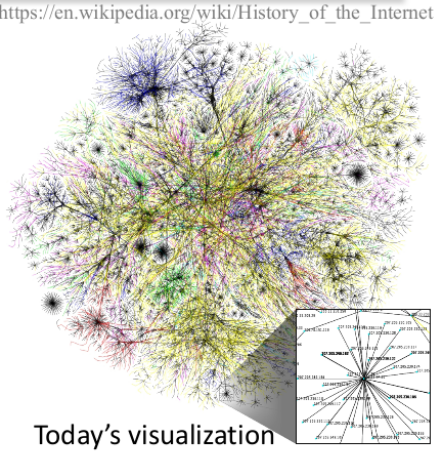
Making Routing Scalable

Source: https://en.wikipedia.org/wiki/History_of_the_Internet

- Our routing study thus far - idealized
  - all routers identical & network "flat"

  … not true in practice

- Scale: with billions of destinations..
  - Can't store all destinations in routing tables!
  - Routing table exchange would swamp links!
- Administrative autonomy
  - Internet = network of networks
  - Each network admin may want to control routing in its own network



Today's visualization of routing paths through a portion of the Internet

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

Intra-AS Routing

- Routing among hosts, routers in same AS ("network")
- All routers in AS must run same intra-domain protocol
- Routers in different AS can run different intra-domain routing protocol
- Gateway router: at "edge" of its own AS, has link(s) to router(s) in other AS'es

Inter-AS Routing

- Routing among AS'es
- Gateways perform inter-domain routing (as well as intra-domain routing)

Interconnected ASes

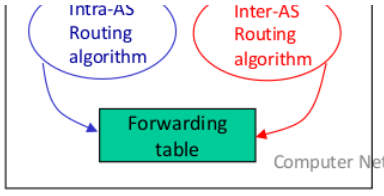
Open Shortest Path First (OSPF)

- "Open": publicly available
- Uses link-state algorithm
  - Link state packet dissemination
  - Topology map at each node
  - Route computation using Dijkstra's algorithm
- Router floods OSPF link-state advertisements to all other routers in entire AS
  - Carried in OSPF messages directly over IP (rather than TCP or UDP)
  - Link state: for each attached link
- IS-IS (Intermediate System to Intermediate System Protocol) routing protocol
  - Nearly identical to OSPF

table configured by both
ıter-AS routing algorithm

routing determine
entries for destinations within AS

- Inter-AS & intra-AS determine entries for external destinations



Achievable via:

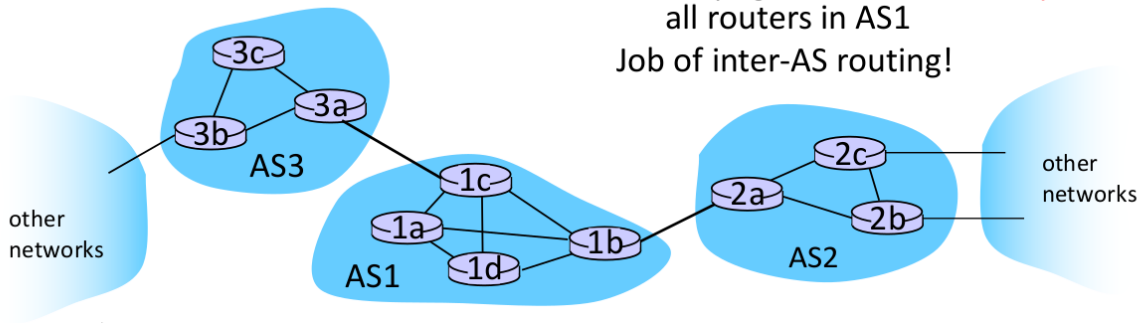
Suppose router in AS1 receives datagram destined outside of AS1:

- Router should forward packet to gateway router, but which one?

AS1 must:

- Learn which destinations are reachable through AS2, which through AS3
- Propagate this reachability info to all routers in AS1

Job of inter-AS routing!

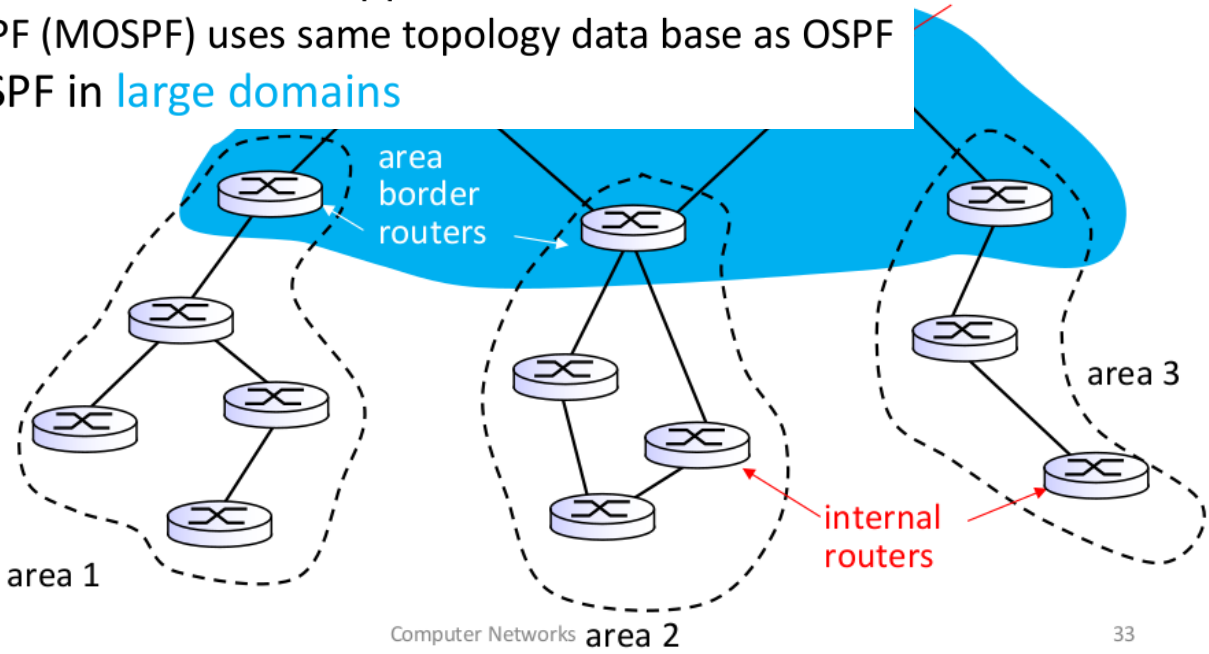


Intra-AS Routing

- Also known as interior gateway protocols (IGP)
- Most common intra-AS routing protocols:
  - RIP: Routing Information Protocol (RIPng: RFC2080)
  - OSPF: Open Shortest Path First (RFC5340) (IS-IS protocol essentially same as OSPF)
  - (E)IGRP: (Enhanced) Interior Gateway Routing Protocol (Cisco proprietary for decades, until 2016, RFC7868)

‘Gatewayrouter’ verbinden zwischen 2 Autonomen Systems wobei jeder Router Einfluss auf forwarding table hat.
→ Schickt 1d zu 2a: Siehe Forwarding Tabelle und sende durch – Das Intra-AS behandelt das Senden von 1d zu 1b:

- Security: all OSPF messages authenticated (to prevent malicious intrusion)
- Multiple same-cost paths allowed (only one path in RIP)
- For each link, multiple cost metrics for different ToS, e.g.:
  - Satellite link cost set low for best effort ToS
  - High for real-time ToS
- Integrated uni- and multi-cast support:
  - Multicast OSPF (MOSPF) uses same topology data base as OSPF
- Hierarchical OSPF in large domains



→ Flooding

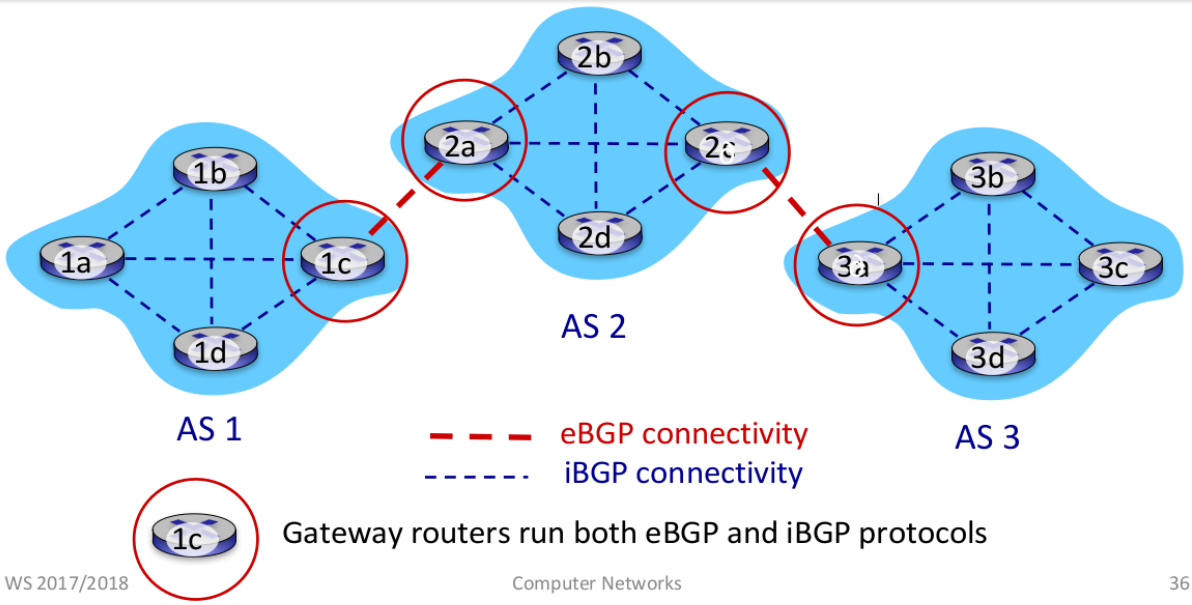
→ Zum Vermeiden des Floodings des gesamten Netzwerkes ↔ Gebe Daten an relevante Router

- Two-level hierarchy: local area, backbone
  - Link-state advertisements only in area
  - Each nodes has detailed area topology; only know direction (shortest path) to networks in other areas
- Area border routers
  - "Summarize" distances to networks in own area
  - Advertise to other area border routers
- Backbone routers
  - Run OSPF routing limited to backbone
- Boundary routers
  - Connect to other AS'es

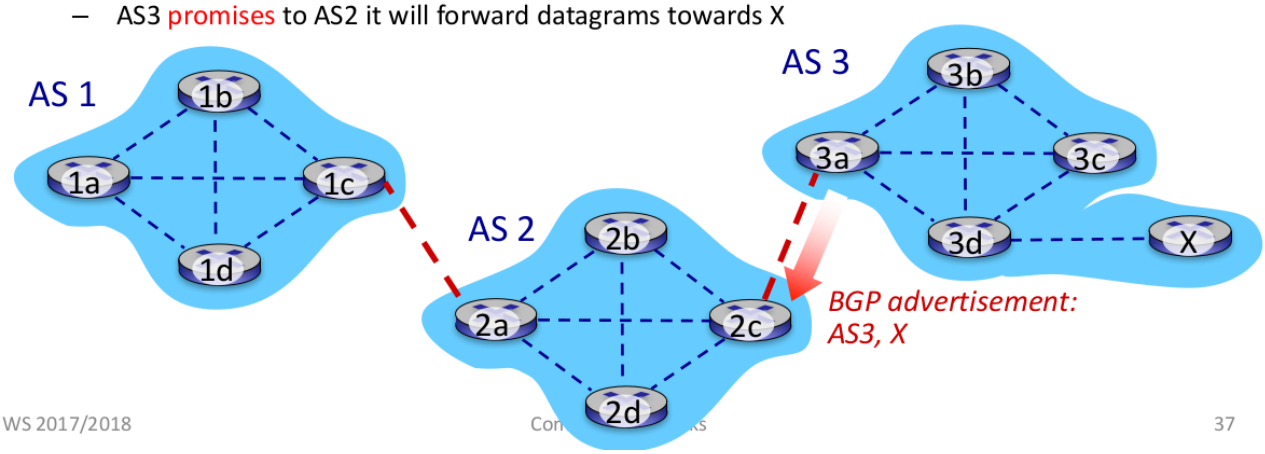
Internet Inter-AS Routing: BGP

- BGP (Border Gateway Protocol) (RFC4271)
  - The de facto/jure inter-domain routing protocol
  - "Glue that holds the internet together"
- BGP provides each AS a means to:
  - eBGP (external or exterior): obtain subnet reachability information from neighboring ASes
  - iBGP (internal or interior): propagate reachability information to all AS-internal routers.
  - Determine "good" routes to other networks based on reachability information and policy
- Allows subnet to advertise its existence to rest of internet
  - "I am here"
- BGP in AT: https://www.vix.at (Vienna Internet eXchange); 141 ASes (Dec 18, 2017)

– Protokoll dass die Knoten des Webs zusammenhält



- BGP session: two BGP routers ("peers") exchange BGP messages over semi-permanent TCP connection (port 179)
  - Advertising paths to different destination network prefixes (BGP is a "path vector" protocol)
- When AS3 gateway router 3a advertises path AS3,X to AS2 gateway router 2c:
  - AS3 promises to AS2 it will forward datagrams towards X

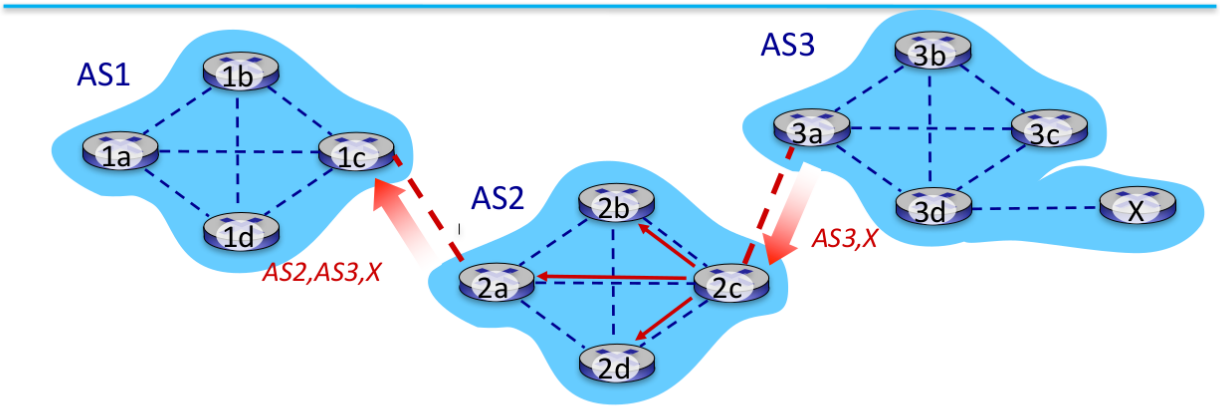


– Extern & Intern-Dataflows
→ Wenn neues Netzwerk/Router dazukommt, so würde man anhand dieses Protokolls das dazuschalten ankündigen – oben: Kontakt von ‘X’ wird weitergeleitet.

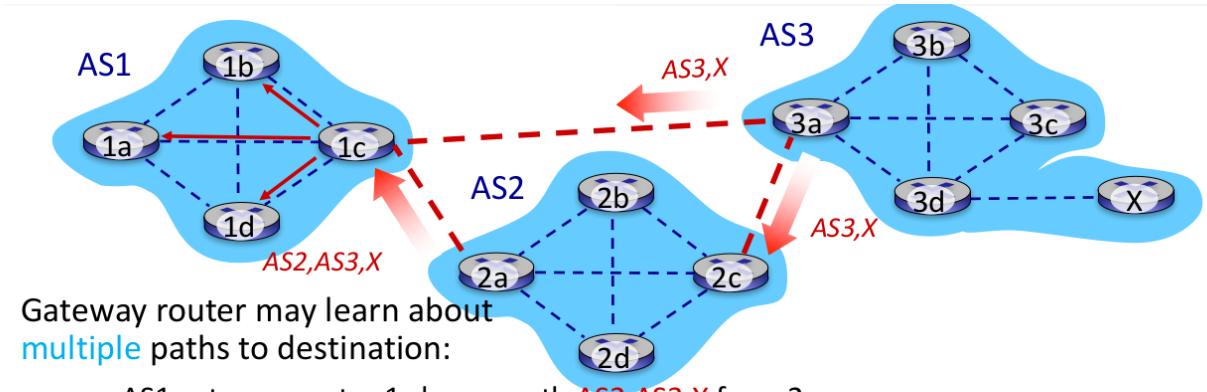
- Advertised prefix includes BGP attributes
  - Prefix + attributes = “route”
- Two important attributes
  - AS-PATH: list of ASes through which prefix advertisement has passed (e.g., AS3; X)
  - NEXT-HOP: indicates specific internal-AS router (IP address) to next-hop AS (e.g., IP address of leftmost interface for router 3a)
- 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

– Für AS1: X ist erreichbar über AS3 und AS2
Samples:

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



Gateway router may learn about multiple paths to destination:

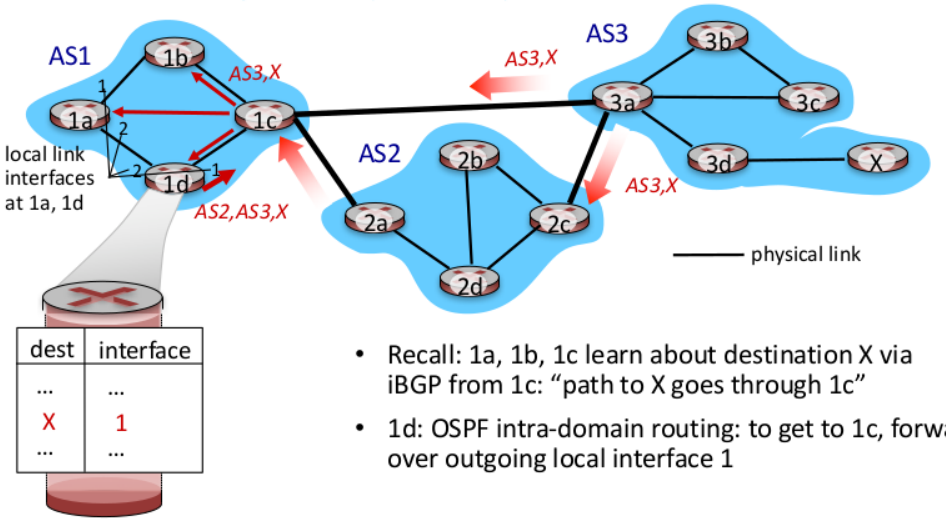
- AS1 gateway router 1c learns path AS2,AS3,X from 2a
- AS1 gateway router 1c learns path AS3,X from 3a
- Based on policy, AS1 gateway router 1c chooses path AS3,X, and advertises path within AS1 via iBGP

- BGP messages 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 message; also used to close connection

- Router may learn about more than one route to destination AS, selects route based on:
  1. Local preference value attribute: policy decision
  2. Shortest AS-PATH
  3. Closest NEXT-HOP router: hot potato routing
  4. Additional criteria based on BGP identifiers

BGP, OSPF, forwarding table entries

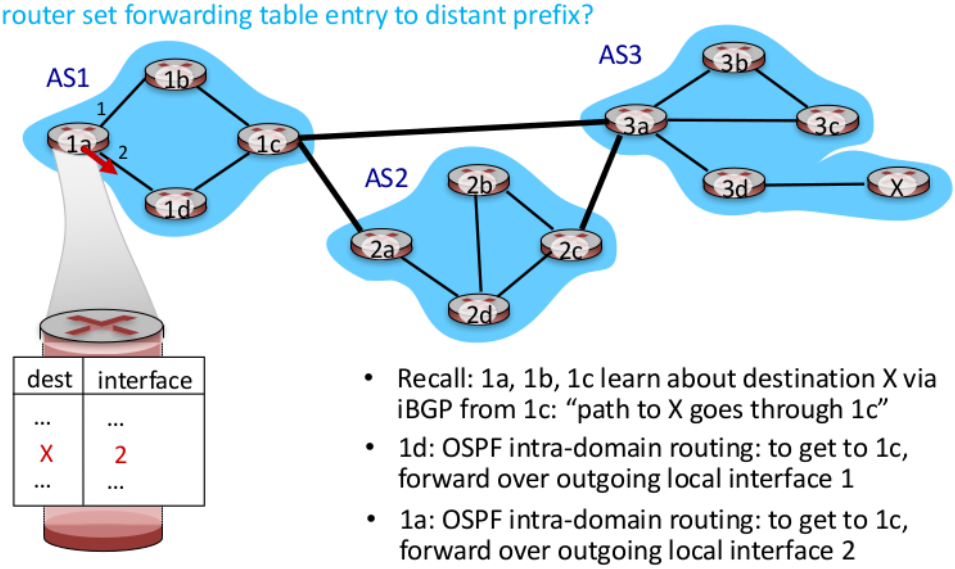
Q: how does router set forwarding table entry to distant prefix?



- Recall: 1a, 1b, 1c learn about destination X via iBGP from 1c: "path to X goes through 1c"
- 1d: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 1

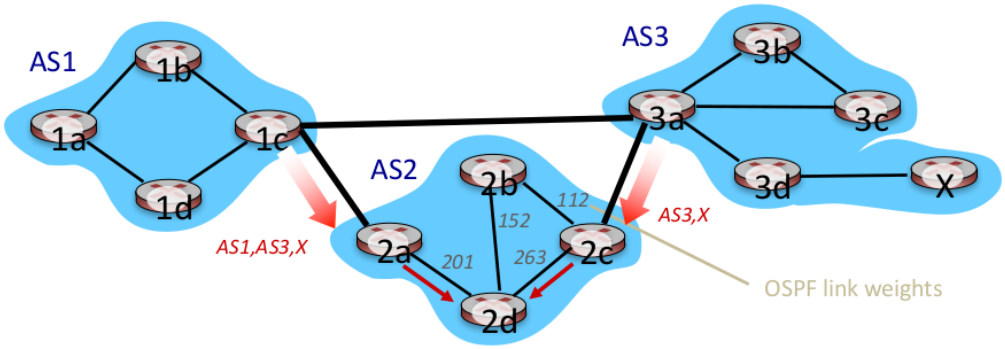
→ Pending on Intranet – Hier: Shortest AS-Path (Siehe AS3,X)

Q: how does router set forwarding table entry to distant prefix?



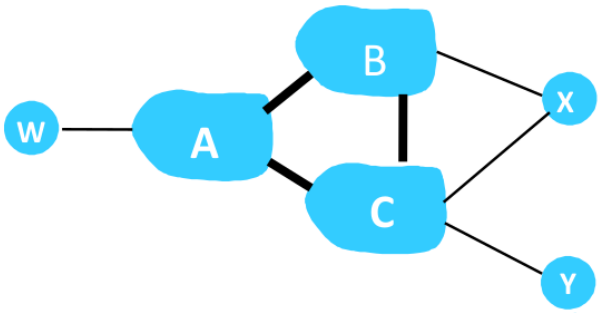
- Recall: 1a, 1b, 1c learn about destination X via iBGP from 1c: "path to X goes through 1c"
- 1d: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 1
- 1a: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 2

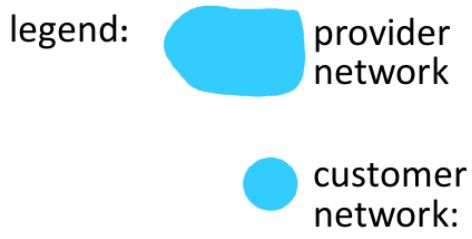
Hot Potato Routing



- 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!

BGP: Achieving Policy via Advertisements





Suppose an ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs)

- A advertises path Aw to B and to C
- B chooses not to advertise BAw to C:
  - B gets no "revenue" for routing CBAw, since none of C, A, w are B's customers
  - C does not learn about CBAw path
- C will route CAw (not using B) to get to w
- A,B,C are provider networks
- X,W,Y are customer (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

→ Packete so schnell wie möglich wieder los werden :: Weitergabe von 2d via 2a oder 2c möglich ↔ Intra nimmt es den Gate mit den niedrigsten Kosten: 2a.

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 no policy decisions needed

Scale

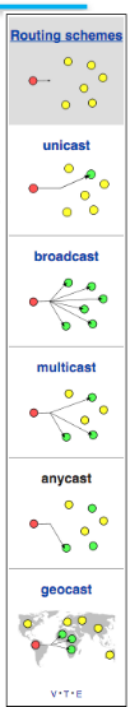
- Hierarchical routing saves table size, reduced update traffic

Performance

- Intra-AS: can focus on performance
- Inter-AS: policy may dominate over performance

BGP to implement IP-Anycast

- Remember terminology
  - Unicast: one-to-one; broadcast: one-to-all; multicast: one-to-many
- Anycast: one-to-one-of-many
  - Single receiver from a group based on some criteria (e.g., nearest according to distance metric)
  - Used in DNS and CDN: e.g., redirect to nearest DNS server or content (replication) server



Software-Defined Networking (SDN)

- Internet network layer: distributed, per-router approach
  - Monolithic routers and different "middleboxes"
  - ~2005: renewed interest to rethink control plane
- SDN key characteristics
  - Flow-based forwarding
  - Separation of data plane and control plane
  - Network control functions: external to data plane switches
  - A programmable network

BITMOVIN

SDN in a Nutshell

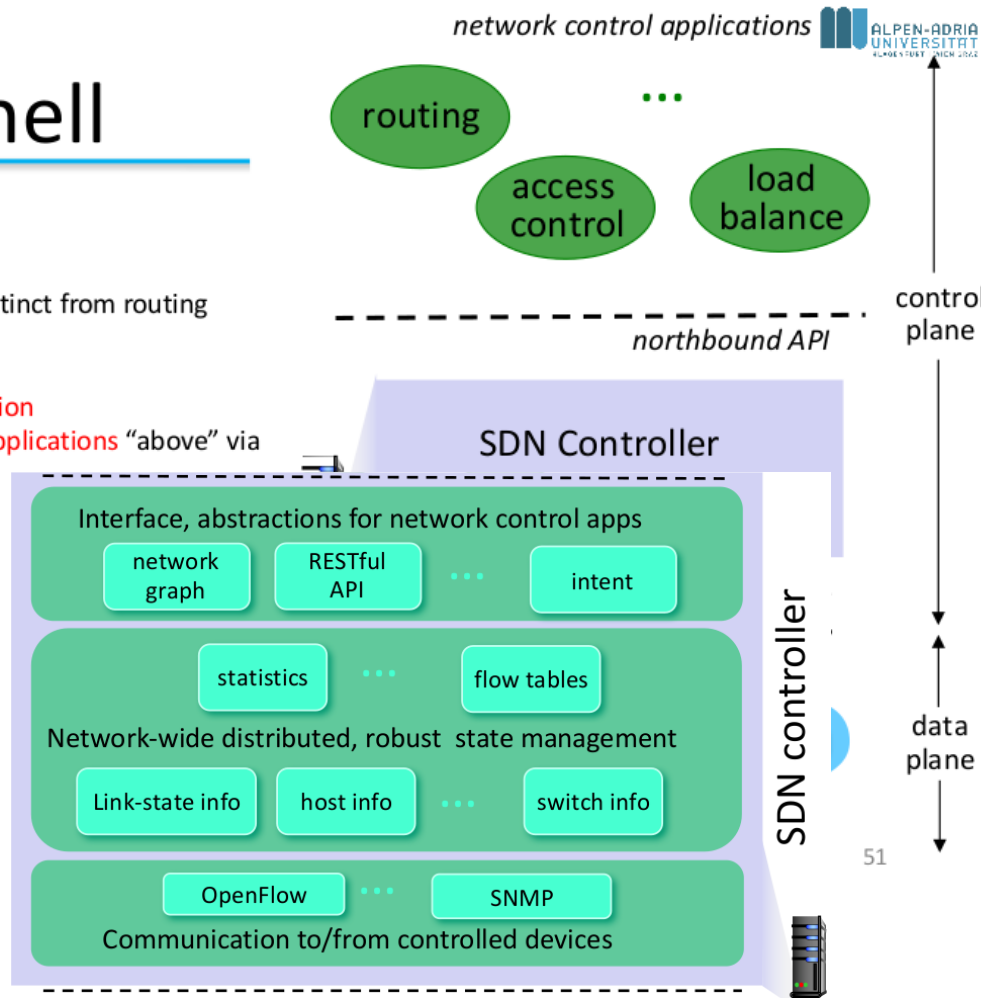
- Network control applications
  - "Brains" of control
  - Unbundled: e.g., by 3rd party distinct from routing vendor or SDN controller
- SDN controller
  - Maintain network state information
  - Interacts with network control applications "above" via northbound API

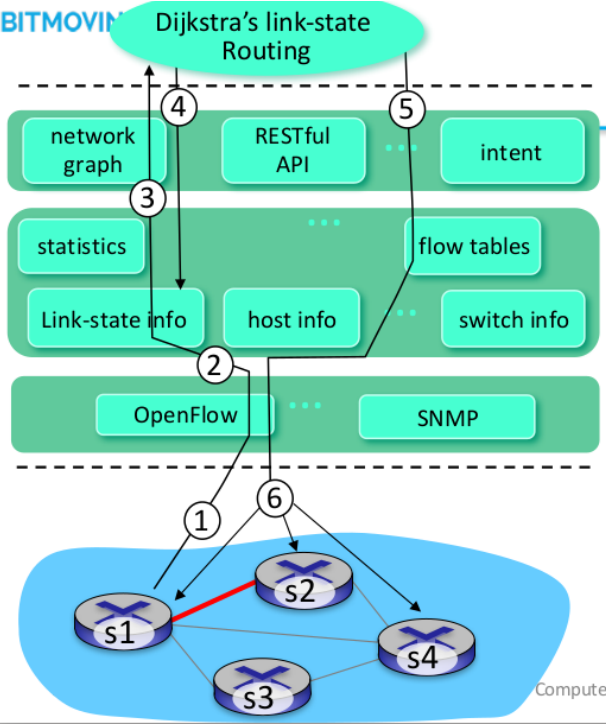
Components of a SDN Controller

Interface layer to network control apps: abstractions API

Network-wide state management layer: state of networks links, switches, services: a distributed database

Communication layer: communicate between SDN controller and controlled switches





SDN: Control/Data Plane Interaction Example

① S1, experiencing link failure using OpenFlow port status message to notify controller

② SDN controller receives OpenFlow message, updates link status info

③ Dijkstra's routing algorithm application has previously registered to be called when ever link status changes. It is called.

④ Dijkstra's routing algorithm access network graph info, link state info in controller, computes new routes

⑤ link state routing app interacts with flow-table-computation component in SDN controller, which computes new flow tables needed

⑥ Controller uses OpenFlow to install new tables in switches that need updating

@Characteristics
– Bsp.: FTP – out-of-band
– SDN Controller muss Data- als auch Controlplane befassen

Network Control and Management

- ICMP: Internet Control Message Protocol (RFC792)
  - Used by hosts and routers to communicate network-level information
  - Error reporting: unreachable host, network, port, protocol
  - Echo request/reply (used by ping)
- Network Management
  - Network management refers to the activities, methods, procedures, and tools that pertain to the operation, administration, maintenance, and provisioning of networked systems
- SNMP: Simple Network Management Protocol
  - Request/response protocol (GET/SET) but also "traps"
  - Widely used for management and monitoring (e.g., Nagios)

* snmpget -v 1 -c public ppsb.itec.uni-klu.ac.at system.sysUpTime.0
= Drucker.
* snmpget -v 1 -c public ppsa.itec.uni-klu.ac.at system.sysUpTime.0
= Drucker von Timse
MIB – Management Information Base ↔ Strukturierungsform der Antwort, standardisiert/global/firmenintern/gerätespezifisch

* snmpget -v 1 -c public posb.itec.uni-klu.ac.at 1.3.6.1.2.1.43.10.2.1.4.1.1 ← Request
SNMPv2-SMI::mib-2.43.10.2.1.4.1.1 = Counter32:58415 ←Antwort
– @Uni: Erreichbar nur via VPN