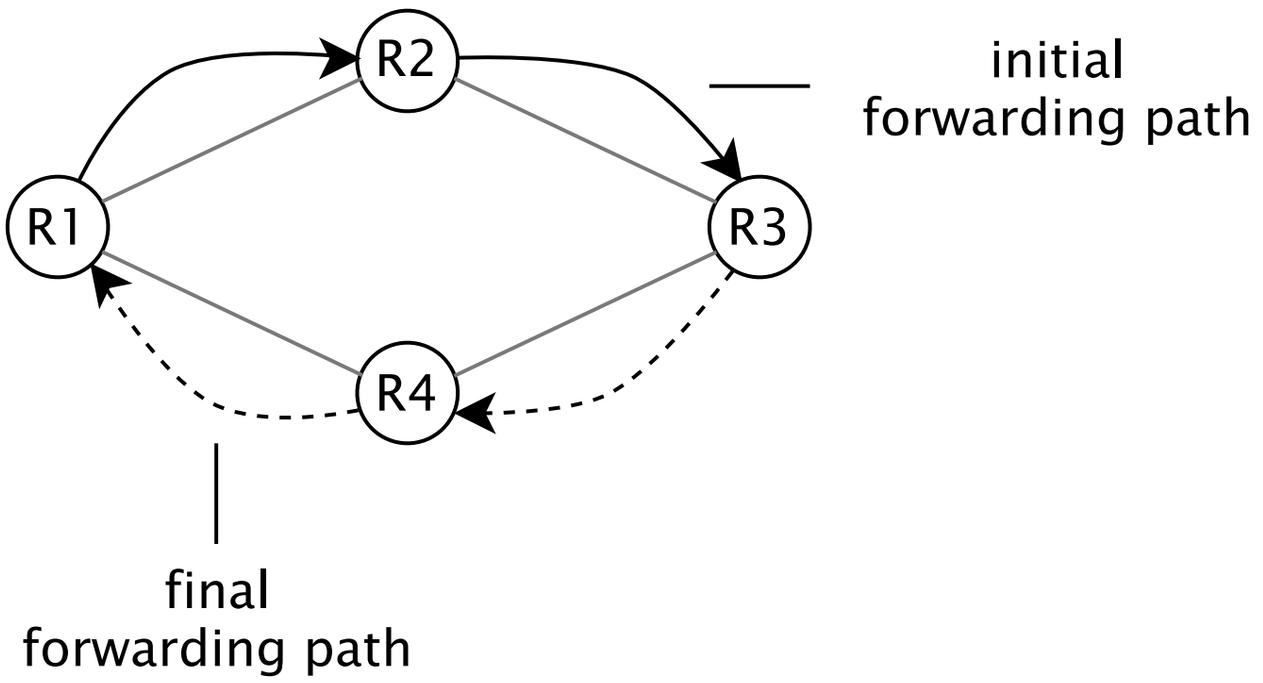
Find an ordering in which to activate the final IGP without causing any forwarding anomalies

problem Find an ordering in which to activate the final IGP without causing any forwarding anomalies

which ones?

There are two types of forwarding anomalies
forwarding loops

forwarding loop — when the actual forwarding path contains a cycle

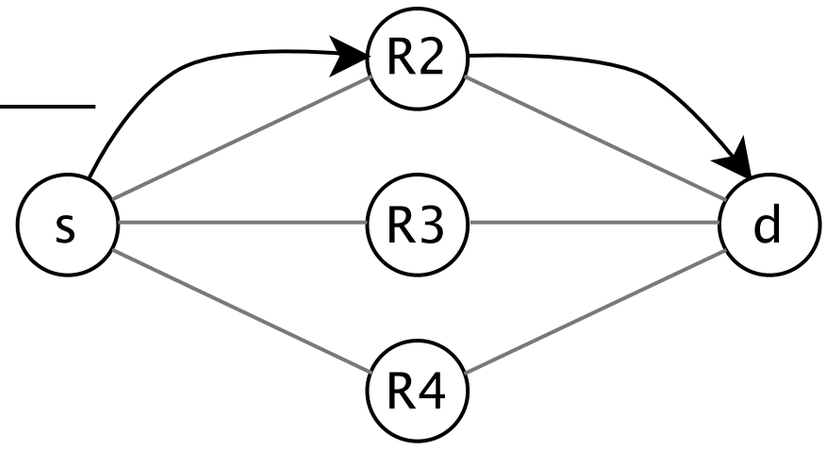


There are two types of forwarding anomalies
forwarding loops and traffic-shifts

traffic shift

when the actual forwarding paths
changes more than once
during the reconfiguration

initial and final
forwarding paths

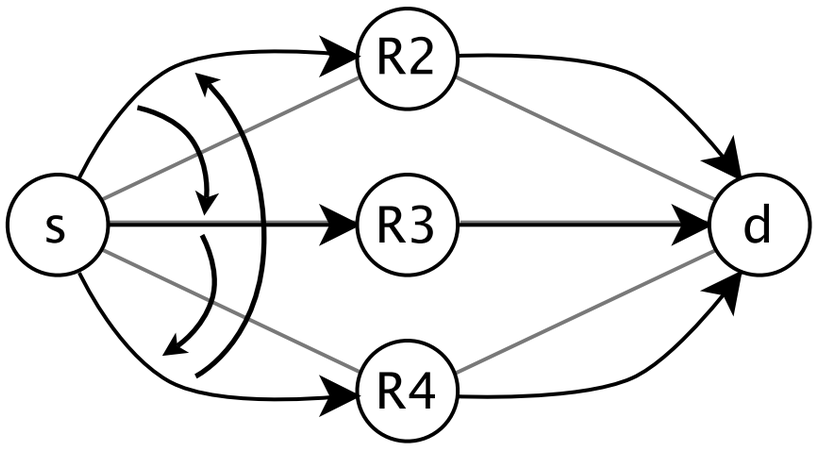


There are two types of forwarding anomalies
forwarding loops and traffic-shifts

traffic shift

when the actual forwarding paths
changes more than once
during the reconfiguration

traffic shifts: 0



Potential forwarding anomalies depend on the type of IGP involved in the reconfiguration

	Link-State	Distance-Vector
Link-State	forwarding loop traffic shift	no forwarding loops traffic shift
Distance-Vector	no forwarding loop traffic shift	no forwarding loops traffic shift

Traffic shift are pervasive and can appear in any reconfiguration case

	Link-State	Distance-Vector
Link-State	forwarding loop **traffic shift**	no forwarding loops **traffic shift**
Distance-Vector	no forwarding loop **traffic shift**	no forwarding loops **traffic shift**

Surprisingly, forwarding loops can appear only in pure link-state scenarios

	Link-State	Distance-Vector
Link-State	**forwarding loop** traffic shift	no forwarding loops traffic shift
Distance-Vector	no forwarding loop traffic shift	no forwarding loops traffic shift

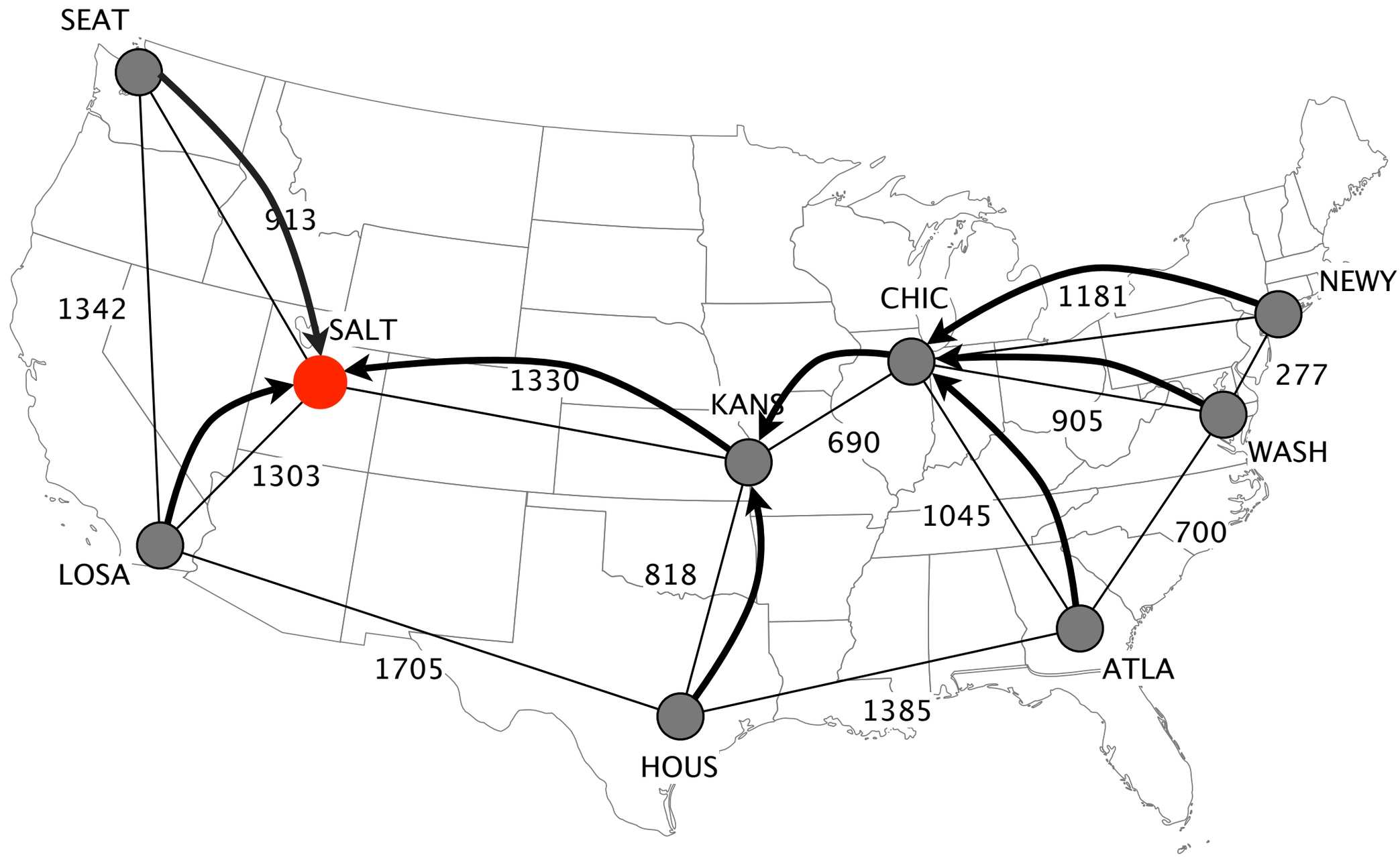
We considered the two most practically relevant scenarios

	Link-State	Distance-Vector
Link-State	forwarding loop traffic shift	no forwarding loops traffic shift
Distance-Vector	no forwarding loop traffic shift	no forwarding loops traffic shift

	Link-State	Distance-Vector
Link-State	forwarding loop traffic shift	no forwarding loops traffic shift
Distance-Vector	no forwarding loop traffic shift	no forwarding loops traffic shift

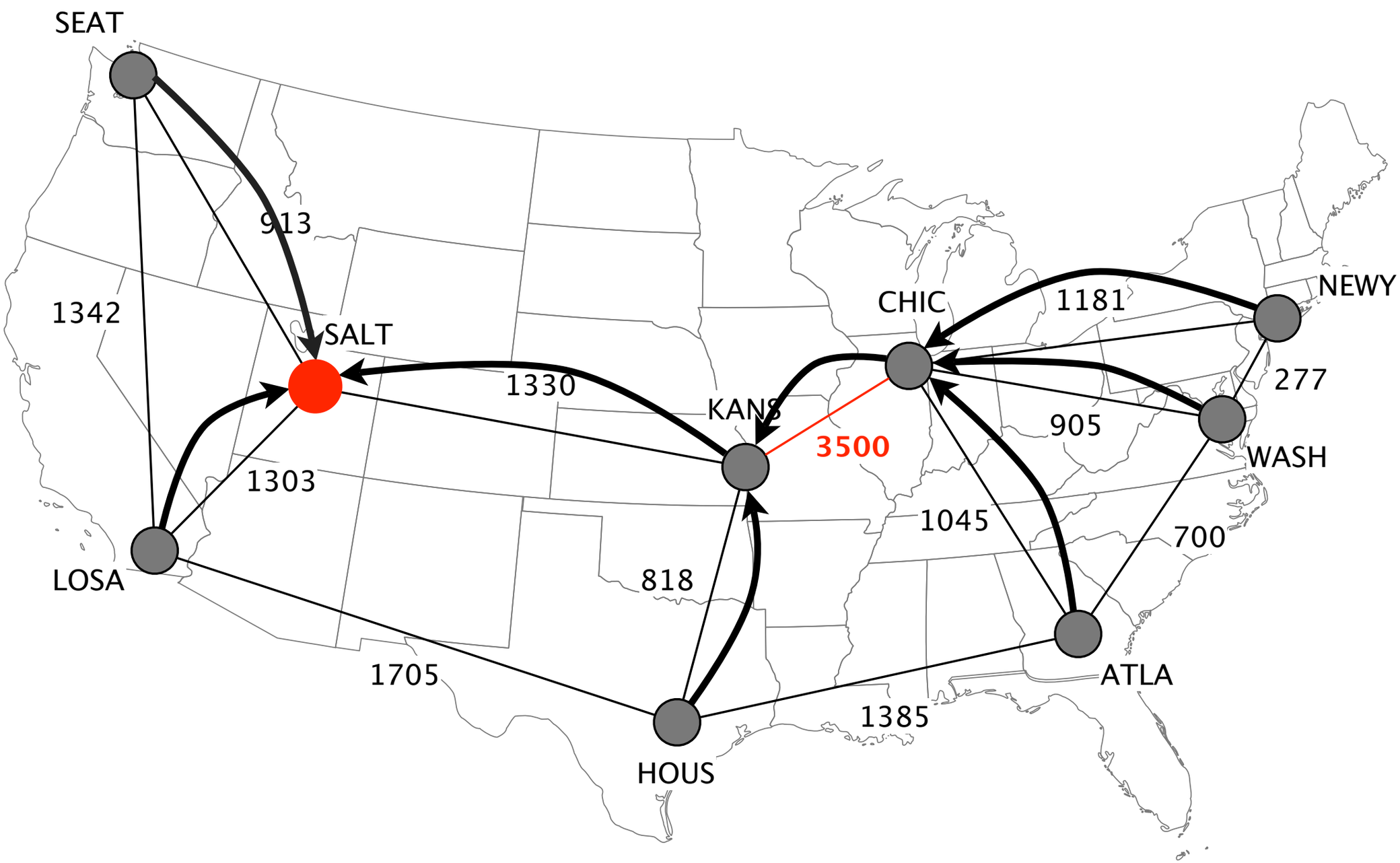
Find an ordering in which to activate the final IGP without causing any forwarding anomalies

Initial forwarding paths towards SALT



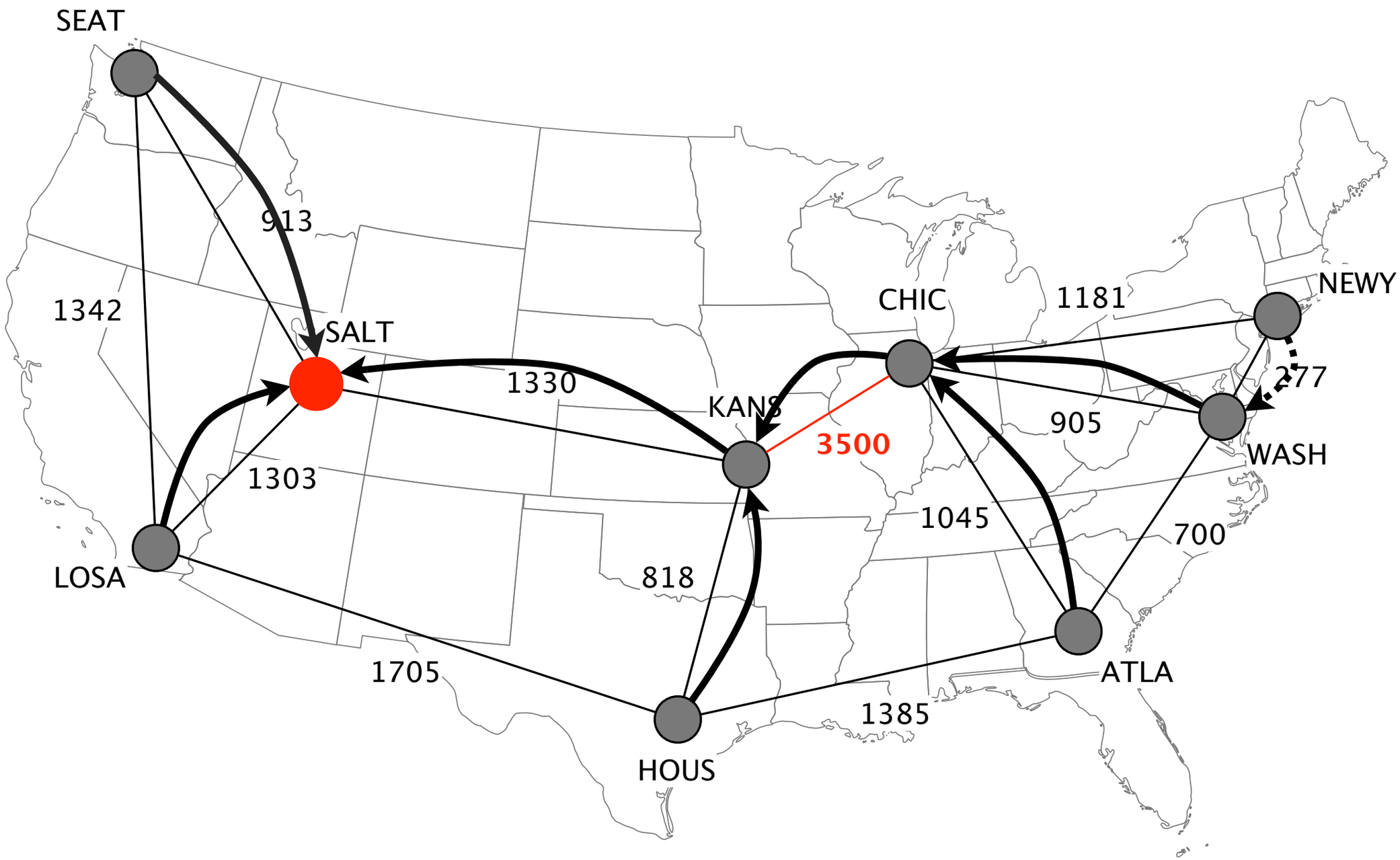
Migrated []

To migrate [NEWY, WASH, CHIC, ATLA, ...]



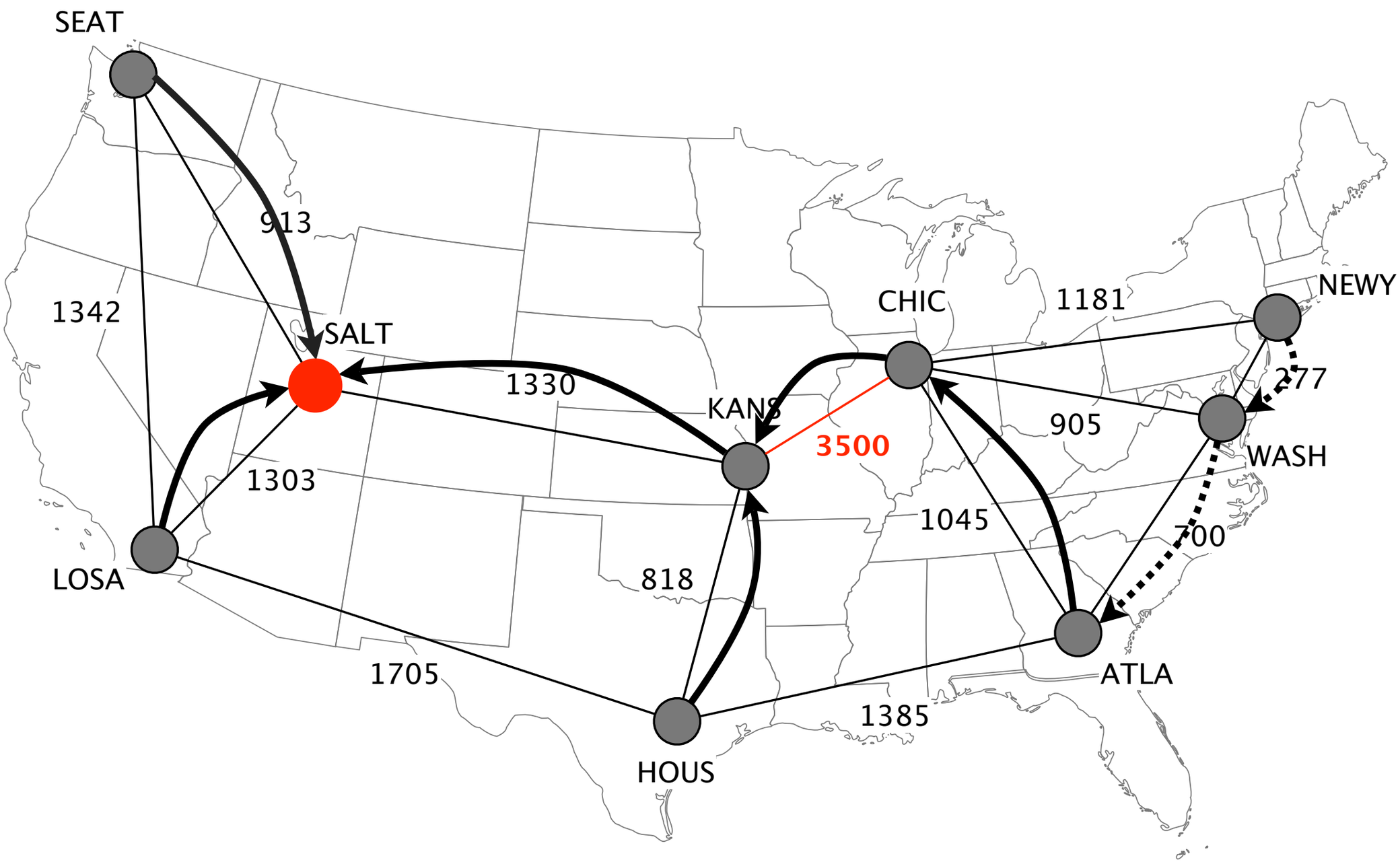
Migrated [NEWY]

To migrate [WASH, CHIC, ATLA, ...]



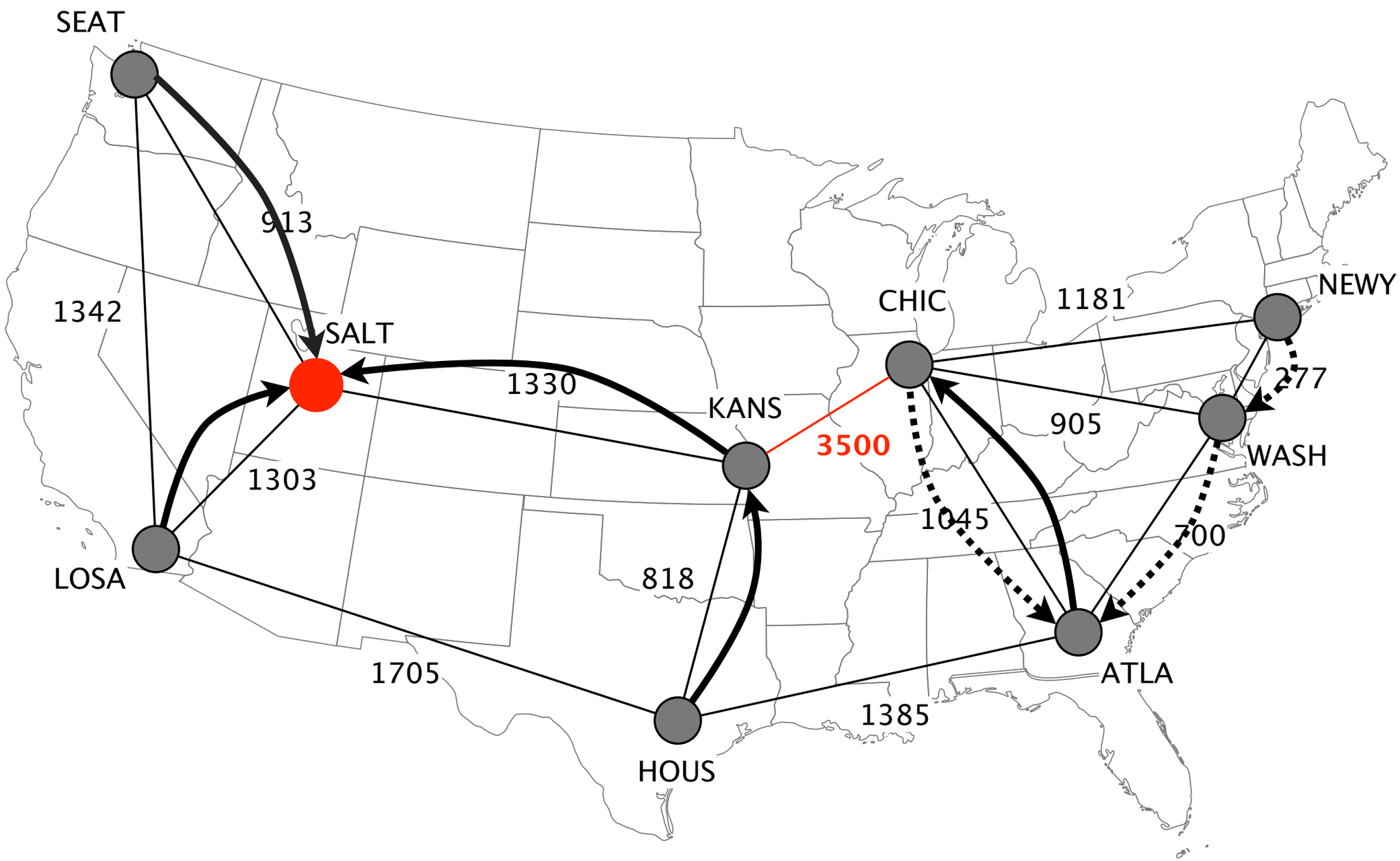
Migrated [NEWY, WASH]

To migrate [CHIC, ATLA, ...]



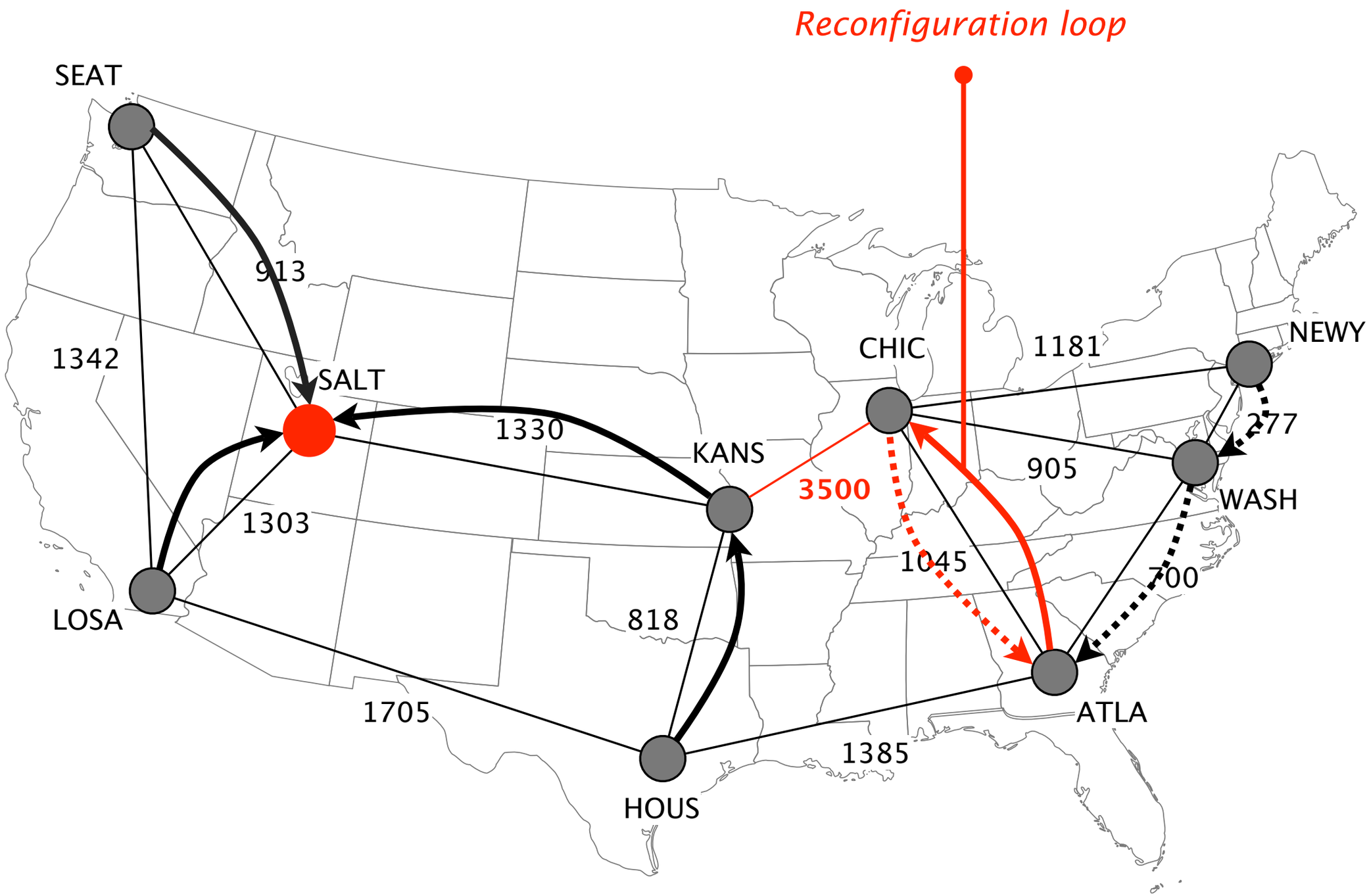
Migrated [NEWY, WASH, CHIC]

To migrate [ATLA, ...]



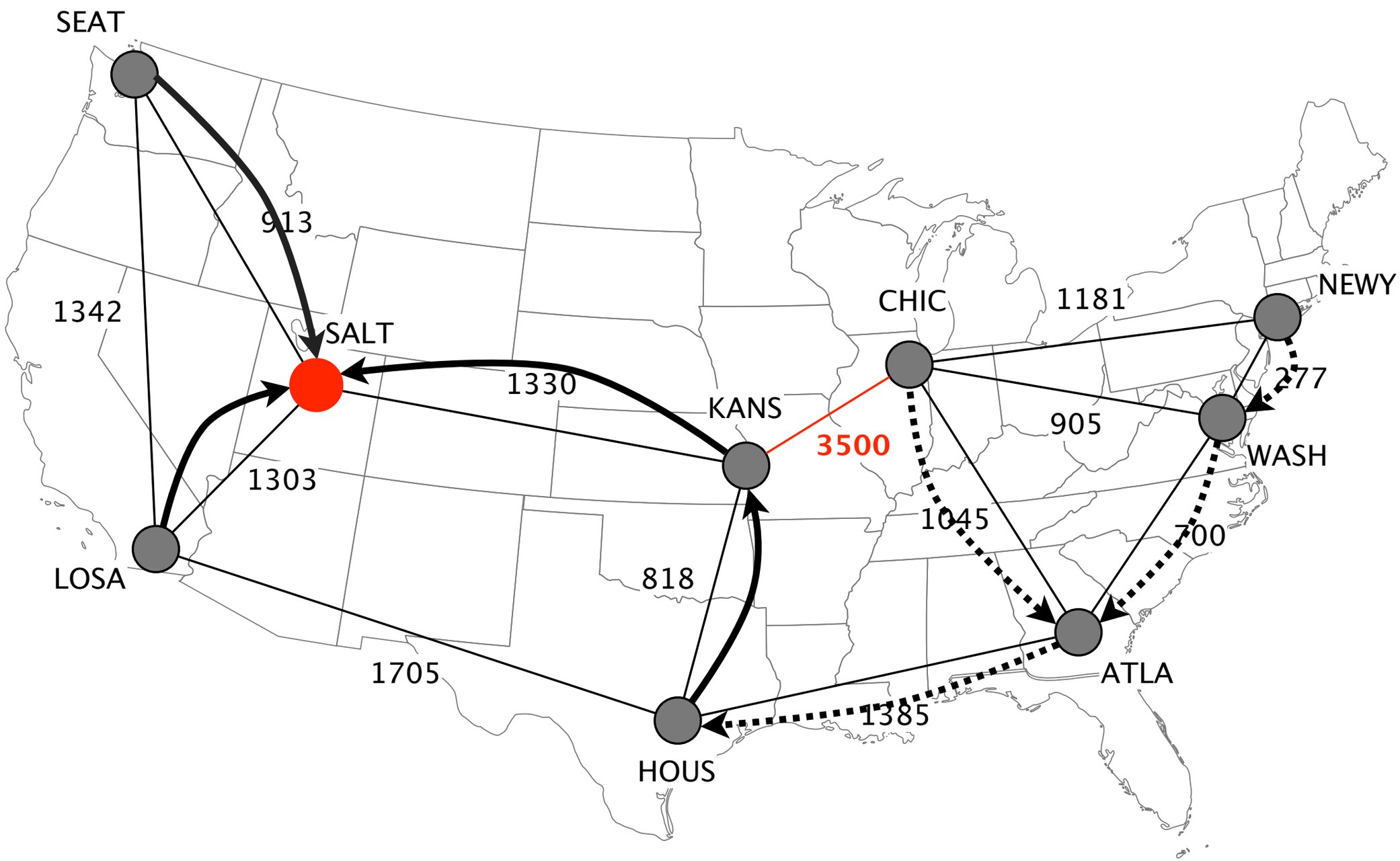
Migrated [NEWY, WASH, CHIC]

To migrate [ATLA, ...]

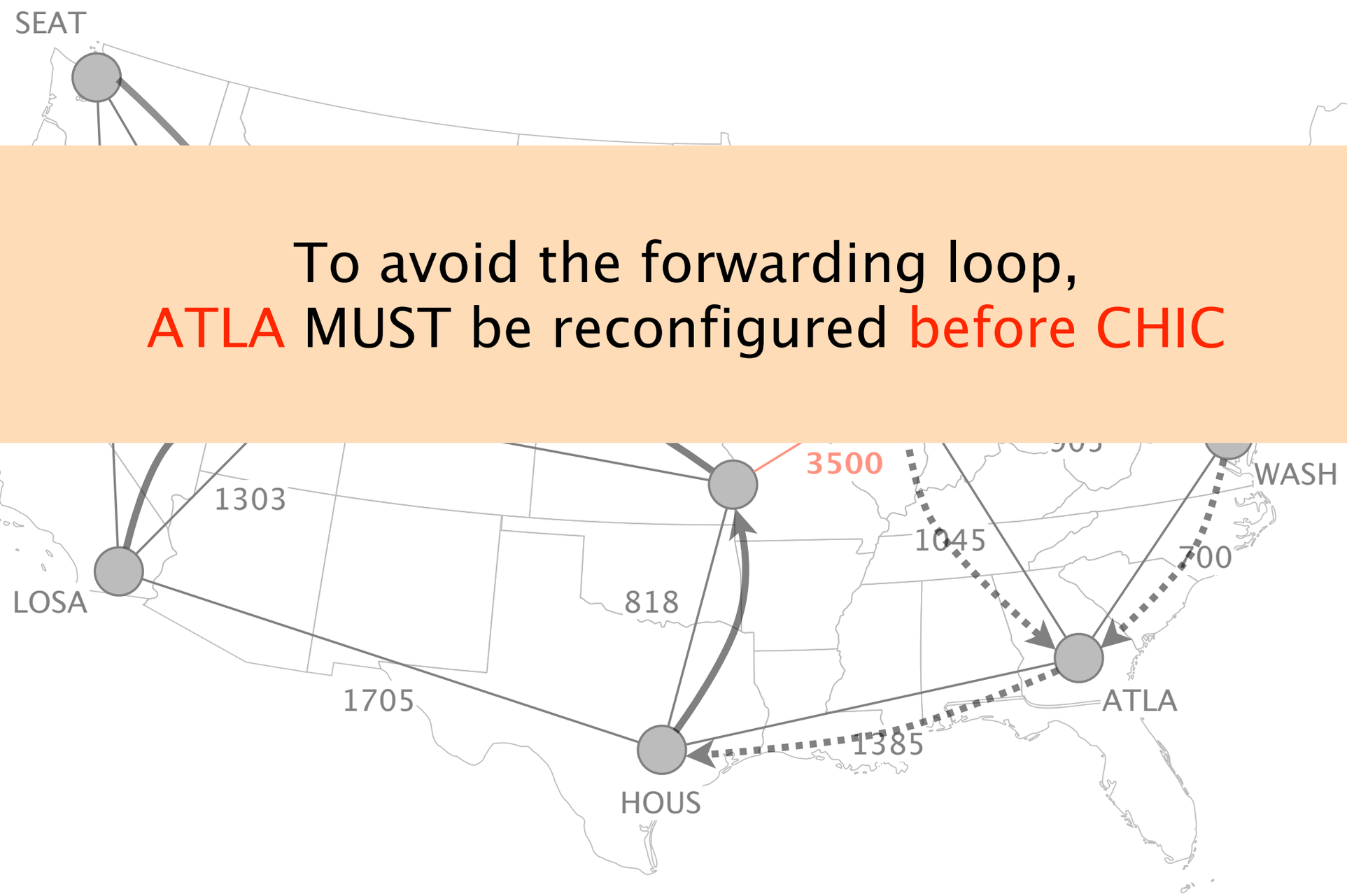


Migrated [NEWY, WASH, CHIC, ATLA]

To migrate []

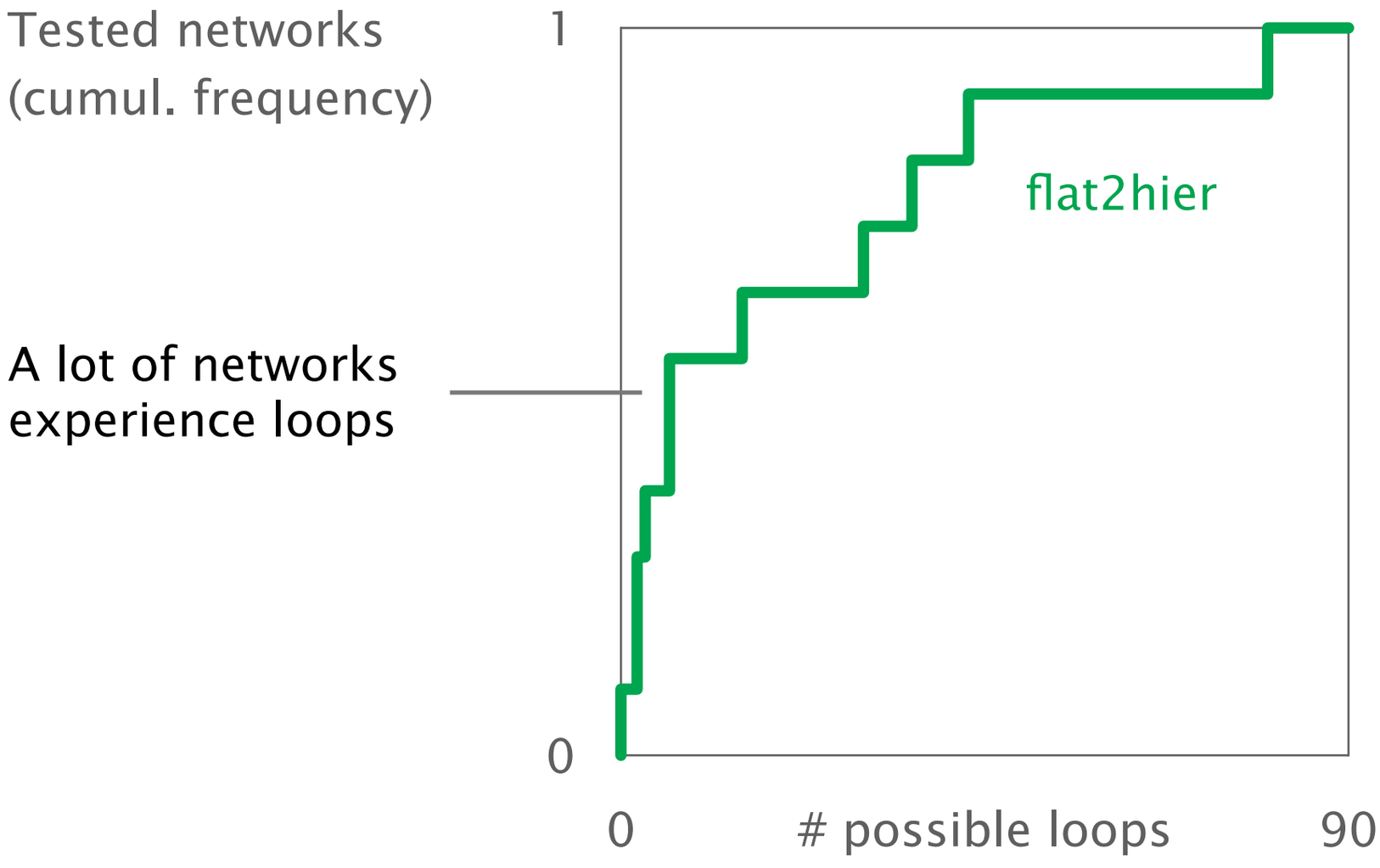


To avoid the forwarding loop,
ATLA MUST be reconfigured before CHIC



Are forwarding loops such a problem?

Numerous forwarding loops can appear in LS to LS reconfigurations



Up to 80 *reconfiguration loops* can arise during an IGP migration

Find an ordering in which to activate the final IGP without causing any forwarding anomalies

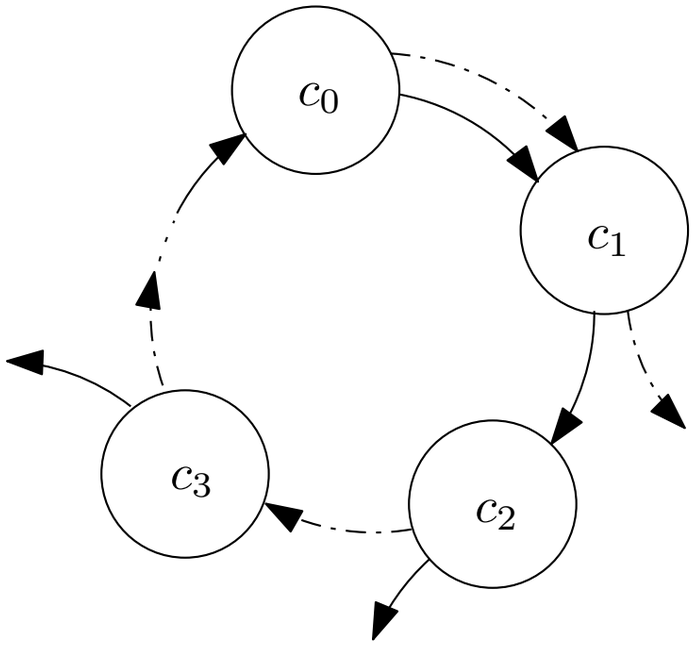
Is it easy to compute?

Find an ordering in which to activate the final IGP without causing any forwarding anomalies

Is it easy to compute?

Does it always exist?

An ordering does not always exist and deciding if one exists is NP-complete



LEGEND:

initial ——▶

final - - -▶

The Enumeration Algorithm [correct & complete]

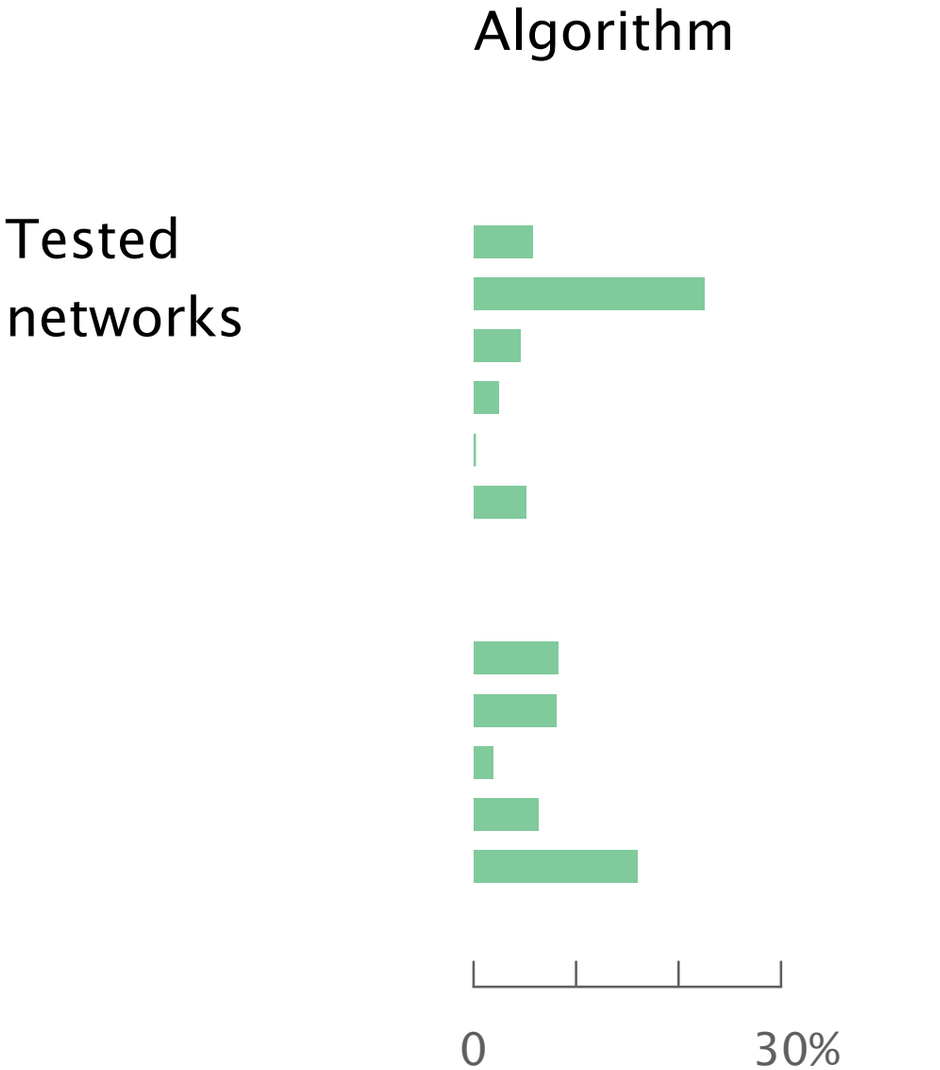
1. Merge the initial and the final forwarding paths
2. For each migration loop in the merged graph,

   Output ordering constraints such that

   at least one router in the initial state

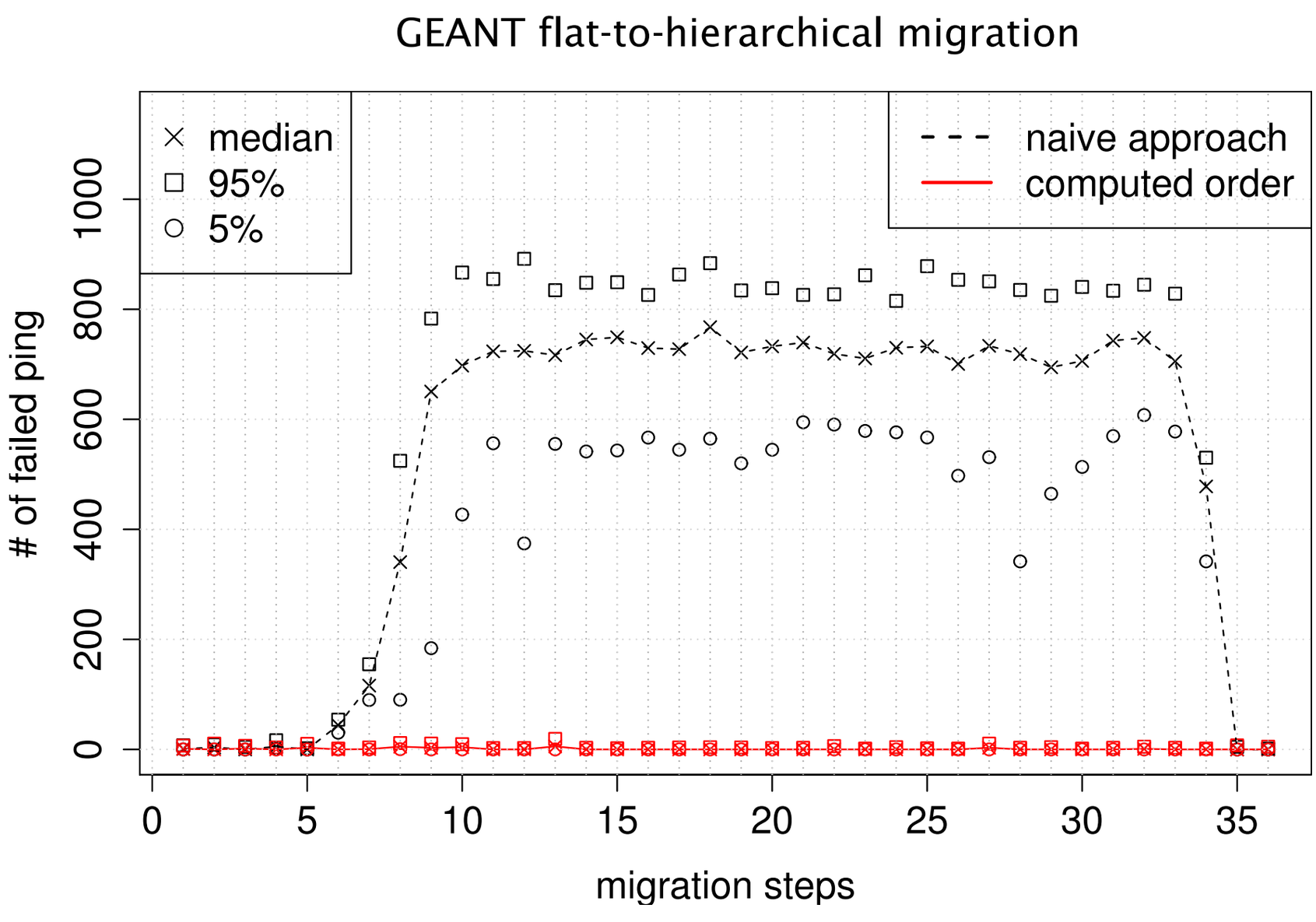
   is migrated before at least one in the final
3. Solve the system by using Linear Programming

But, in nearly all tested scenarios, the algorithm has found an ordering



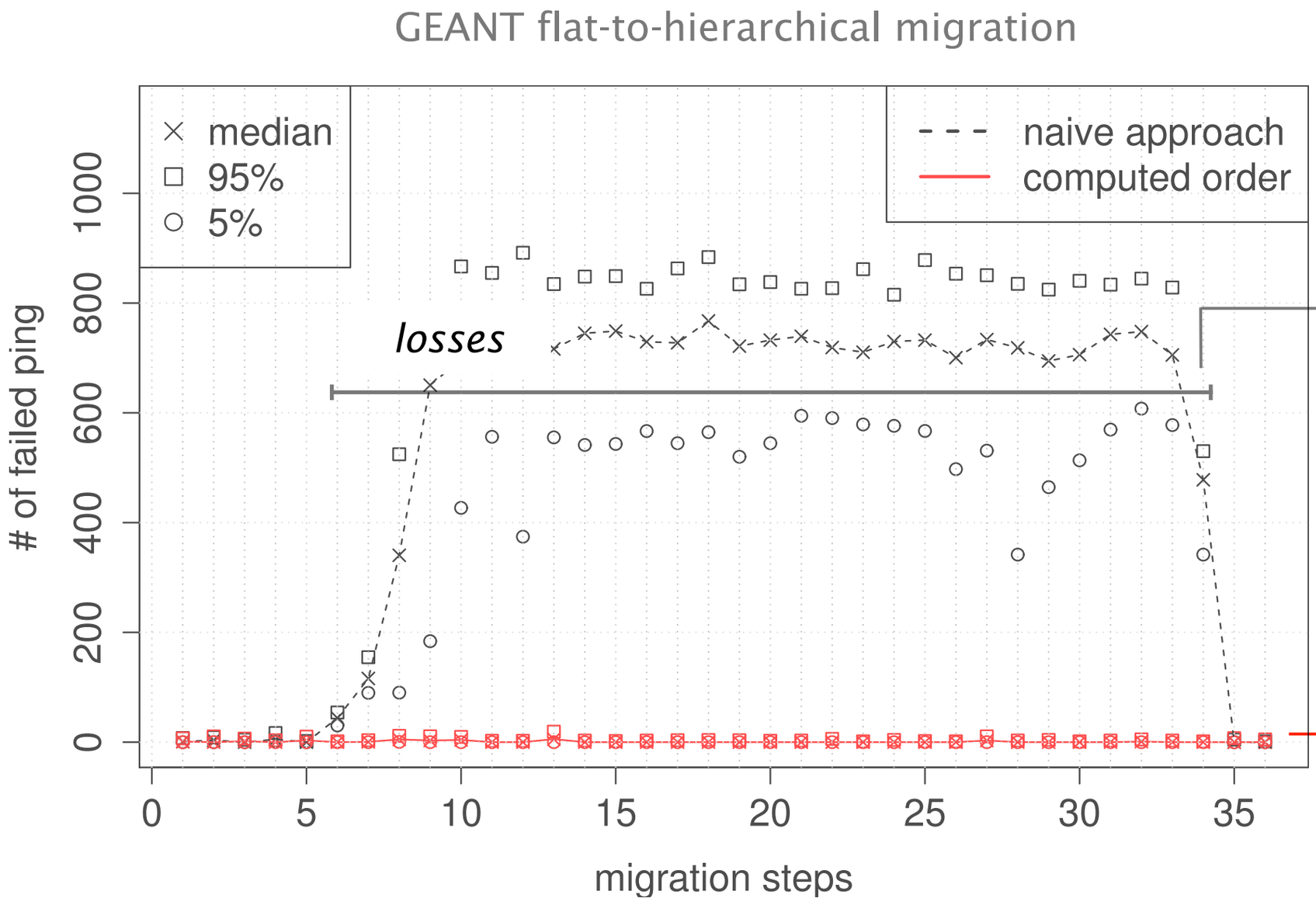
More than 20% of the routers might be involved in the ordering

Using our framework, we were able to achieve lossless reconfiguration



Average results (50 repetitions) computed on 700+ pings per step from every router to 5 problematic destinations

By following the computed ordering, lossless IGP reconfiguration are possible



Traffic is lost for more than 80% of the process

No loss occurs with our framework

Average results (50 repetitions) computed on 700+ pings per step from every router to 5 problematic destinations

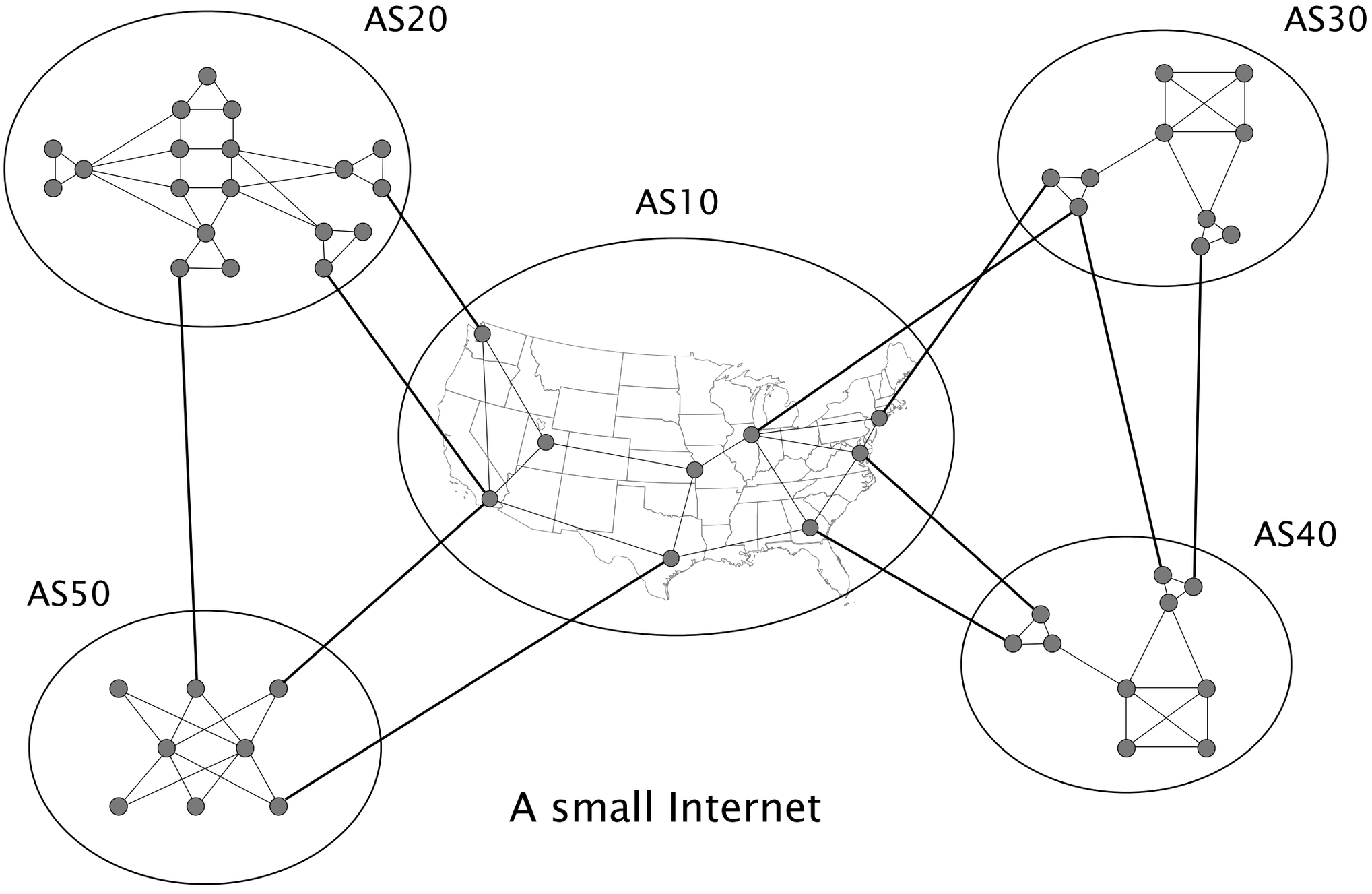
Methods and Techniques for Disruption-free Network Reconfiguration

IGP reconfiguration

BGP reconfiguration

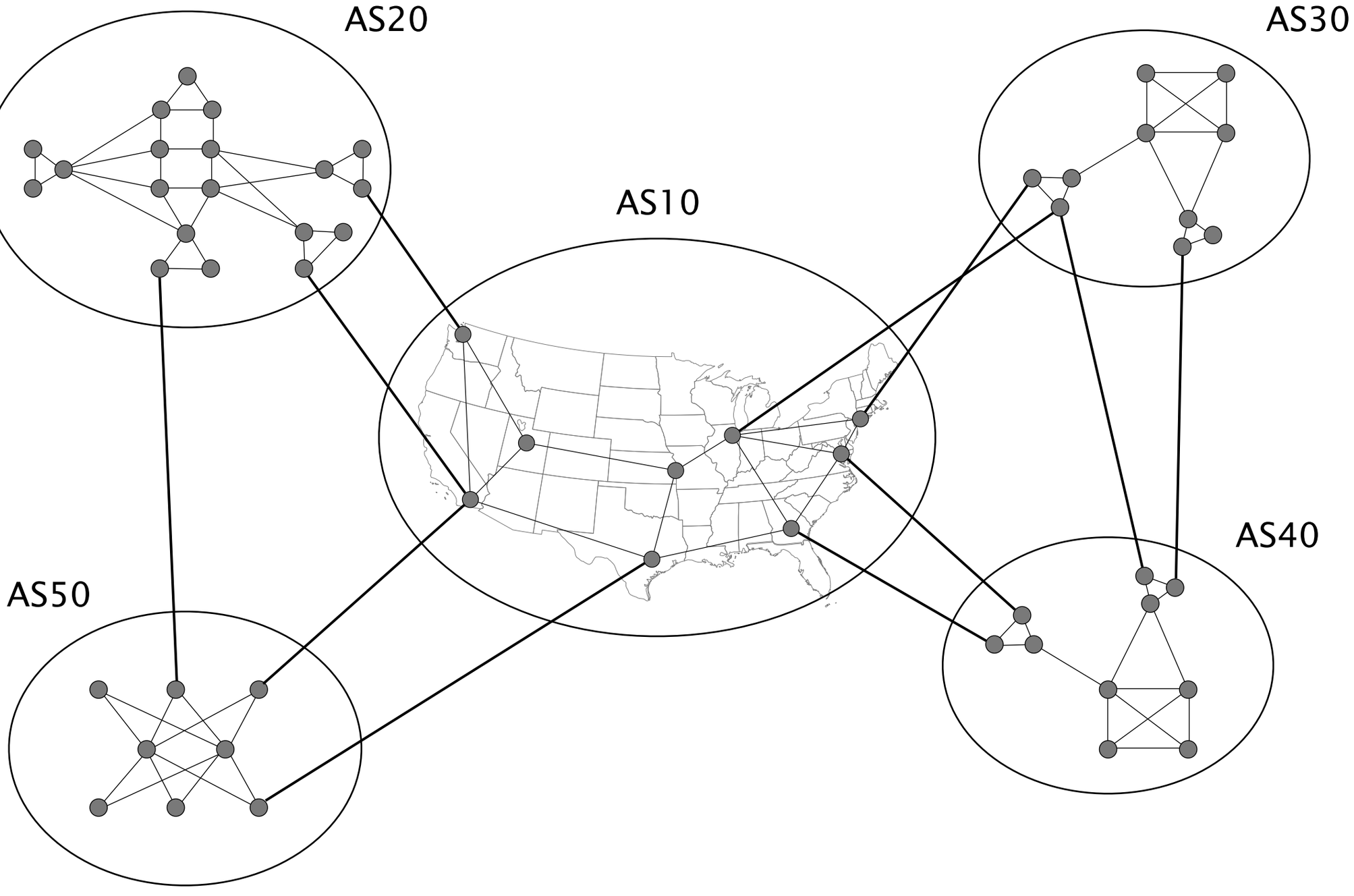
Principles

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

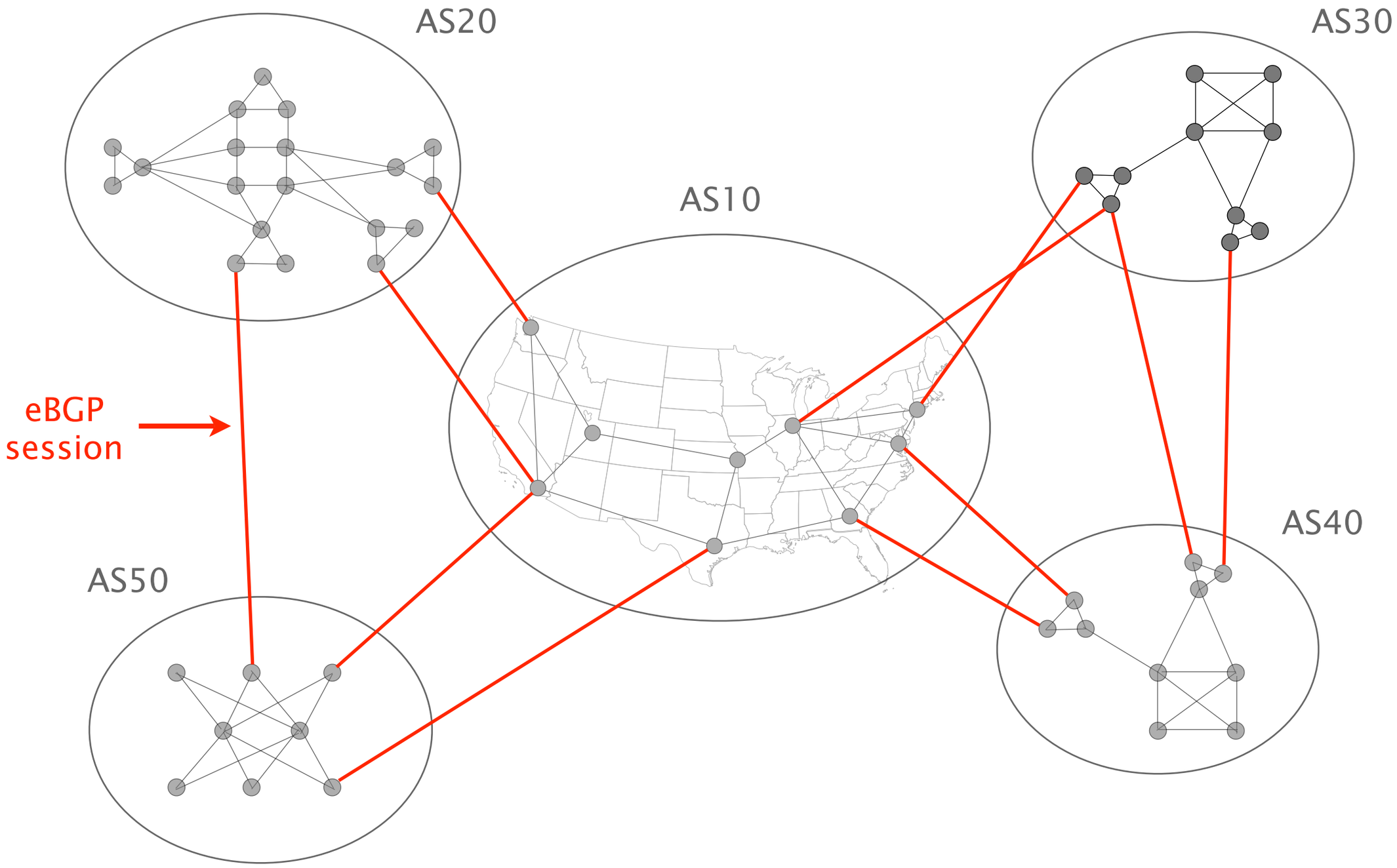


A small Internet

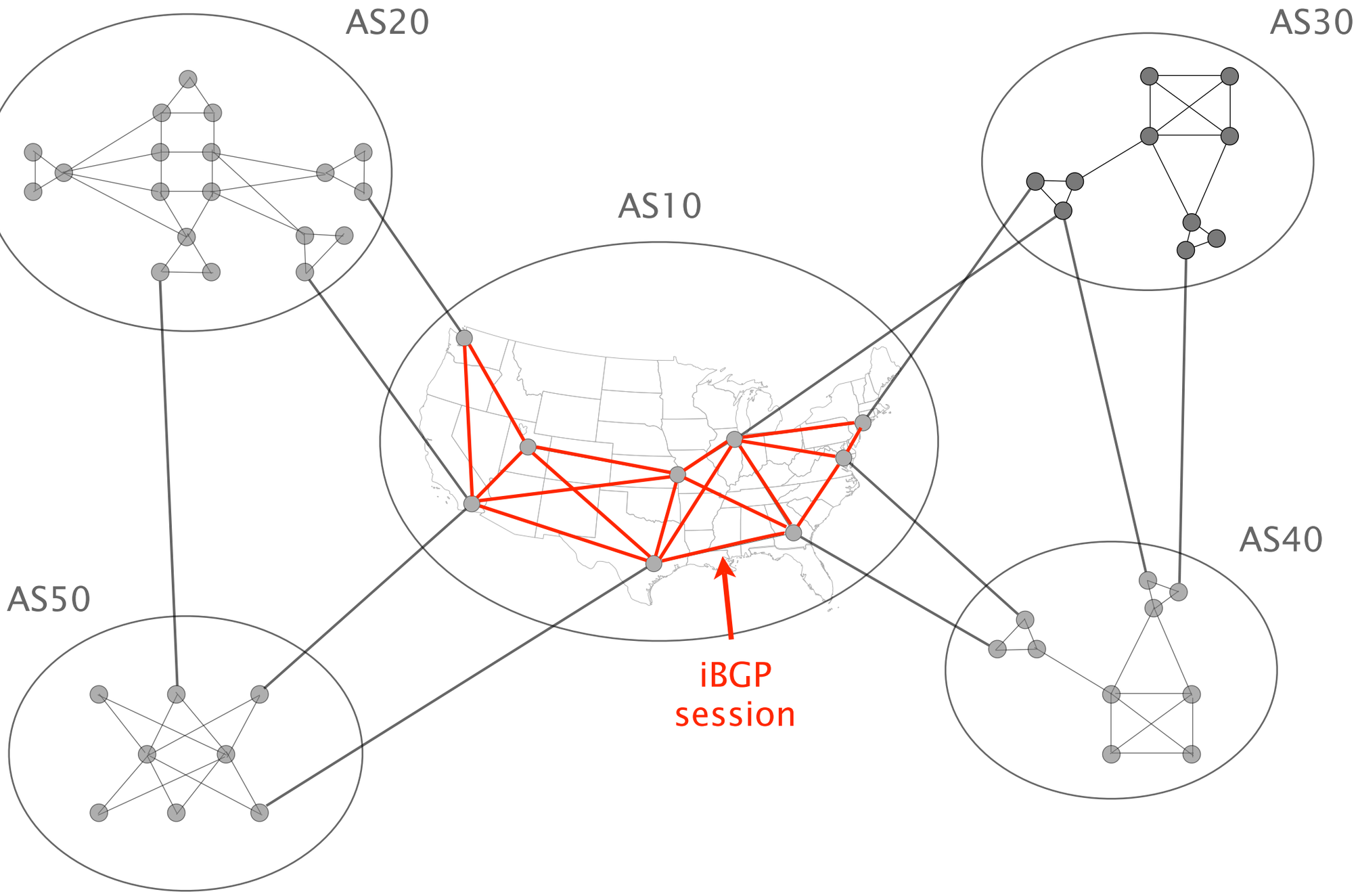
BGP comes in two flavors



external BGP (eBGP) exchanges reachability information between ASes



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



Each flavor of a BGP configuration can be changed

Typical reconfiguration scenarios consist in

iBGP	Add sessions
	Remove sessions
	Change type (e.g., turn a router into a route-reflector)
eBGP	Add sessions
	Remove sessions
	Modify policies (e.g., turn a client into a peer)

Reconfiguring BGP can be disruptive

Reconfiguring BGP (*) can lead to

- signaling anomalies [Griffin02]
- forwarding anomalies [Griffin02]
- dissemination anomalies [INFOCOM12]

or any combination of those

(*) [Guichard00, Smith10, Herrero10]

Reconfiguring BGP can be disruptive

Reconfiguring BGP (*) can lead to

- signaling anomalies
- forwarding anomalies
- dissemination anomalies

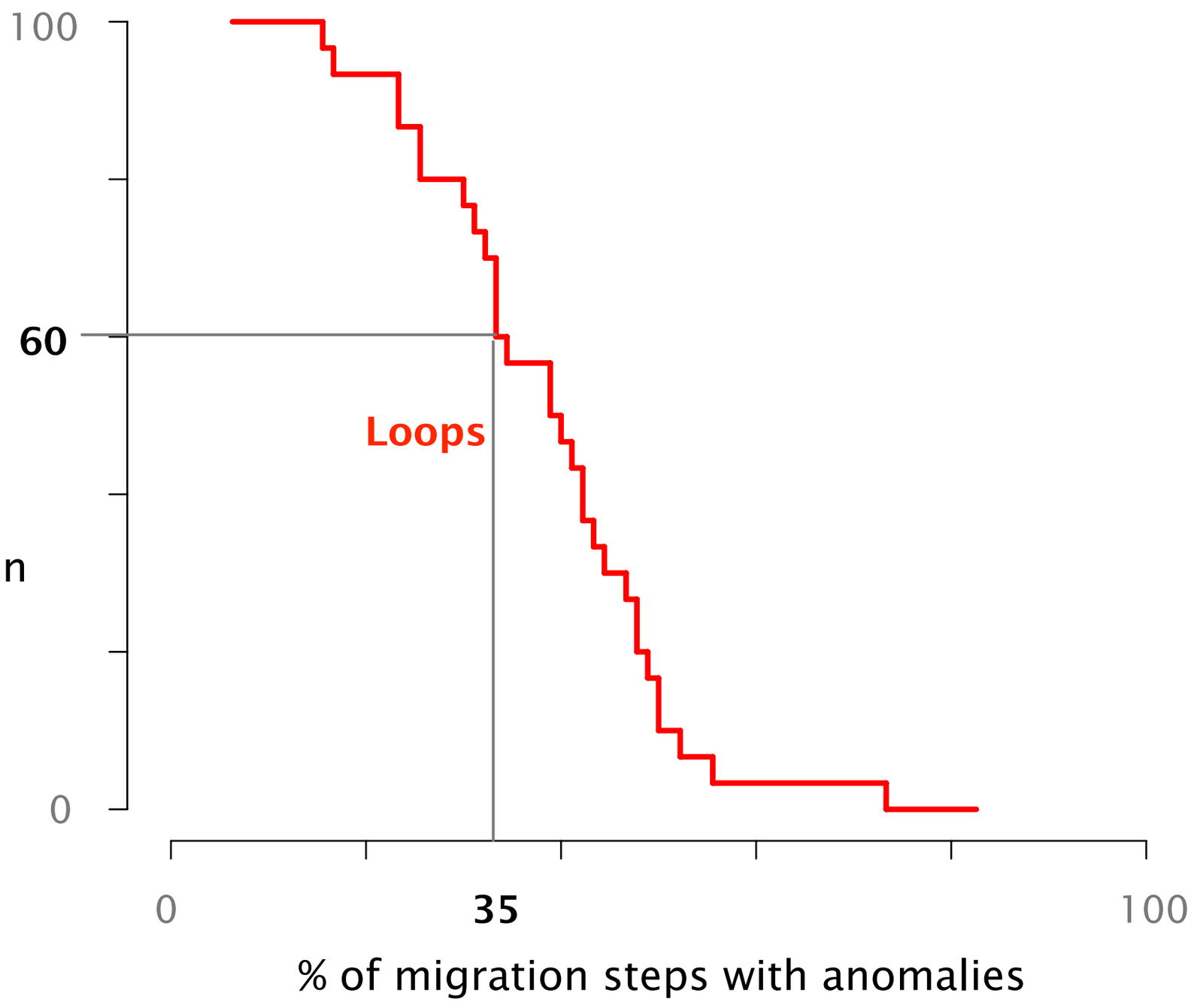
How many ?

or any combination of those

(*) [Guichard00, Smith10, Herrero10]

Best practices do not work

Tier1 experiments
(cumul. frequency)

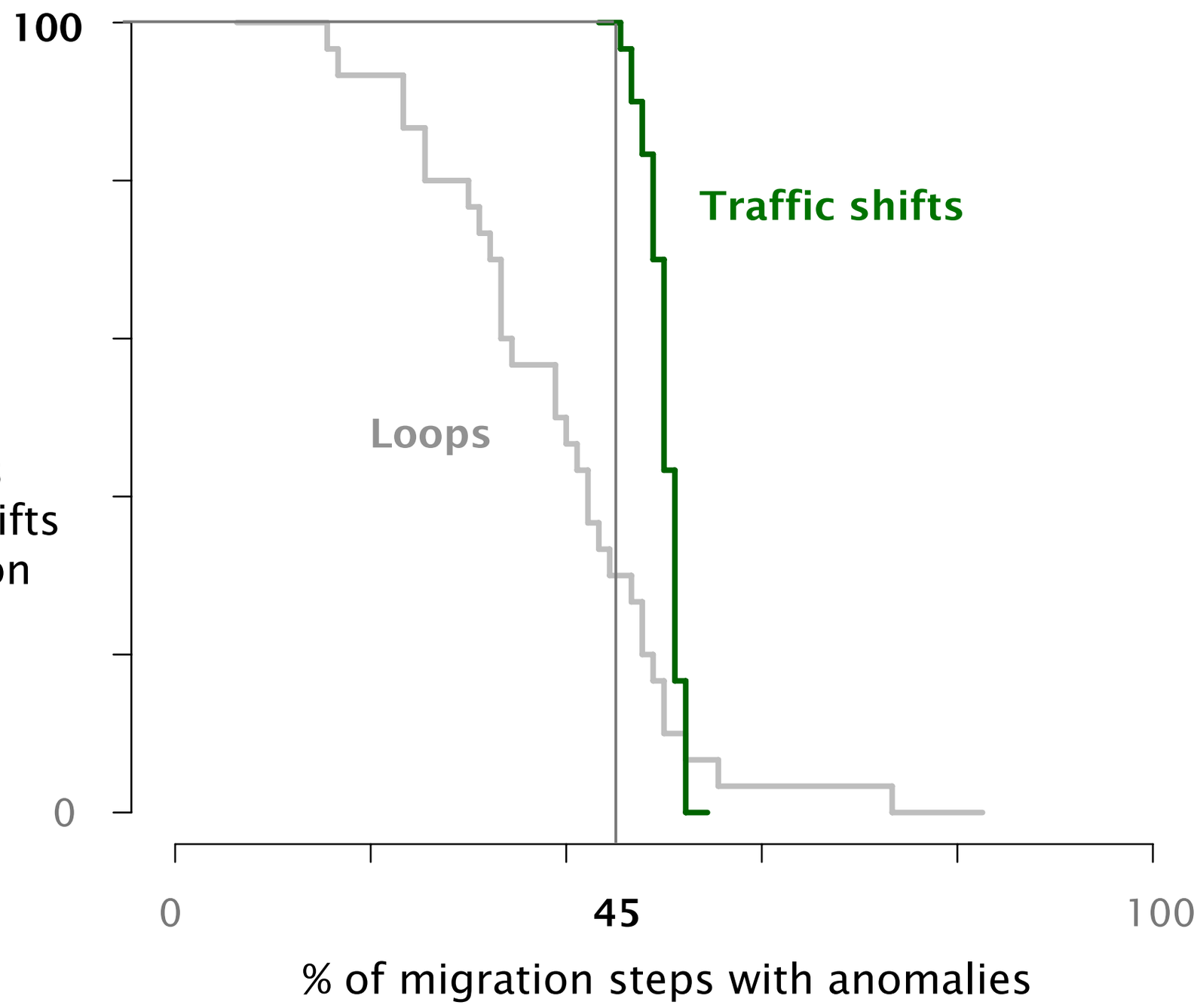


60% of the experiments
were subject to loops
for > **35%** of the migration

Best practices do not work

Tier1 experiments
(cumul. frequency)

100% of the experiments were subject to traffic shifts for > **40%** of the migration



Just like IGPs, finding an anomaly-free ordering is hard

Deciding if an anomaly-free ordering exists is at least NP-hard

It might even be harder

Just like IGPs, finding an anomaly-free ordering is hard and might not exist

Deciding if an anomaly-free ordering exists is at least NP-hard

It might even be harder

Due to contradictory constraint,
anomaly-free ordering might not exist

Anomalies are guaranteed to appear, no matter what

But unlike IGPs, an algorithmic approach is not viable

There are way more BGP destinations than IGP ones

two orders of magnitude (i.e., 450.000 vs 1000s)

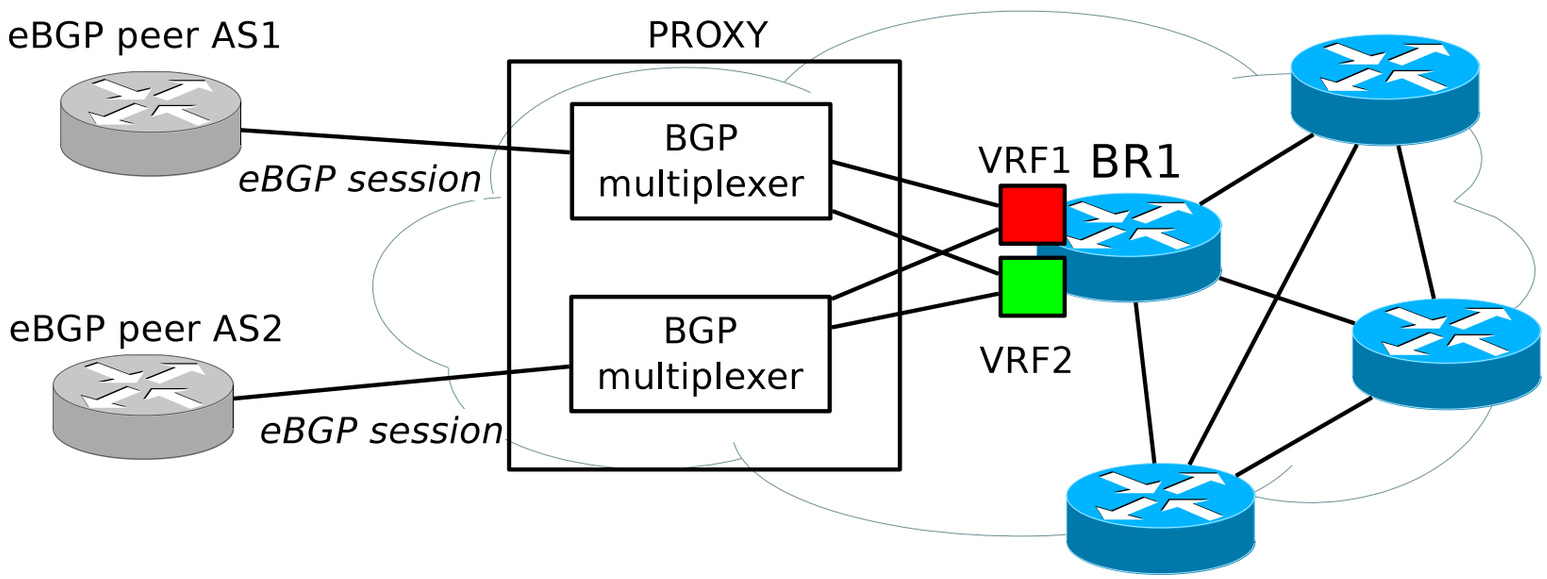
BGP destinations can be announced from any subset of nodes

while IGP destinations are usually announced from 1 node

Local changes can have *remote* impact

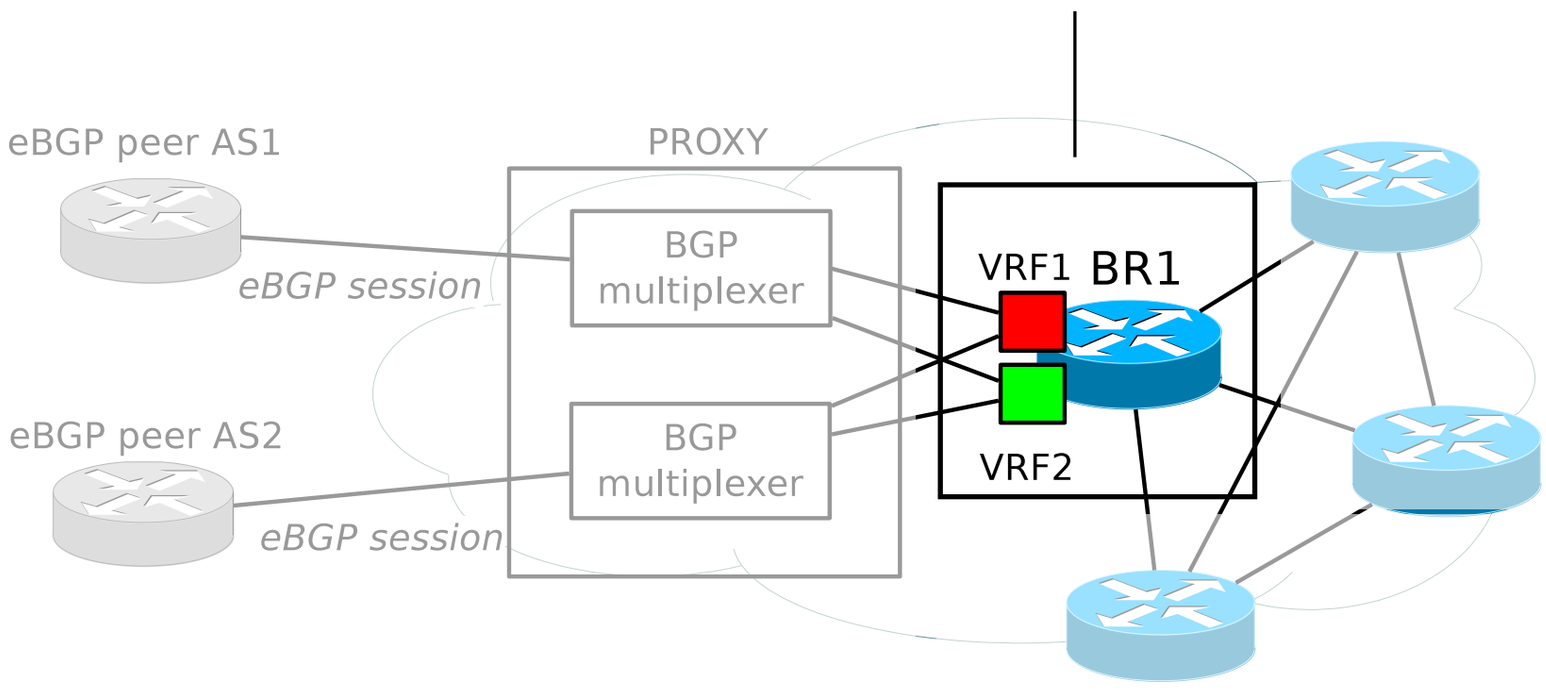
meaning we must them into account as well

To circumvent the inherent complexity,
we developed a reconfiguration framework



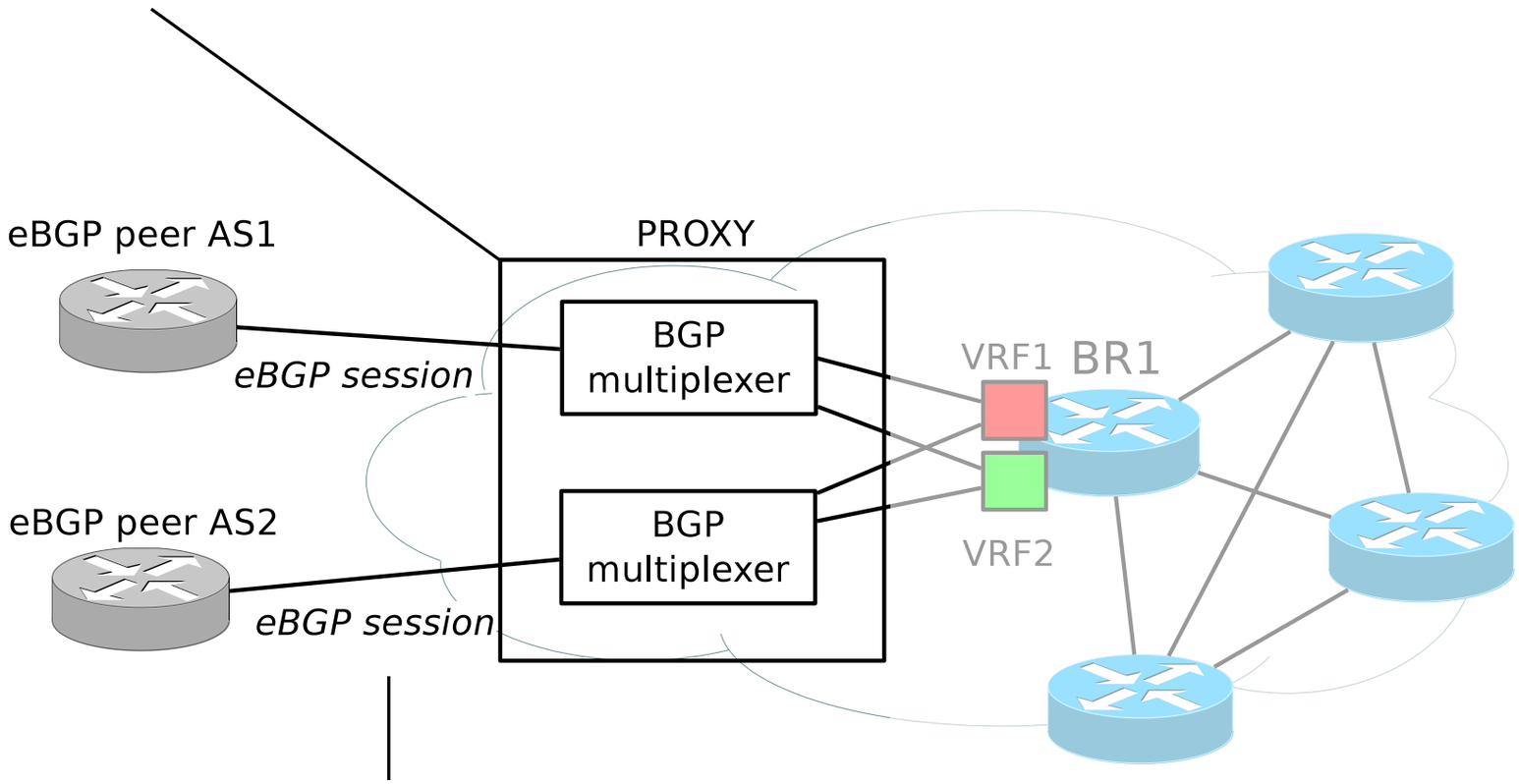
To circumvent the inherent complexity, we developed a reconfiguration framework

By leveraging specific technologies (L3VPNs), routers can maintain different BGP routing planes



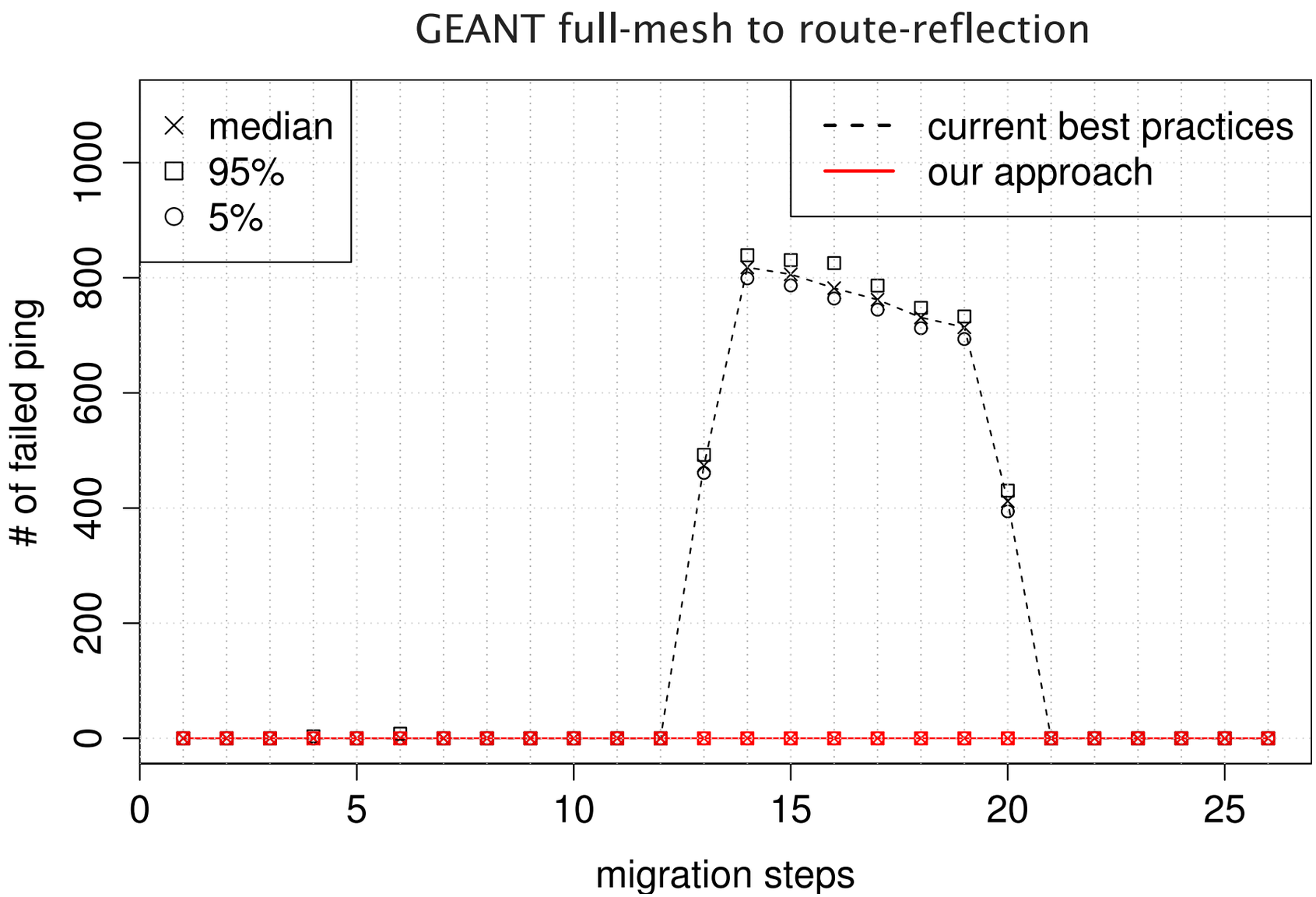
To circumvent the inherent complexity, we developed a reconfiguration framework

A proxy distributes BGP updates to the different routing planes



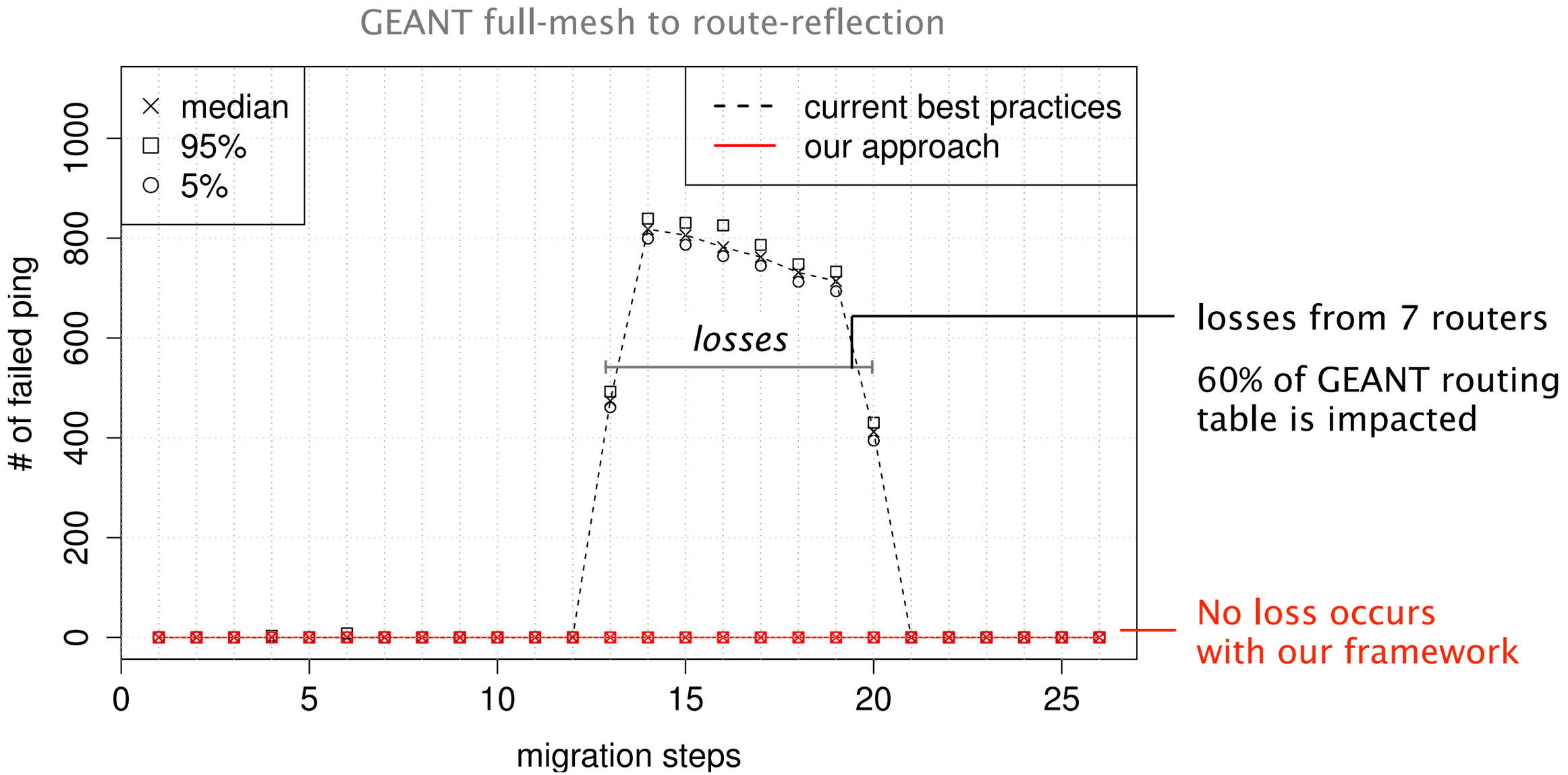
Our framework is completely transparent for neighboring router

Our reconfiguration framework enables lossless reconfiguration



Average results (30 repetitions) computed on 120+ pings per step from every router to 16 summary prefixes

Our reconfiguration framework enables lossless reconfiguration



Average results (30 repetitions) computed on 120+ pings per step from every router to 16 summary prefixes

Methods and Techniques for Disruption-free Network Reconfiguration

IGP reconfiguration

BGP reconfiguration

Principles

Disruption-free reconfiguration building blocks

FIB updates *must* be atomic

absolute requirement

Local reconfiguration should only have local impact

enable a fine control of the reconfiguration process

Computing a configuration outcome should be efficient

facilitate the computation of a reconfiguration ordering

Disruption-free reconfiguration building blocks

Allow multiple independent protocol instances

separate the initial from the final configuration

Rely on encapsulation

ensure forwarding correctness

Avoid protocol dependencies

avoid side effects

Disruption-free reconfiguration building blocks

Allow multiple independent protocol instances

separate the initial from the final configuration

Rely on encapsulation

ensure forwarding correctness

Avoid protocol dependencies

avoid side effects

when applicable

Methods and Techniques for Disruption-free Network Reconfiguration

IGP reconfiguration

BGP reconfiguration

Principles

High-level overview of the contributions

Provide a deep theoretical and practical understanding of routing reconfiguration problems

Bring flexibility to network management

regularly move to the best network-wide configuration

Development of a complete reconfiguration framework

which works in today's networks (come talk to me)

Publications

IGP reconfiguration

[SIGCOMM11] Laurent Vanbever, Stefano Vissicchio, Cristel Pelsser, Pierre Francois and Olivier Bonaventure. Seamless Network-Wide IGP Migrations. In *ACM SIGCOMM Conference*, 2011

[TON12a] Laurent Vanbever, Stefano Vissicchio, Cristel Pelsser, Pierre Francois and Olivier Bonaventure. Lossless Migrations of Link-State IGPs. In *IEEE/ACM Transactions on Networking*, 2012. (To appear).

BGP reconfiguration

[INFOCOM12] Stefano Vissicchio, Luca Cittadini, Laurent Vanbever and Olivier Bonaventure. iBGP Deceptions: More Sessions, Fewer Routes. In *IEEE INFOCOM*, 2012

[TON12b] Stefano Vissicchio, Laurent Vanbever, Cristel Pelsser, Luca Cittadini, Pierre Francois and Olivier Bonaventure. Improving Network Agility with Seamless BGP Reconfigurations. In *IEEE/ACM Transactions on Networking*, 2012. (To appear).

Methods and Techniques for Disruption-free Network Reconfiguration

Laurent Vanbever

inl.info.ucl.ac.be/lvanbeve

Towards flexible networks

with seamless reconfiguration