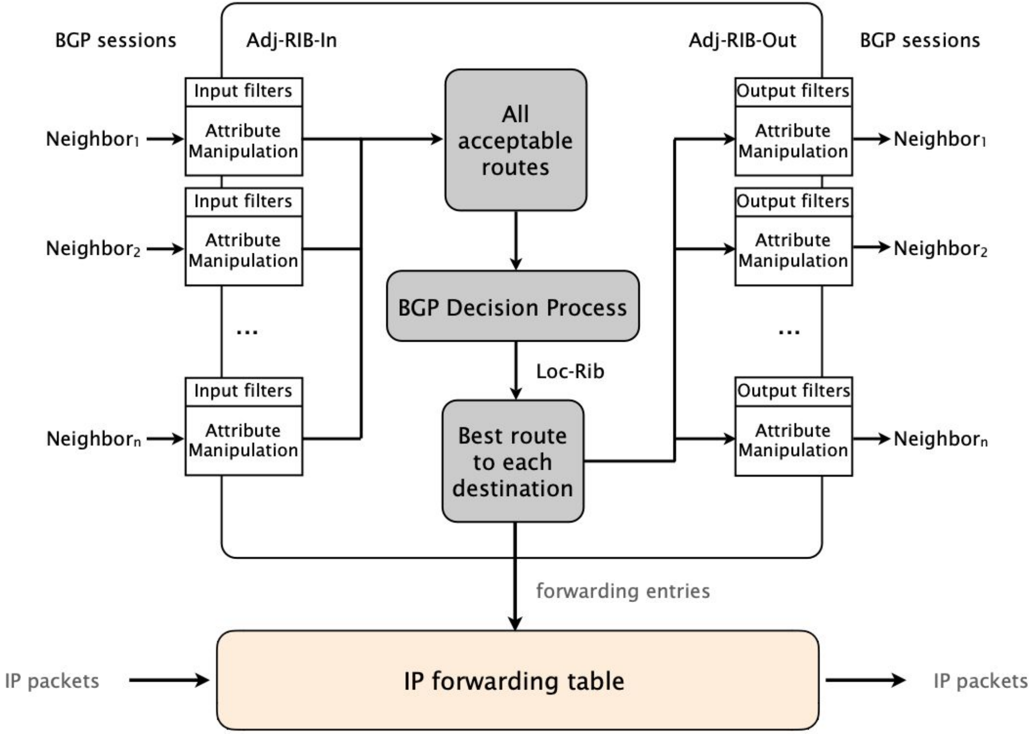
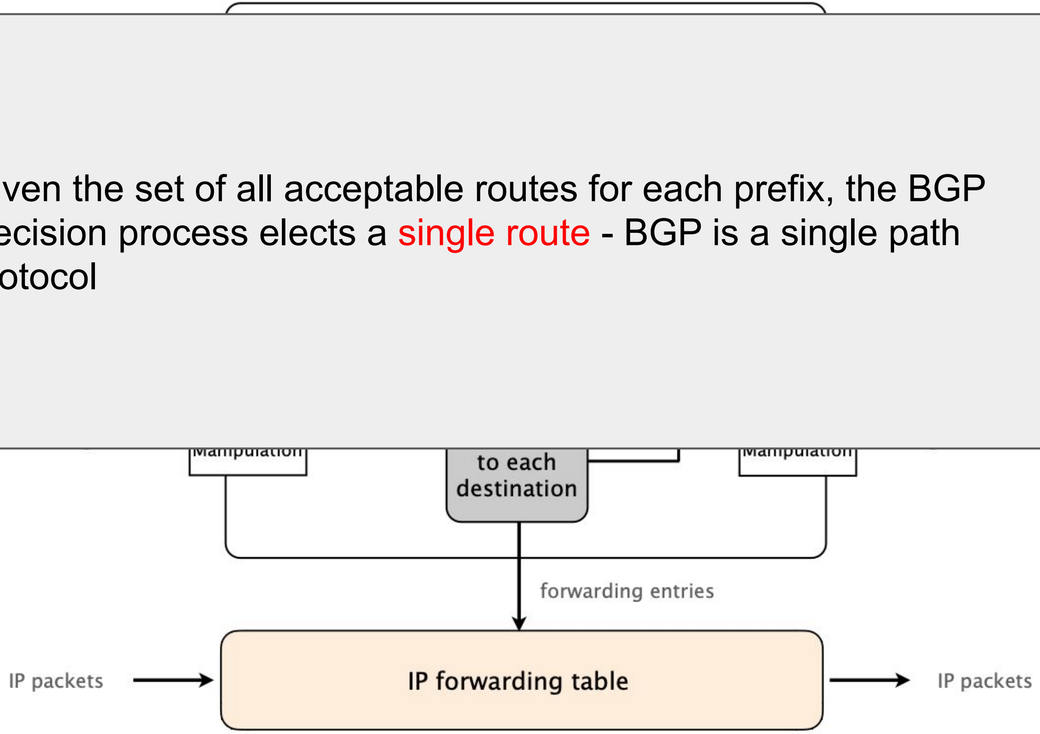
Each BGP router processes UPDATEs according to a precise pipeline



Each BGP router processes UPDATEs according to a precise pipeline

Given the set of all acceptable routes for each prefix, the BGP Decision process elects a single route - BGP is a single path protocol



Each BGP router processes UPDATEs according to a precise pipeline

Prefer routes…

with higher LOCAL-PREF

with shorter AS-PATH length

with lower MED

learned via eBGP instead of iBGP

with lower IGP metric to the next-hop

with smaller egress IP address (tie-break)

Each BGP router processes UPDATEs according to a precise pipeline

Prefer routes...

These two steps aim at directing traffic
as quickly as possible out of the AS (early exit routing)

learned via eBGP instead of iBGP

with lower IGP metric to the next-hop

with smaller egress IP address (tie-break)

In Reality

Rules for route selection in priority order

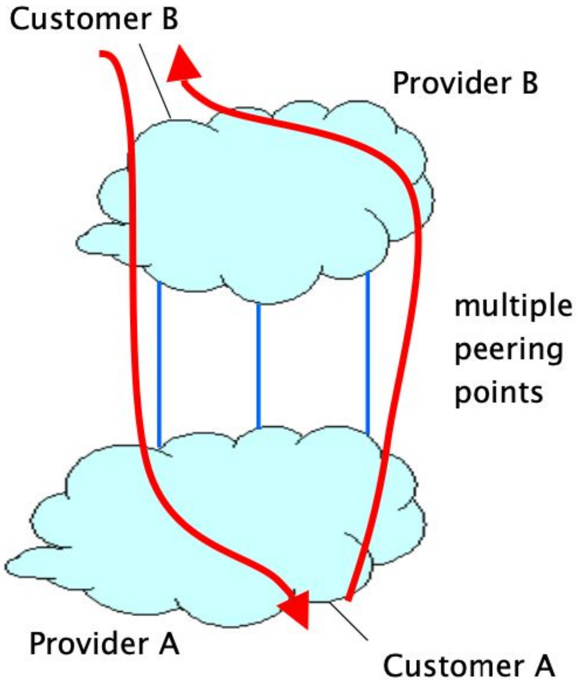
1. Make or save money (send to customer > peer > provider)
2. Maximize performance (smallest AS path length)
3. Minimize use of my network bandwidth ("hot potato")

ASes are selfish

They dump traffic
as soon as possible
to someone else

This leads to asymmetric routing

Traffic does not flow on
the same path
in both directions



Typical BGP Export Policy

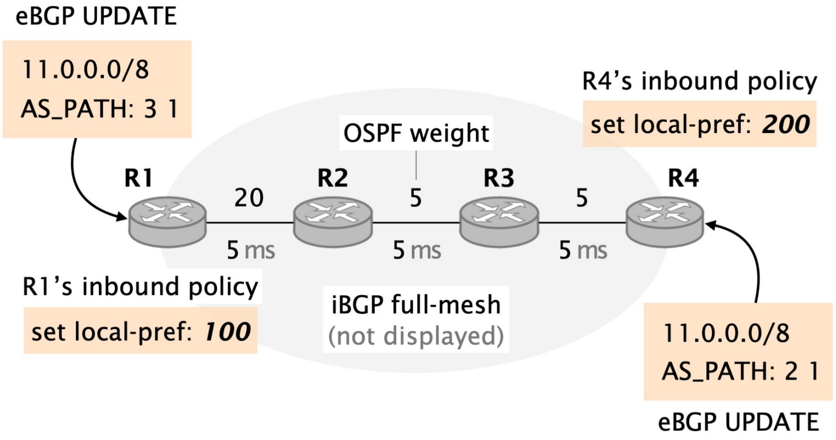
Destination prefix advertised by…	Export route to…
Customer	Everyone (providers, peers, other customers)
Peer	Customers
Provider	Customers

Question

Consider the BGP network composed of 4 routers depicted in Figure on the right. Two of these routers, R1 and R4 are egress routers and maintain eBGP sessions with external neighbors.

R1 is configured to associate a local-preference of 100 to externally-learned routes, while R4 is configured to associate a local-preference of 200 to externally-learned routes.

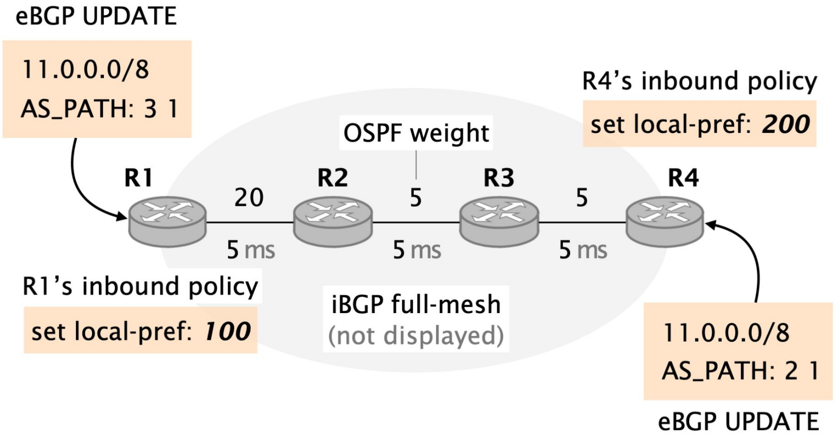
R2 and R3 are internal routers. All four routers are connected in an iBGP full-mesh. OSPF is used as intra-domain routing protocol. The link weights are indicated in the figure, e.g. the (R1, R2) link is configured with a weight of 20. The Figure also indicates the propagation delay for each link (e.g., it takes 5ms for a packet to propagate between R1 and R2).



A simple BGP network learning external routes via eBGP on $R1$ and $R4$.

Question

Considering the above configuration, indicate the next-hop used by each router in the steady state, i.e., once the network has fully converged. Use the keyword “external” to indicate that an edge router is forwarding outside of the domain. Note that we are not looking for the BGP next-hop but rather the next-hop a packet would take when being forwarded.



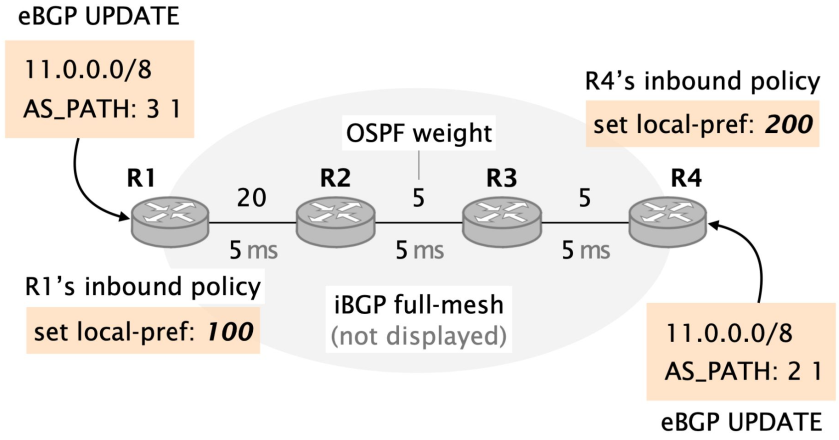
A simple BGP network learning external routes via eBGP on $R1$ and $R4$.

Question

Considering the above configuration, indicate the next-hop used by each router in the steady state, i.e., once the network has fully converged. Use the keyword “external” to indicate that an edge router is forwarding outside of the domain. Note that we are not looking for the BGP next-hop but rather the next-hop a packet would take when being forwarded.

Solution: Since the externally-learned route at R4 has a higher local-preference than the one at R1 (200 vs. 100), all routers select the route from R4. We get the following next-hops:

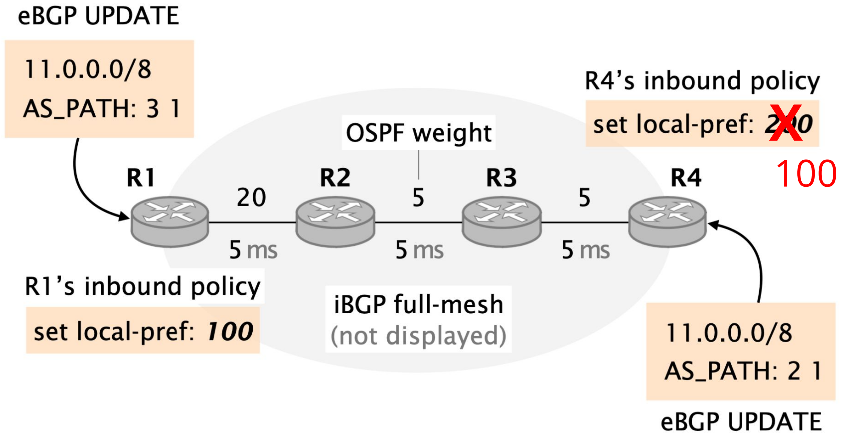
- R1: R2
- R2: R3
- R3: R4
- R4: <external>



A simple BGP network learning external routes via eBGP on $R1$ and $R4$.

Question

It turns out that the network operator changed her mind. This time, she configures R4 to associate a local-preference of 100 to externally-learned routes (i.e. the same local-preference value as on R1). Indicate the next-hop used by each router in the steady state (once the network has fully converged). Again use the keyword “external” to indicate that an egress router is forwarding outside of the domain.



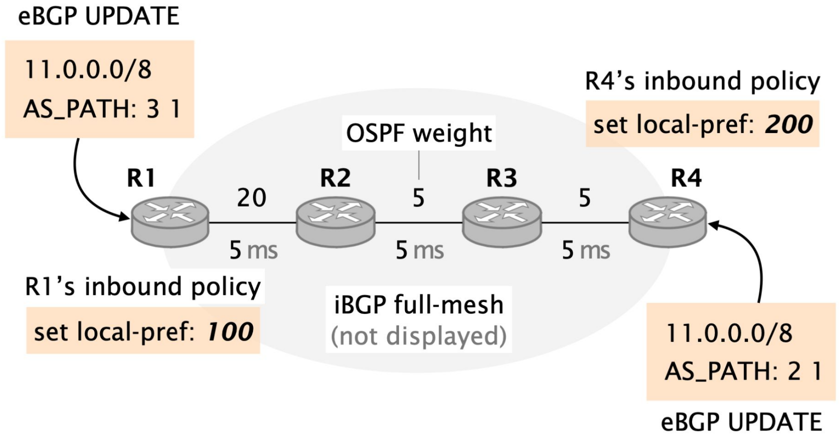
A simple BGP network learning external routes via eBGP on $R1$ and $R4$.

Question

It turns out that the network operator changed her mind. This time, she configures R4 to associate a local-preference of 100 to externally-learned routes (i.e. the same local-preference value as on R1). Indicate the next-hop used by each router in the steady state (once the network has fully converged). Again use the keyword “external” to indicate that an egress router is forwarding outside of the domain.

Solution: Since both externally-learned routes are now equally preferred, the routers consider the next criteria in the decision process. Finally, they will select the route with the lower IGP metric to the BGP next-hop. We get the following next-hops:

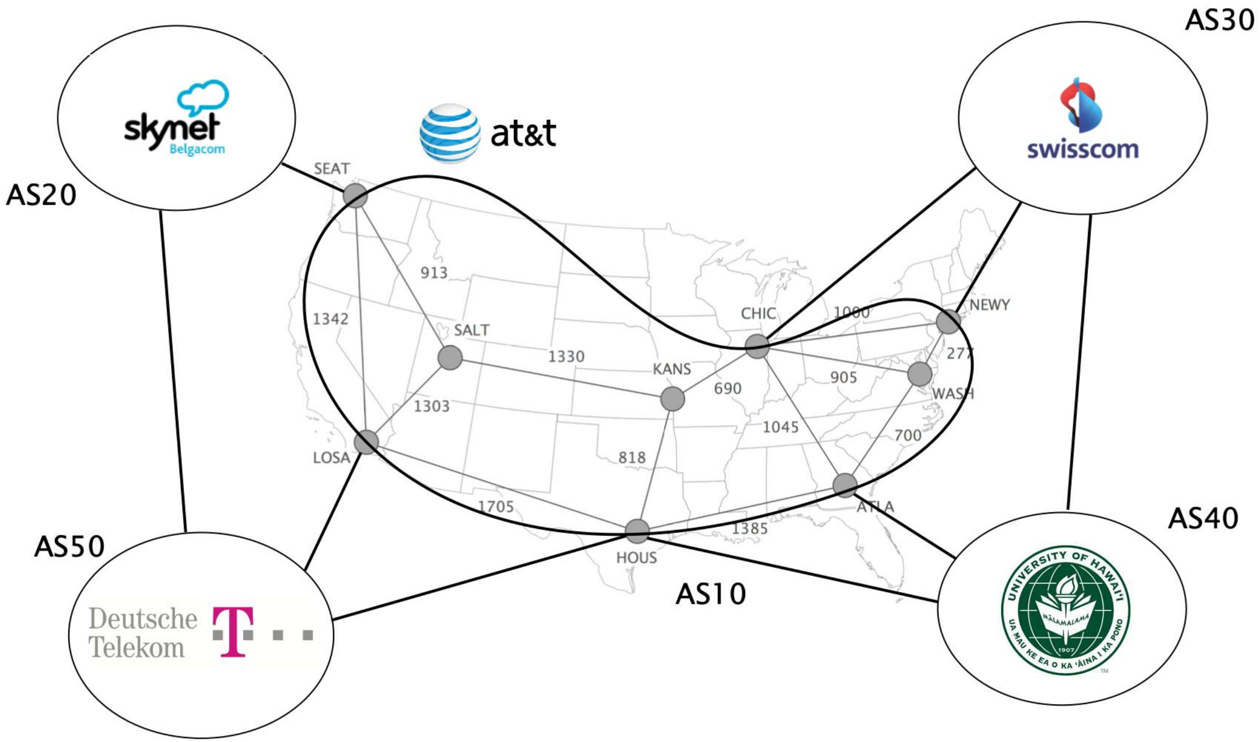
- R1: <external>
- R2: R3
- R3: R4
- R4: <external>



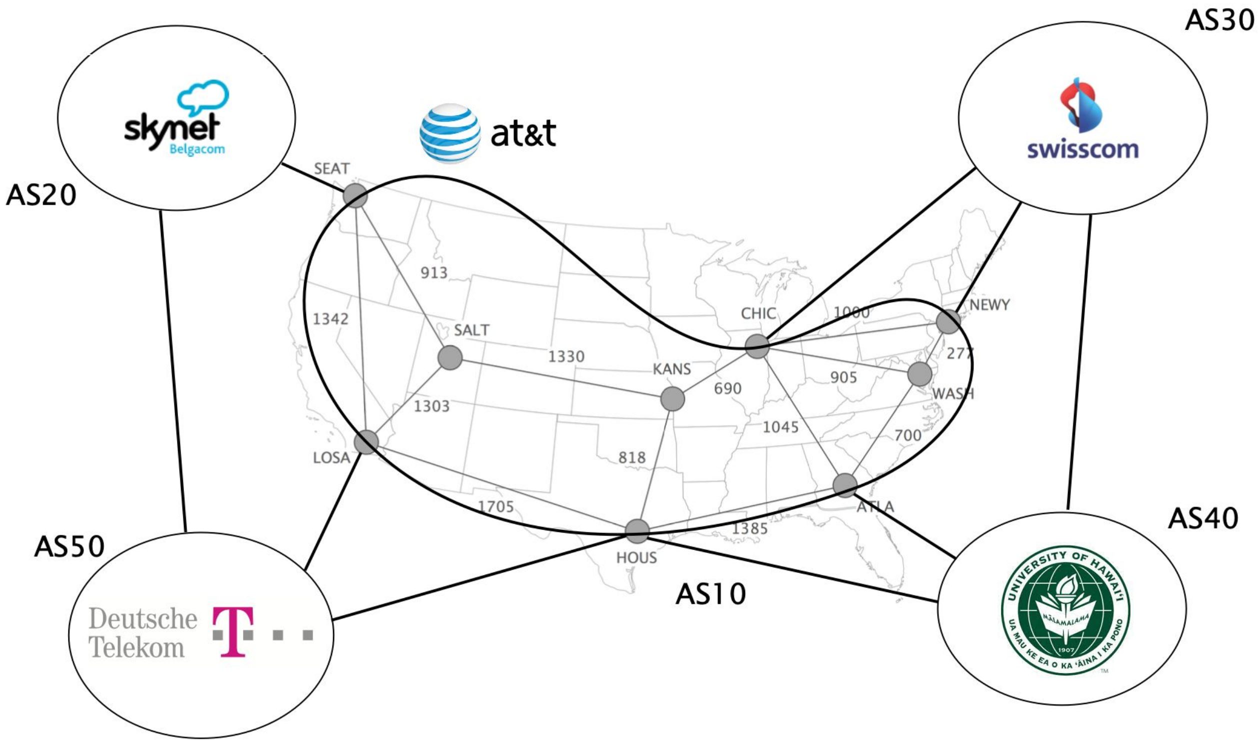
A simple BGP network learning external routes via eBGP on $R1$ and $R4$.

BGP Policies

The Internet topology is shaped according to business relationships



2 ASes only connect if they have a business relationship

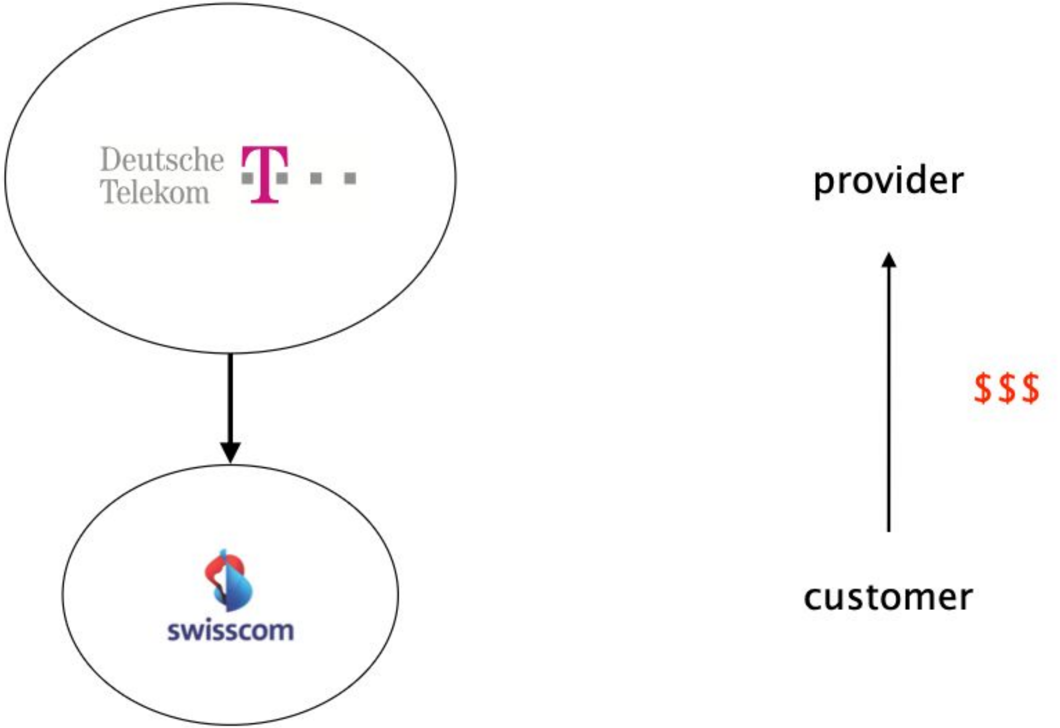


2 Main business relationships today

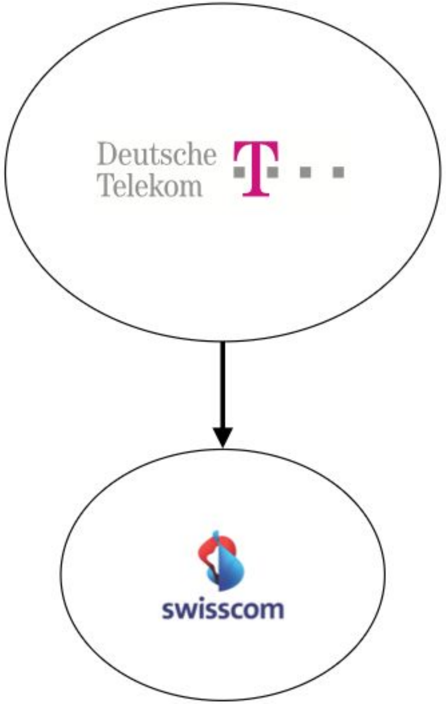
There are 2 main business relationships today:

- customer/provider
- peer/peer

Customers pay providers to get Internet connectivity



The amount paid is based on peak usage, usually according to the 95th percentile rule



Every 5 minutes, DT

records the # of bytes sent/received

At the end of the month, DT

- sorts all values in decreasing order
- removes the top 5% values
- bills wrt highest remaining value

Most ISPs discounts traffic unit price when pre-committing to certain volume

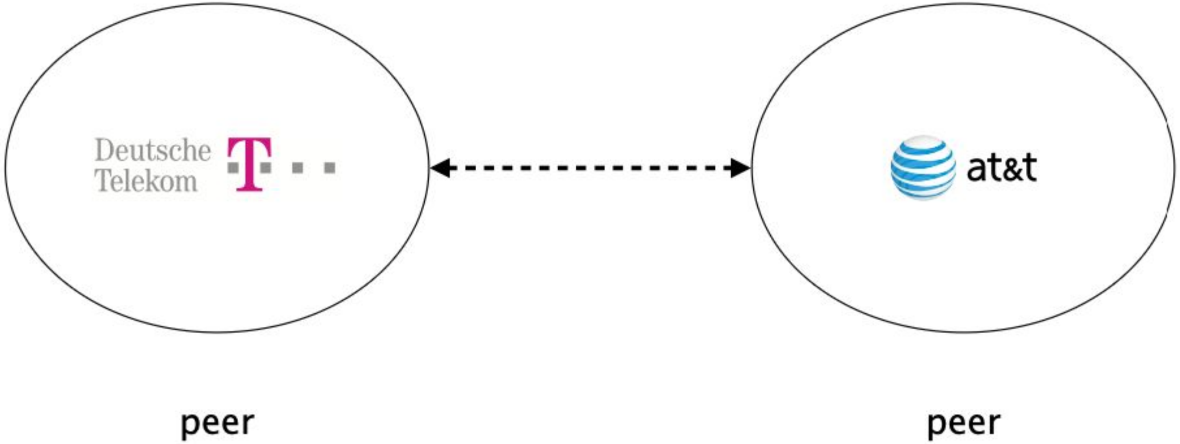
commit		unit price ($)	Minimum monthly bill ($/month)
10	Mbps	12	120
100	Mbps	5	500
1	Gbps	3.50	3,500
10	Gbps	1.20	12,000
100	Gbps	0.70	70,000

Examples taken from The 2014 Internet Peering Playbook

Internet Transit Prices have been continuously declining during the last 20 years

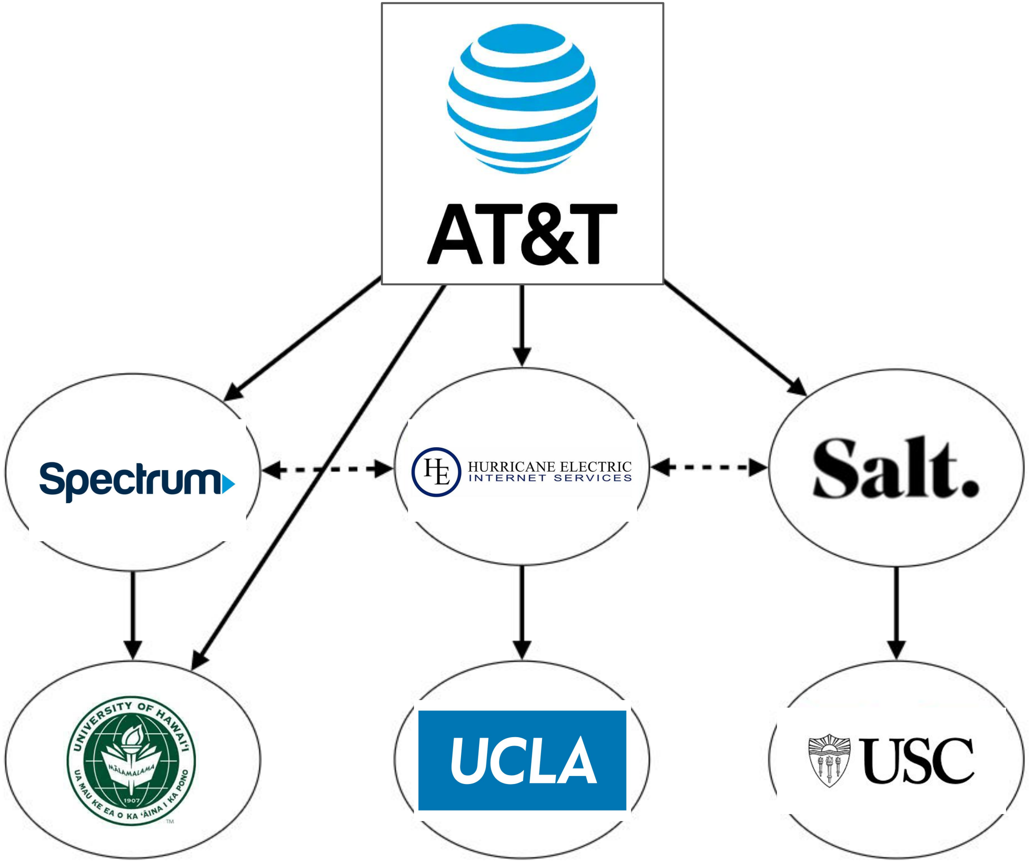
Internet Transit Pricing (1998-2015)			
Source: http://DrPeering.net			
Year	**Internet Transit Price**		**% decline**
1998	$1,200.00	per Mbps	
1999	$800.00	per Mbps	33%
2000	$675.00	per Mbps	16%
2001	$400.00	per Mbps	41%
2002	$200.00	per Mbps	50%
2003	$120.00	per Mbps	40%
2004	$90.00	per Mbps	25%
2005	$75.00	per Mbps	17%
2006	$50.00	per Mbps	33%
2007	$25.00	per Mbps	50%
2008	$12.00	per Mbps	52%
2009	$9.00	per Mbps	25%
2010	$5.00	per Mbps	44%
2011	$3.25	per Mbps	35%
2012	$2.34	per Mbps	28%
2013	$1.57	per Mbps	33%
2014	$0.94	per Mbps	40%
2015	$0.63	per Mbps	33%

Peers don’t pay each other for connectivity, they do it out of common interest

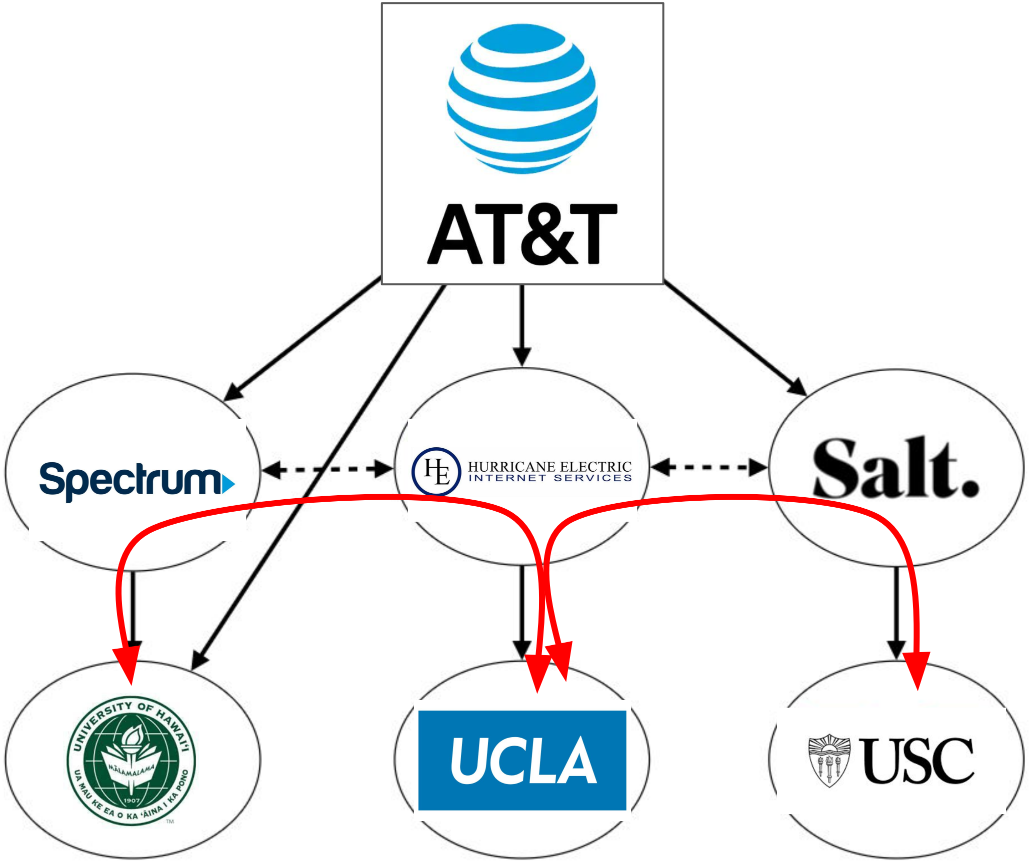


DT and ATT exchange *tons* of traffic.

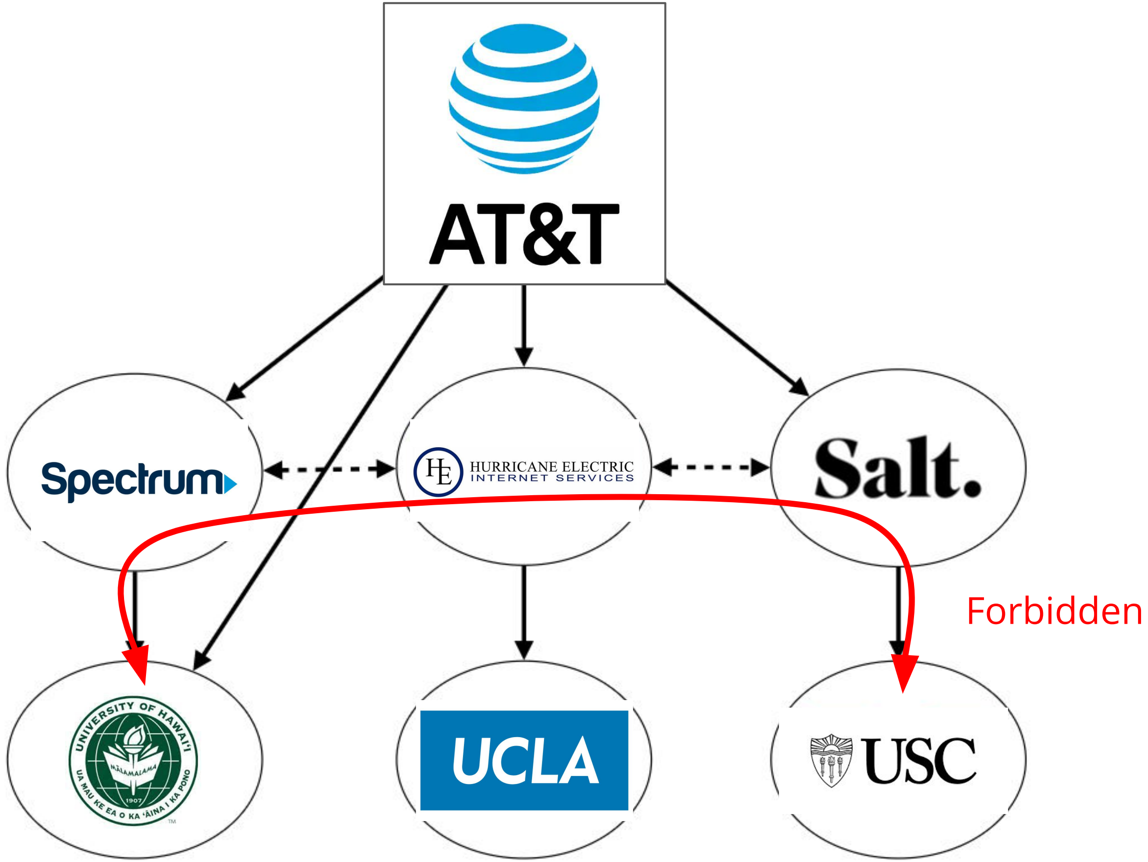
they save money by directly connecting to each other


AT&T
Spectrum
HURRICANE ELECTRIC
INTERNET SERVICES
Salt.
UNIVERSITY OF HAWAI'I
UCLA
USC

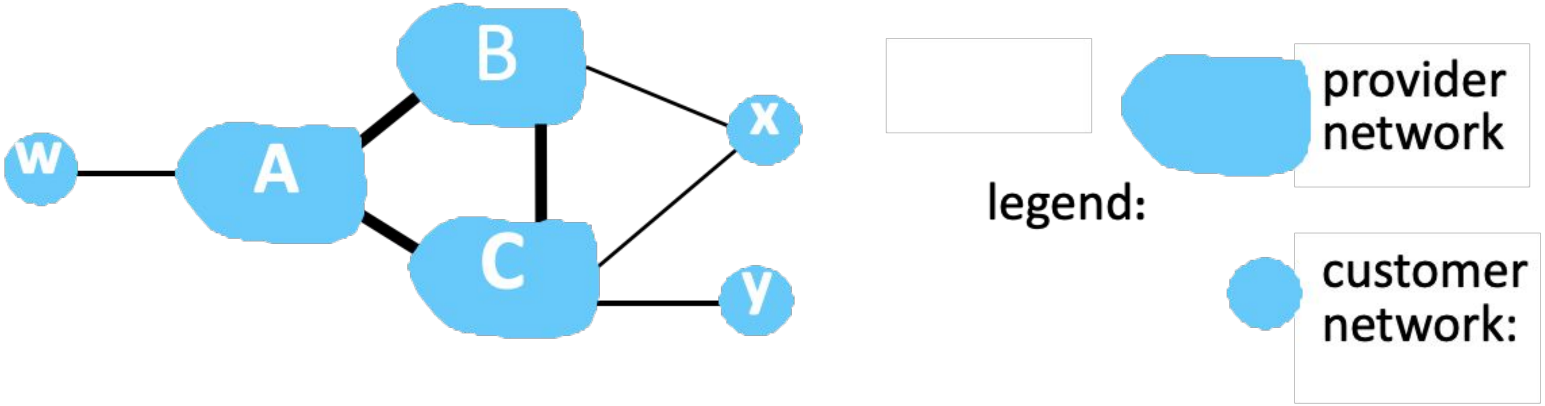
Providers transit traffic for their customers



Providers DO NOT transit traffic between each other

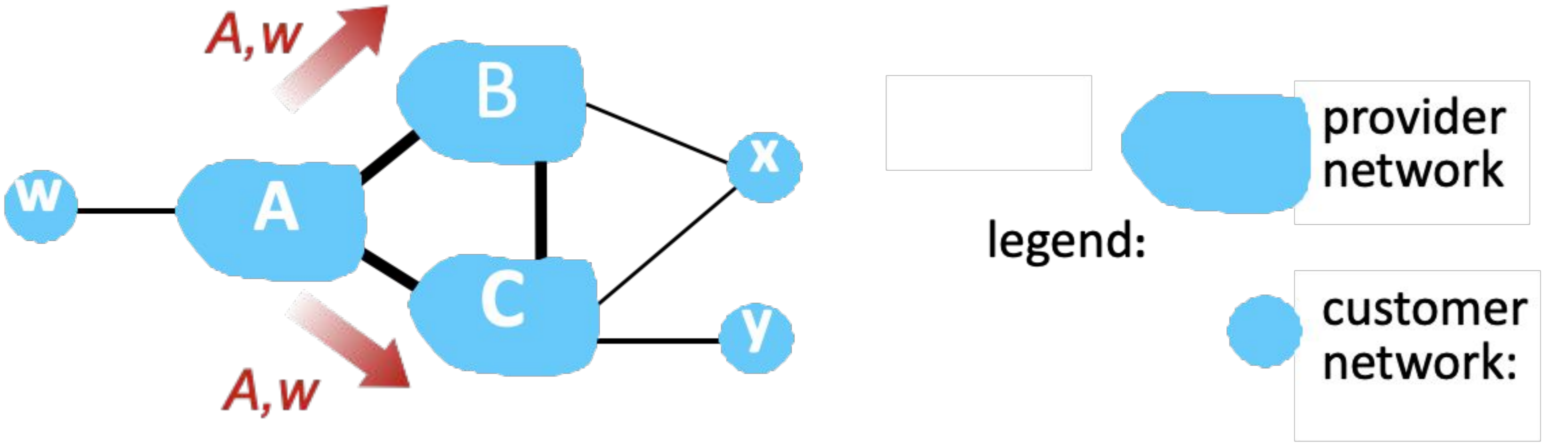


Providers DO NOT transit traffic between each other



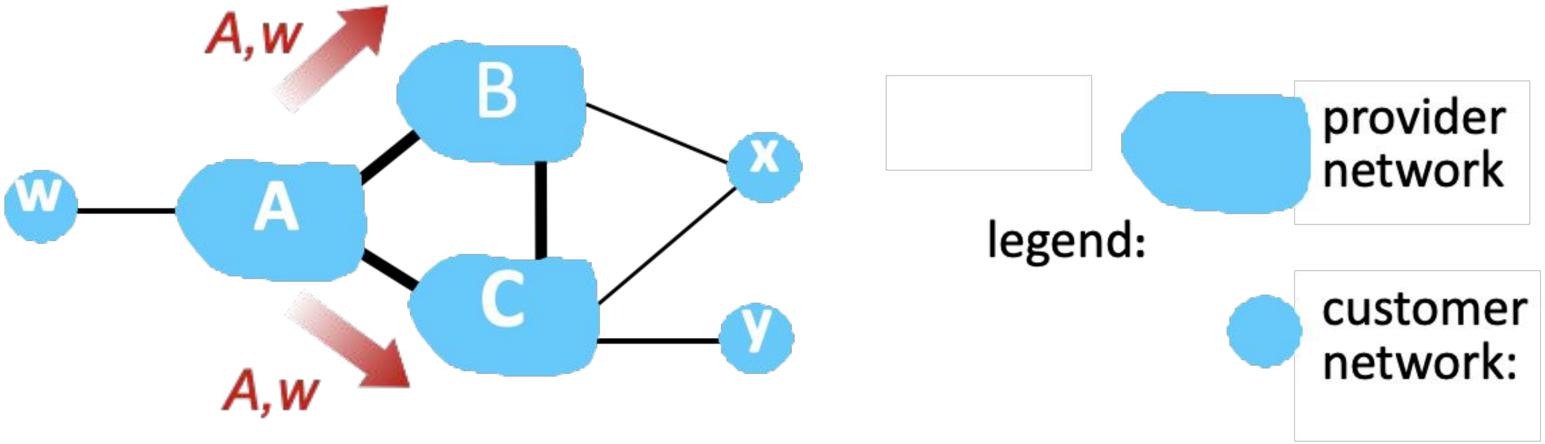
ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical “real world” policy)

Providers DO NOT transit traffic between each other



ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical “real world” policy)

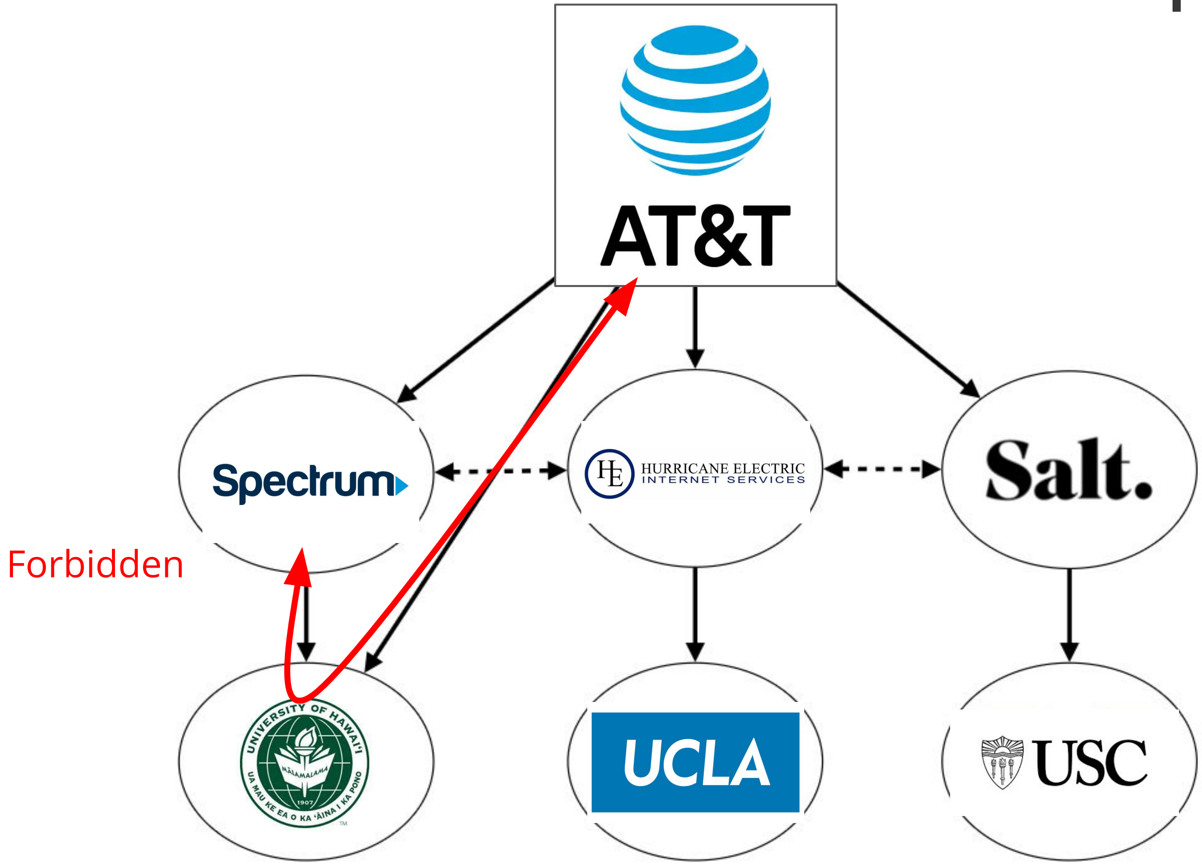
Providers DO NOT transit traffic between each other



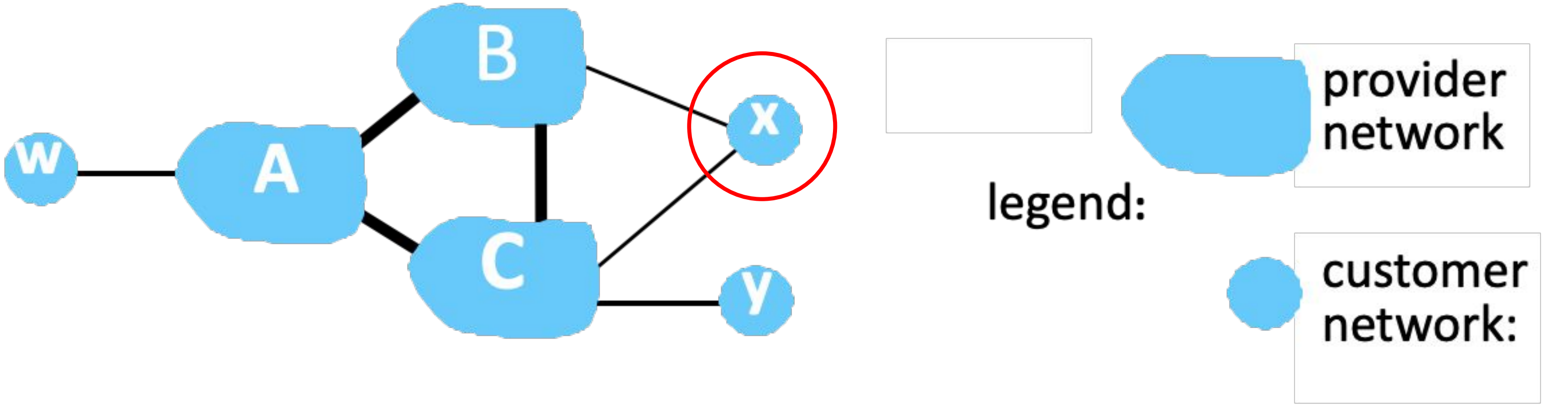
ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A advertises path Aw to B and to C
- B chooses not to advertise B-Aw to C!
  - B gets no "revenue" for routing C-B-Aw, since none of C, A, w are B's customers
  - C does not learn about C-B-Aw path
- C will route C-Aw (not using B) to get to w

Customers DO NOT transit traffic between their providers



Customers DO NOT transit traffic between their providers



ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A,B,C are provider networks
- x,w,y are customers (of provider networks)
- x is dual-homed: attached to two networks
- policy to enforce: x does not want to route from B to C via x
  - .. so x will not advertise to B a route to C