Introduction to Network Layer Control Plane: Intra- and Inter-AS Routing Protocols

November 7, 2024

Min Suk Kang

Associate Professor

School of Computing/Graduate School of Information Security

lab3 open: SDN & routing protocol

NetS&P Research Lab @ KAIST

KAIST

Making routing scalable

our routing study thus far - idealized

- all routers identical
- network “flat”

… not true in practice

scale: billions of destinations:

- can’t store all destinations in routing tables!
- routing table exchange would swamp links!

administrative autonomy:

- Internet: a network of networks
- each network admin may want to control routing in its own network

Internet approach to scalable routing

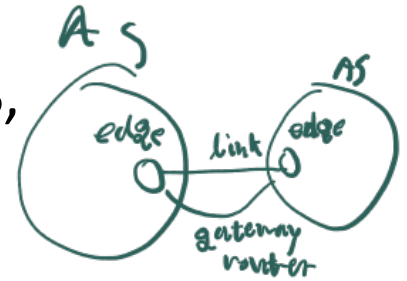
aggregate routers into regions known as "autonomous systems" (AS) (a.k.a. "domains")

intra-AS (aka "intra-domain"): routing among *within same AS ("network")*

- all routers in AS must run same intra-domain protocol
- routers in different AS can run different intra-domain routing protocols
- gateway router: at "edge" of its own AS, has link(s) to router(s) in other AS'es

inter-AS (aka "inter-domain"): routing *among* AS'es

- gateways perform inter-domain routing (as well as intra-domain routing)

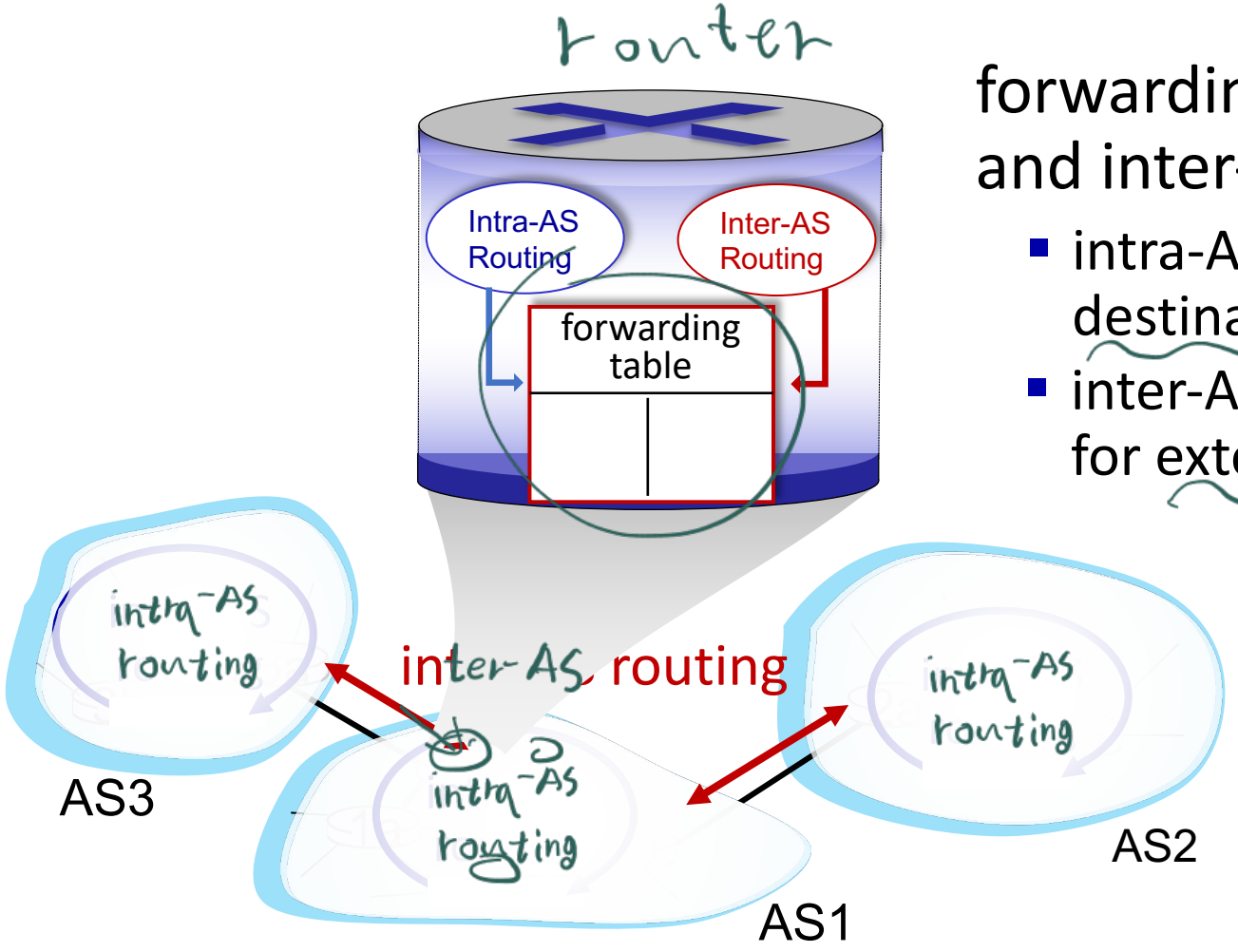


Autonomous System (AS) Numbers

- Each AS identified by an ASN number
  - 16-bit values
  - 64512 – 65535 are reserved
- Currently, there are ~60000 ASNs
  - AT&T: 5074, 6341, 7018, …
  - Sprint: 1239, 1240, 6211, 6242, …
  - AS1781 Korea Advanced Institute of Science and Technology
  - Google 15169, 36561 (formerly YT), + others
  - Facebook 32934
  - North America ASs → ftp://ftp.arin.net/info/asn.txt

AS 제공사

Interconnected ASes



forwarding table configured by intra- and inter-AS routing algorithms

- intra-AS routing determine entries for destinations within AS
- inter-AS & intra-AS determine entries for external destinations

↳ AS 내부/외부 routing algorithm 역할

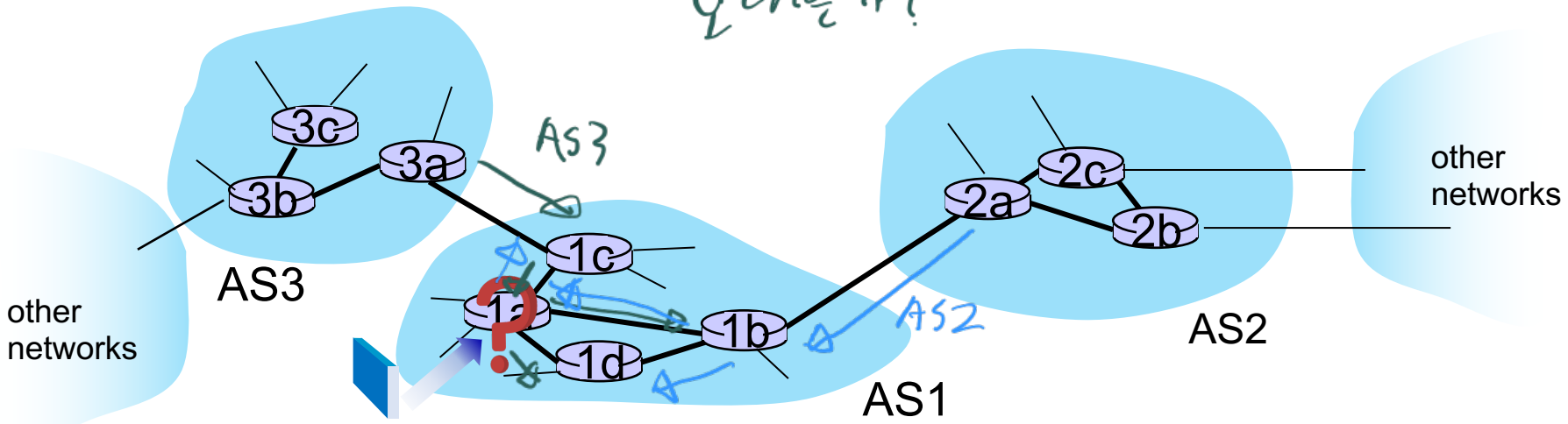
Inter-AS routing: a role in intradomain forwarding

- suppose router in AS1 receives datagram destined outside of AS1:
  - router should forward packet to gateway router in AS1, but which one?

AS1 inter-domain routing must:

1. learn which destinations reachable through AS2, which through AS3
2. propagate this reachability info to all routers in AS1

AS2 및 AS3를 통해 도달가능한 목적지 확인 & 목적지 정보를 AS1의 모든 router에 전파



Inter-AS routing: routing within an AS

most common intra-AS routing protocols:

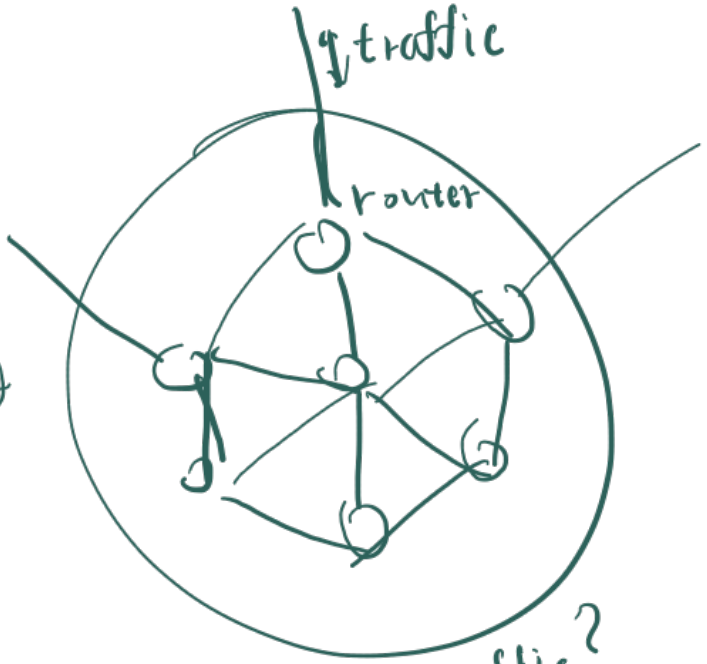
- RIP: Routing Information Protocol [RFC 1723]
  - classic DV: DVs exchanged every 30 secs
  - no longer widely used
- EIGRP: Enhanced Interior Gateway Routing Protocol
  - DV based
  - formerly Cisco-proprietary for decades (became open in 2013 [RFC 7868])
- OSPF: Open Shortest Path First [RFC 2328]
  - link-state routing
  - IS-IS protocol (ISO standard, not RFC standard) essentially same as OSPF

OSPF (Open Shortest Path First) routing

- "open": publicly available
- classic link-state
  - each router floods OSPF link-state advertisements (directly over IP rather than using TCP/UDP) to all other routers in entire AS
  - multiple link costs metrics possible: bandwidth, delay
  - each router has full topology, uses Dijkstra's algorithm to compute forwarding table
- *security:* all OSPF messages authenticated (to prevent malicious intrusion)

Setting Link Costs for OSPF

- Admin has powerful knobs for his/her network



control traffic?
⇒ Set link cost
(traffic에 비례하여 cost 지정)

- But is link cost an effective control knob?

너의 비슷한 cost의 두 경로가 있는 경우,
minor traffic에도 너무 sensitive하게
반응할 수 있다. (oscillation)
일반적으로는 유리X

Network layer: "control plane" roadmap

- introduction
- routing protocols
- intra-ISP routing: OSPF
- routing among ISPs: BGP
- SDN control plane
- Internet Control Message Protocol

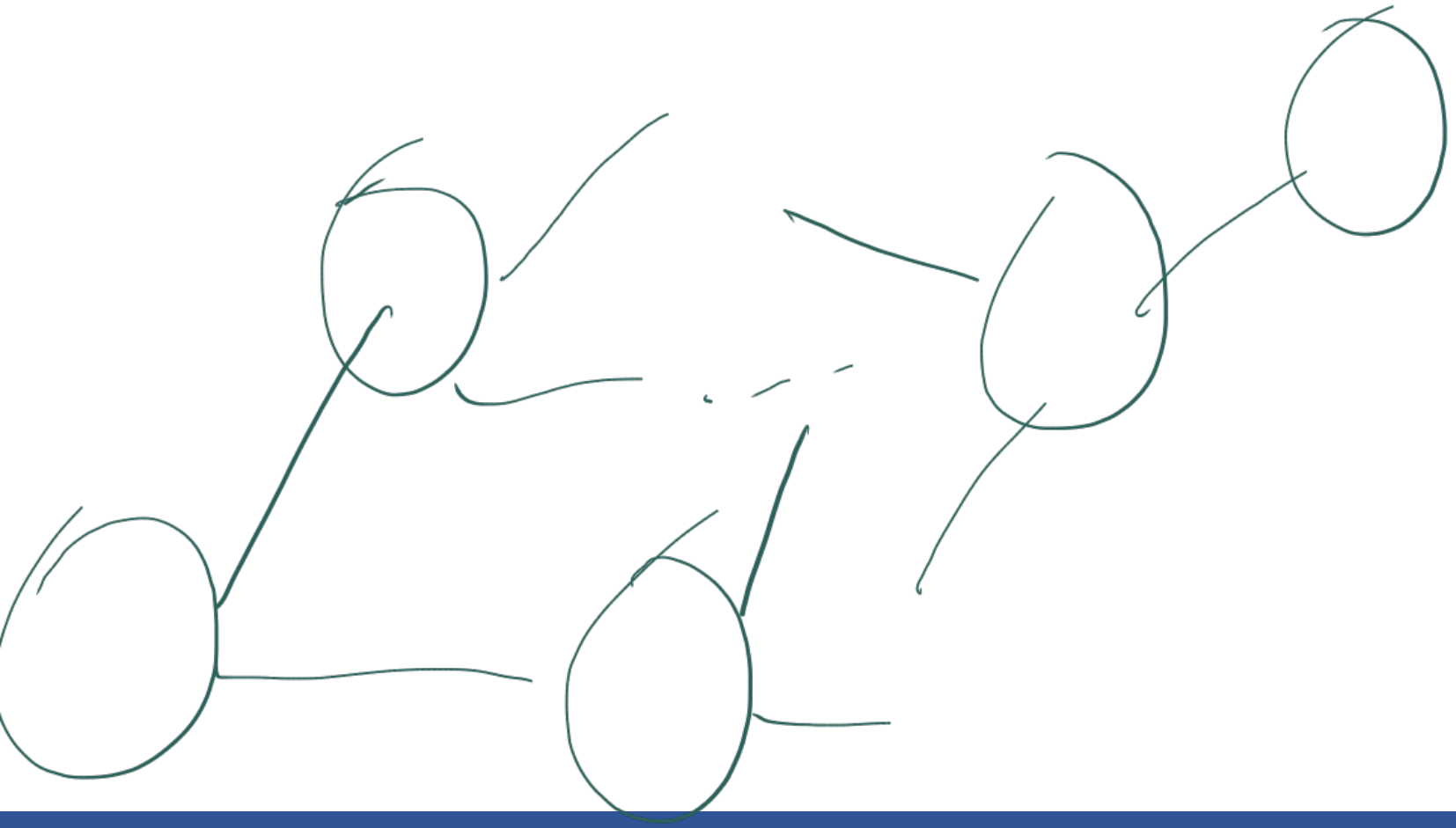
inter-AS protocol

- network management, configuration
  - SNMP
  - NETCONF/YANG

Can’t we run LS or DV for inter-AS routing?

link state

distance vector

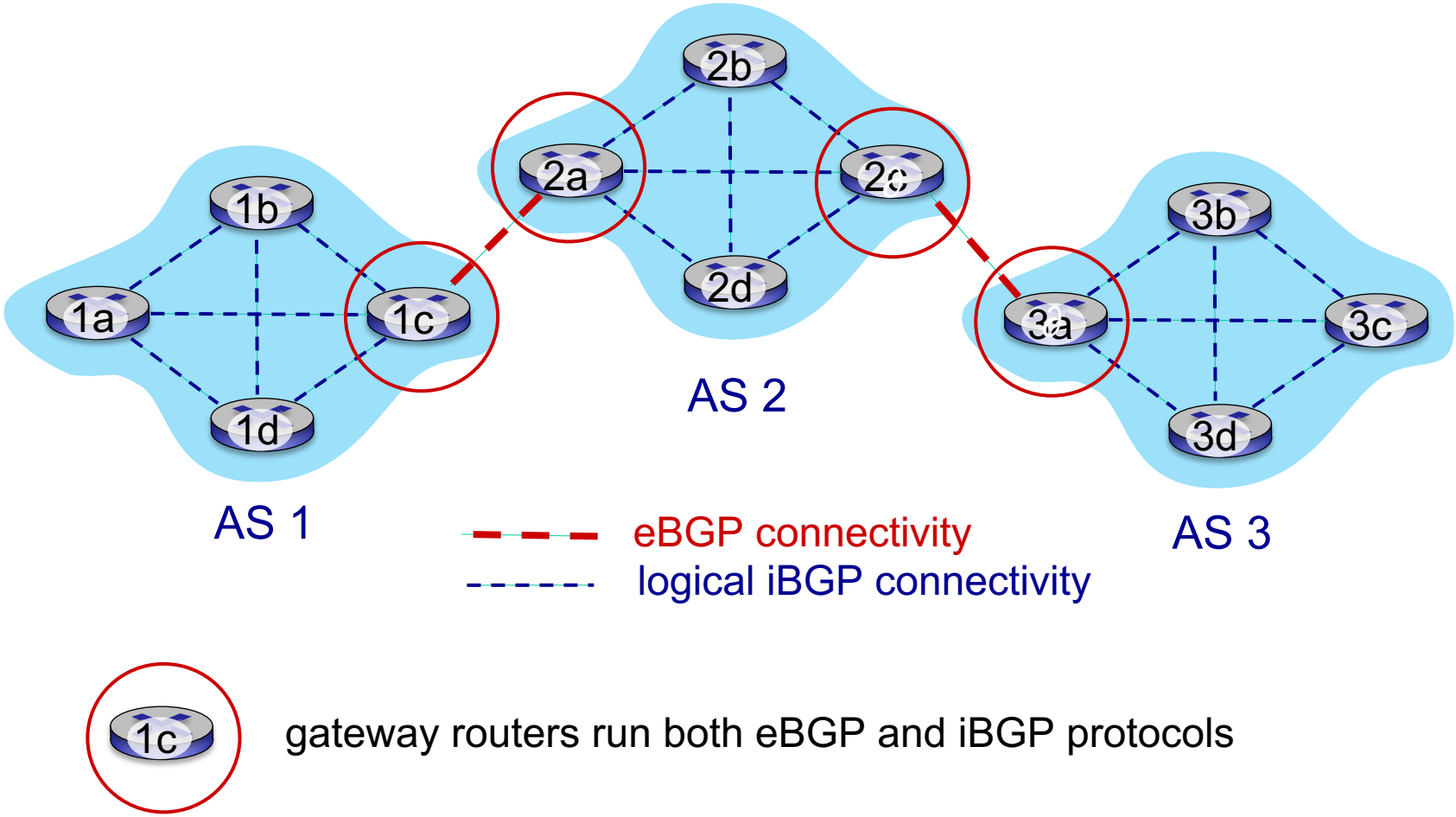


- LS: 고성능 라우터를 internet의 전체적 topology 파악 불가 사용X
- DV: not expressive enough (Dn(x) 정보만 저장하므로 얻을 수 있는 정보↓)

Internet inter-AS routing: BGP

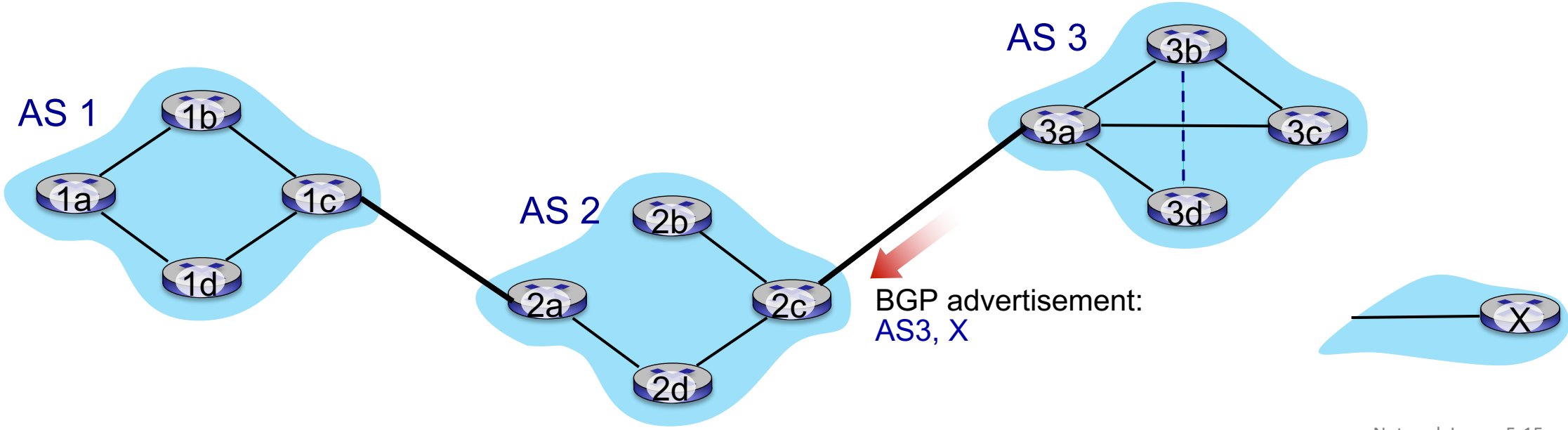
- BGP (Border Gateway Protocol): *the* de facto inter-domain routing protocol
  - "glue that holds the Internet together"
- allows subnet to advertise its existence, and the destinations it can reach, to rest of Internet: *"I am here, here is who I can reach, and how"*
- BGP provides each AS a means to:
  - eBGP: obtain subnet reachability information from neighboring ASes
  - iBGP: propagate reachability information to all AS-internal routers.
  - determine "good" routes to other networks based on reachability information and *policy*

eBGP, iBGP connections



BGP basics

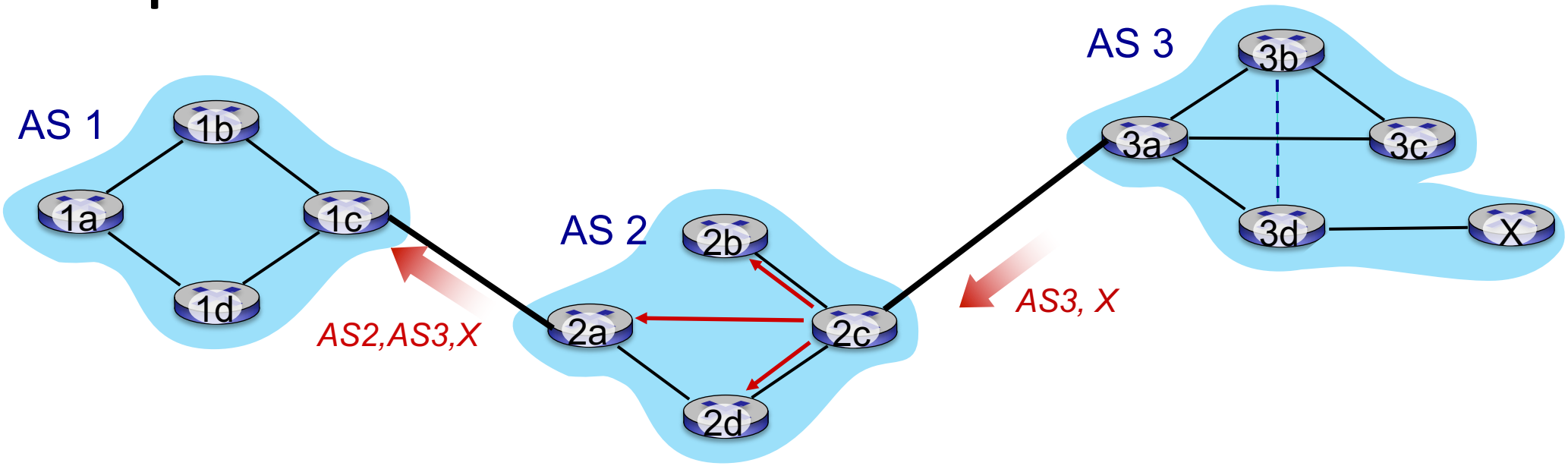
- BGP session: two BGP routers ("peers") exchange BGP messages over semi-permanent TCP connection:
  - advertising *paths* to different destination network prefixes (BGP is a "path vector" protocol)
- when AS3 gateway 3a advertises path AS3,X to AS2 gateway 2c:
  - AS3 *promises* to AS2 it will forward datagrams towards X



Path attributes and BGP routes

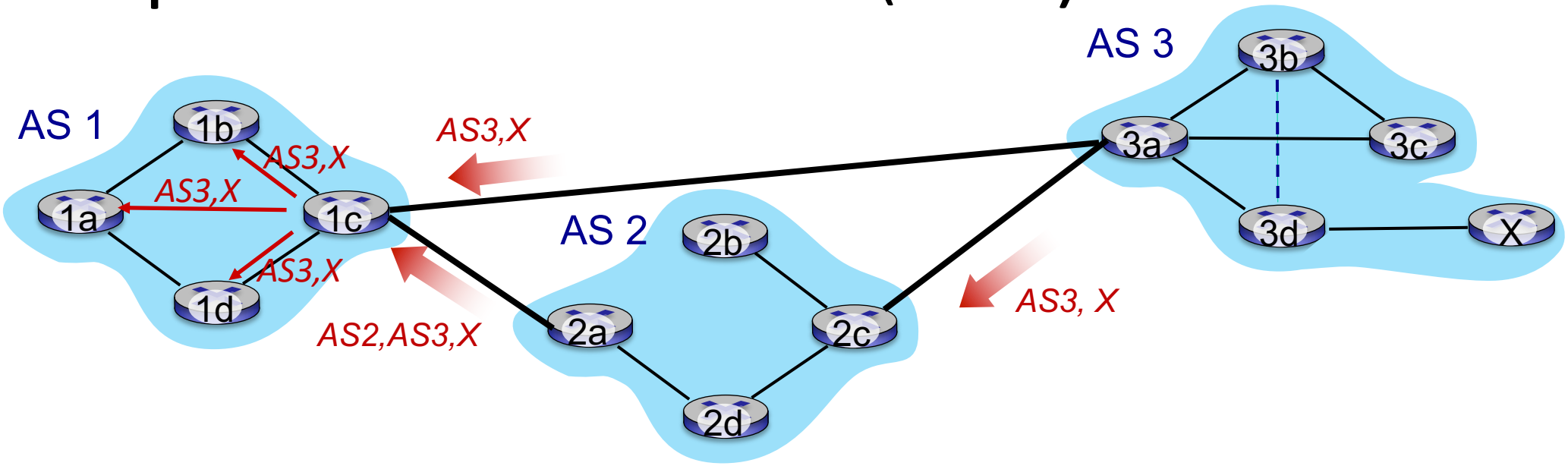
- BGP advertised route: prefix + attributes
  - prefix: destination being advertised
  - two important attributes:
    - AS-PATH: list of ASes through which prefix advertisement has passed
    - NEXT-HOP: indicates specific internal-AS router to next-hop AS
- policy-based routing:
  - gateway receiving route advertisement uses *import policy* to accept/decline path (e.g., never route through AS Y).
  - AS policy also determines whether to *advertise* path to other other neighboring ASes

BGP path advertisement



- AS2 router 2c receives path advertisement AS3,X (via eBGP) from AS3 router 3a
- based on AS2 policy, AS2 router 2c accepts path AS3,X, propagates (via iBGP) to all AS2 routers
- based on AS2 policy, AS2 router 2a advertises (via eBGP) path AS2, AS3, X to AS1 router 1c

BGP path advertisement (more)



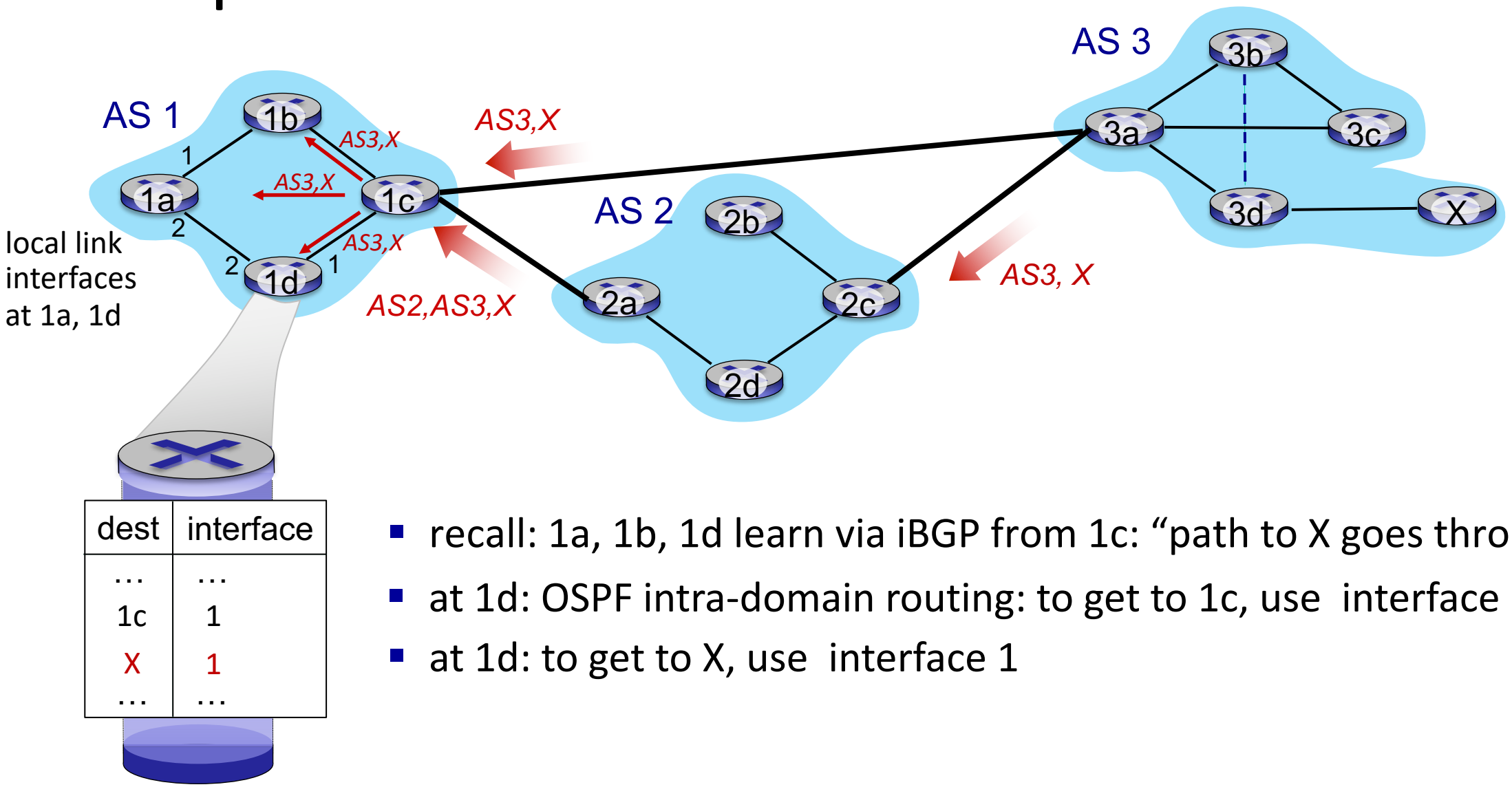
gateway router may learn about multiple paths to destination:

- AS1 gateway router 1c learns path *AS2,AS3,X* from 2a
- AS1 gateway router 1c learns path *AS3,X* from 3a
- based on *policy,* AS1 gateway router 1c chooses path *AS3,X* and advertises path within AS1 via iBGP

BGP messages

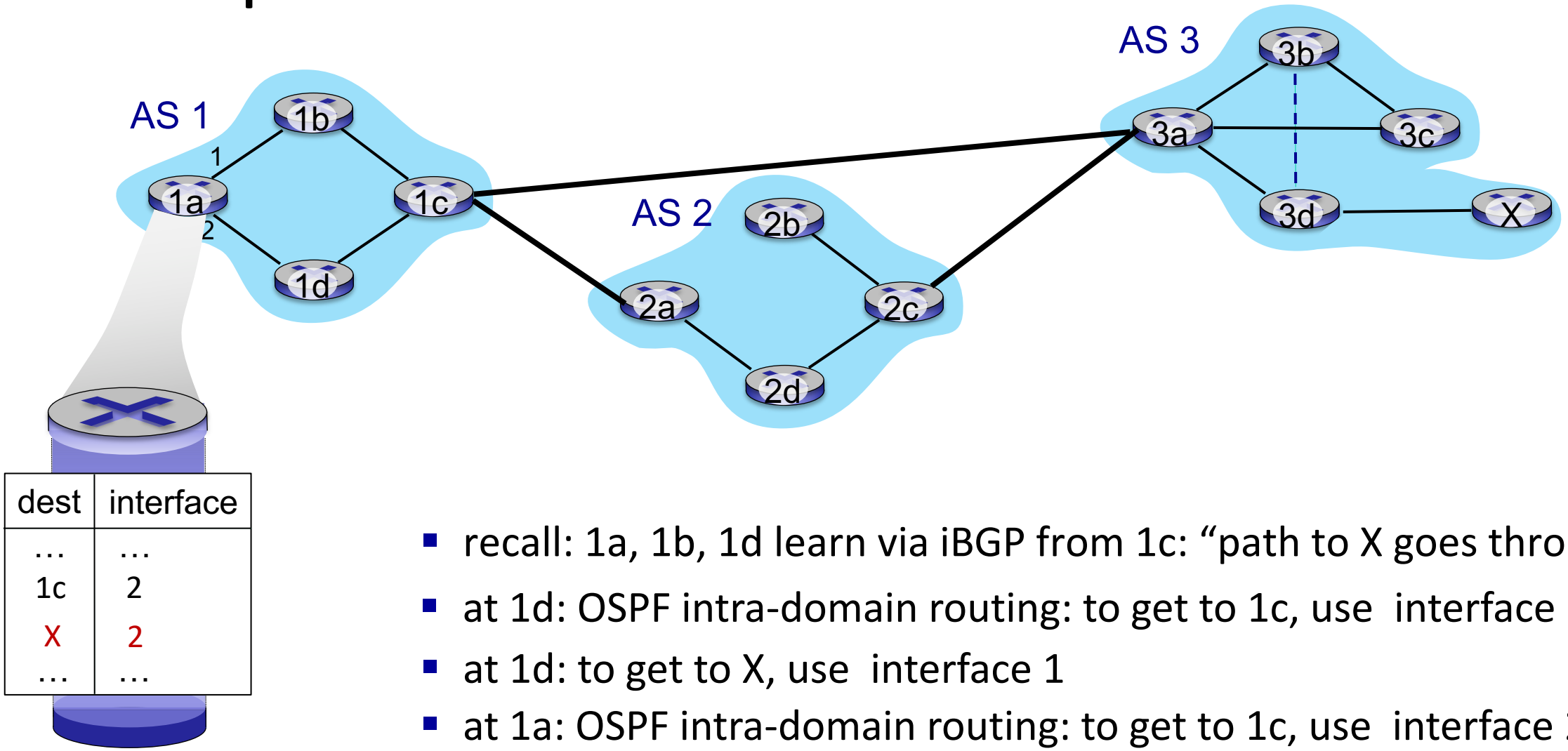
- BGP messages exchanged between peers over TCP connection
- BGP messages:
  - OPEN: opens TCP connection to remote BGP peer and authenticates sending BGP peer
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE: keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: reports errors in previous msg; also used to close connection

BGP path advertisement



- recall: 1a, 1b, 1d learn via iBGP from 1c: “path to X goes through 1c”
- at 1d: OSPF intra-domain routing: to get to 1c, use interface 1
- at 1d: to get to X, use interface 1

BGP path advertisement



- recall: 1a, 1b, 1d learn via iBGP from 1c: “path to X goes through 1c”
- at 1d: OSPF intra-domain routing: to get to 1c, use interface 1
- at 1d: to get to X, use interface 1
- at 1a: OSPF intra-domain routing: to get to 1c, use interface 2
- at 1a: to get to X, use interface 2

Why different Intra-, Inter-AS routing ?

policy:

- inter-AS: admin wants control over how its traffic routed, who routes through its network
- intra-AS: single admin, so policy less of an issue

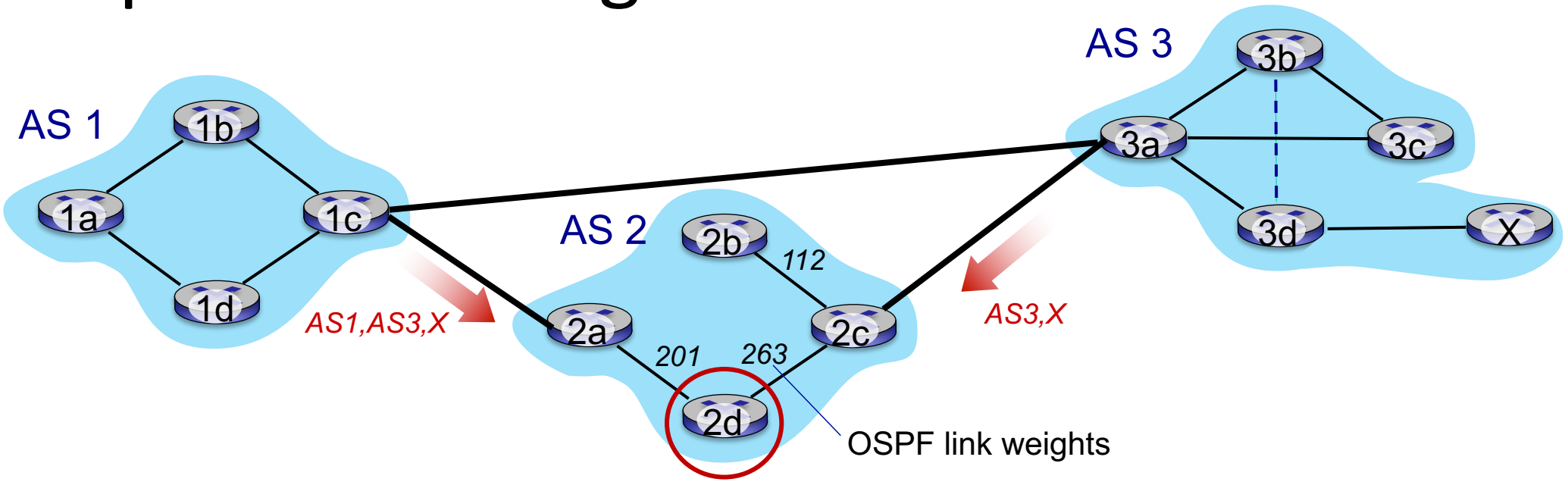
scale:

- hierarchical routing saves table size, reduced update traffic

performance:

- intra-AS: can focus on performance
- inter-AS: policy dominates over performance

Hot potato routing

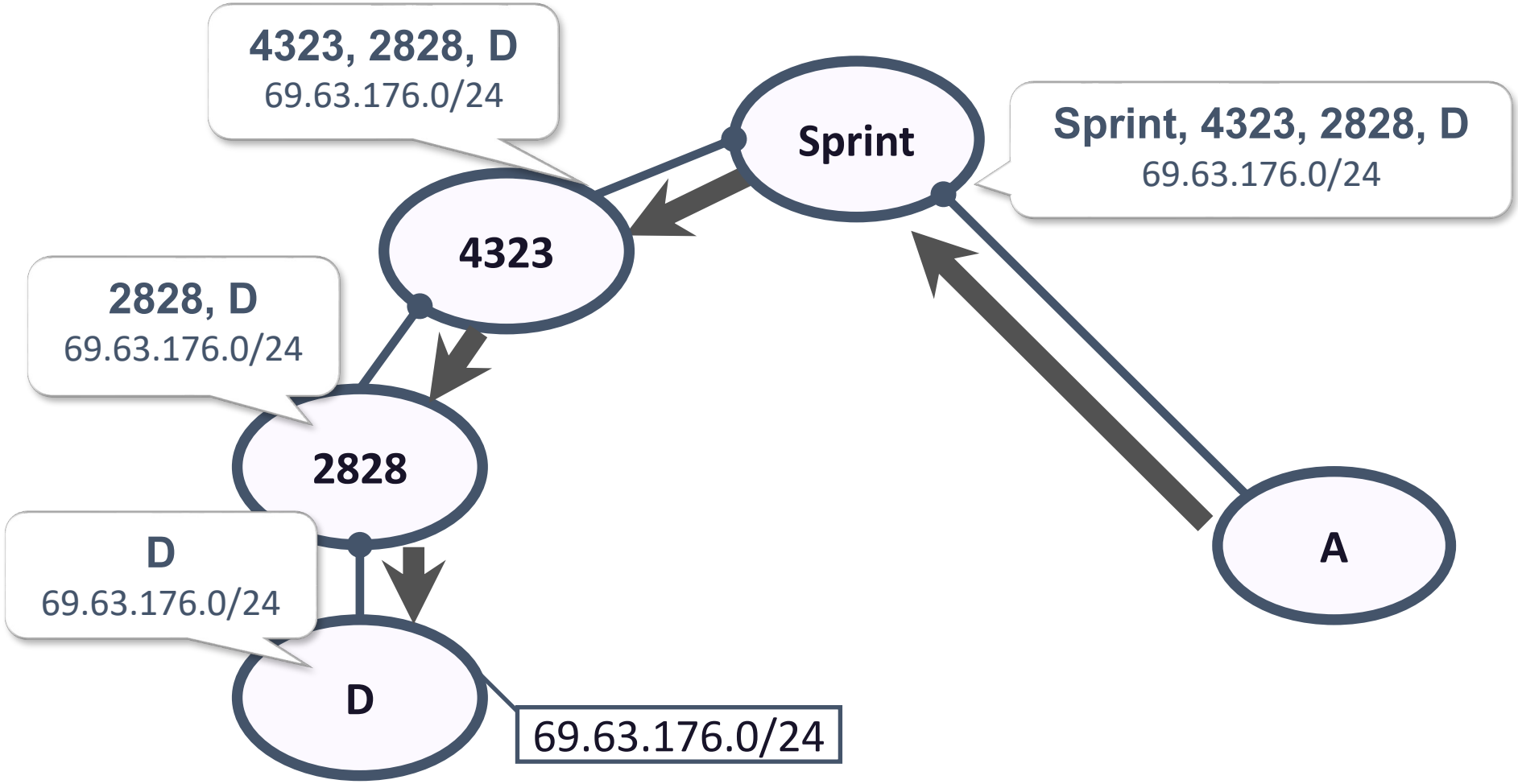


- 2d learns (via iBGP) it can route to X via 2a or 2c
- hot potato routing: choose local gateway that has least *intra-domain* cost (e.g., 2d chooses 2a, even though more AS hops to *X*): don't worry about inter-domain cost!

Border Gateway Protocol

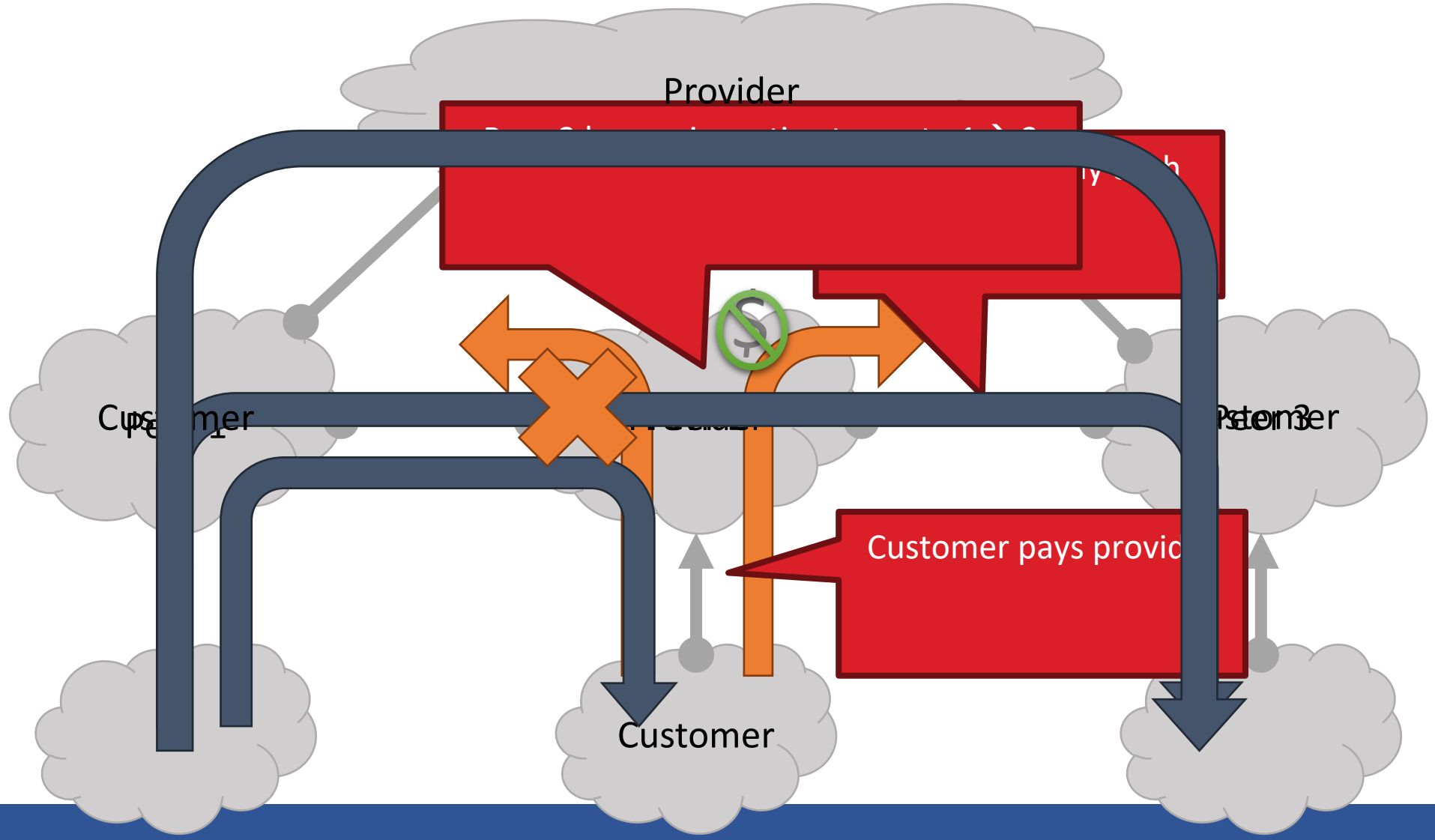
- Border Gateway Protocol (BGP)
  - De facto inter-domain protocol of the Internet
  - Uses a **path vector** routing
    - =/= link state routing
      - flood the routing table to entire network
    - =/= distance vector routing
      - advertise the cost but not actual paths
  - Policy based routing protocol
- Relatively simple protocol, but…
  - Complex, manual configuration
  - Entire world sees advertisements
    - Errors can screw up traffic globally
  - Policies driven by economics
    - Not by performance (e.g. shortest paths)

BGP

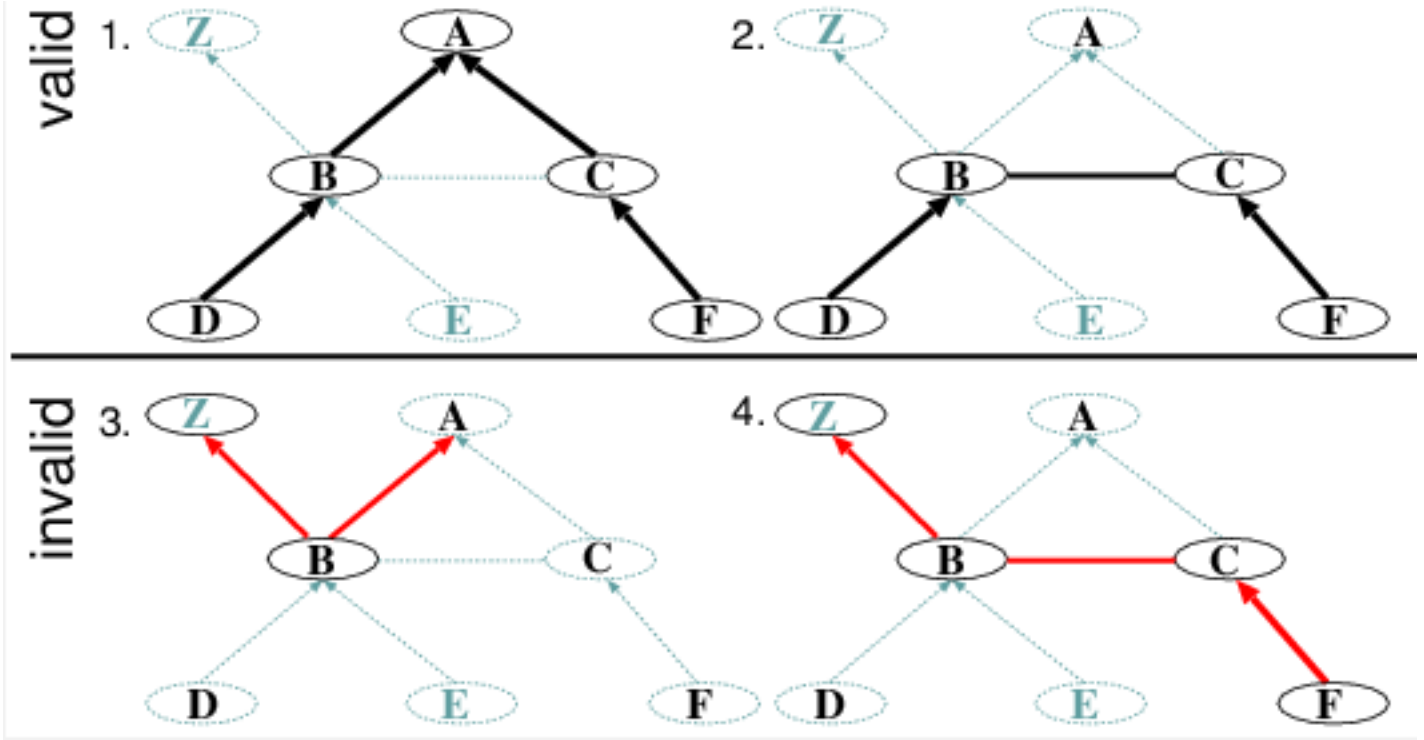

4323, 2828, D
69.63.176.0/24
Sprint
Sprint, 4323, 2828, D
69.63.176.0/24
4323
2828, D
69.63.176.0/24
2828
D
69.63.176.0/24
A
D
69.63.176.0/24

25

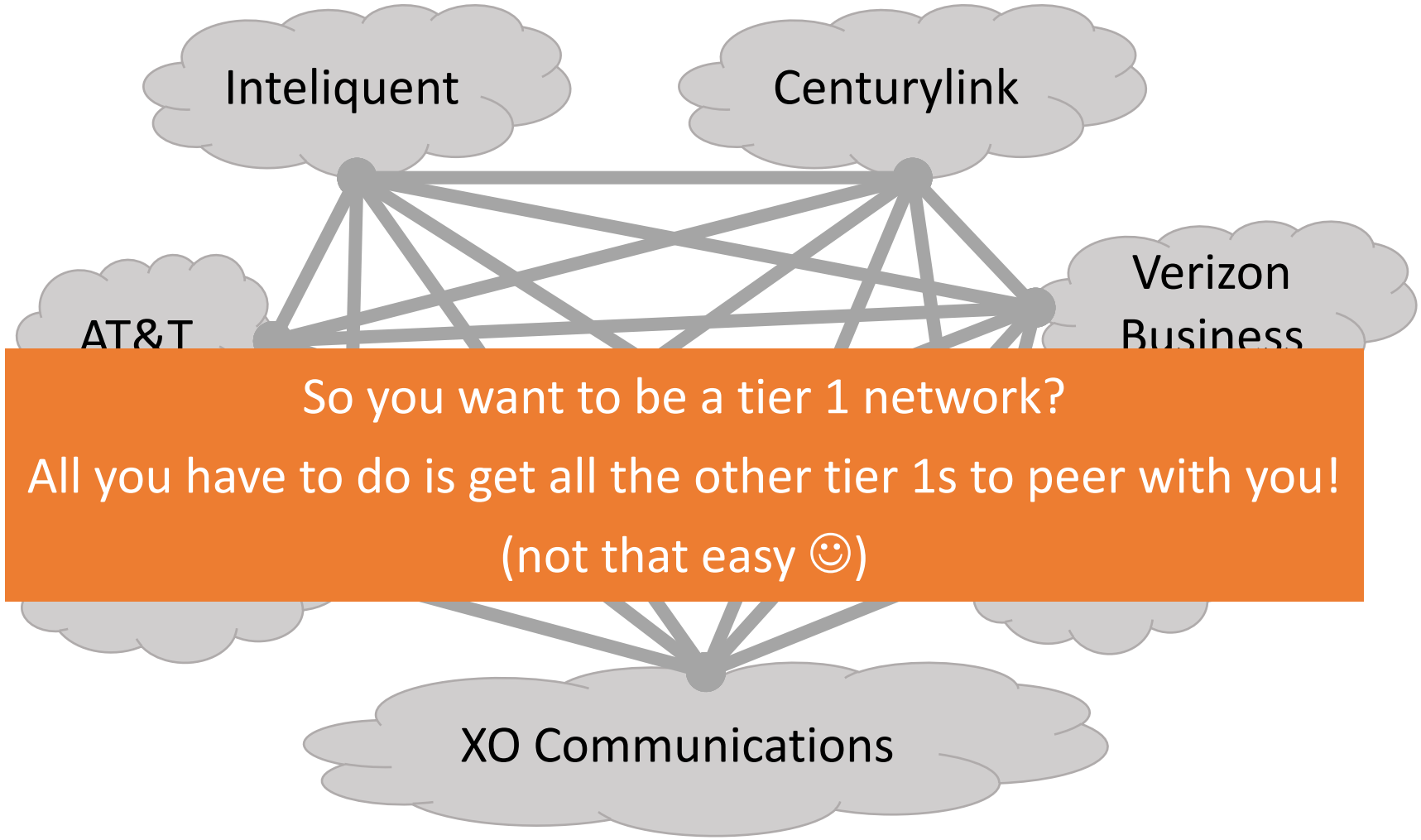
BGP's Business Relationships



BGP paths: valids and invalids



Tier-1 ISP Peering

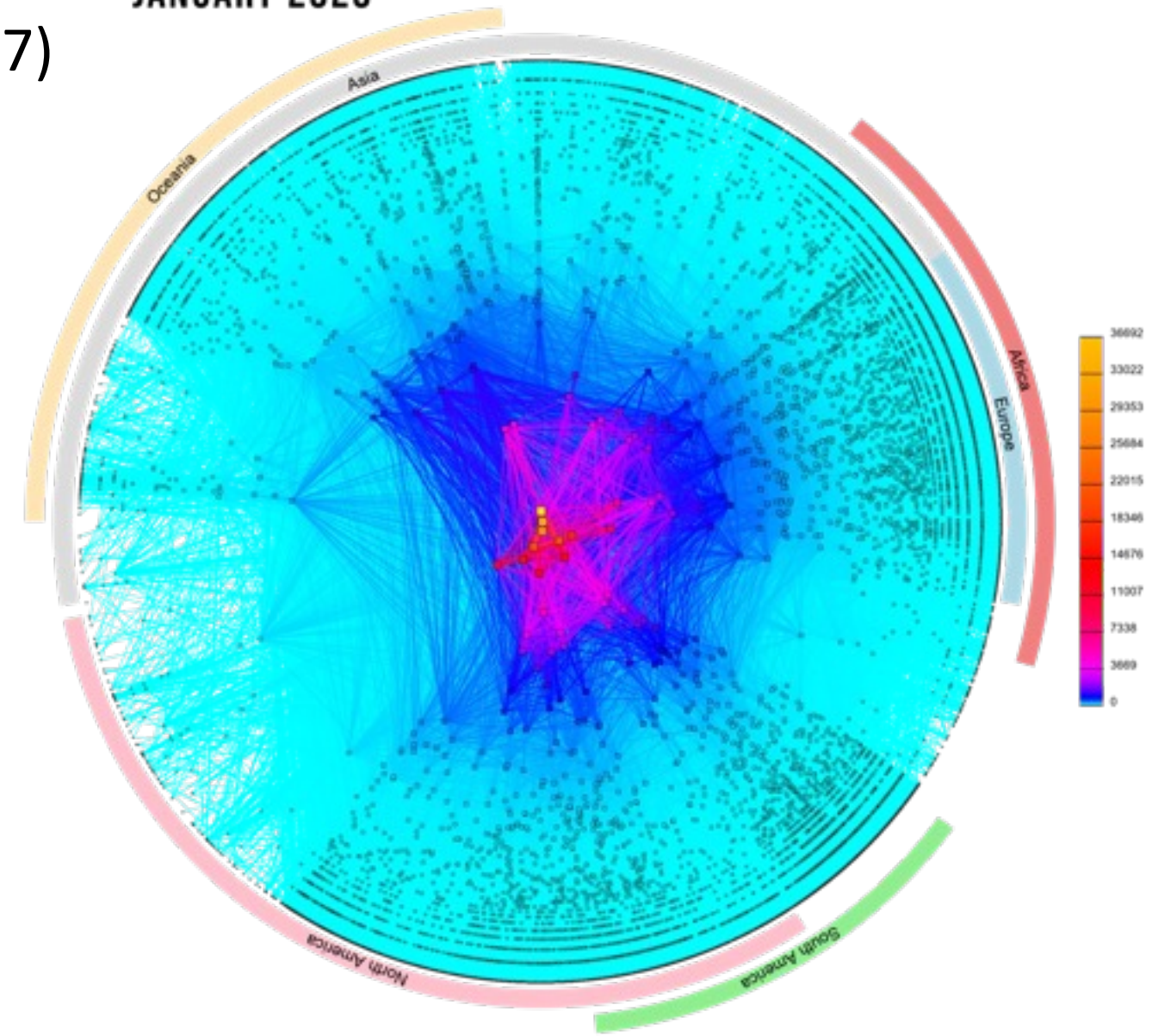


So you want to be a tier 1 network?

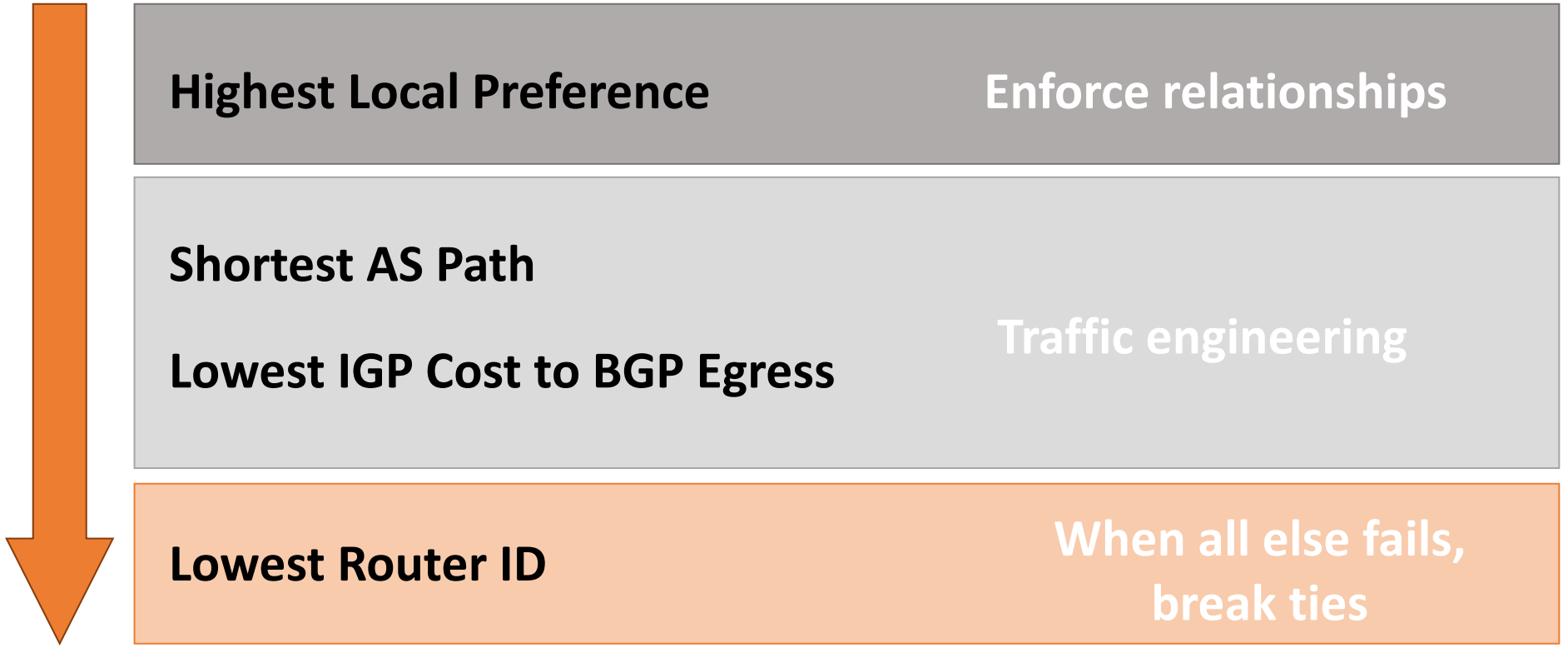
All you have to do is get all the other tier 1s to peer with you!

(not that easy ☺)

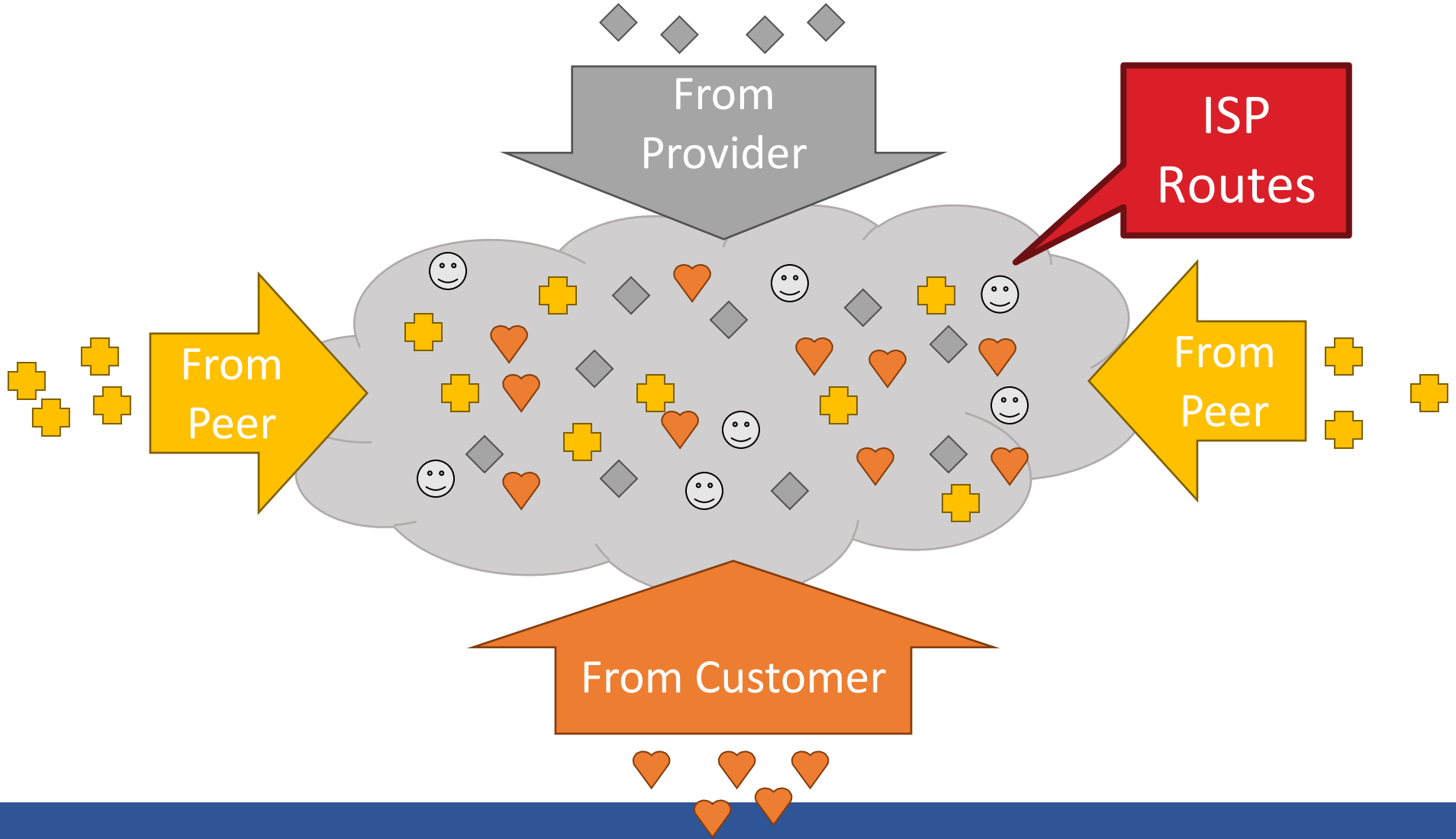
AS level topology by CAIDA (2017)



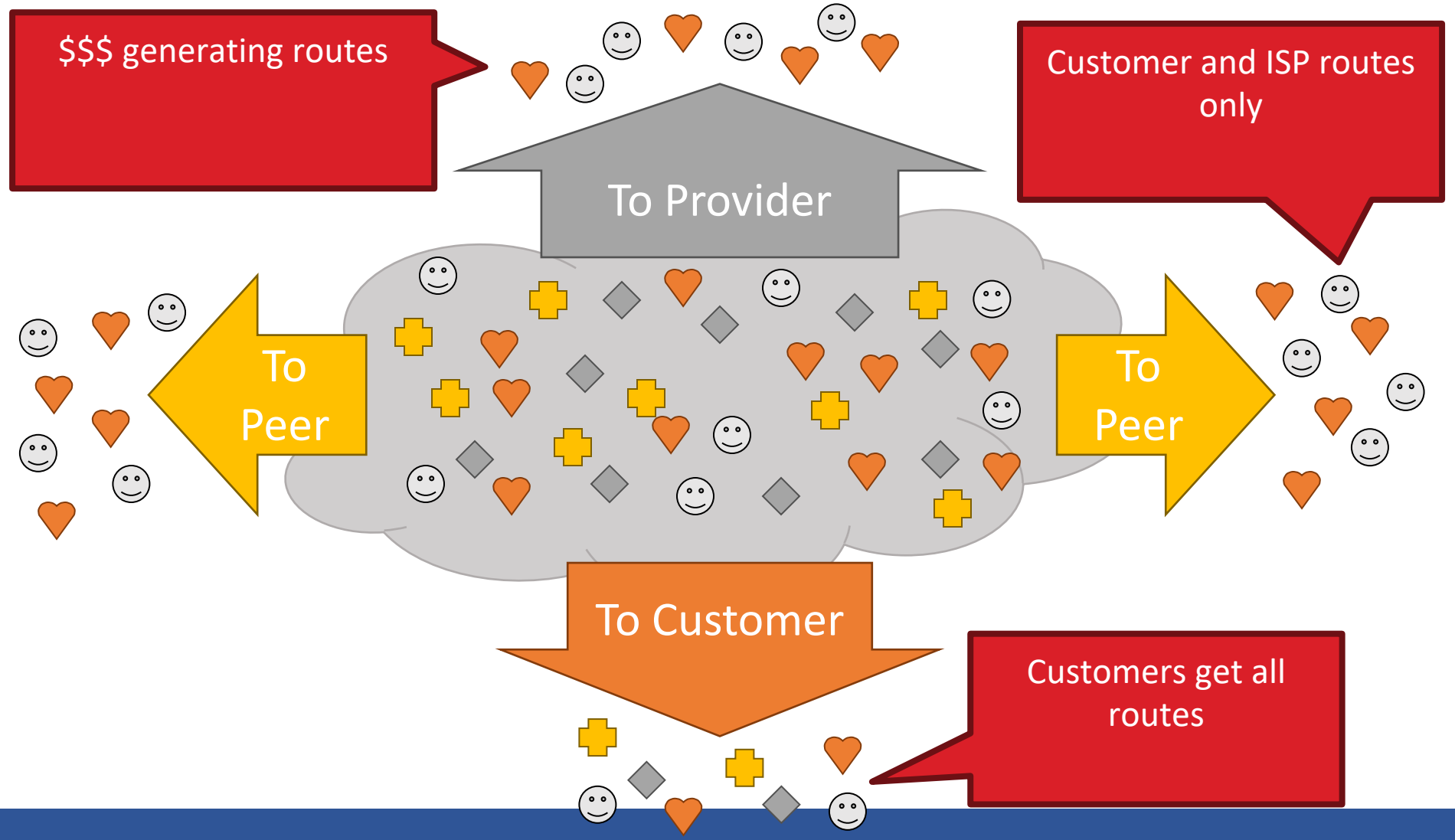
Route Selection Summary



Importing Routes



Exporting Routes



BGP: Robustness?

- Robustness: Is BGP robust enough to prevent/avoid mistakes?

Business All Industry Technology Tr

Ministry finds KT work
incorrect command as
outage

Ministry to review introducing measures fo

By Kim Da-sol

Published : Oct 29, 2021 - 17:22 Updated : Oct 29, 2021 - 17:23

An incorrect command input -- in fact, forge
routing setting command to the router dev
outage in less than a minute.

Authorities found that the network path inf
sent to border gateway protocol (BGP), was
which is used for internal path setting at KT

The ministry's investigation team also foun
features in place to prevent the spread of w

SOC PRIME

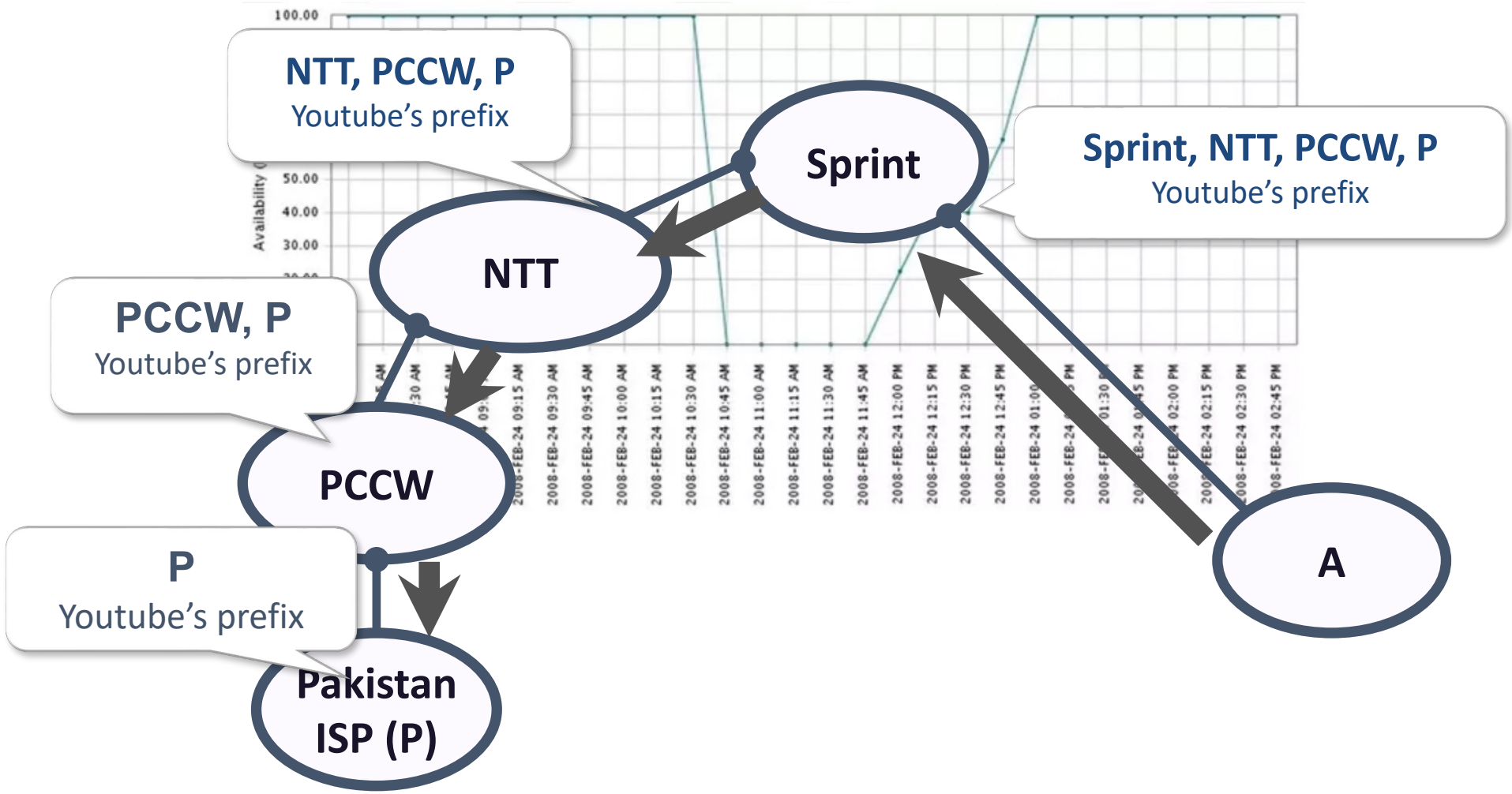
What Is BGP and How Its Failure Took Facebook Down?

WRITTEN BY
Eugene Tkachenko

October 08, 2021 · 4 min read

On October 4, 2021, Facebook – and all the major services Facebook

Pakistan Youtube Outage Event (2008)



Interception in real world

S2W

Feb 17 · 18 min read · Listen

Post Mortem of KlaySwap Incident through BGP Hijacking | EN

Author: S2W TALON with eyez (Lead by Sojun Ryu)

> *Last Modified : 2022.02.16.*

TIMELINE

Preliminary Preparation

The first transaction in the attacker's Account
2021.06.29 08:31

Test Factory Contract created
2022.01.05 08:49

Malicious API server domain created
2022.01.06 07:58

Certificate of malicious API server domain enabled
2022.01.06 08:02

Factory Contract for theft created
2022.01.07 02:49

BGP Hijacking

BGP Hijacking started (211.249.216.0/21)
2022.02.03 10:04

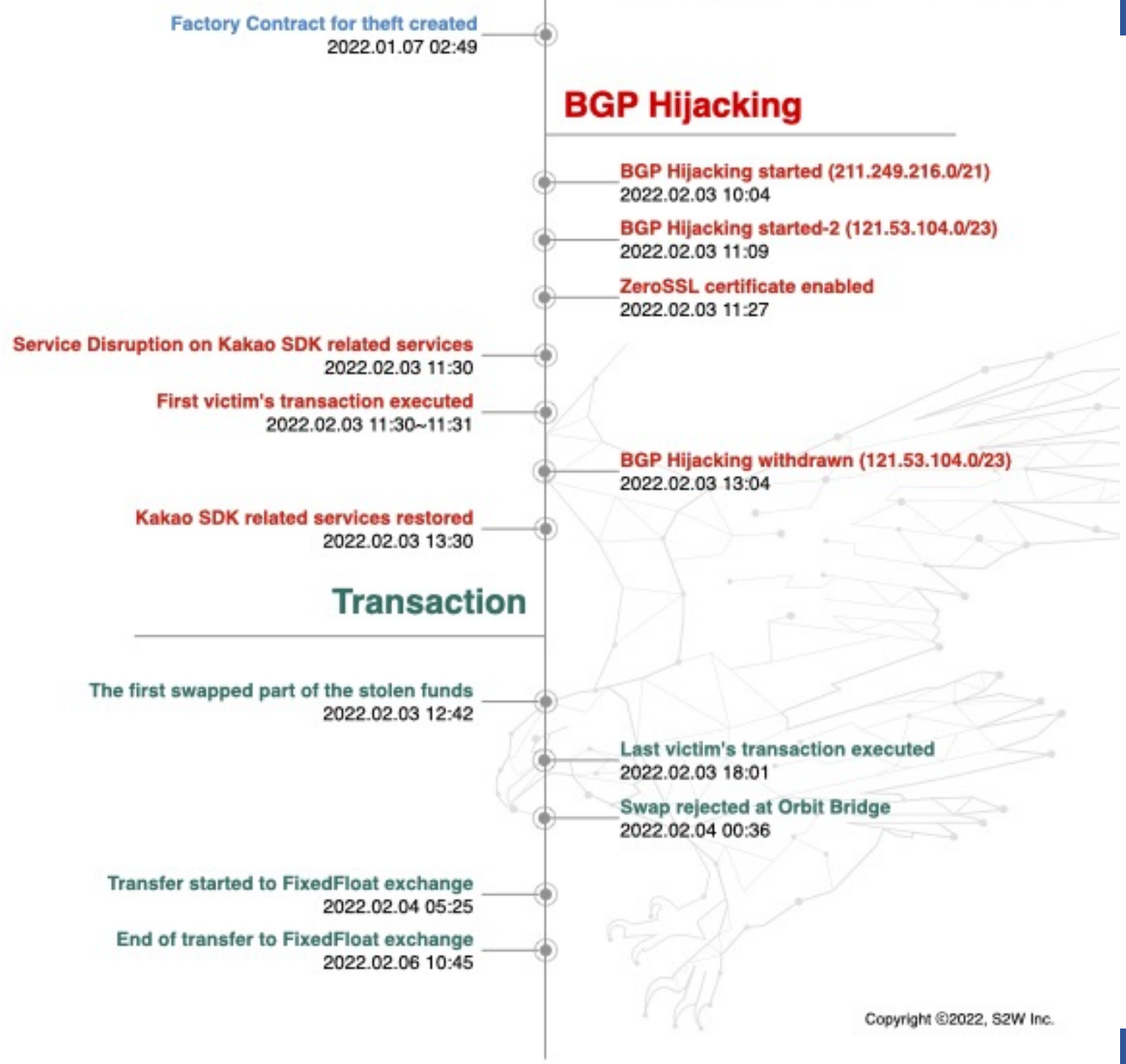
BGP Hijacking started-2 (121.53.104.0/23)
2022.02.03 11:09

ZeroSSL certificate enabled
2022.02.03 11:27

Service Disruption on Kakao SDK related services
2022.02.03 11:30

First victim's transaction executed
2022.02.03 11:30~11:31

BGP Hijacking withdrawn (121.53.104.0/23)
2022.02.03 13:04


Factory Contract for theft created
2022.01.07 02:49
BGP Hijacking
BGP Hijacking started (211.249.216.0/21)
2022.02.03 10:04
BGP Hijacking started-2 (121.53.104.0/23)
2022.02.03 11:09
ZeroSSL certificate enabled
2022.02.03 11:27
Service Disruption on Kakao SDK related services
2022.02.03 11:30
First victim's transaction executed
2022.02.03 11:30~11:31
BGP Hijacking withdrawn (121.53.104.0/23)
2022.02.03 13:04
Kakao SDK related services restored
2022.02.03 13:30
Transaction
The first swapped part of the stolen funds
2022.02.03 12:42
Last victim's transaction executed
2022.02.03 18:01
Swap rejected at Orbit Bridge
2022.02.04 00:36
Transfer started to FixedFloat exchange
2022.02.04 05:25
End of transfer to FixedFloat exchange
2022.02.06 10:45
Copyright ©2022, S2W Inc.

Search

Write

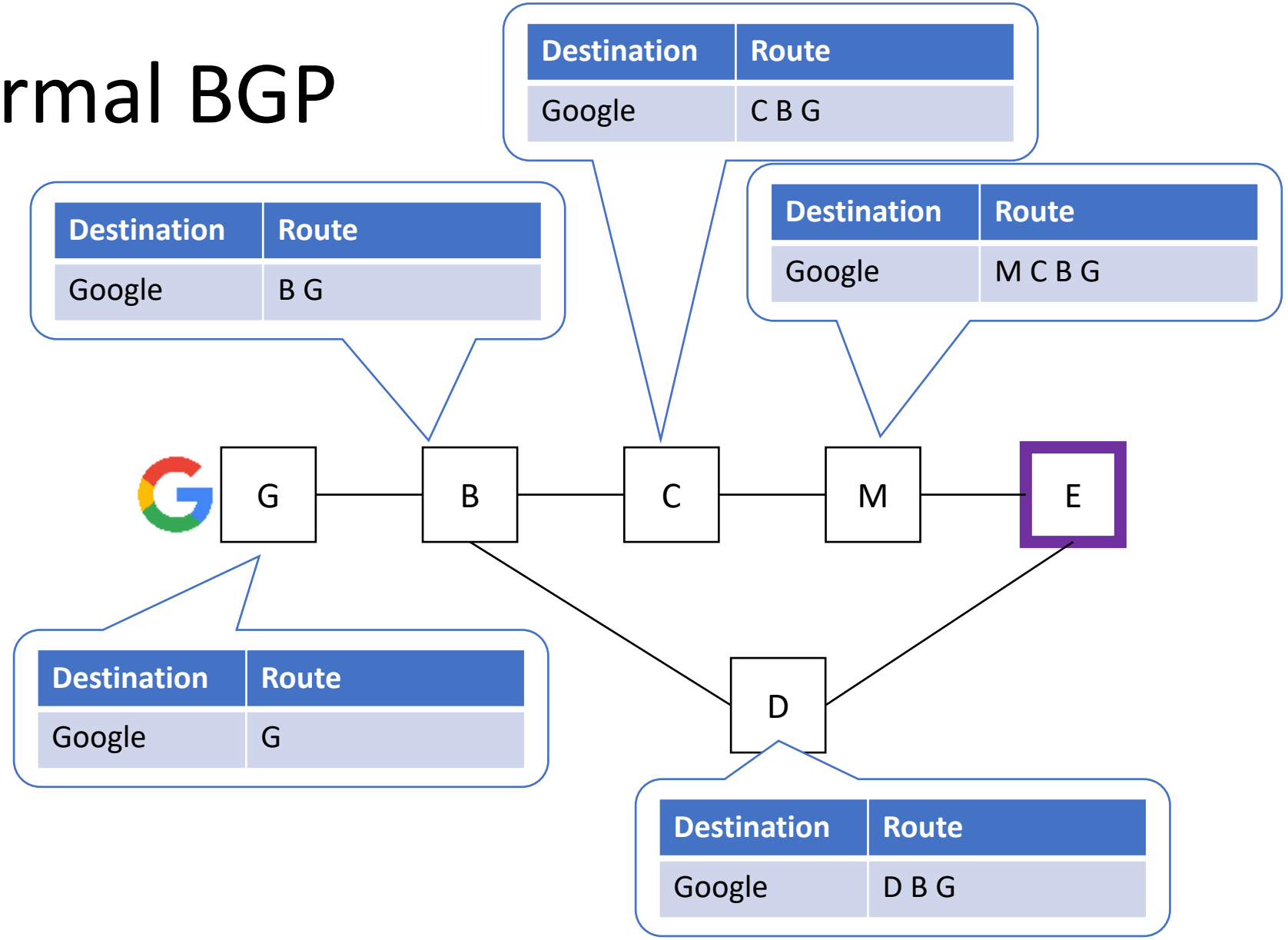
Truth Behind the Celer Network cBridge cross-chain bridge incident: BGP hijacking

SlowMist ·

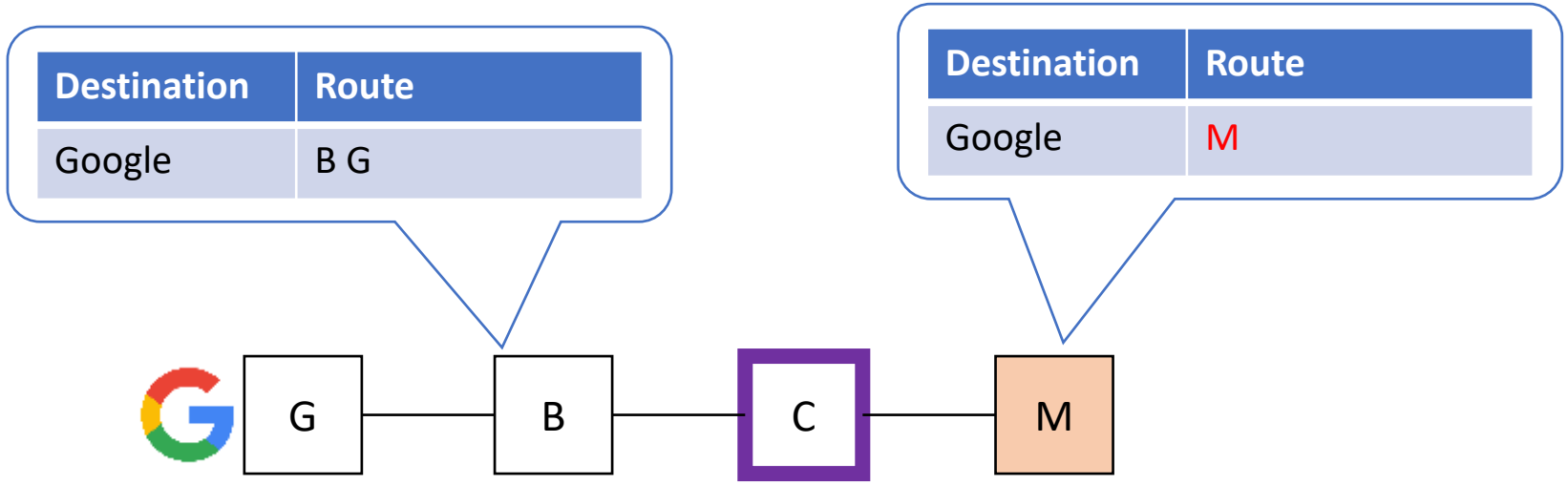
Published in Coinmonks · 8 min read · Aug 20, 2022

59

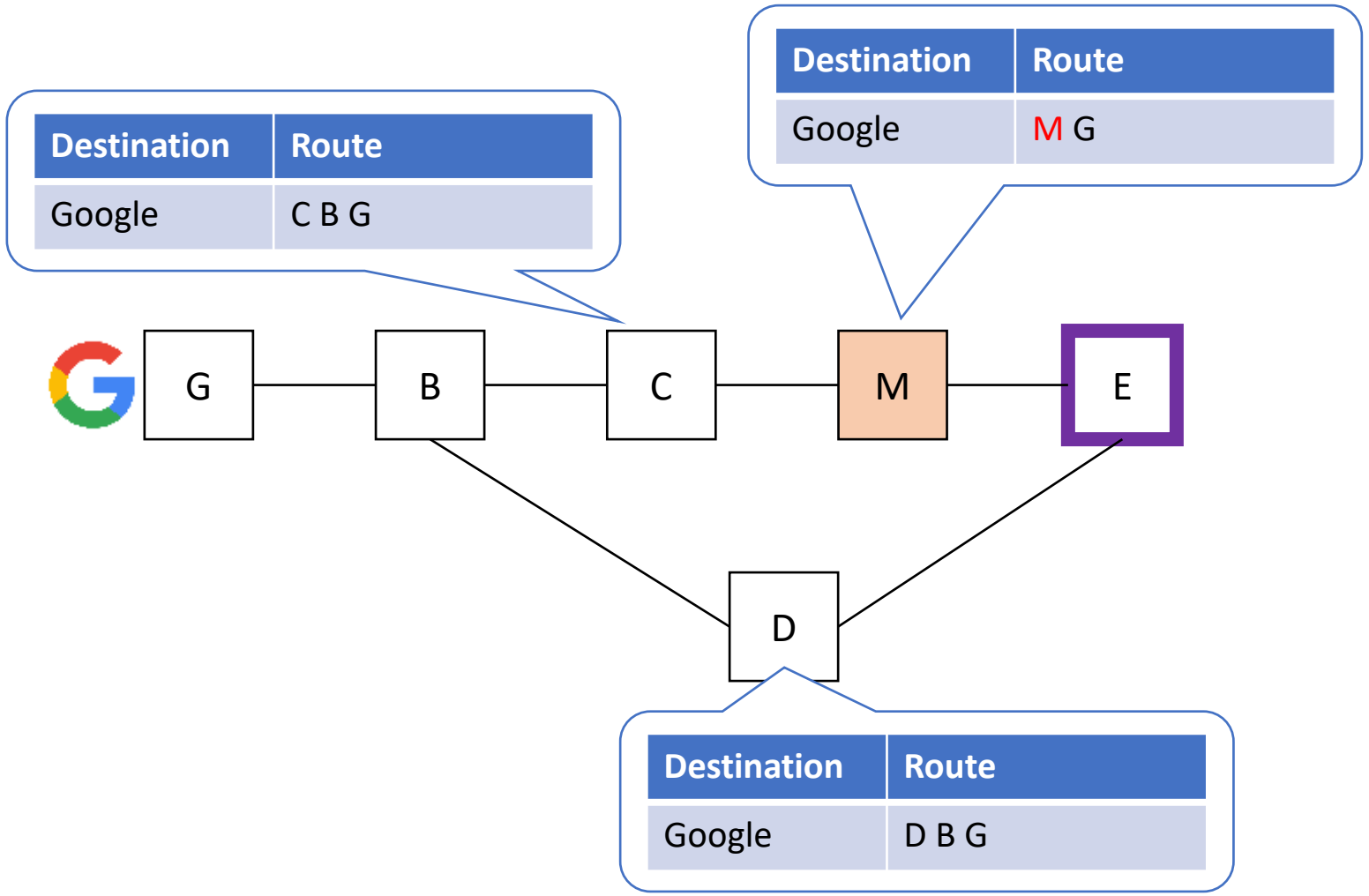
Normal BGP


Destination
Route
Google
C B G
Destination
Route
Google
B G
Destination
Route
Google
M C B G
G
B
C
M
E
D
Destination
Route
Google
G
Destination
Route
Google
D B G

1) Prefix hijacking

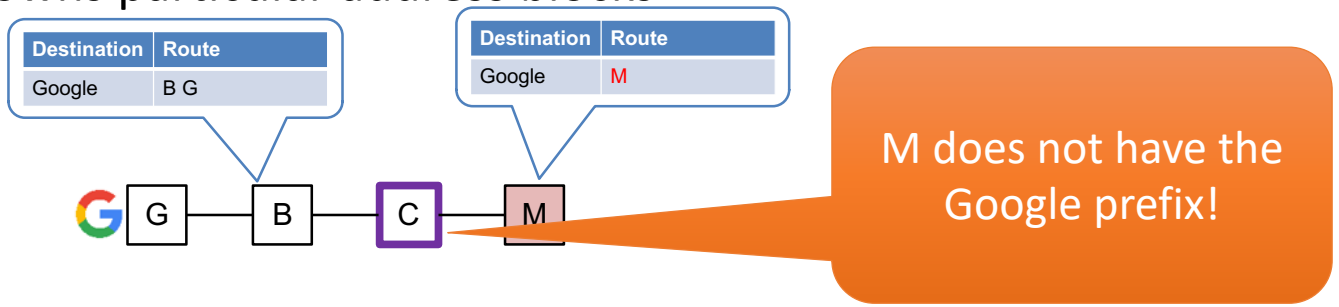


2) One-hop prefix hijacking

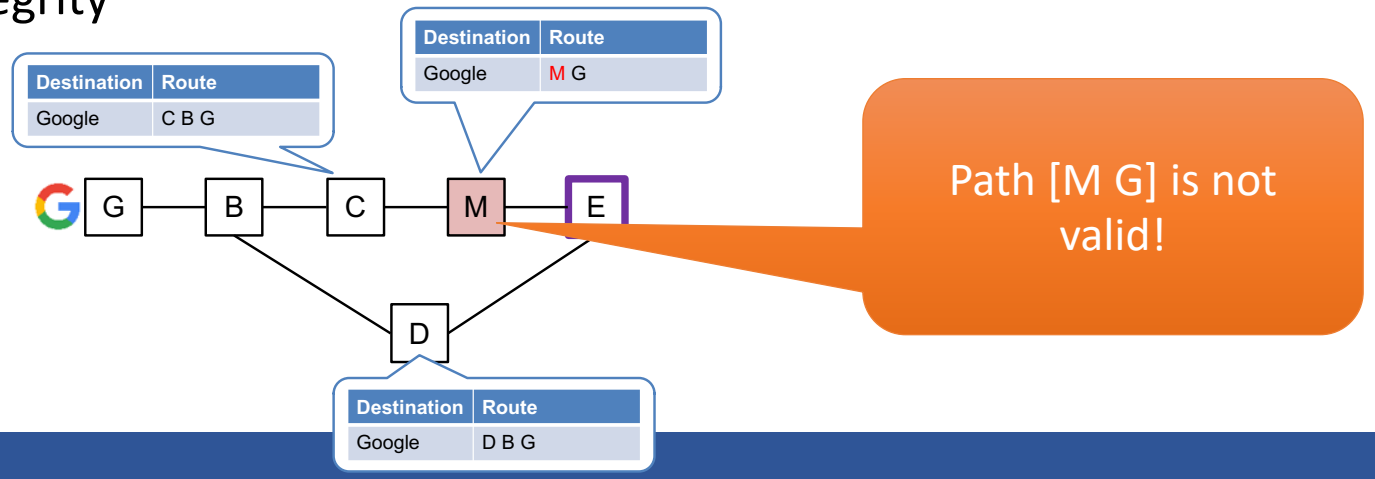


RPKI and BGPSEC

- Resource Public Key Infrastructure (RPKI)
  - Ensures particular AS owns particular address blocks



- BGPSEC
  - Provides AS path integrity



Next…

- *Chapter 5.5 The SDN Control Plane*