CS 43: Computer Networks

BGP

November 23, 2026



Adapted from Slides by: Kurose & Ross, D. Choffnes, K. Webb, J. Rexford, Laurent Vanbever

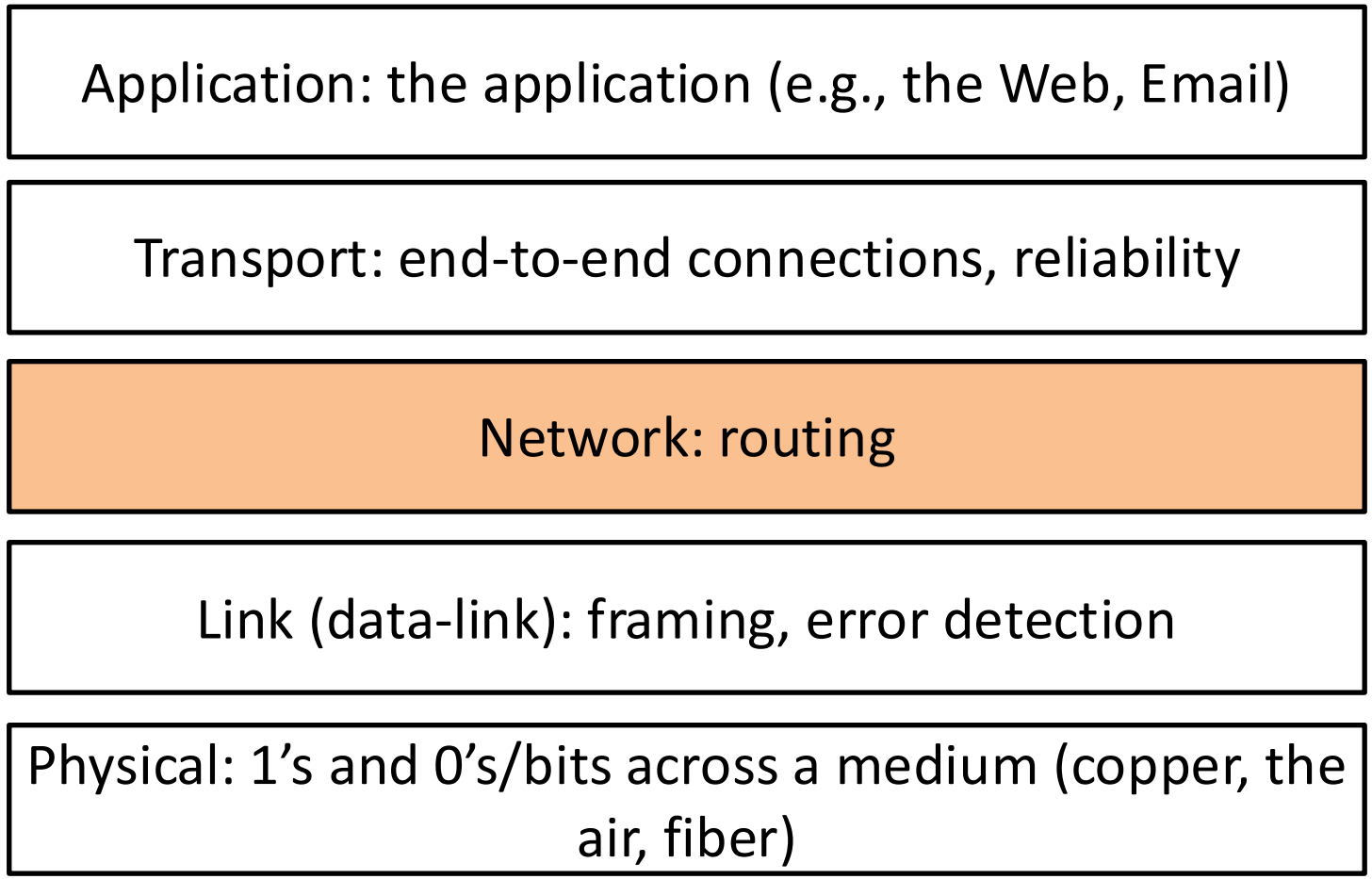
Reading Quiz
(graded for correctness)

Put devices away

Note: Reading quiz slides will have an orange border.

End of Reading Quiz!

The Network Layer!



Distance Vector Algorithm & Bellman-Ford equation

Let $d_x(y)$ be the cost of the least-cost path known by x to reach y

Each node bundles these distances into one message (called a vector) that it repeatedly sends to all its neighbors

until convergence

Distance Vector Algorithm & Bellman-Ford equation

Let $d_x(y)$ be the cost of the least-cost path known by x to reach y

Each node bundles these distances into one message (called a vector) that it repeatedly sends to all its neighbors

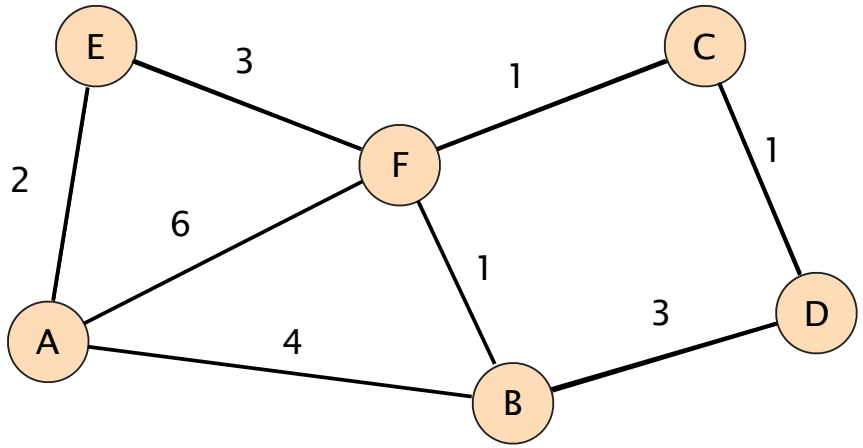
until convergence

Each node updates its distances based on neighbors' vectors:

$d_x(y) = \min\{ c(x,v) + d_v(y) \}$ over all neighbors v

Optimum 1-hop path

A			B		
Dst	Cst	Hop	Dst	Cst	Hop
A	0	A	A	4	A
B	4	B	B	0	B
C	∞	-	C	∞	-
D	∞	-	D	3	D
E	2	E	E	∞	-
F	6	F	F	1	F

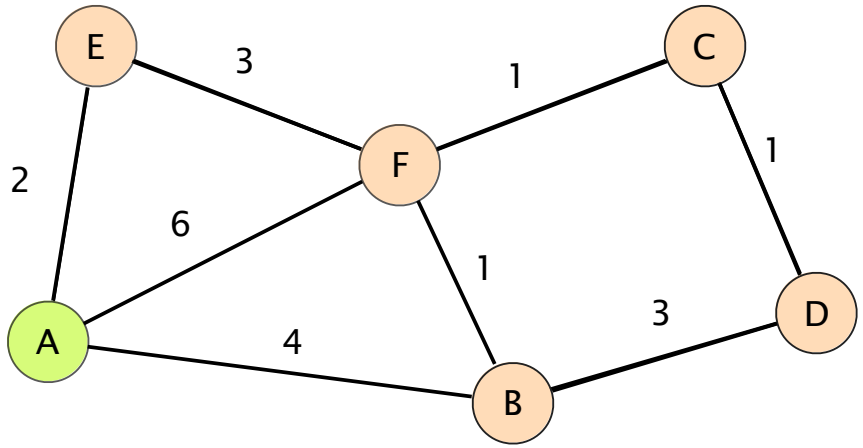


C			D			E			F		
Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop
A	∞	-	A	∞	-	A	2	A	A	6	A
B	∞	-	B	3	B	B	∞	-	B	1	B
C	0	C	C	1	C	C	∞	-	C	1	C
D	1	D	D	0	D	D	∞	-	D	∞	-
E	∞	-	E	∞	-	E	0	E	E	3	E
F	1	F	F	∞	-	F	3	F	F	0	F

Adapted from Laurent Vanbever, ETH Zurich

Optimum 1-hop path

A			B		
Dst	Cst	Hop	Dst	Cst	Hop
A	0	A	A	4	A
B	**4**	**B**	B	0	B
C	∞	-	C	∞	-
D	∞	-	D	3	D
E	**2**	**E**	E	∞	-
F	**6**	**F**	F	1	F

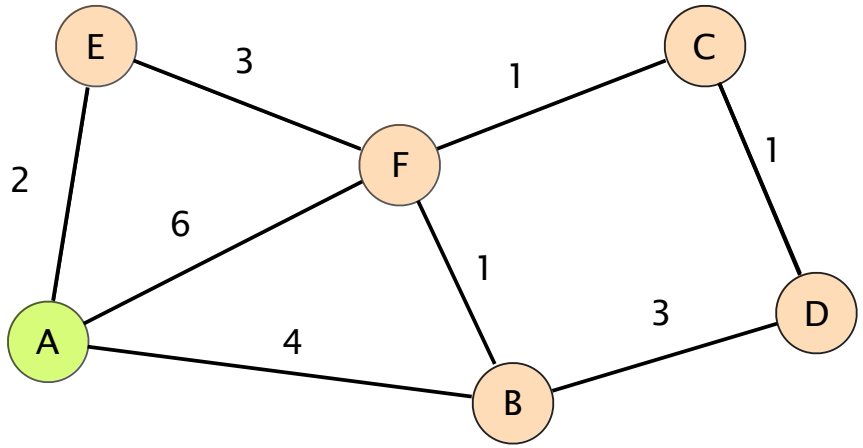


C			D			E			F		
Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop
A	∞	-	A	∞	-	A	2	A	A	6	A
B	∞	-	B	3	B	B	∞	-	B	1	B
C	0	C	C	1	C	C	∞	-	C	1	C
D	1	D	D	0	D	D	∞	-	D	∞	-
E	∞	-	E	∞	-	E	0	E	E	3	E
F	1	F	F	∞	-	F	3	F	F	0	F

Adapted from Laurent Vanbever, ETH Zurich

Optimum 2-hops path

A			B		
Dst	Cst	Hop	Dst	Cst	Hop
A	0	A	A	4	A
B	4	B	B	0	B
C	**7**	**F**	C	2	F
D	**7**	**B**	D	3	D
E	2	E	E	4	F
F	**5**	**E**	F	1	F

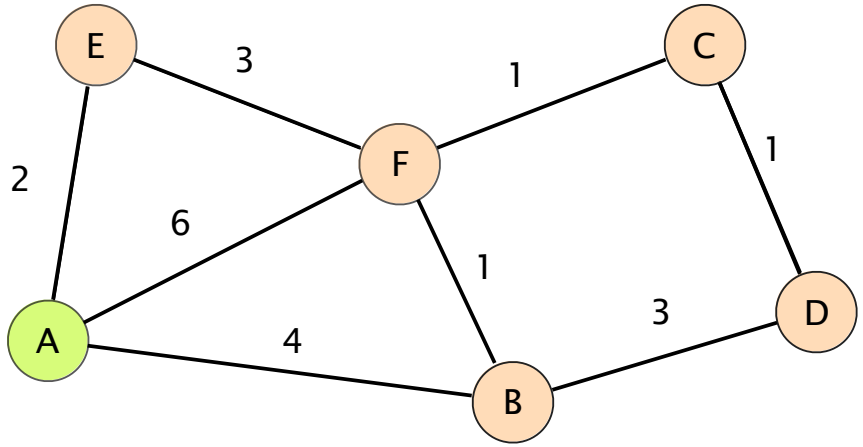


C			D			E			F		
Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop
A	7	F	A	7	B	A	2	A	A	5	B
B	2	F	B	3	B	B	4	F	B	1	B
C	0	C	C	1	C	C	4	F	C	1	C
D	1	D	D	0	D	D	∞	-	D	2	C
E	4	F	E	∞	-	E	0	E	E	3	E
F	1	F	F	2	C	F	3	F	F	0	F

Adapted from Laurent Vanbever, ETH Zurich

Optimum 3-hops path

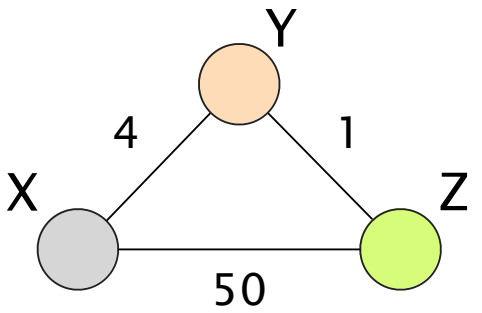
A			B		
Dst	Cst	Hop	Dst	Cst	Hop
A	0	A	A	4	A
B	4	B	B	0	B
C	6	E	C	2	F
D	7	B	D	3	D
E	2	E	E	4	F
F	5	E	F	1	F



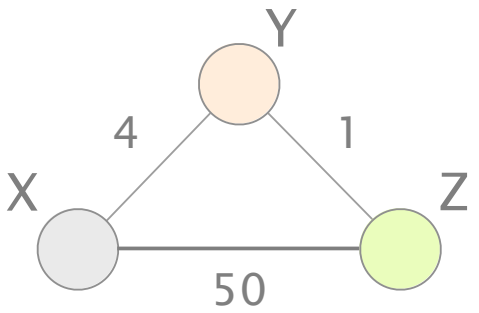
C			D			E			F		
Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop	Dst	Cst	Hop
A	6	F	A	7	B	A	2	A	A	5	B
B	2	F	B	3	B	B	4	F	B	1	B
C	0	C	C	1	C	C	4	F	C	1	C
D	1	D	D	0	D	D	5	F	D	2	C
E	4	F	E	5	C	E	0	E	E	3	E
F	1	F	F	2	C	F	3	F	F	0	F

Let’s consider the convergence process after a link cost change

Consider the following network



Consider the following network



Y state

dest.	via X	via Z
X	4	6

Y reaches X directly

Z state

dest.	via X	via Y
X	50	5

Z reaches X via Y

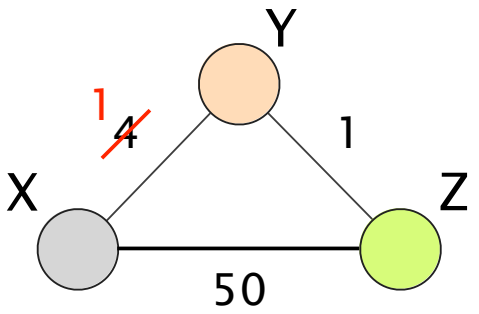
Each “via” entry = cost of the link to that neighbor + what that neighbor last advertised.

The green entry (the smallest) is what the router advertises next.

t = 0

(X,Y) weight changes

from 4 to 1



time t=0

Y state

dest.	via X	via Z
X	4	6

Z state

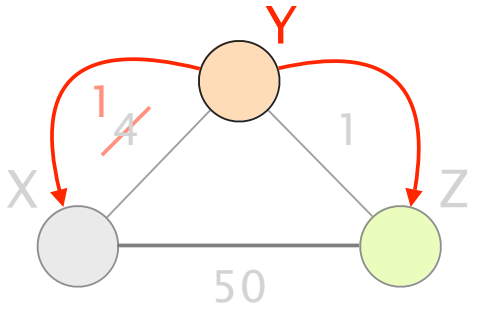
dest.	via X	via Y
X	50	5

Node detects local cost change, update their vectors, and notify their neighbors if it has changed

t = 1

Y updates its vector,

sends it to X and Z



t=0

Y state

dest.	via X	via Z
X	4	6

t=1

dest.	via X	via Z
X	1	6

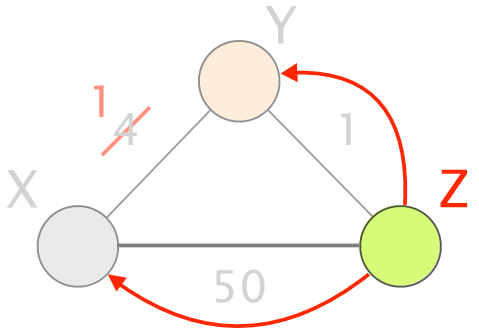
Z state

dest.	via X	via Y
X	50	5

t = 2

Z updates its vector,

sends it to X and Y

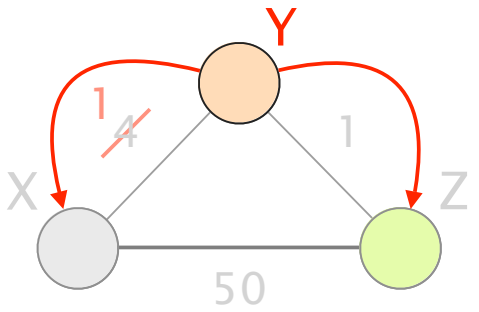


	t=0		t=1		t=2	
Y state	dest.	via	dest.	via		
		X Z		X Z		
	X	4 6	X	1 6		
Z state	dest.	via			dest.	via
		X Y				X Y
	X	50 5			X	50 2

t = 3

Y updates its vector,

sends it to X and Z



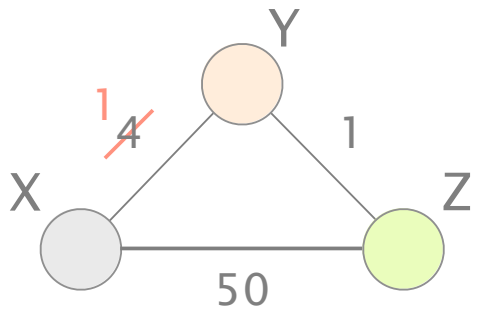
	t=0		t=1		t=2		t=3	
Y state	dest.	via	dest.	via			dest.	via
		X Z		X Z				X Z
	X	4 6	X	1 6			X	1 3
Z state	dest.	via			dest.	via		
		X Y				X Y		
	X	50 5			X	50 2		

Adapted from Laurent Vanbever, ETH Zurich

t > 3

no one moves anymore

network has converged!



	t=0	t=1	t=2	t>3
Y state	dest. via X Z X 4 6	dest. via X Z X 1 6		dest. via X Z X 1 3
Z state	dest. via X Y X 50 5		dest. via X Y X 50 2	dest. via X Y X 50 2

Adapted from Laurent Vanbever, ETH Zurich

The algorithm terminates
after 3 iterations

Good news travel fast!

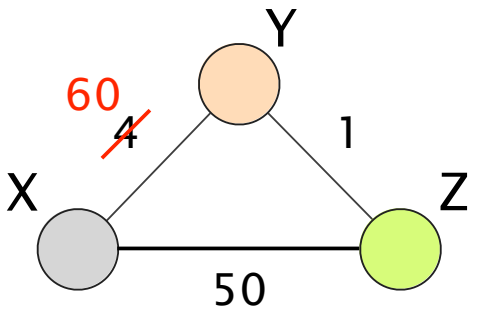
Good news travel fast!

What about bad ones?

t = 0

(X,Y) weight changes

from 4 to 60



time t=0

Y state

dest.	via X	via Z
X	4	6

Z state

dest.	via X	via Y
X	50	5

The X–Y link cost changes from 4 to 60. Y detects the change and recomputes. What is Y's new cost to X?

A. 4

B. 6

C. 50

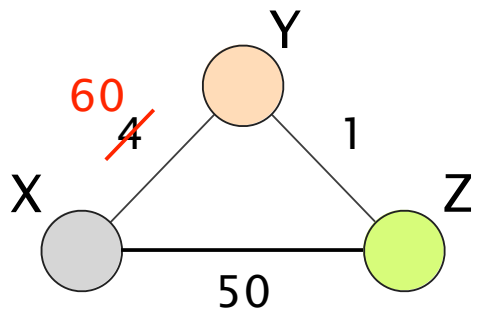
D. 51

E. 60

t = 0

(X,Y) weight changes

from 4 to 60



time t=0

Y state

dest.	via X	via Z
X	4	6

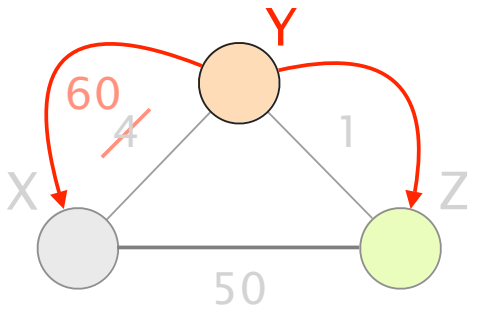
Z state

dest.	via X	via Y
X	50	5

t = 1

Y updates its vector,

sends it to X and Z



t=0

t=1

Y state

dest.	via X	via Z
X	4	6

dest.	via X	via Z
X	60	6

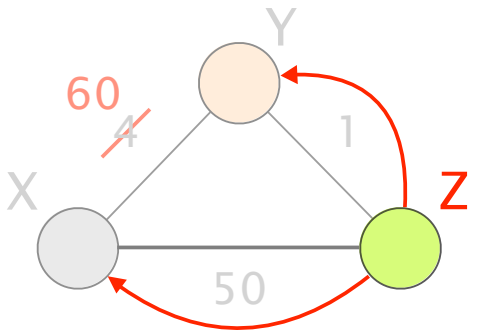
Z state

dest.	via X	via Y
X	50	5

t = 2

Z updates its vector,

sends it to X and Y



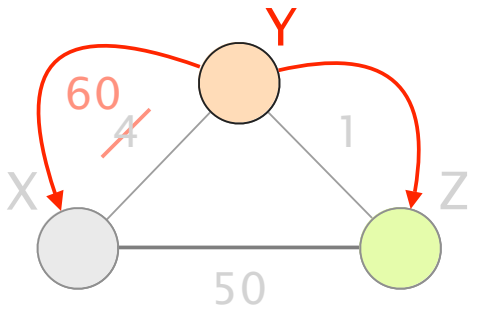
	t=0	t=1	t=2
Y state	dest. via X Z	dest. via X Z	
	X 4 6	X 60 6	
Z state	dest. via X Y		dest. via X Y
	X 50 5		X 50 7

Adapted from Laurent Vanbever, ETH Zurich

t = 3

Y updates its vector,

sends it to X and Z

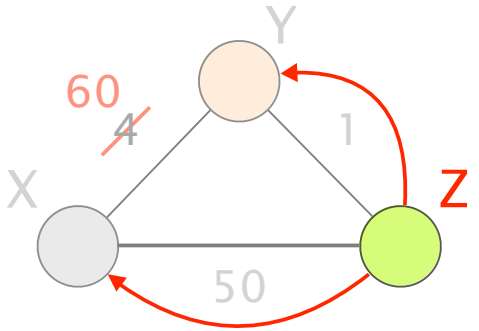


	t=0	t=1	t=2	t=3
Y state	dest. via X Z	dest. via X Z		dest. via X Z
	X 4 6	X 60 6		X 60 8
Z state	dest. via X Y		dest. via X Y	
	X 50 5		X 50 7	

t = 4

Z updates its vector,

sends it to X and Y…

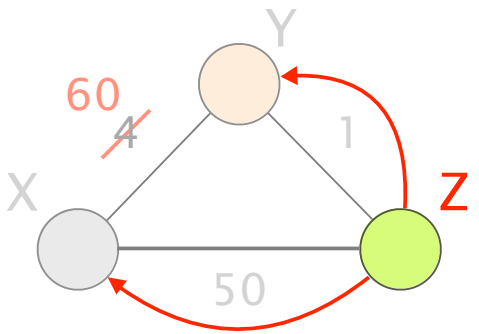


t=4

Y
state

Z
state

dest.	via	
	X	*Y*
X	50	9



t=4

Y state

dest.	via X	via Z
X	60	51

… many iterations later …

Z state

dest.	via X	via Y
X	50	9

dest.	via X	via Y
X	50	52

This problem is known as count-to-infinity, a type of routing loop

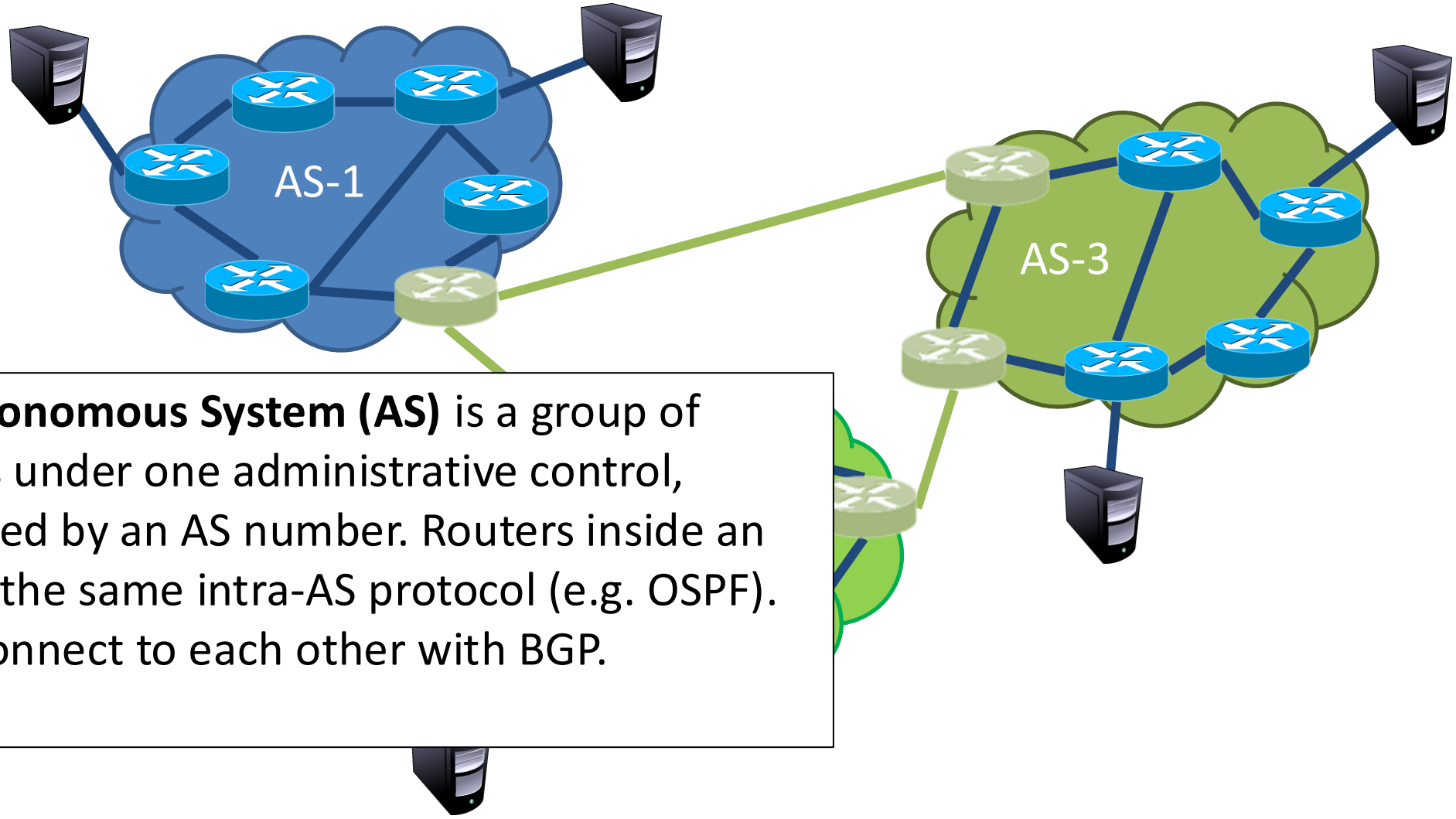
Count-to-infinity leads to very slow convergence

what if the cost had changed from 4 to 9999?

Routers don’t know "when neighbors use them"

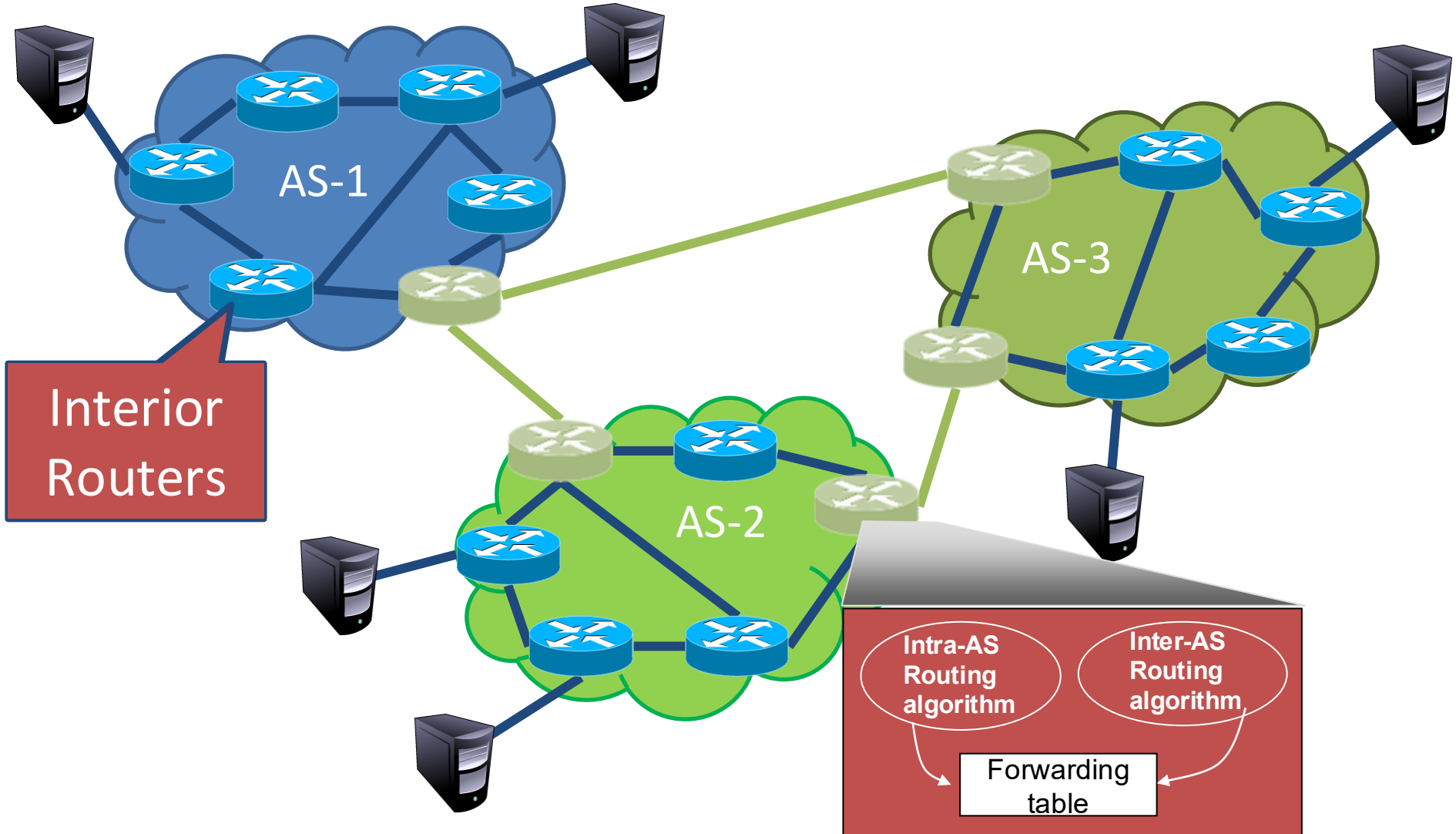
Z does not know that Y has switched to use itself

Hierarchical routing: Autonomous Systems

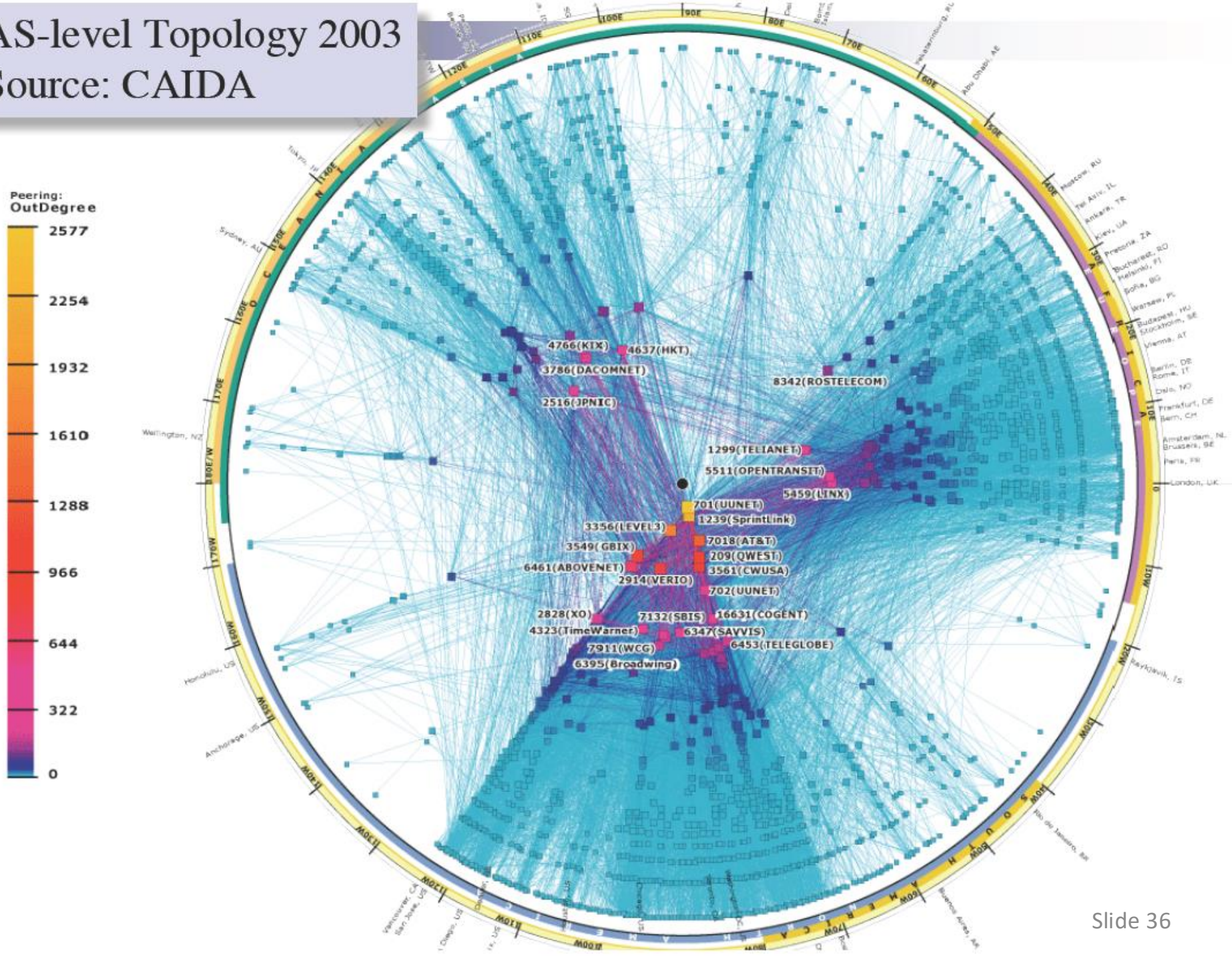


An **Autonomous System (AS)** is a group of routers under one administrative control, identified by an AS number. Routers inside an AS run the same intra-AS protocol (e.g. OSPF). ASes connect to each other with BGP.

Hierarchical routing: Interconnected ASes



AS-level Topology 2003
Source: CAIDA



BGP needs to solve three key challenges:

scalability, privacy and policy enforcement

There is a huge # of networks and prefixes

1M prefixes, >70,000 networks, millions (!) of routers

Networks don't want to divulge internal topologies

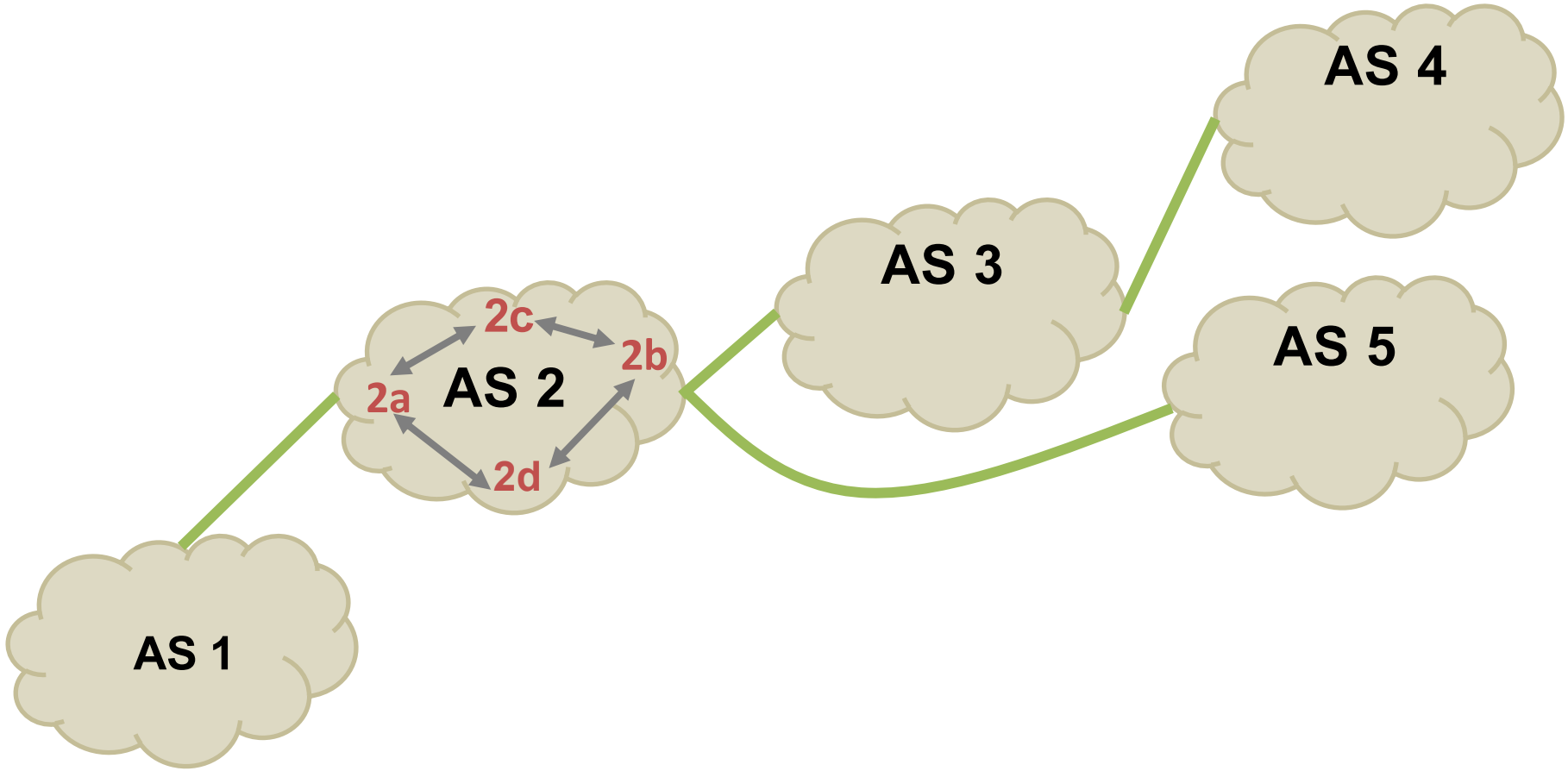
or their business relationships

Networks need to control where to send and receive traffic

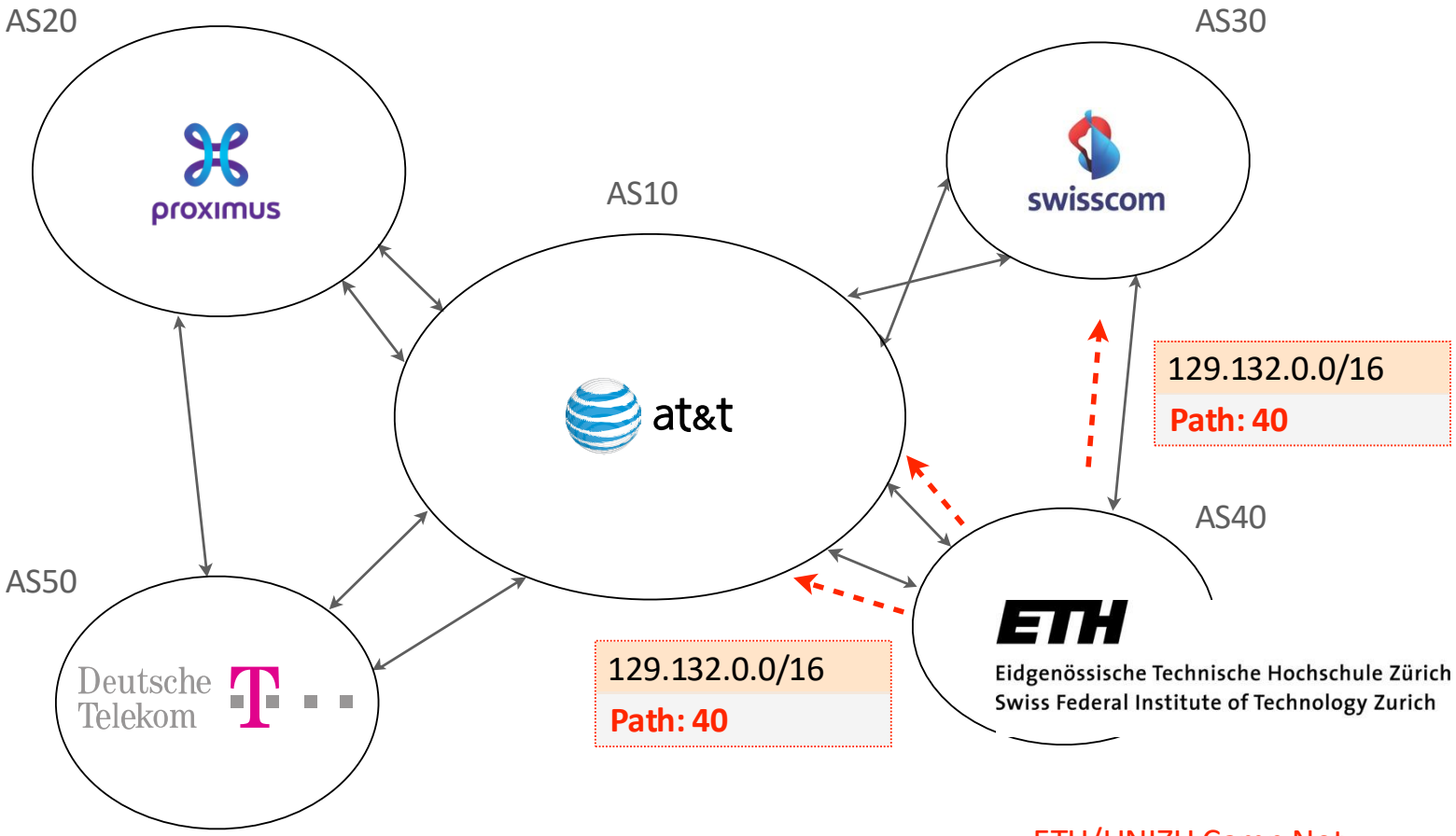
without an Internet-wide notion of a link cost metric

Path Vector Protocol

- Key idea: advertise the entire path
  - Distance vector: send *distance metric* per dest d
  - Path vector: send the *entire path* for each dest d



BGP announcements carry complete path information instead of distances

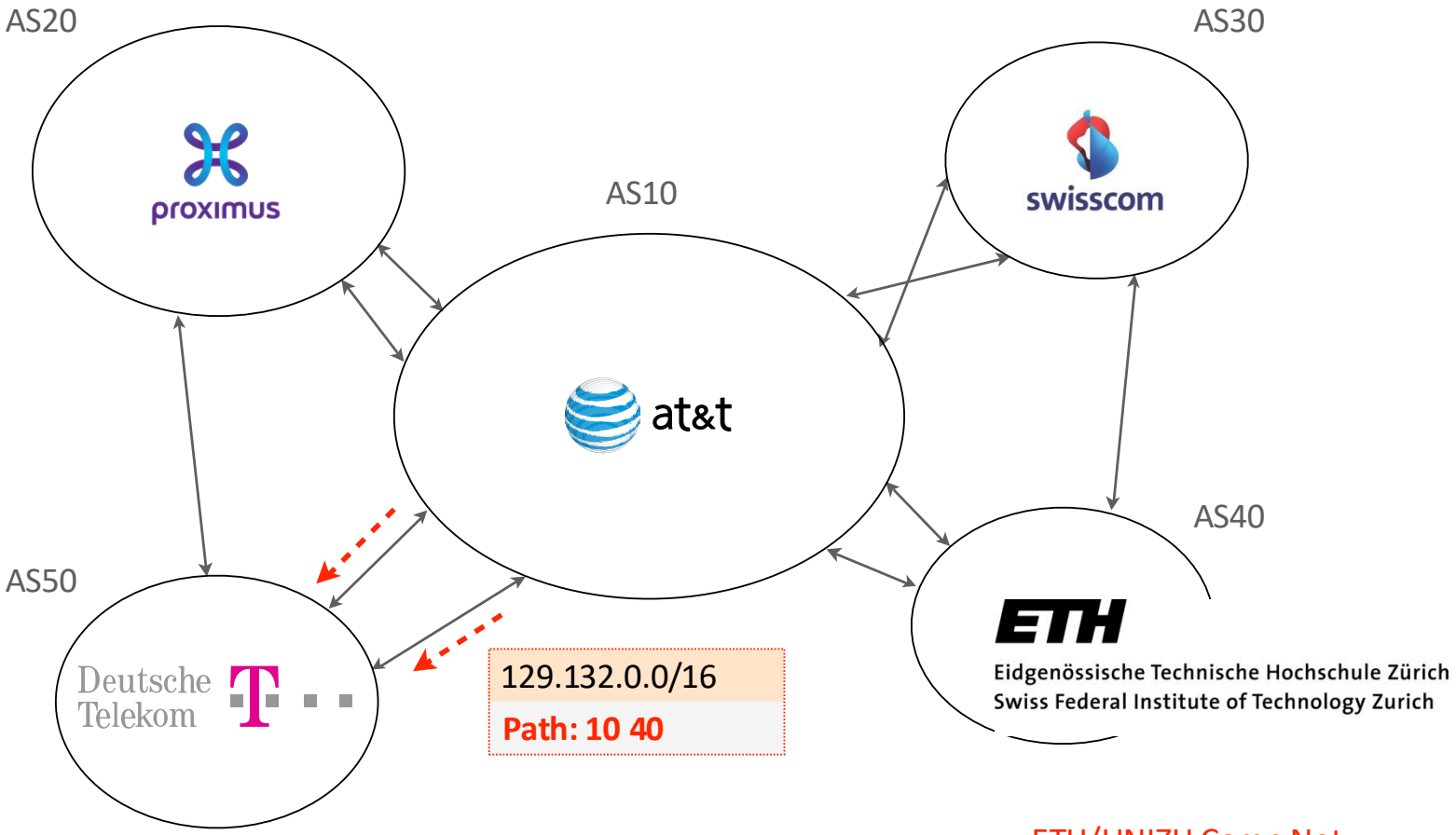


ETH/UNIZH Camp Net

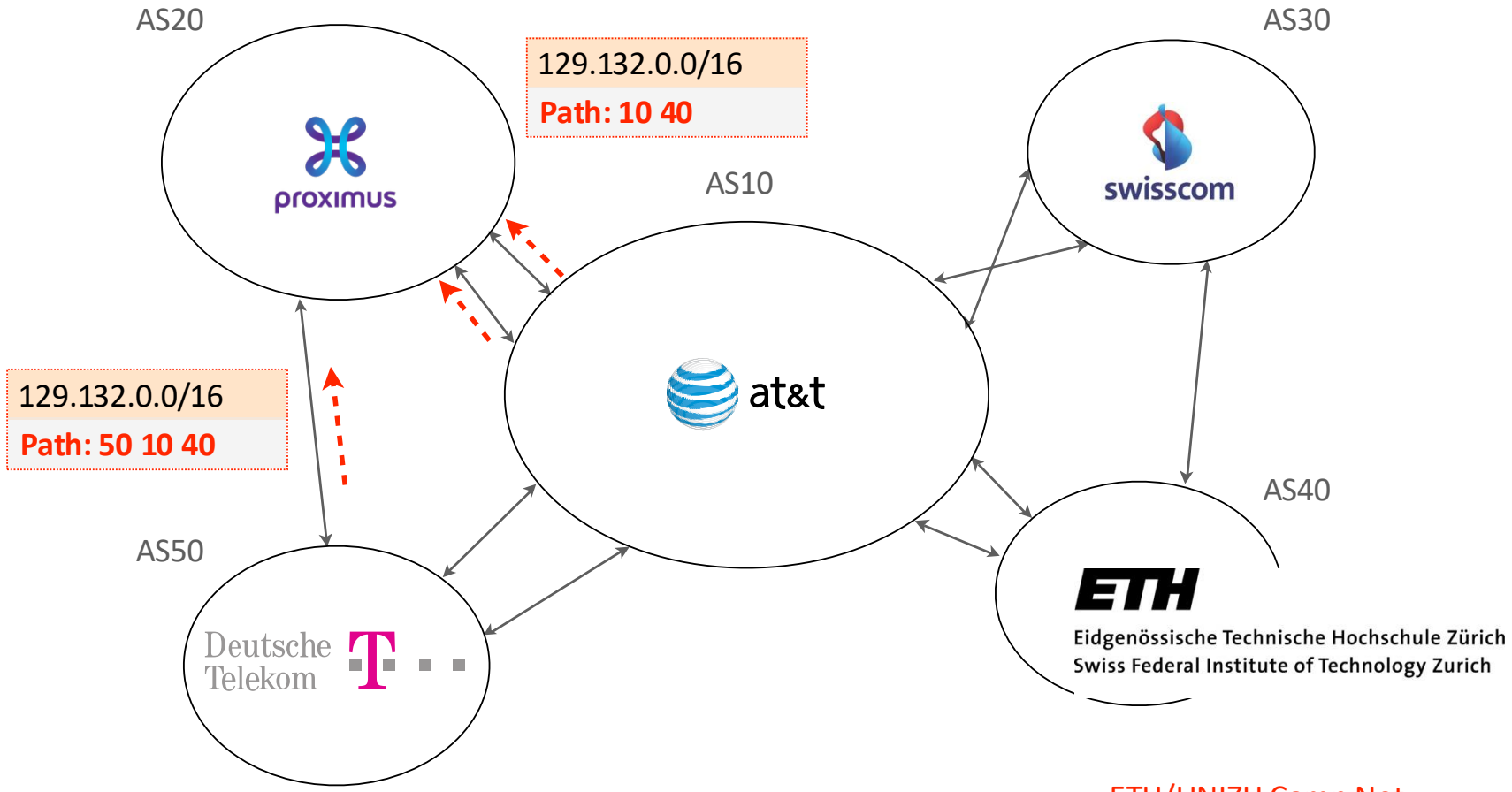
Adapted from Laurent Vanbever, ETH Zurich

Each AS appends itself to the path

when it propagates announcements



ETH/UNIZH Camp Net

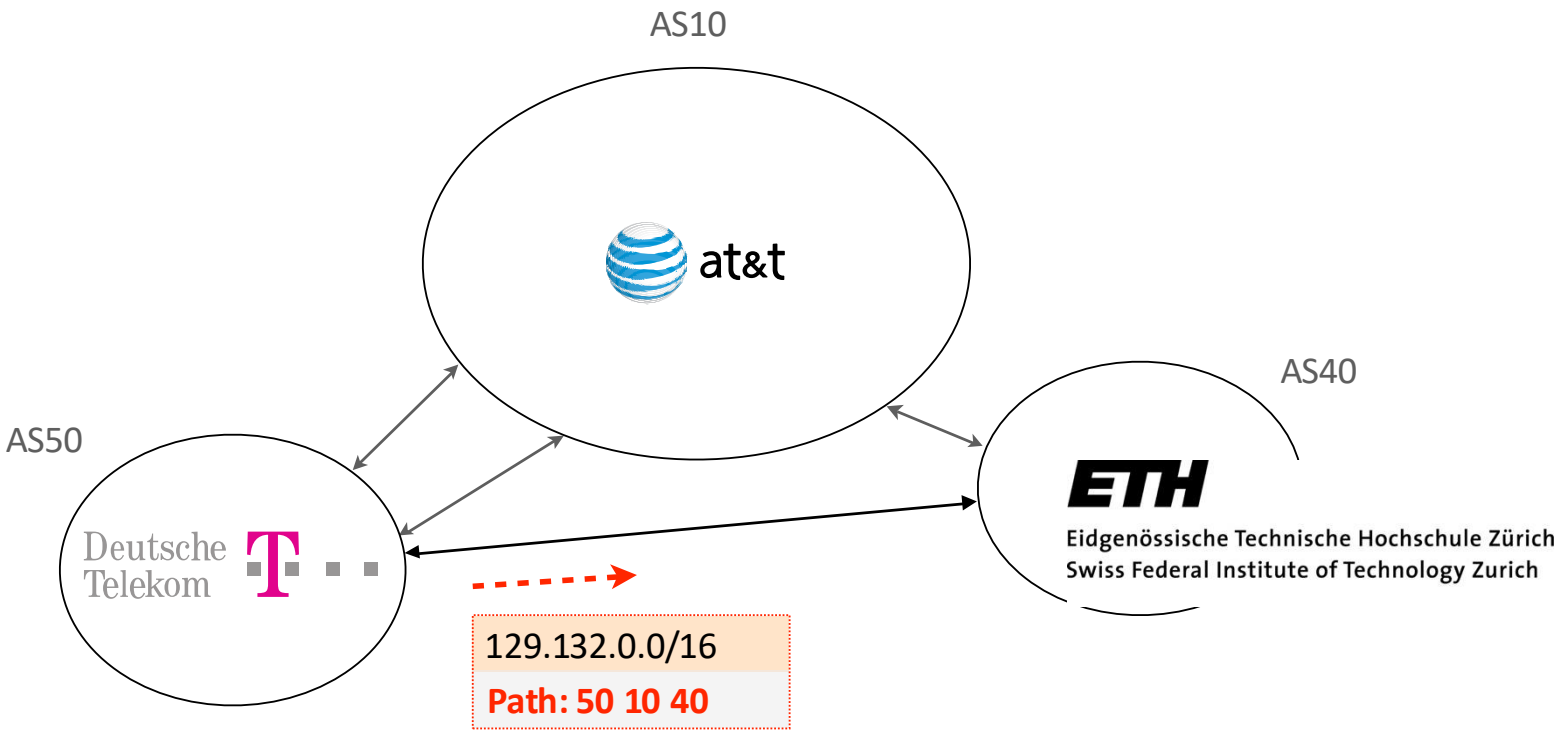


ETH/UNIZH Camp Net

Adapted from Laurent Vanbever, ETH Zurich

Complete path information enables ASes to easily detect a loop

ETH sees itself in the path and discards the route



Adapted from Laurent Vanbever, ETH Zurich

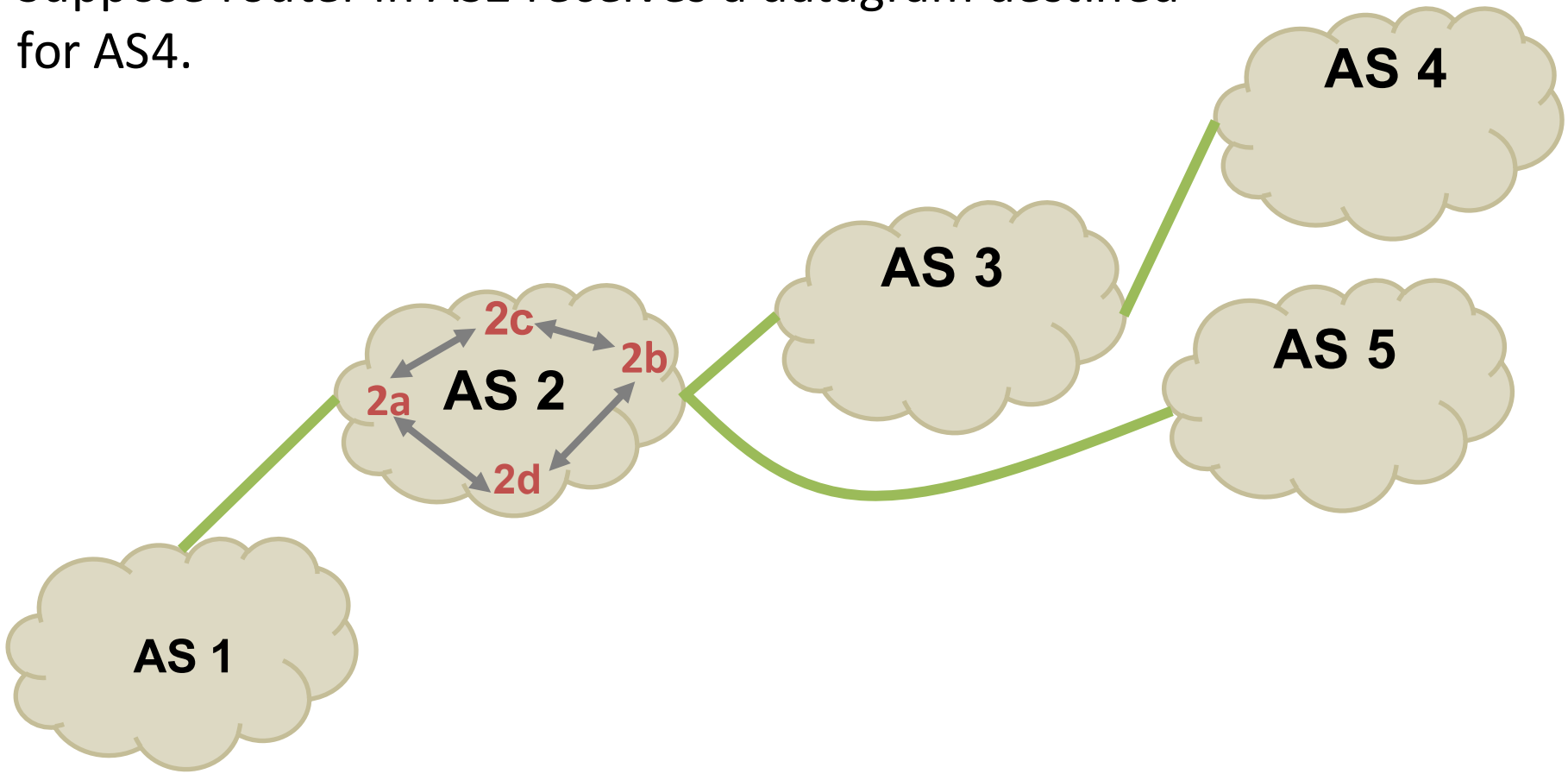
Life of a BGP router is made of three consecutive steps

while true:

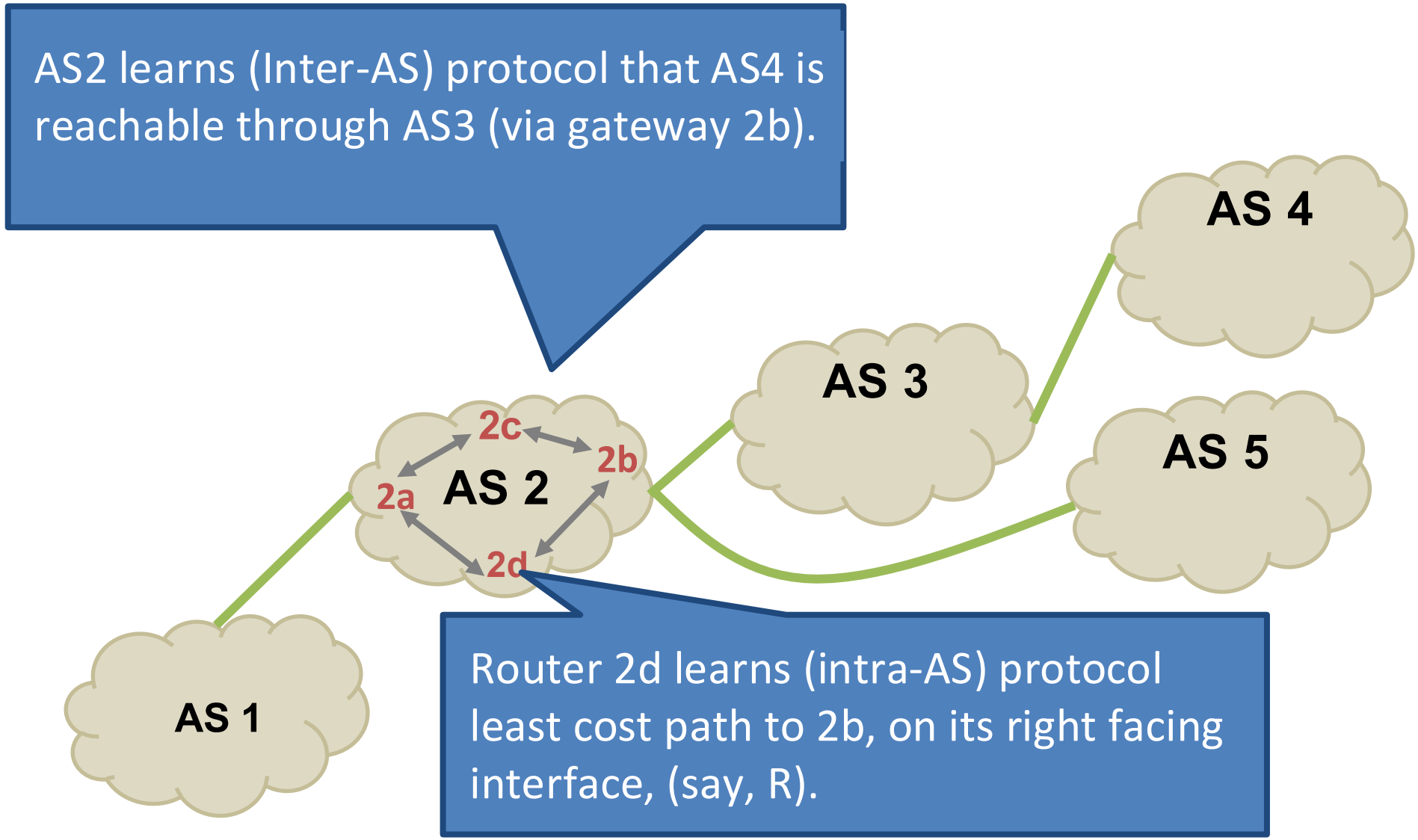
- receive routes from my neighbors
- select one best route for each prefix
- export the best route to my neighbors

Building the forwarding table in router 2d, for path to AS4

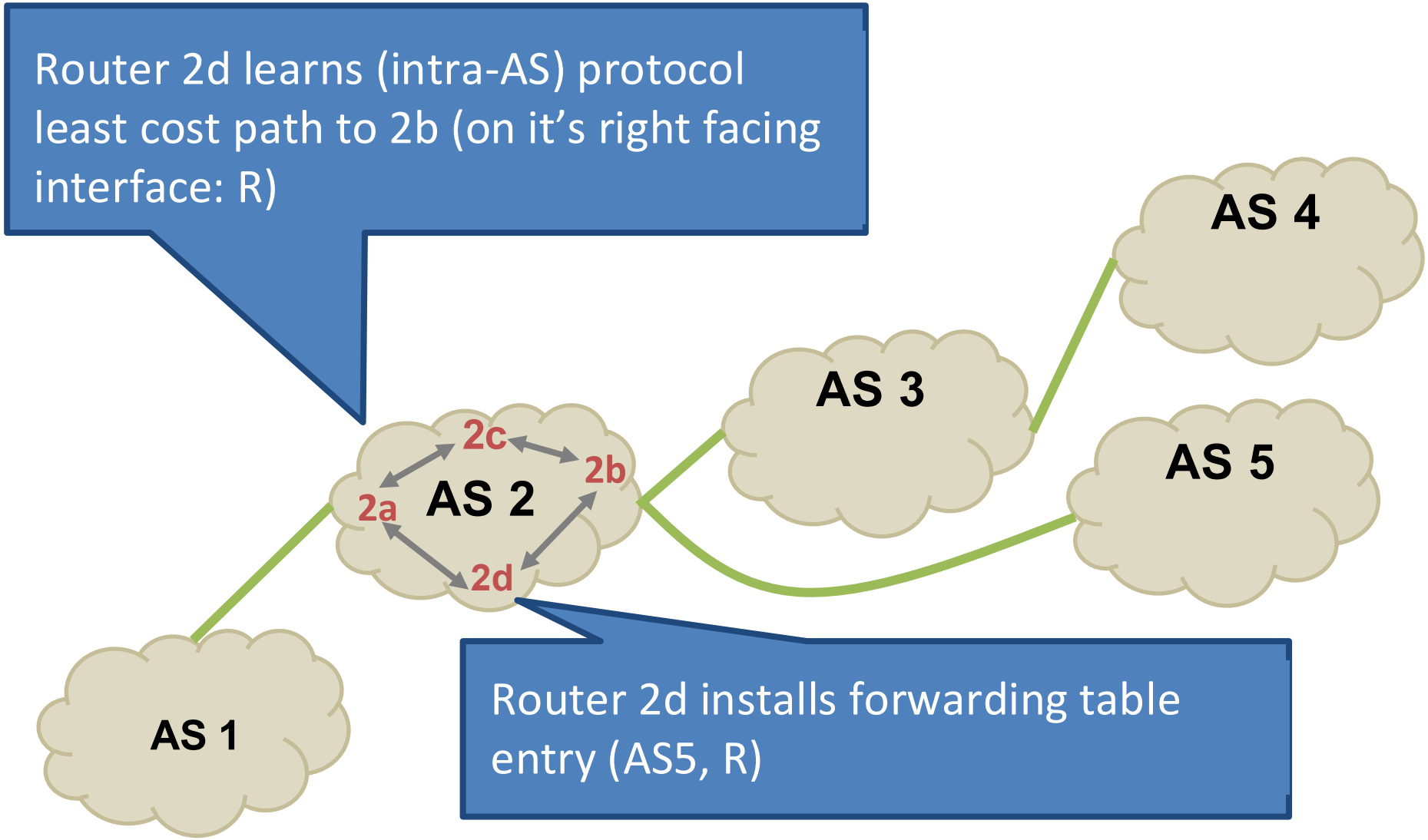
Suppose router in AS2 receives a datagram destined for AS4.



Building the forwarding table in router 2d, for path to AS4



Building the forwarding table in router 2d, for path to AS 5

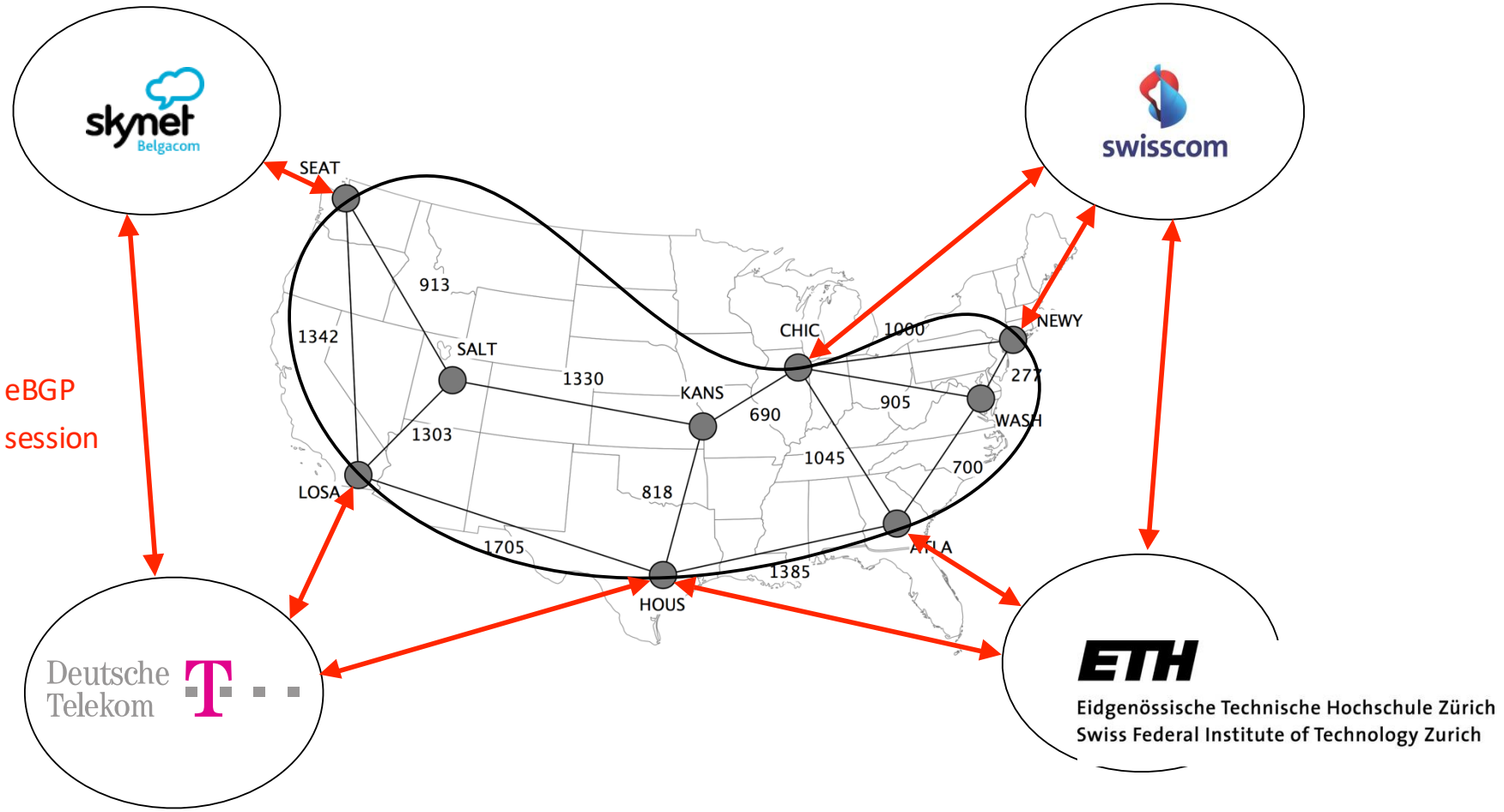


On the wire, BGP is a rather simple protocol composed of four basic messages

type	used to…
OPEN	establish TCP-based BGP sessions
NOTIFICATION	report unusual conditions
UPDATE	inform neighbor of a new best route a change in the best route the removal of the best route
KEEPALIVE	inform neighbor that the connection is alive

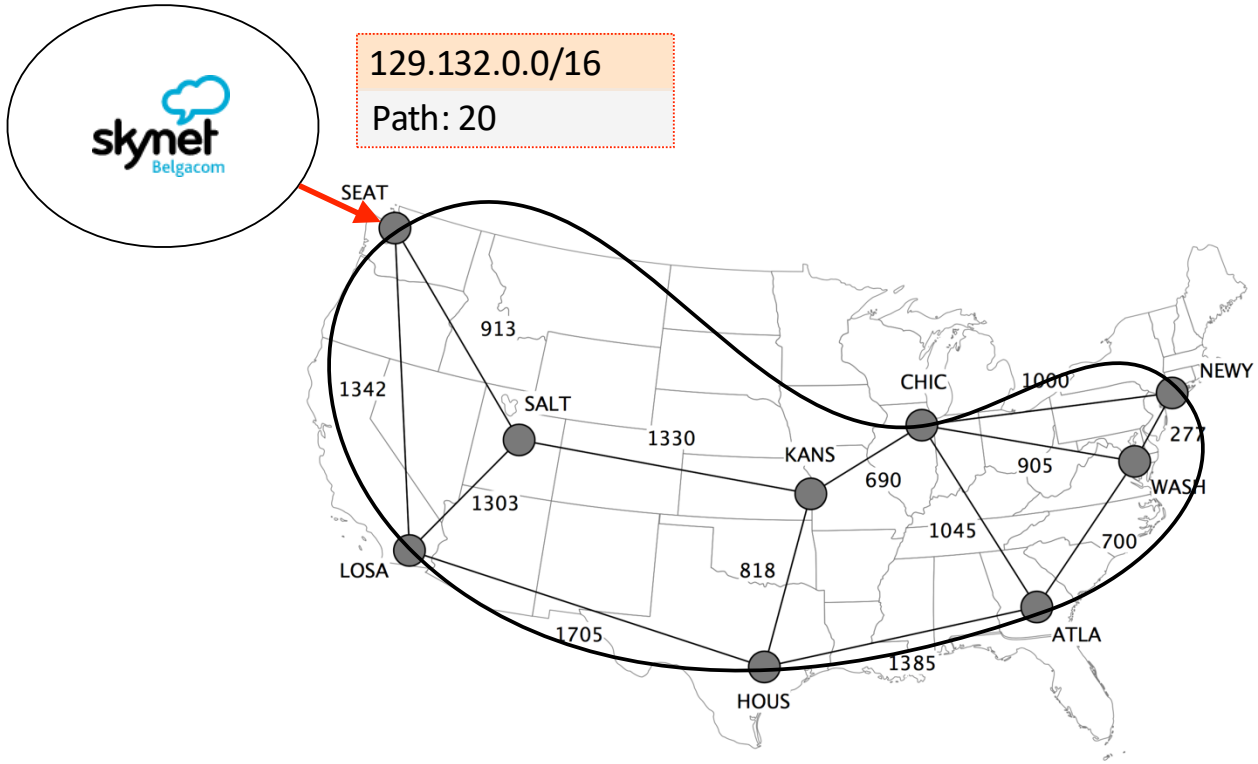
Adapted from Laurent Vanbever, ETH Zurich

external BGP (eBGP) sessions connect border routers in different ASes



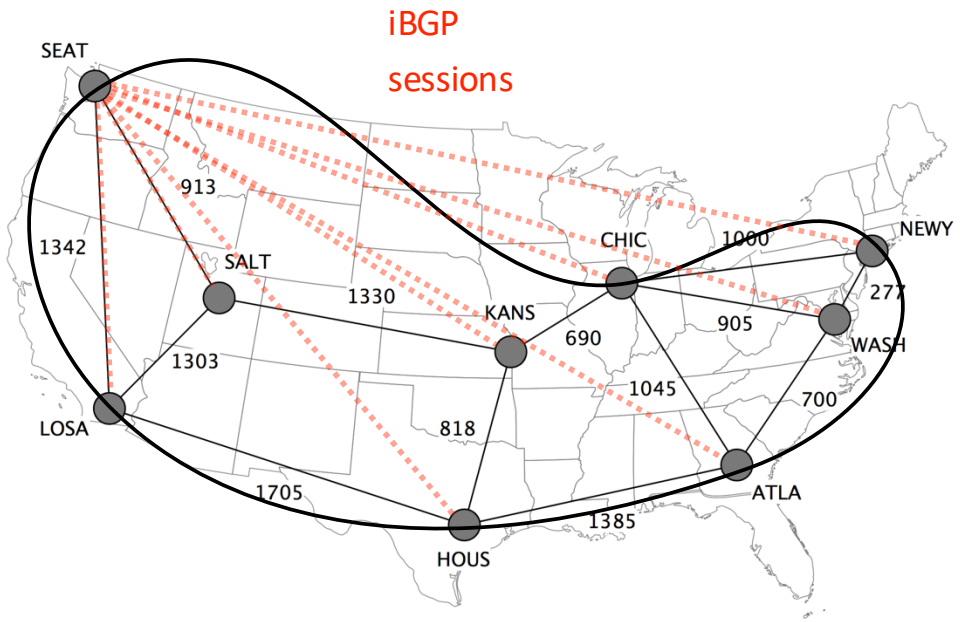
Adapted from Laurent Vanbever, ETH Zurich

Border routers use eBGP sessions to learn routes to external destinations

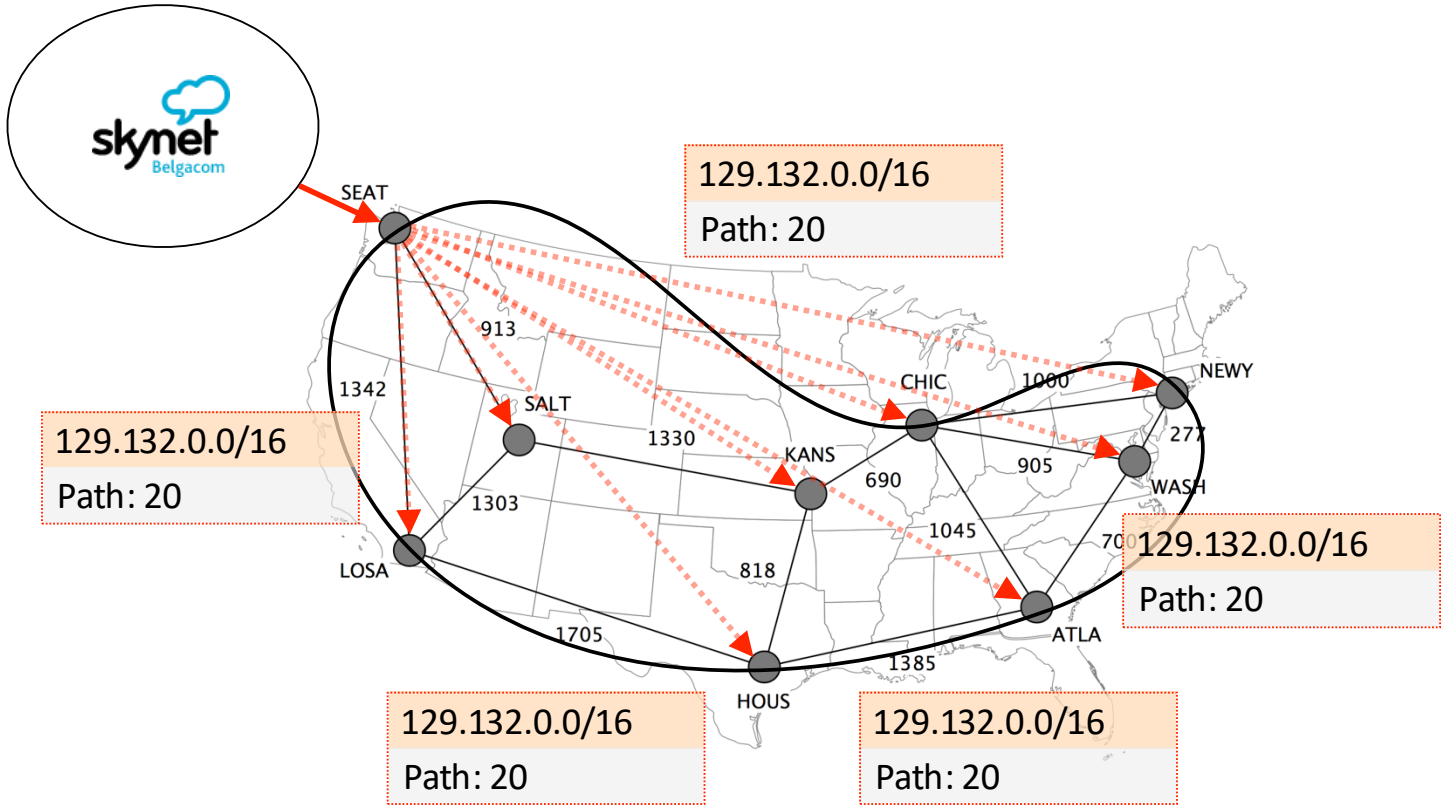


internal BGP (iBGP) sessions connect

all routers within the AS

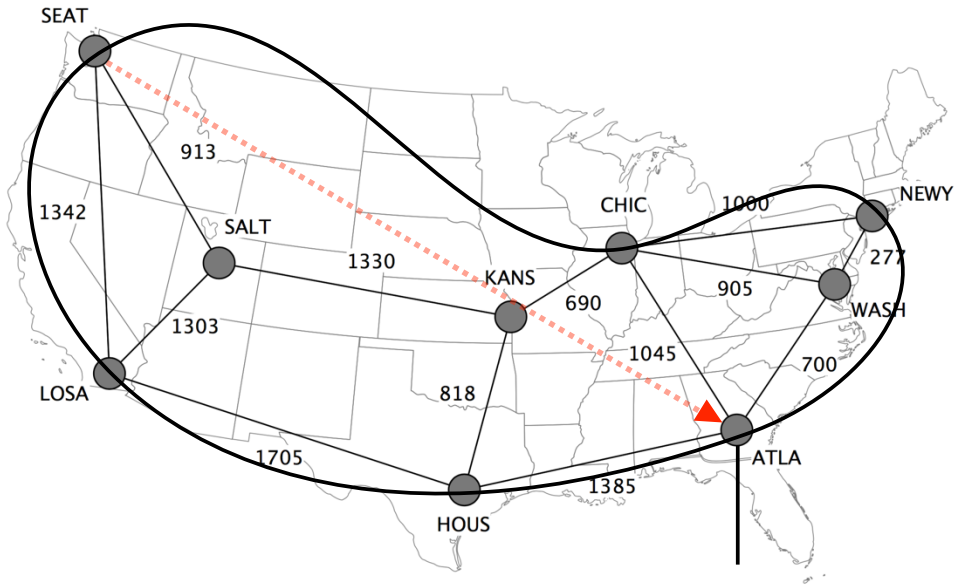


Border routers use iBGP sessions to disseminate externally-learned routes internally



129.132.0.0/16

Path: 20



I can reach "129.132/16" via SEAT,

internal NH is CHIC

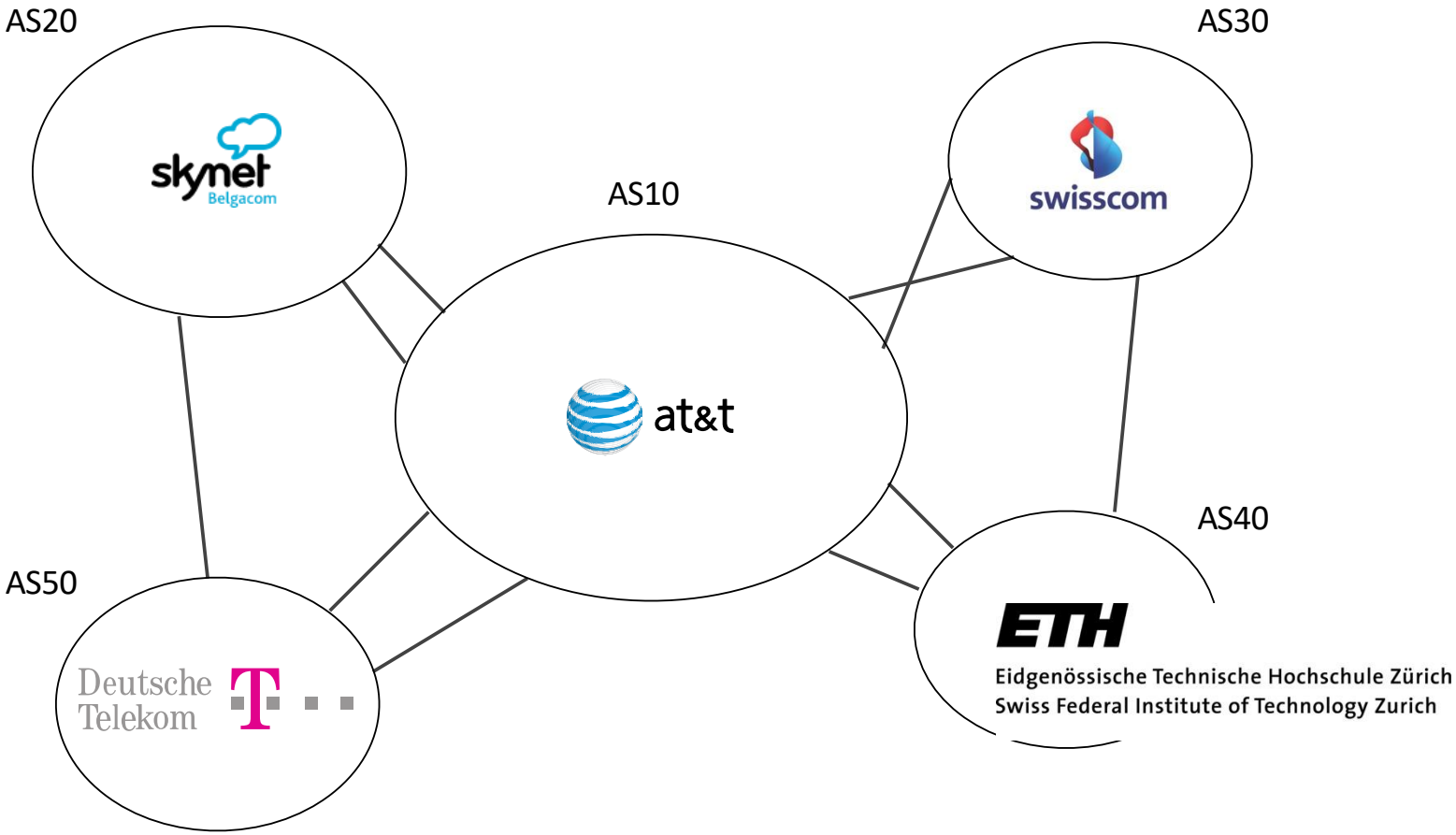
learned via IGP (*e.g.*, OSPF)

Adapted from Laurent Vanbever, ETH Zurich

Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol):
  The de facto inter-domain routing protocol
- BGP provides each AS a means to:
  - external BGP: obtain subnet reachability information from neighboring ASs.
  - internal BGP: propagate reachability information to all AS-internal routers.
  - determine "good" routes to other networks based on reachability information and policy.
- Allows a subnet to advertise its prefix to the rest of the Internet

The Internet topology is shaped according to business relationships



Adapted from Laurent Vanbever, ETH Zurich

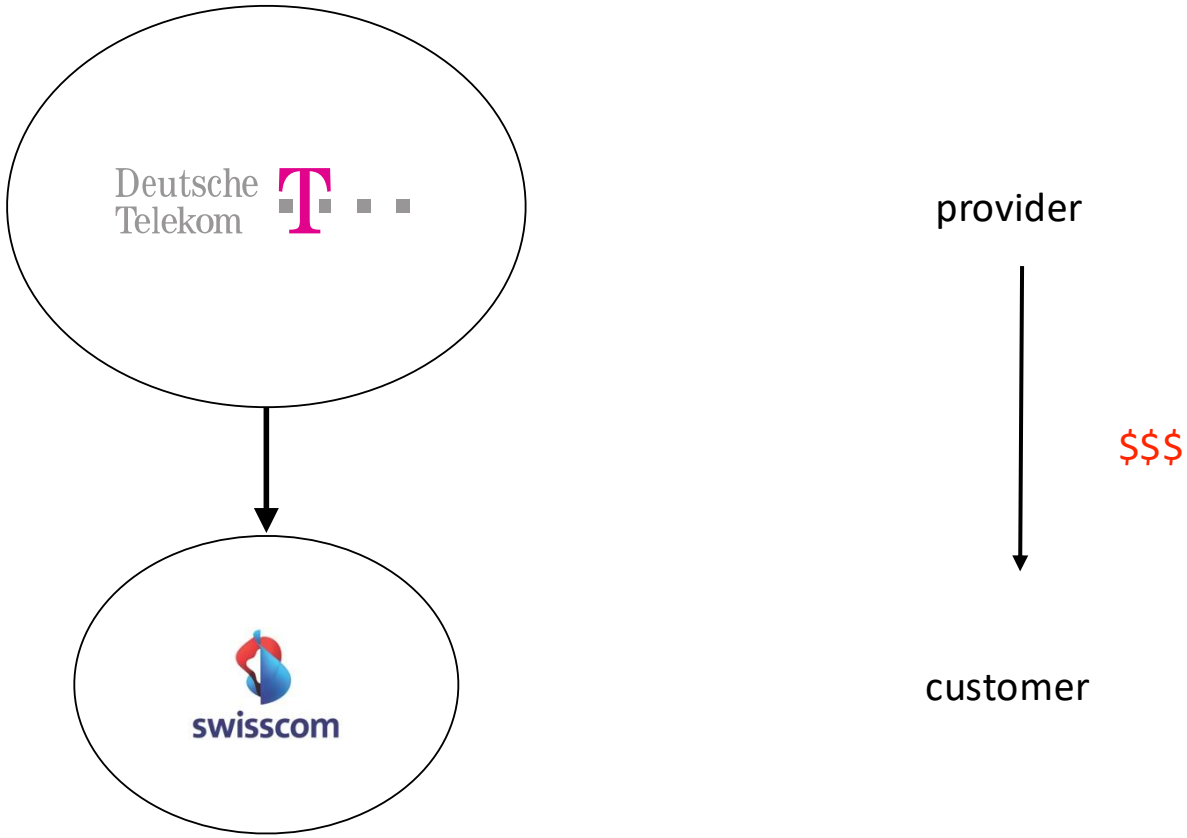
Intuition

2 ASes connect only if they have a business relationship

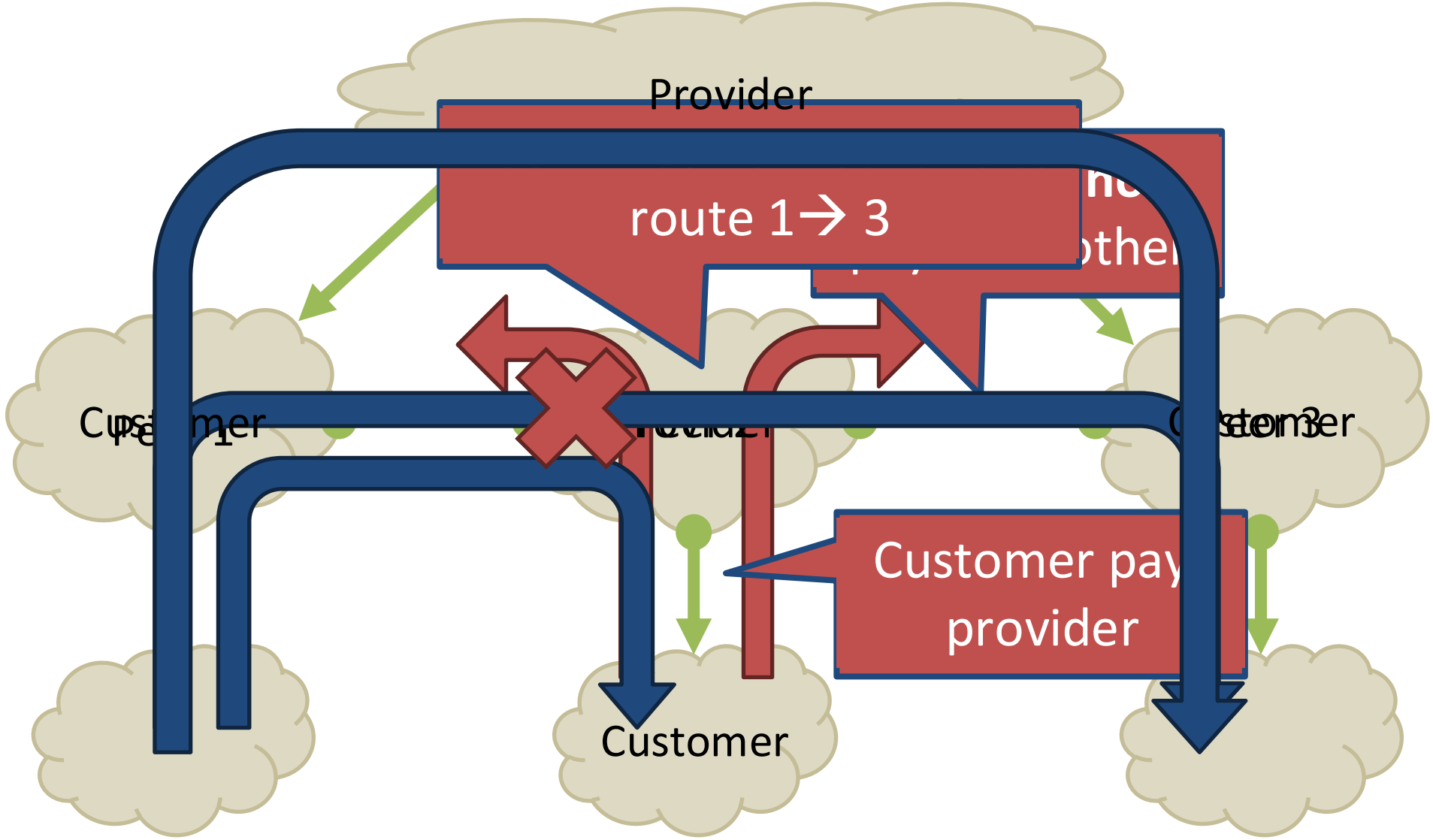
BGP is a “follow the money” protocol

Customers pay providers

to get Internet connectivity

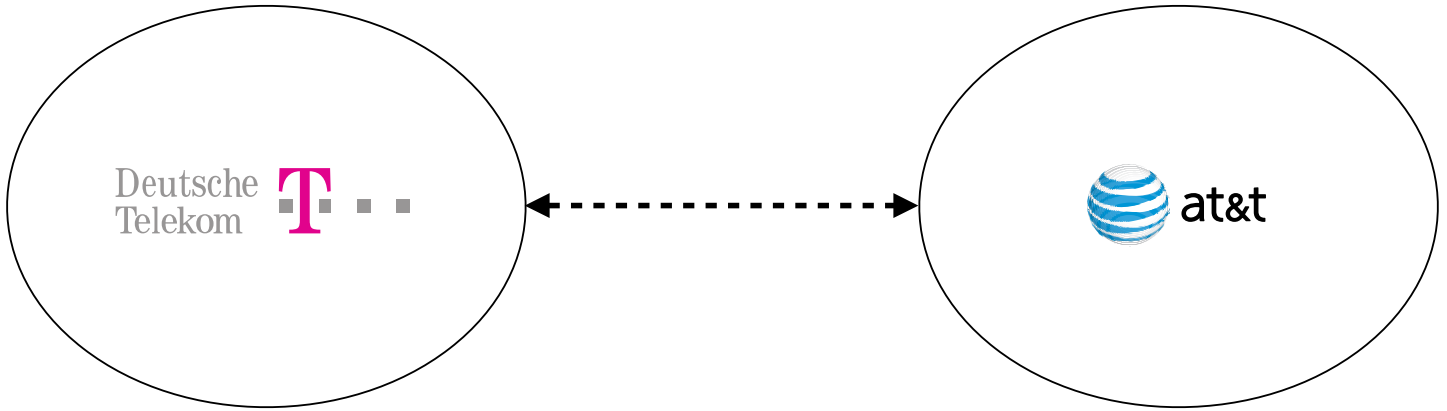


BGP Relationships



Peers don’t pay each other for connectivity,

they do it *out of common interest*

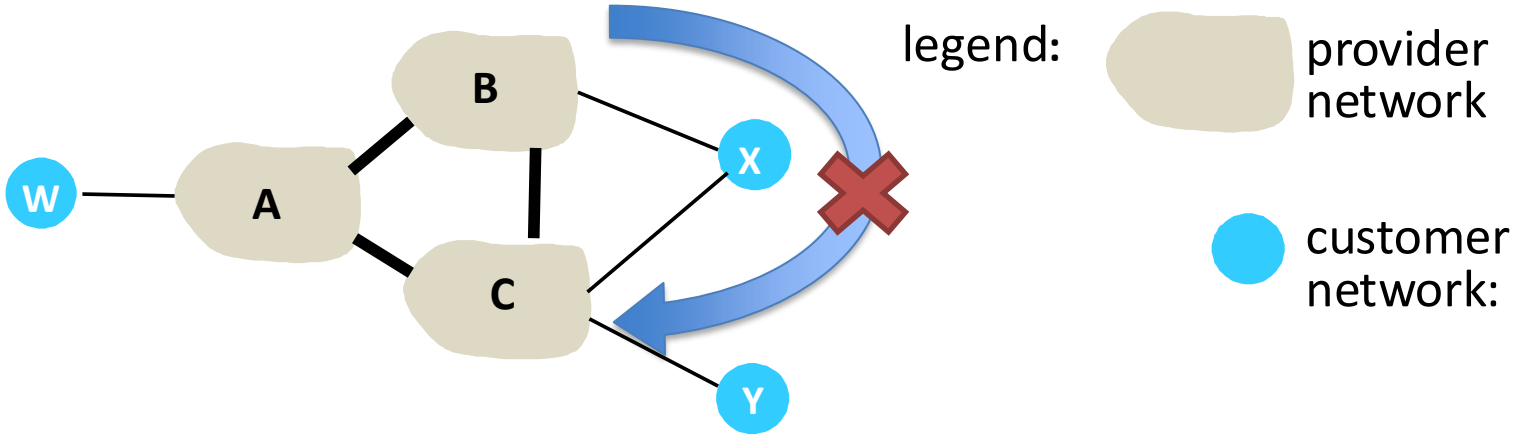


peer peer

DT and ATT exchange *tons* of traffic.

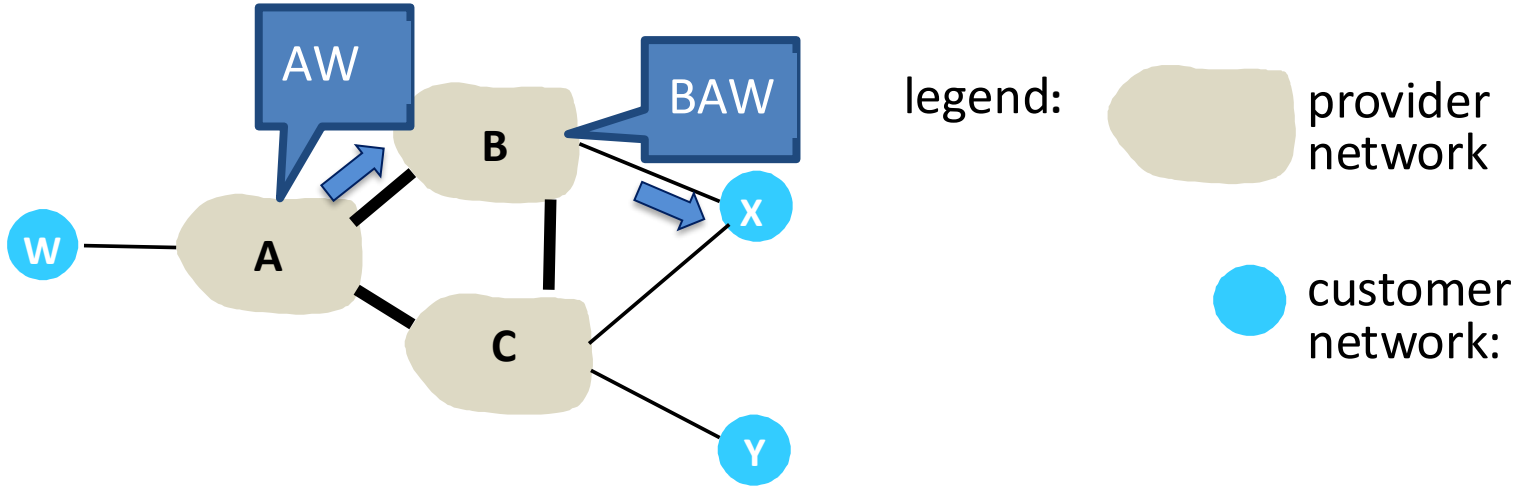
they save money by directly connecting to each other

BGP routing policy



- A,B,C are provider networks
- X,W,Y are customers of the providers
- X is dual-homed: attached to two networks (B and C)
  - X does not want to route from B via X to C
  - .. so X will not advertise to B a route to C

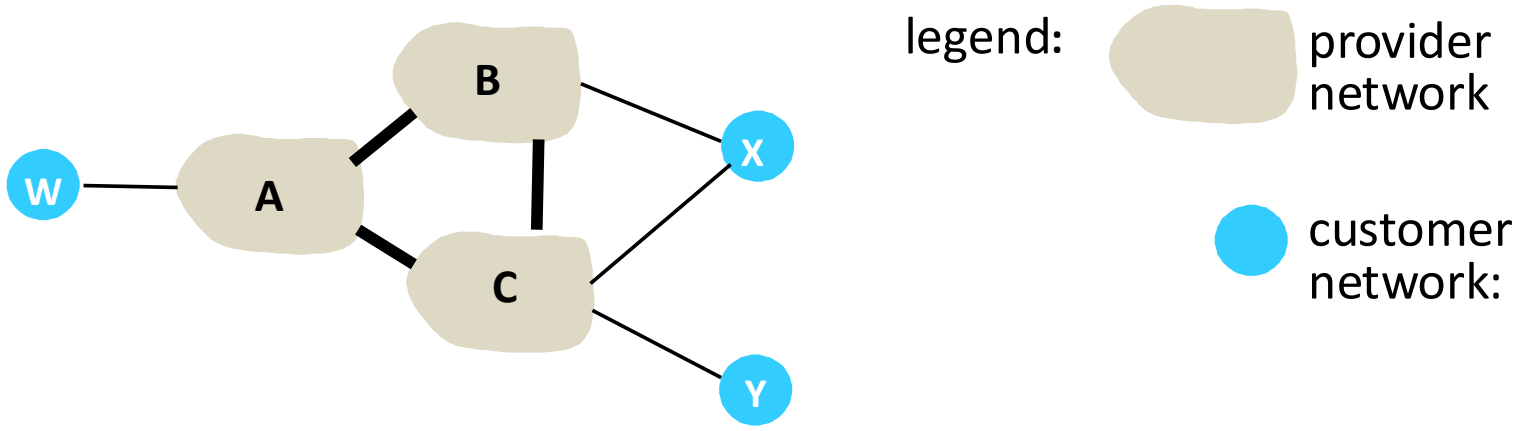
BGP routing policy

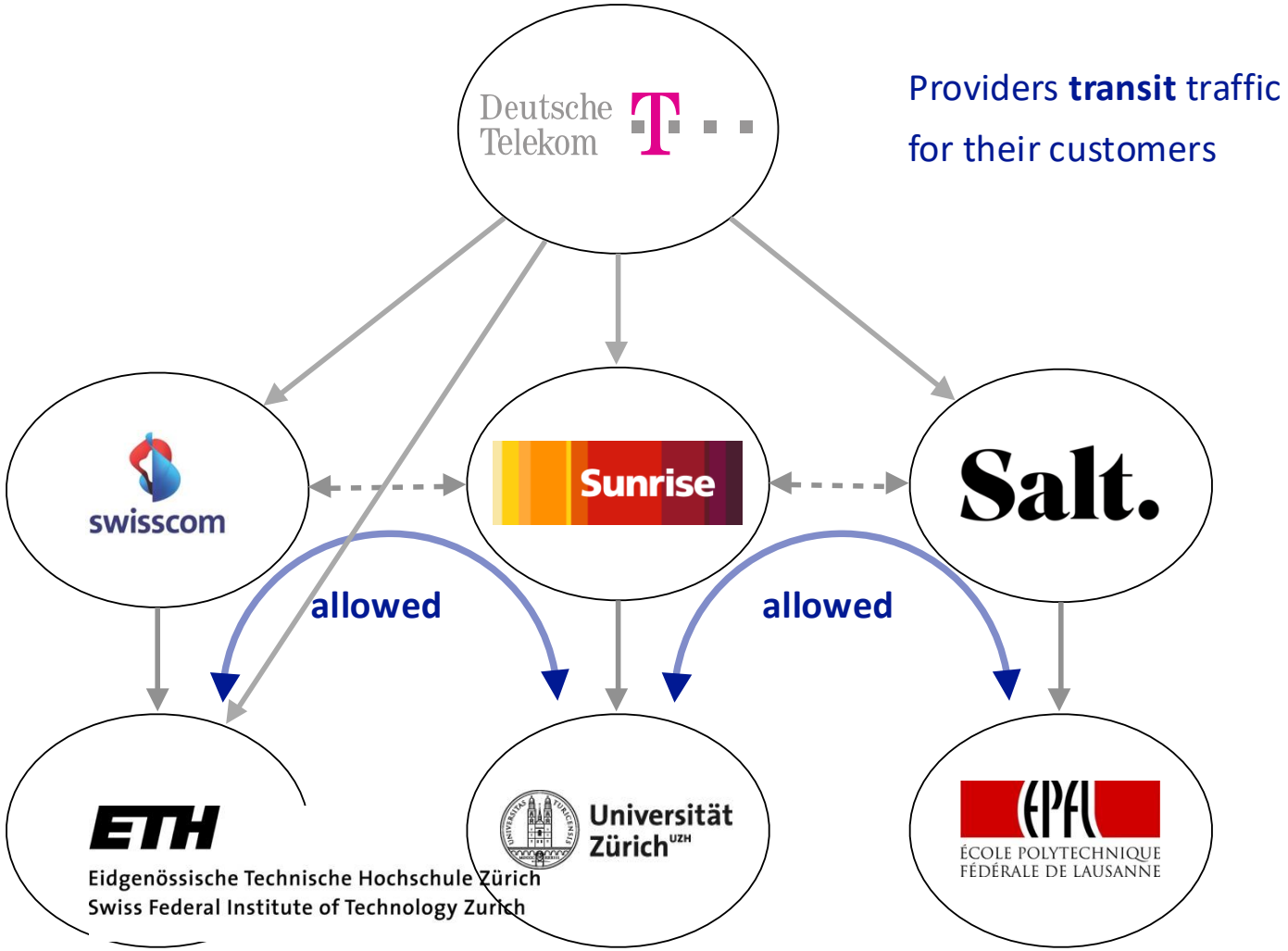


- A advertises path AW to B
- B advertises path BAW to X

B learned a route to w from its peer A. Which neighbors should B advertise it to? (B's neighbors are A, its peer; C, its peer; x, its customer.)

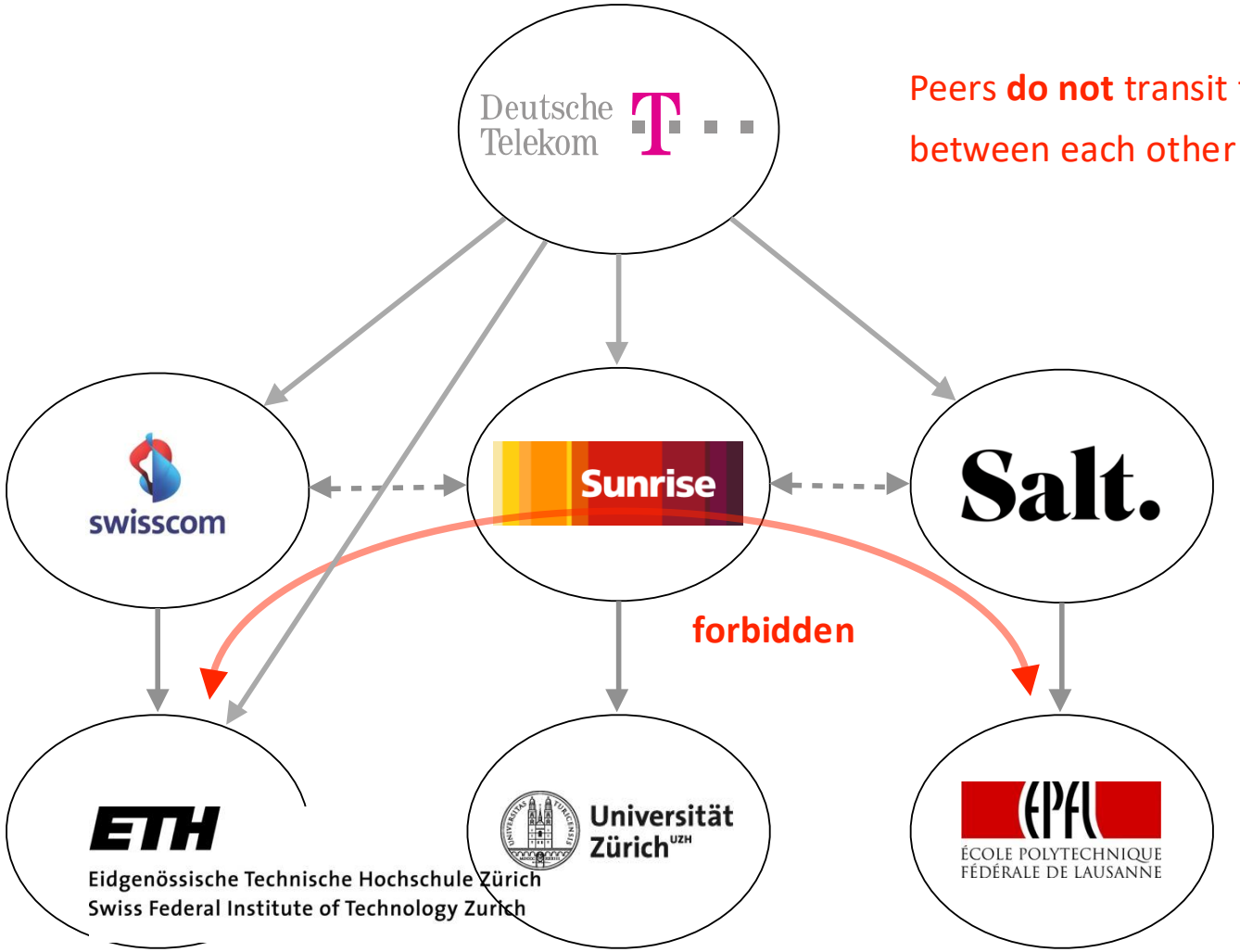
A. Nobody
B. Only x
C. x and C
D. A and C
E. x, A, and C



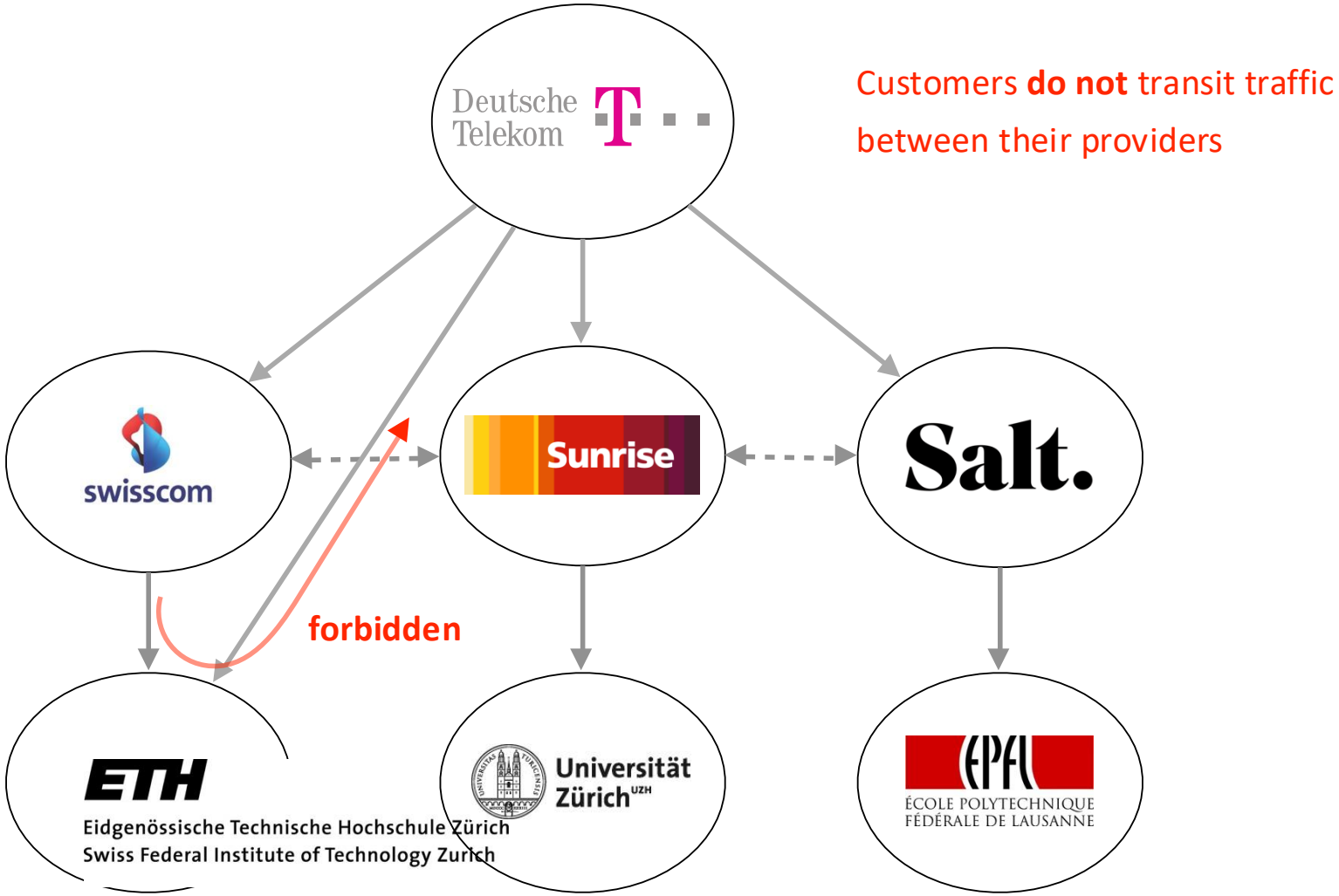


Providers **transit** traffic for their customers

Adapted from Laurent Vanbever, ETH Zurich



Adapted from Laurent Vanbever, ETH Zurich

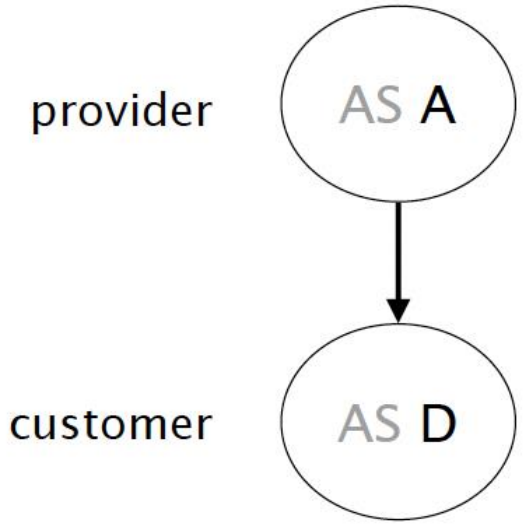


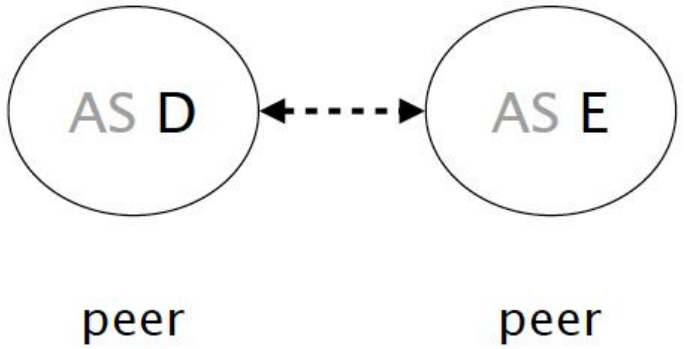
Customers **do not** transit traffic between their providers

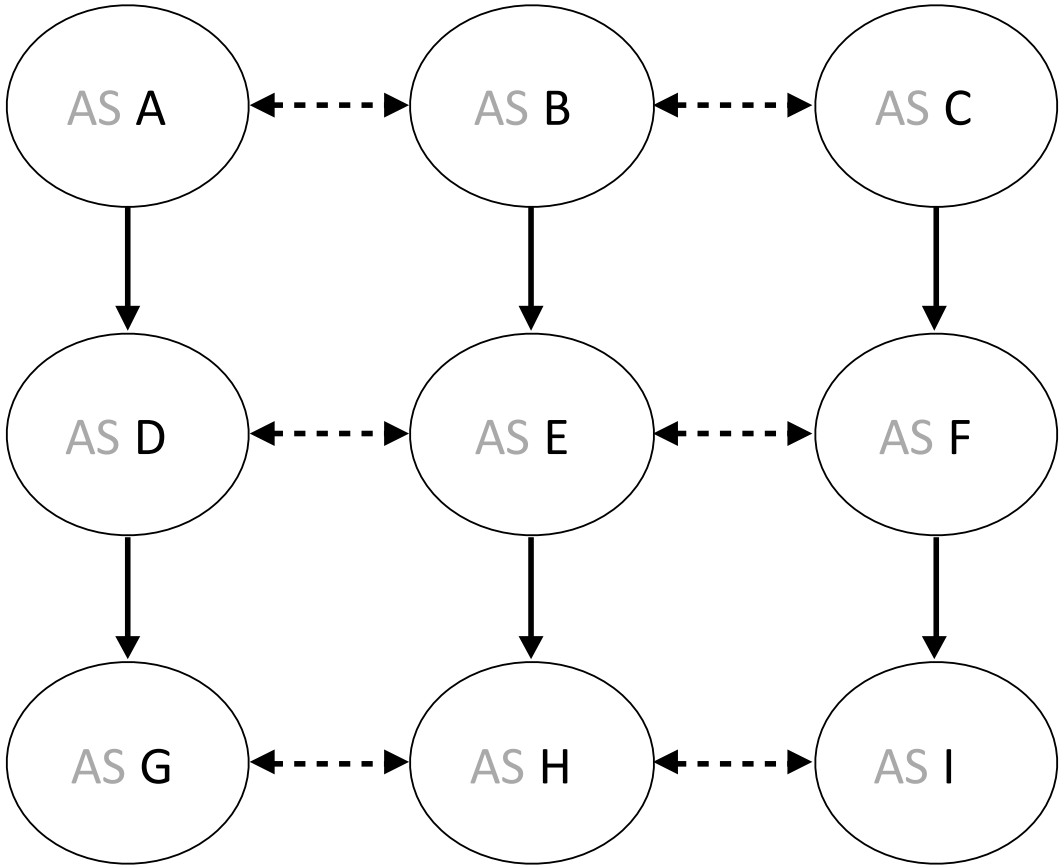
Adapted from Laurent Vanbever, ETH Zurich

Routes coming from peers and providers are only propagated to customers

		send to customer	peer	provider
	customer	✓	✓	✓
from	peer	✓	-	-
	provider	✓	-	-



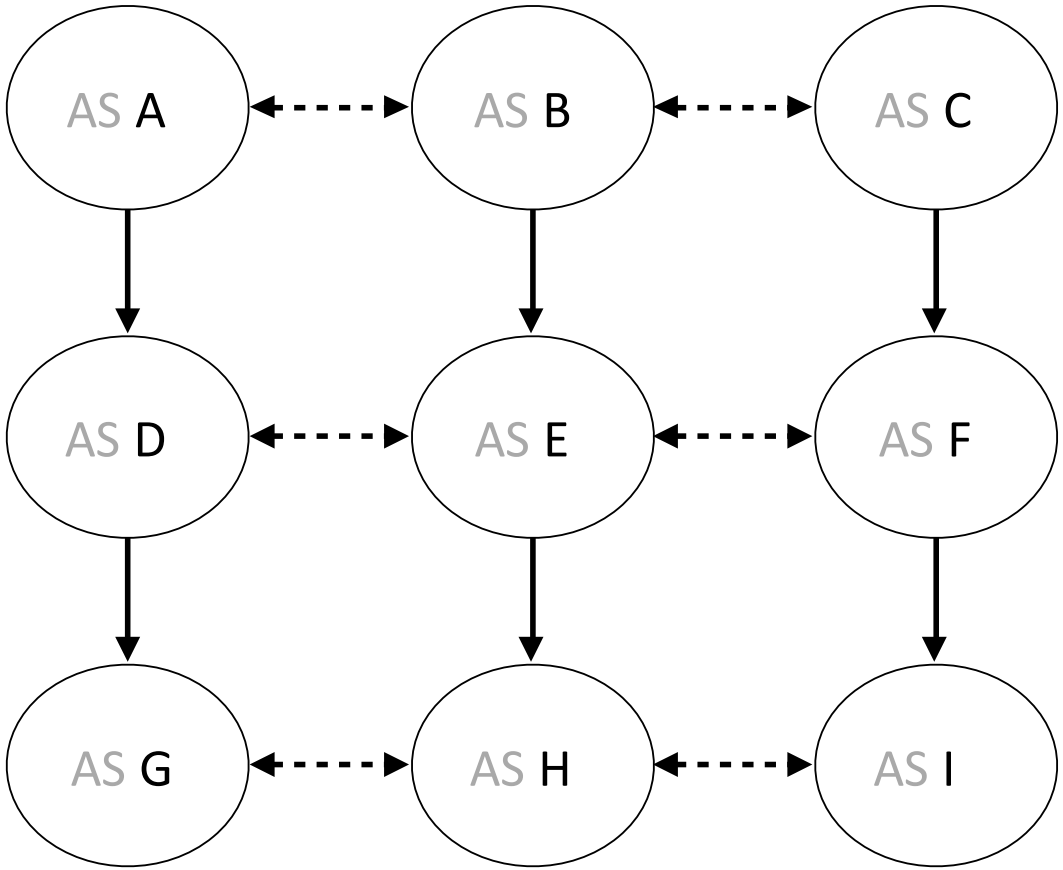




Adapted from Laurent Vanbever, ETH Zurich

A path is valid if every AS on it is following the export rules. Which of these paths from G to H is valid?

1. G, H
2. G,D,E,H
3. G,D,A,B,E,H
4. G,D,E,F,I



Summary

- As we've seen before (DNS), a hierarchy can help manage state storage constraints.
  - intra-AS routing: lots of info about local routes
  - inter-AS routing: less info about far away routes
- BGP: the inter-AS routing protocol for the Internet
  - Decisions often contractual
- BGP advertises AS prefixes, including:
  - entire path of ASes along the way
  - which border router heard the advertisement (Next Hop)