Communication Networks

Spring 2022

Laurent Vanbever

nsg.ee.ethz.ch

ETH Zürich (D-ITET)

May 30 2022

Materials inspired from Scott Shenker, Jennifer Rexford, and Ankit Singla

Last week on

Communication Networks

DNS

Web

google.ch ⟷ 172.217.16.131

(the end)

http://www.google.ch

(the beginning)

DNS

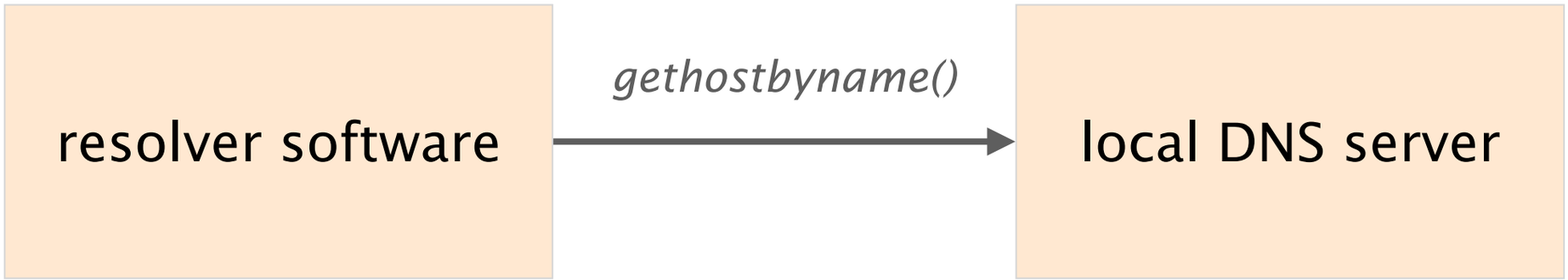
Web

google.ch ⟷ 172.217.16.131

(the end)

Records	Name	Value
A	hostname	IP address
NS	domain	DNS server name
MX	domain	Mail server name
CNAME	alias	canonical name
PTR	IP address	corresponding hostname

Using DNS relies on two components



trigger resolution process

send request to local DNS server

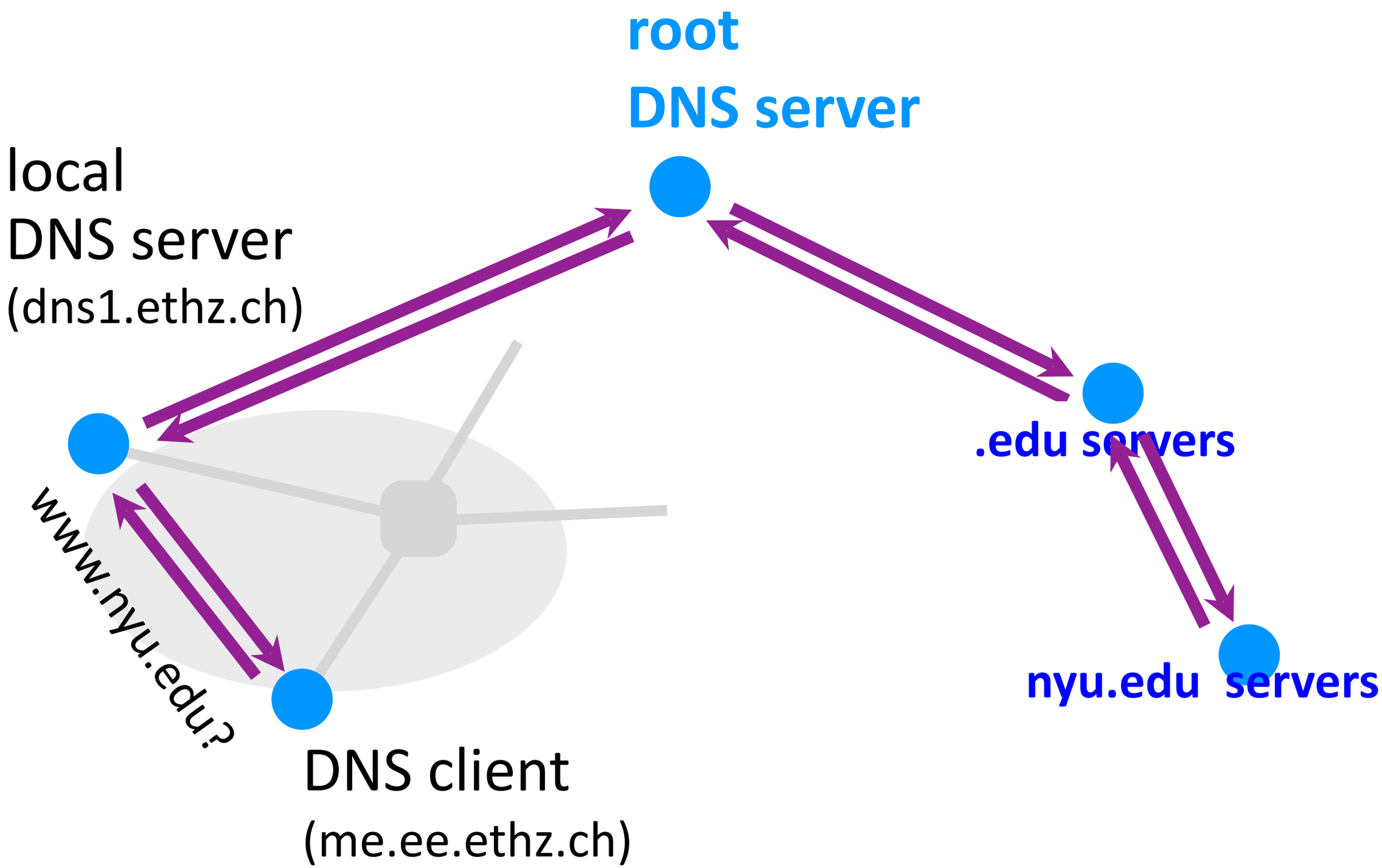
usually, near the endhosts

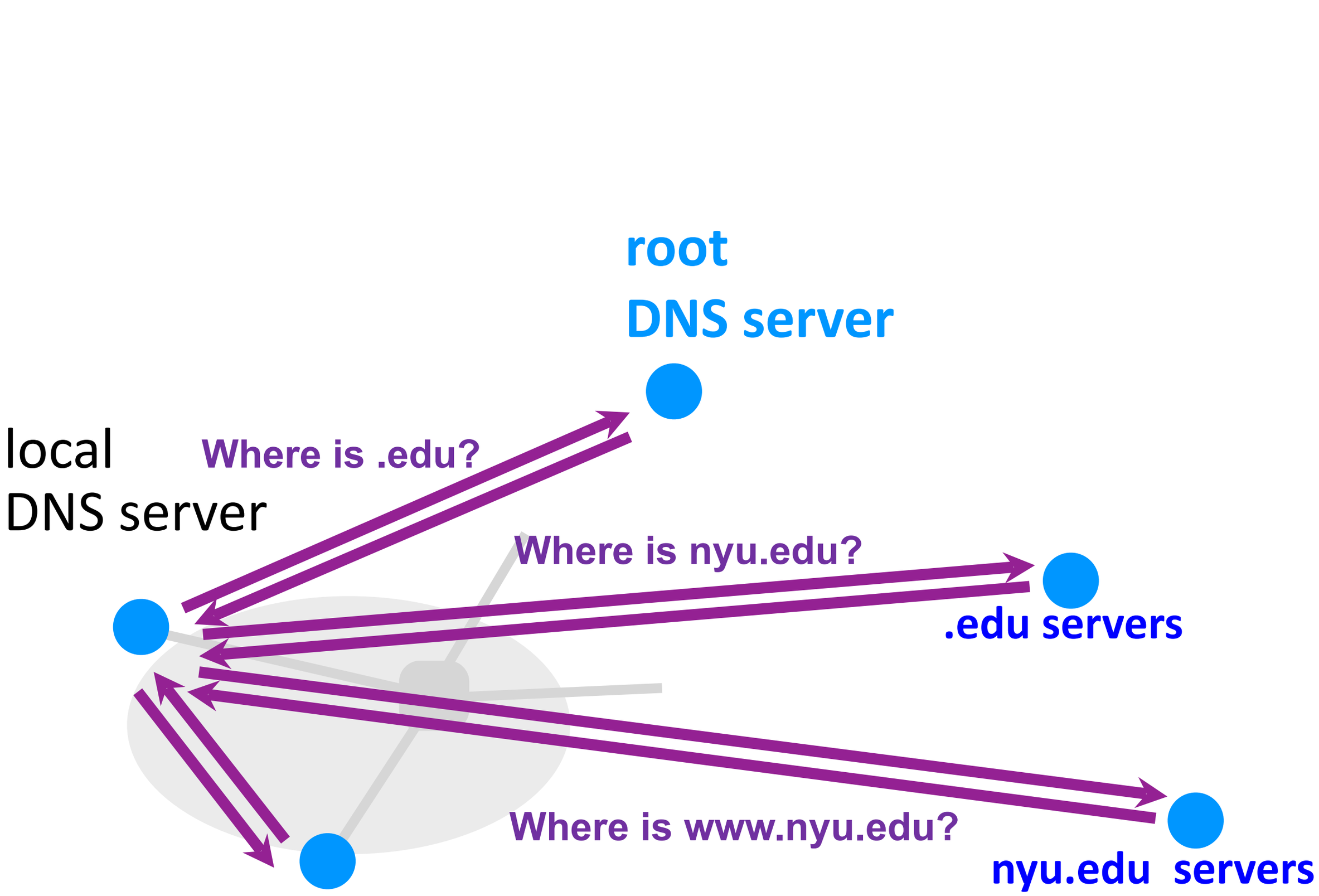
configured statically (resolv.conf)

or dynamically (DHCP)

DNS resolution can either be

recursive or iterative


root
DNS server
local
DNS server
(dns1.ethz.ch)
.edu servers
www.nyu.edu?
nyu.edu servers
DNS client
(me.ee.ethz.ch)


root
DNS server
local
DNS server
Where is .edu?
Where is nyu.edu?
.edu servers
Where is www.nyu.edu?
nyu.edu servers
DNS client
(me.ee.ethz.ch)

DNS

Web

http://www.google.ch

(the beginning)

The WWW is made of three key components

Infrastructure	Content	Implementation
Clients/Browser	Objects files, pictures, videos, …	URL: name content
Servers	*organized* in	HTTP: transport content
Proxies	Web sites a collection of objects	

A Uniform Resource Locator (URL)
refers to an Internet ressource

protocol://hostname[:port]/directory_path/resource

HTTP is a rather simple synchronous request/reply protocol

HTTP is layered over a bidirectional byte stream

typically TCP, but QUIC is ramping up

HTTP is text-based (ASCII)

human readable, easy to reason about

HTTP is stateless

it maintains *no info* about past client requests

Today on

Communication Networks

Web

http://www.google.ch

(the end)

Course recap

The life of Internet packets

(a streamed movie)

Web

Course recap

http://www.google.ch

(the end)

Web

Course recap

The life of Internet packets

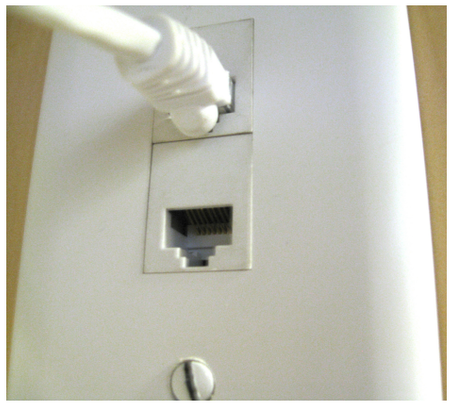
(a streamed movie)

Communication Networks

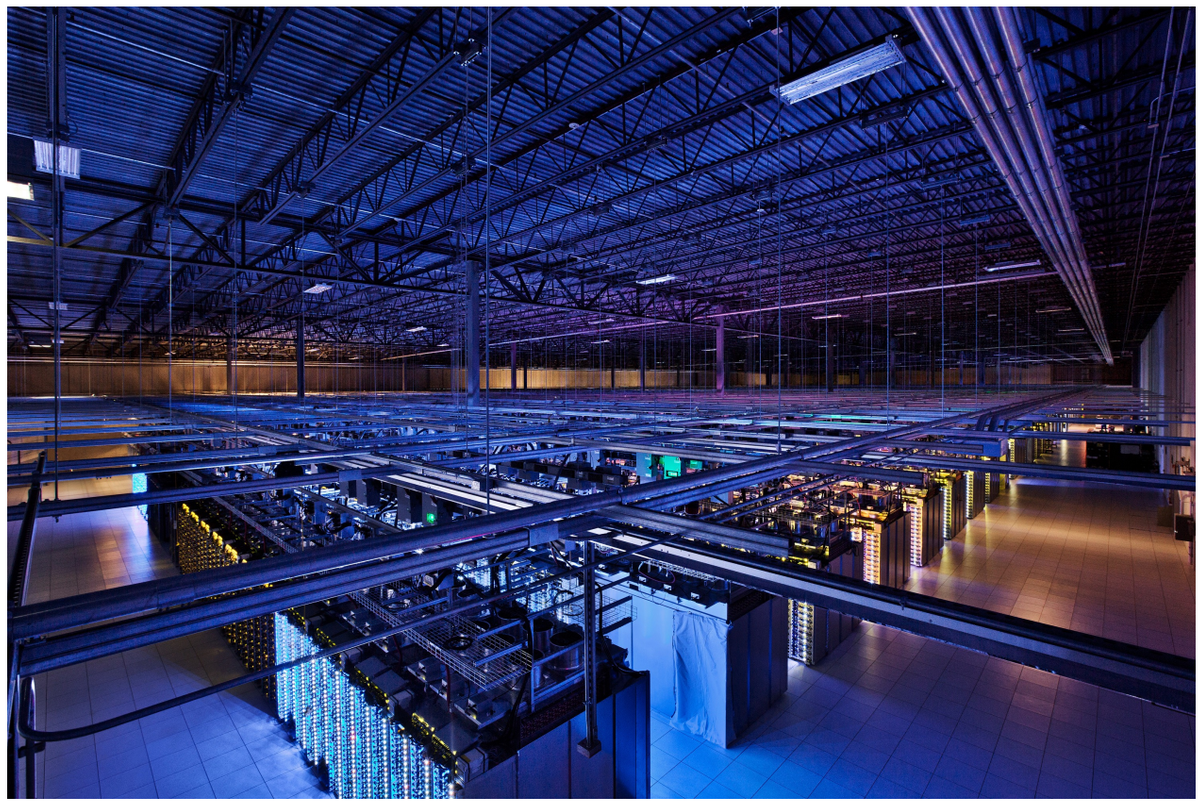
So what?!

Knowledge

Understand how the Internet works and why



from your
network plug…



…to the largest data-centers out there

Let's do a quick recap of the lecture by dissecting

"The life of a few packets" together

Our goal: watch a video on `my.video.com`

A destination outside of our local network

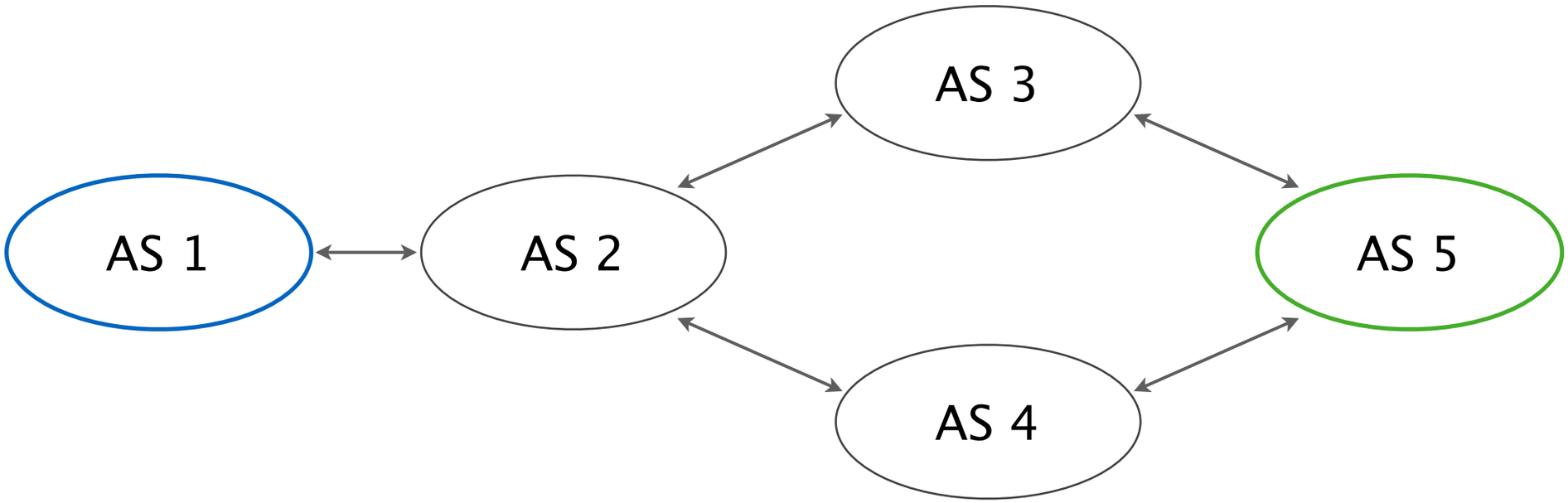
We consider a new host with clean state

I.e., network-wise nothing is configured/known

Which packets do we need to achieve that?

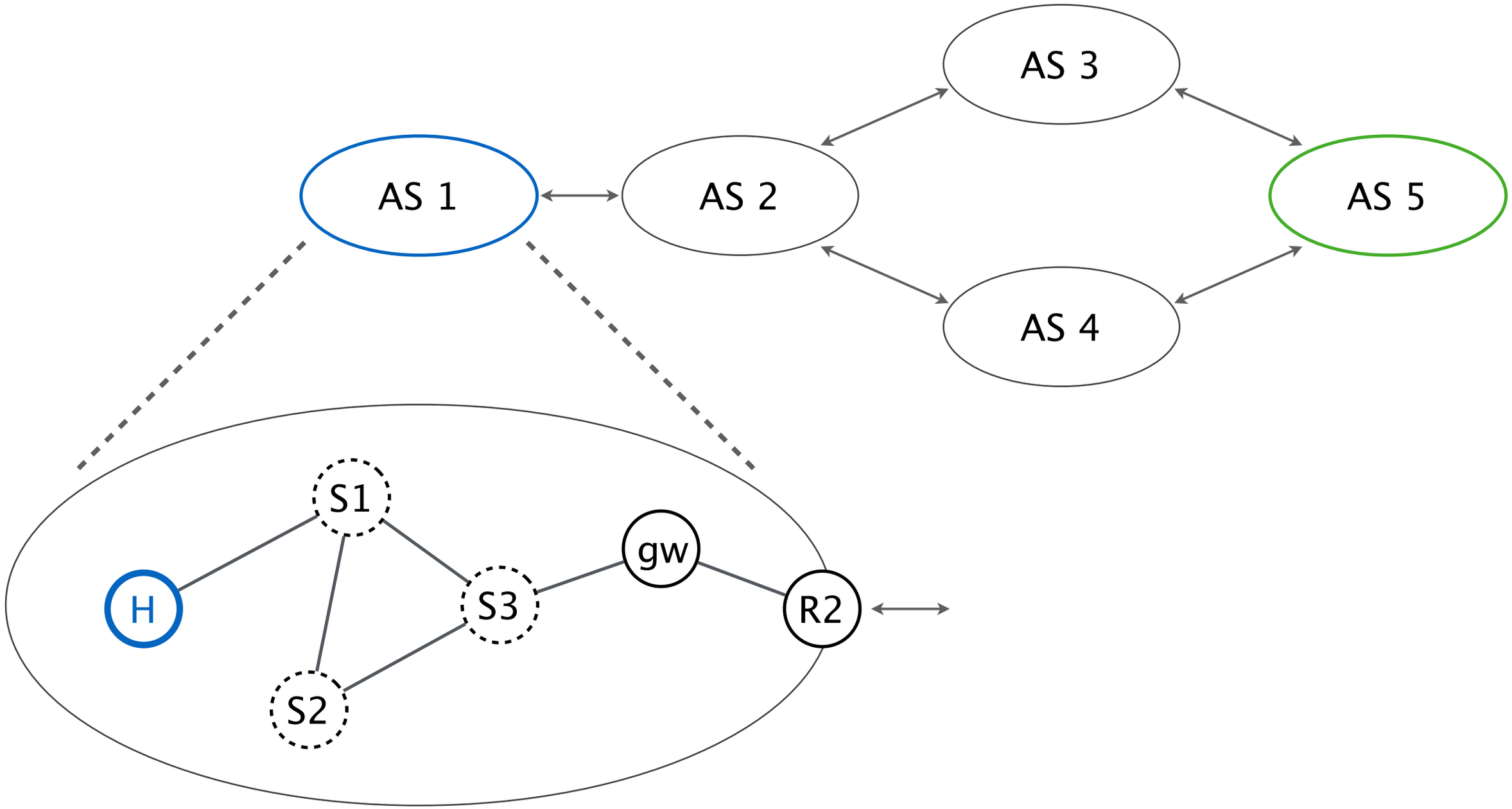
Our host belongs to AS 1,

`my.video.com` belongs to AS 5



Our host belongs to AS 1,

`my.video.com` belongs to AS 5



Problem: Who and where am I?

DHCP

The Dynamic Host Configuration Protocol provides:

- an IP address
- the corresponding IP prefix
- the IP of the default gateway
- DNS server to use
- (many other options)

Manual

Alternatively, we can manually configure the host

You did that extensively during the routing project

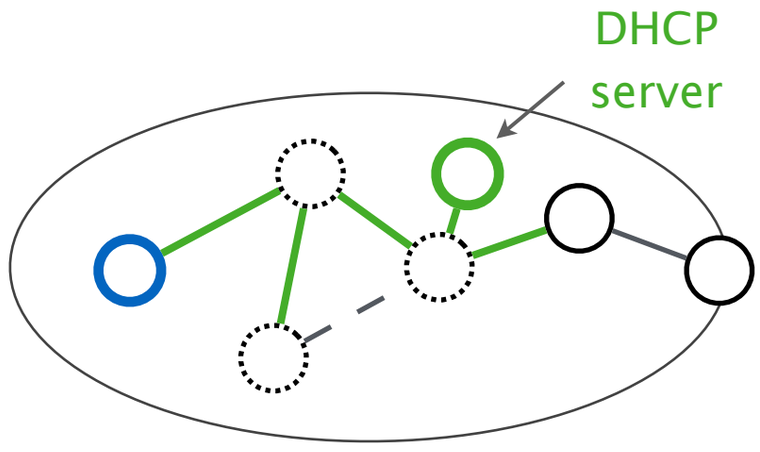
DHCP works within a broadcast domain

(i.e. a local L2 network)

src MAC: host H's MAC

dst MAC: ff:ff:ff:ff:ff:ff

DHCP discovery:
I want an IP



src MAC: host H's MAC

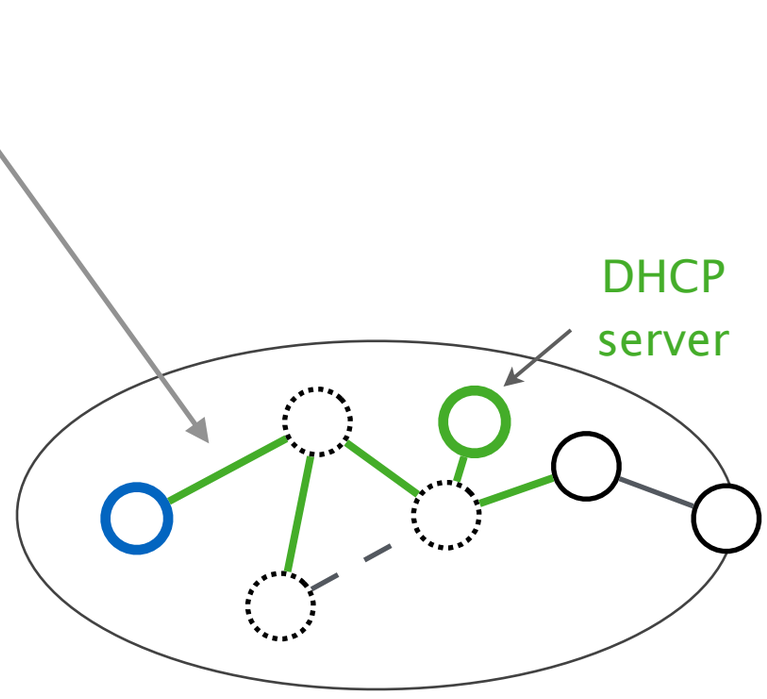
dst MAC: ff:ff:ff:ff:ff:ff

DHCP discovery:

I want an IP

Broadcasted along the layer 2

Spanning Tree computed by the switches



The DHCP server unicasts its answer back to the sender

src MAC: MAC of DHCP

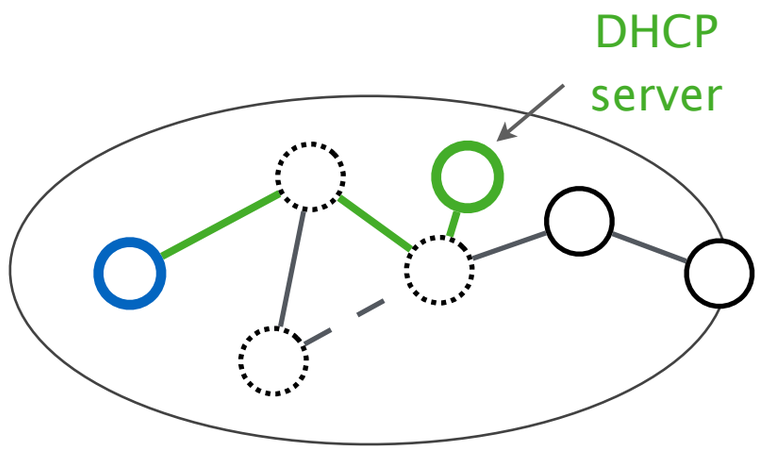
dst MAC: host H's MAC

DHCP offer:

Use 192.168.1.20/24

Default gw: 192.168.1.1

DNS server: 192.168.1.2



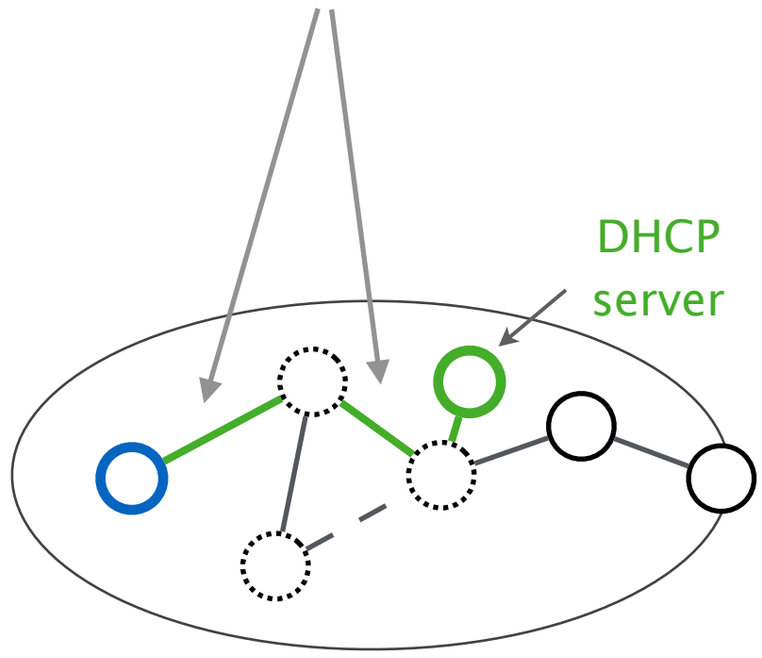
The DHCP server unicasts its answer back to the sender

src MAC: MAC of DHCP

dst MAC: host H's MAC

DHCP offer:
Use 192.168.1.20/24
Default gw: 192.168.1.1
DNS server: 192.168.1.2

The switches have learned over which physical ports they can reach the MAC of H



These slides show a simplified version of DHCP, see exercise 3 for more details

Problem: Who is `my.video.com`?

DNS

The Domain Name System translates names to IPs

The opposite is also possible

Resource Records

A DNS server stores records for different resources

For example domains, mail servers, aliases…

Manual

Alternatively, we can directly provide the IP

But normally we do not know the IPs of external domains

Here, we'll consider that the DNS server is located in the local L2 network

Here, we'll consider that the DNS server is located in the local L2 network

We can also use external DNS servers

e.g. Google's

src MAC: host H's MAC

dst MAC: **???**

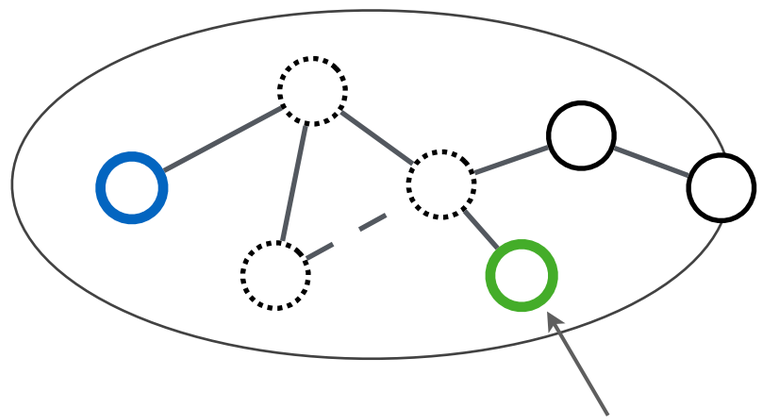
- - - - - - - - - - -

src IP: 192.168.1.20

dst IP: 192.168.1.2

- - - - - - - - - - -

DHCP query:
What is the IP of
`my.video.com`?



src MAC: host H's MAC

dst MAC: ???

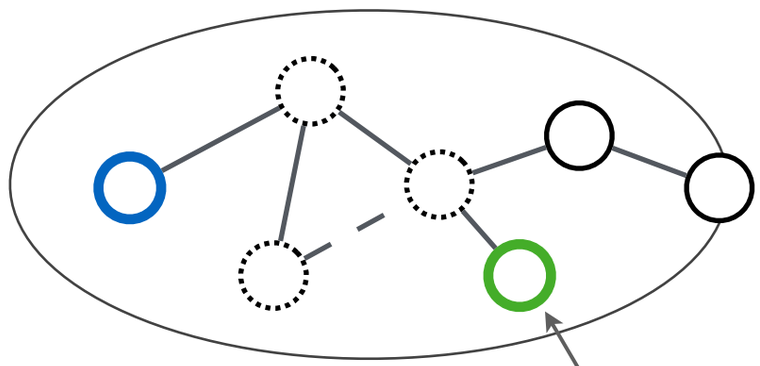
What is the MAC address of the DNS server?

src IP: 192.168.1.20

dst IP: 192.168.1.2

Known via DHCP

DHCP query:
What is the IP of
`my.video.com`?



DNS server
192.168.1.2

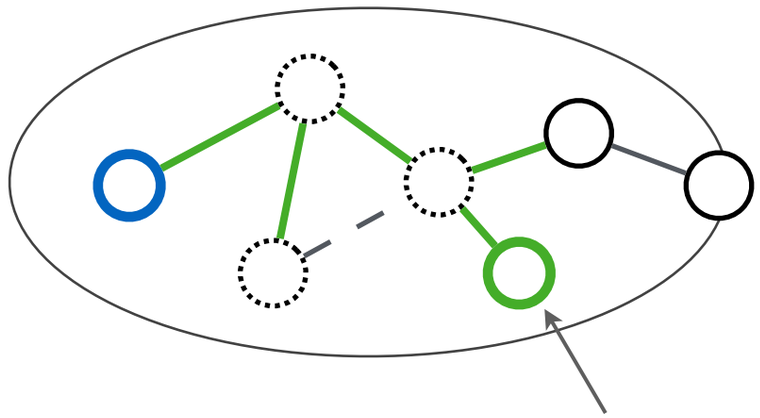
Problem: How to reach destinations in the same layer 2 network?

ARP	The Address Resolution Protocol discovers MACs of IPs Only works inside one layer 2 network
ARP tables	Hosts cache ARP replies in their local ARP table Entries will eventually expire
Manual	Alternatively, we can populate the ARP table statically

Our host performs an ARP request for the IP of the DNS server

src MAC: host H's MAC
dst MAC: ff:ff:ff:ff:ff:ff

ARP request:
Who has 192.168.1.2
tell 192.168.1.20

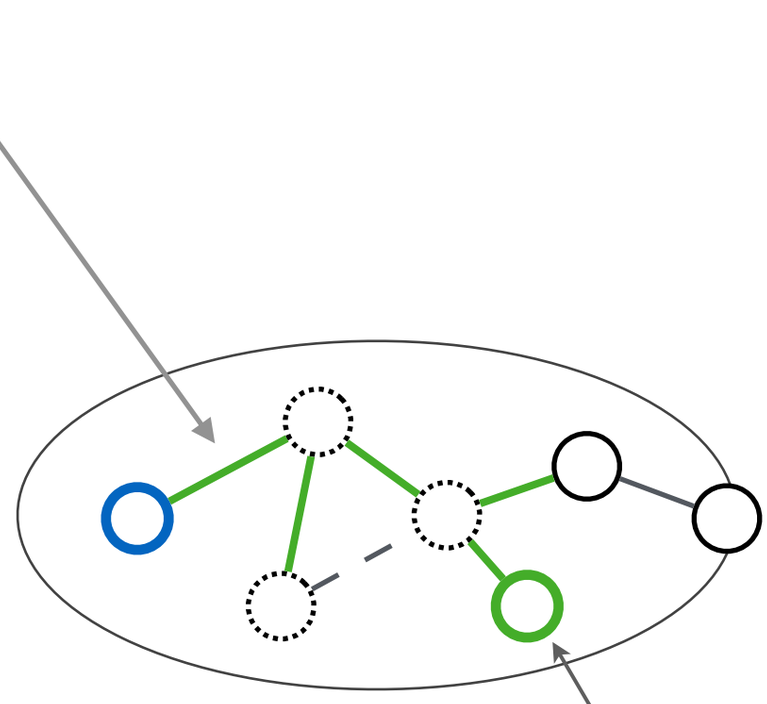


Our host performs an ARP request for the IP of the DNS server

src MAC: host H's MAC
dst MAC: ff:ff:ff:ff:ff:ff

ARP request:
Who has 192.168.1.2
tell 192.168.1.20

Broadcasted along the layer 2
Spanning Tree computed by the switches



DNS server
192.168.1.2

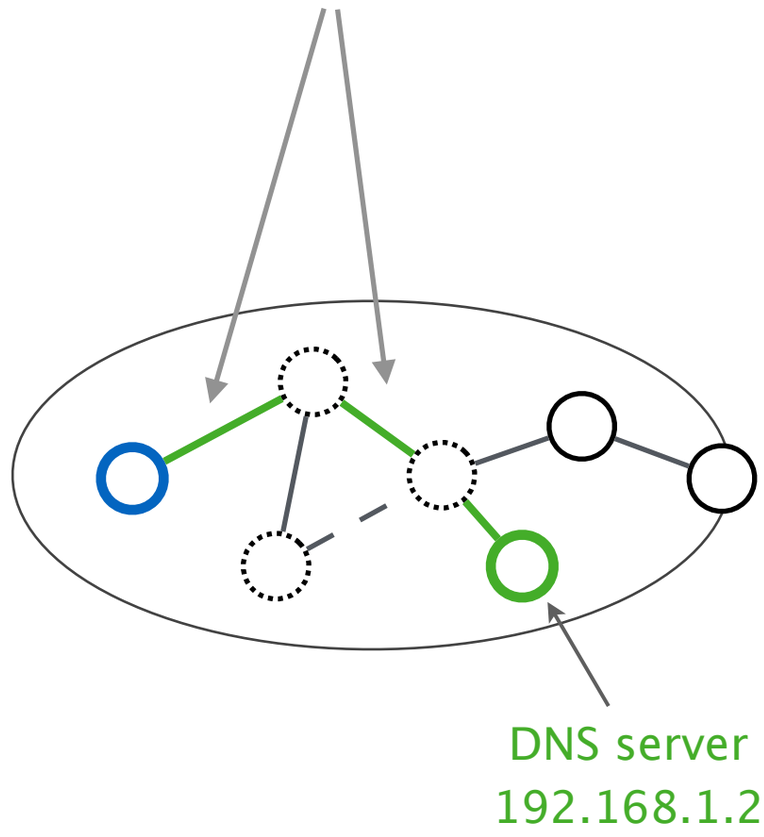
The DNS server unicasts its MAC address

src MAC: MAC of DNS

dst MAC: host H's MAC

ARP reply:
192.168.1.20 is at
<MAC of DNS server>

The switches have learned over which physical ports they can reach the MAC of H



We can finally perform our DNS query

(not shown in detail)

The DNS server might contact other name servers

depending on what is in its cache

We have seen two resolution strategies:

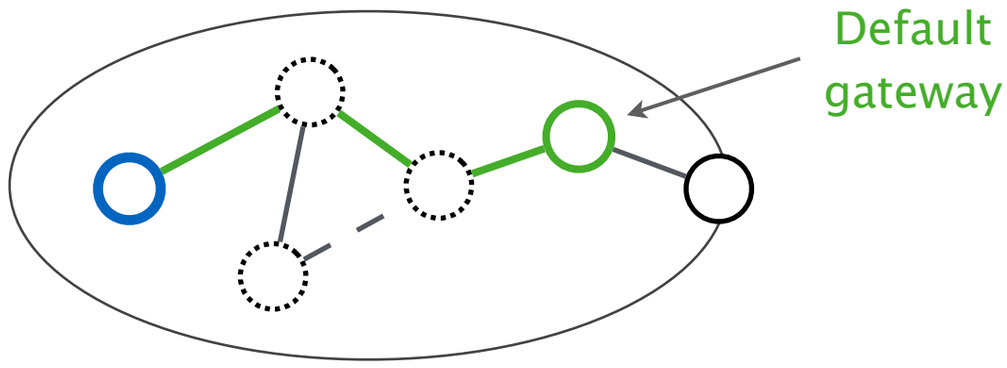
- *recursive*, by offloading it to other servers
- *iterative*, by iteratively querying the "next servers"

In our example, `my.video.com` has the IP: **5.6.7.8**

Problem: How to reach destinations outside of our local network?

Default gateway	We send the packets to our default gateway Known via DHCP (or statically configured)
Routers	The default gateway is normally a layer-3 router For example your "Internet box" at home
How to reach the gateway?	Already solved, we use **ARP** to find the MAC address Then forwarded over the layer 2 network

Our host can finally send
a first packet towards `my.video.com`



src MAC: host H's MAC
dst MAC: <MAC of gw>

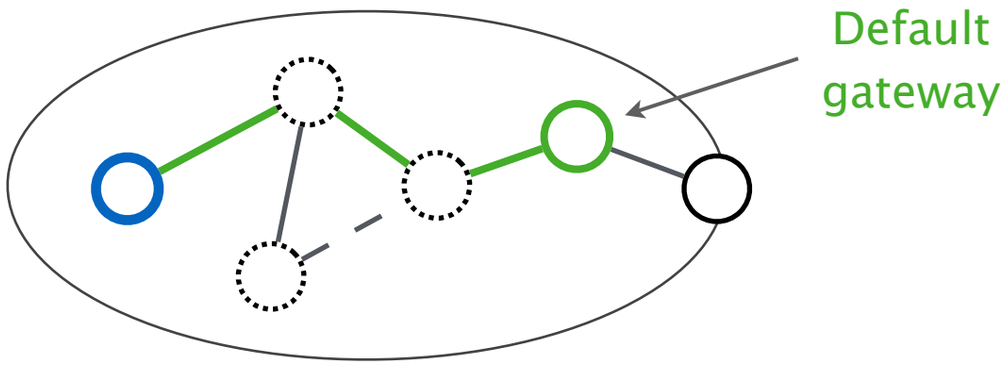
- - - - - - - - - -

src IP: 192.168.1.20
src port: 5555
dst IP: **5.6.7.8**
dst port: **554**

- - - - - - - - - -

TCP SYN

Our host can finally send a first packet towards `my.video.com`



src MAC: host H's MAC
dst MAC: <MAC of gw> → From ARP request for 192.168.1.1

src IP: 192.168.1.20
src port: 5555 → Randomly selected source port
dst IP: **5.6.7.8**
dst port: **443** → HTTP-based streaming (the default nowadays)

TCP SYN → TCP-based data transmission

Problem: How to reach external destinations using a private IP as source address?

NAT

Network Address Translation solves this problem

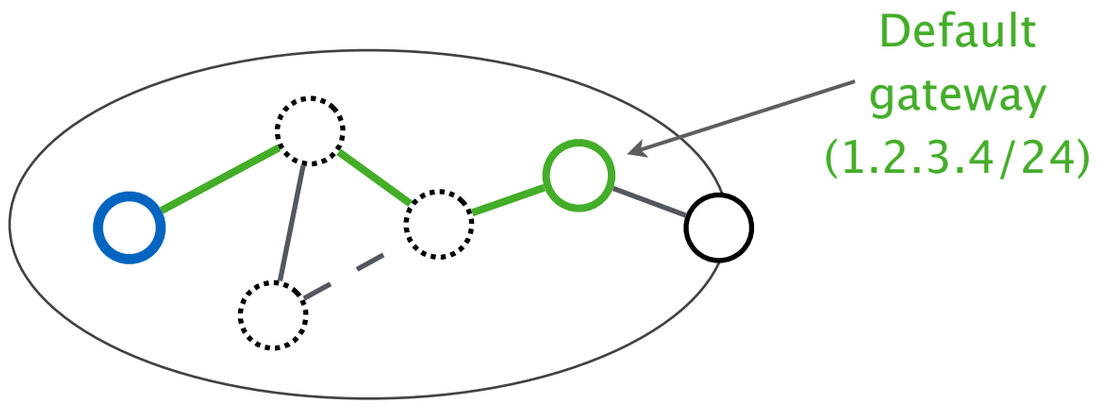
A single public IP is shared between hosts

Benefits

NAT has multiple benefits:

- "solution" to the IPv4 address depletion
- better privacy and anonymization
- hosts not reachable from the outside

Here, we'll consider that
the default gateway performs NAT



src MAC: host H's MAC
dst MAC: <MAC of gw>

src IP: **192.168.1.20**
src port: **5555**
dst IP: 5.6.7.8
dst port: 554

TCP SYN

192.168.1.20:5555

1.2.3.4:7744

Mapping stored
in NAT table

src MAC: <MAC of gw>
dst MAC: **???**

src IP: **1.2.3.4**
src port: **7744**
dst IP: 5.6.7.8
dst port: 554

TCP SYN

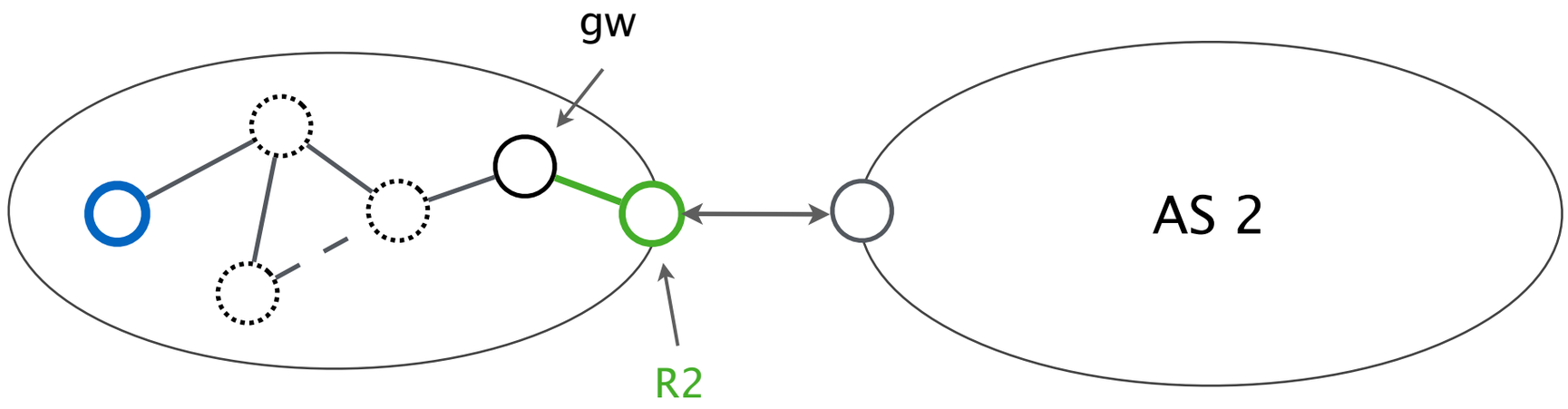
Problem: How to reach external destinations outside of our AS?

BGP	Inter-domain routing using the Border Gateway Protocol
	A path-vector protocol
Forwarding	Based on the best-matching prefix (longest match)
	One next hop for each prefix
iBGP & eBGP	Two versions of BGP to distribute routes
	eBGP distributes routes between ASes

Our packet is forwarded over multiple hops based on best-matching BGP routes

forwarding table:

prefix	next-hop
5.6.7.0/24	R2



src MAC: <MAC of gw>
dst MAC: <MAC of R2>

- - - - - - - - - - -

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

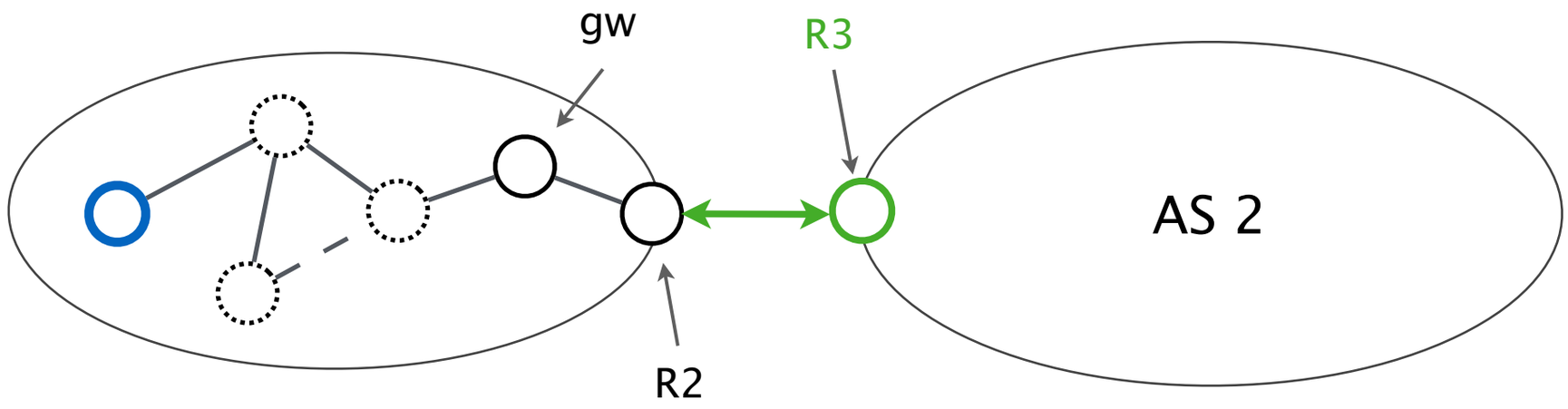
- - - - - - - - - - -

TCP SYN

Our packet is forwarded over multiple hops based on best-matching BGP routes

forwarding table:

prefix	next-hop
5.6.7.0/24	R3



src MAC: <MAC of gw>
dst MAC: <MAC of R2>

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

src MAC: <MAC of R2>
dst MAC: <MAC of R3>

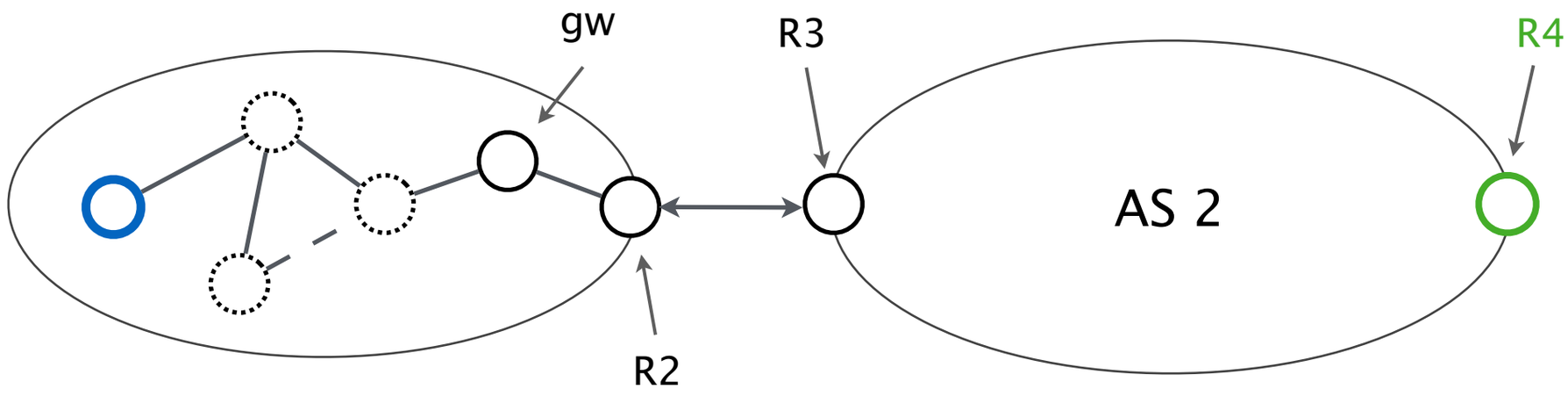
src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

Finally, we reach another AS

forwarding table:

prefix	next-hop
5.6.7.0/24	R4



src MAC: <MAC of gw>
dst MAC: <MAC of R2>

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

src MAC: <MAC of R2>
dst MAC: <MAC of R3>

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

src MAC: <MAC of R3>
dst MAC: ???

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

Problem: How to reach next hops which are not directly connected?

IGP

Forwarding information from Interior Gateway Protocols

Used for intra-domain routing

Two types

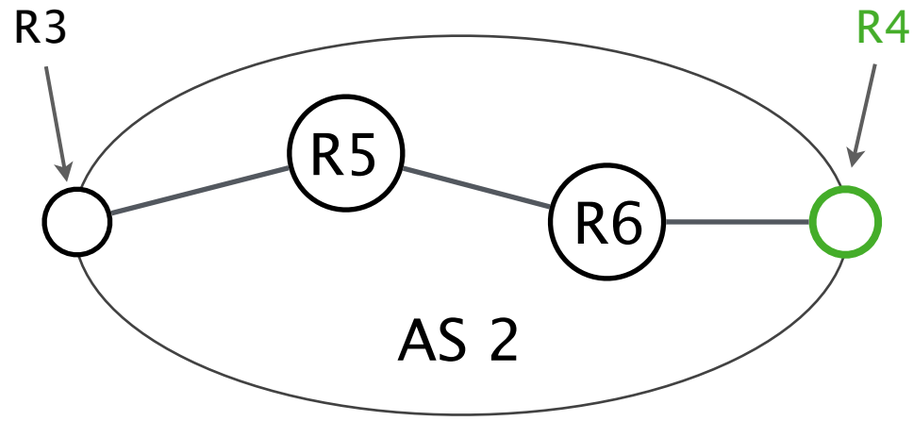
We saw two different types of protocols:

- link-state protocols (e.g., OSPF)
- distance-vector protocols (e.g., RIP)

Using the shortest IGP path, our packet reaches R4

forwarding table:

prefix	next-hop
5.6.7.0/24	R4



src MAC: <MAC of R3>
dst MAC: <MAC of R5>

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

src MAC: <MAC of R5>
dst MAC: <MAC of R6>

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

src MAC: <MAC of R6>
dst MAC: <MAC of R4>

src IP: 1.2.3.4
src port: 7744
dst IP: 5.6.7.8
dst port: 554

TCP SYN

Skipping a few similar steps,

our packet finally reaches the `my.video.com` server

src MAC: <MAC of R12>

dst MAC: <MAC of S>

- - - - - - - - - - -

src IP: 1.2.3.4

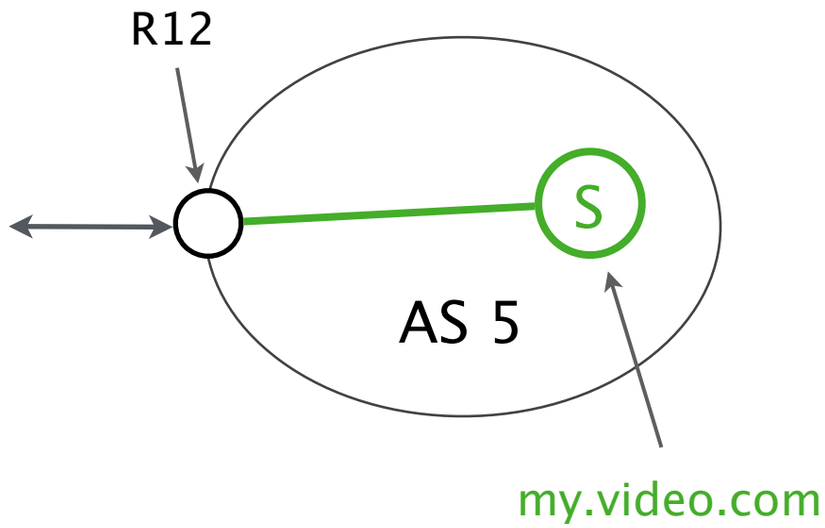
src port: 7744

dst IP: 5.6.7.8

dst port: 554

- - - - - - - - - - -

TCP SYN



Problem: How does the server know
to which application the packet belongs?

Dst port	The virtual ports identify the target application Completely different than physical ports on a device
Well-known	Ports in the range 0-1023 For example our video streaming port 554
Ephemeral	Most ports in the range 1024-65535 For example our source port(s): 7744 (5555 before NAT)

The server answers back with a SYN+ACK packet, which can take a different return path towards H

pkt created by S

src MAC: <MAC of S>
dst MAC: <MAC of R12>

- - - - - - - - - -

src IP: **5.6.7.8**
src port: **554**
dst IP: **1.2.3.4**
dst port: **7744**

- - - - - - - - - -

TCP SYN+ACK

The server answers back with a SYN+ACK packet, which can take a different return path towards H

pkt created by S

src MAC: <MAC of S>
dst MAC: <MAC of R12>

---

src IP: **5.6.7.8**
src port: **554**
dst IP: **1.2.3.4**
dst port: **7744**

TCP SYN+ACK

pkt received by H (after NAT)

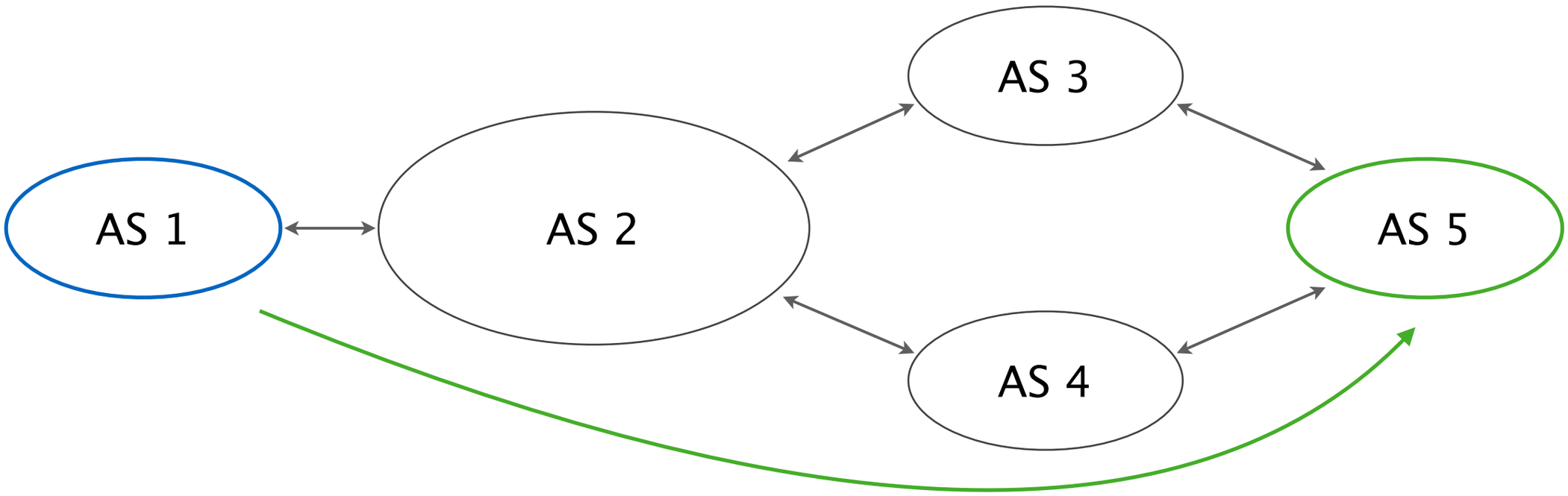
src MAC: <MAC of gw>
dst MAC: <MAC of H>

---

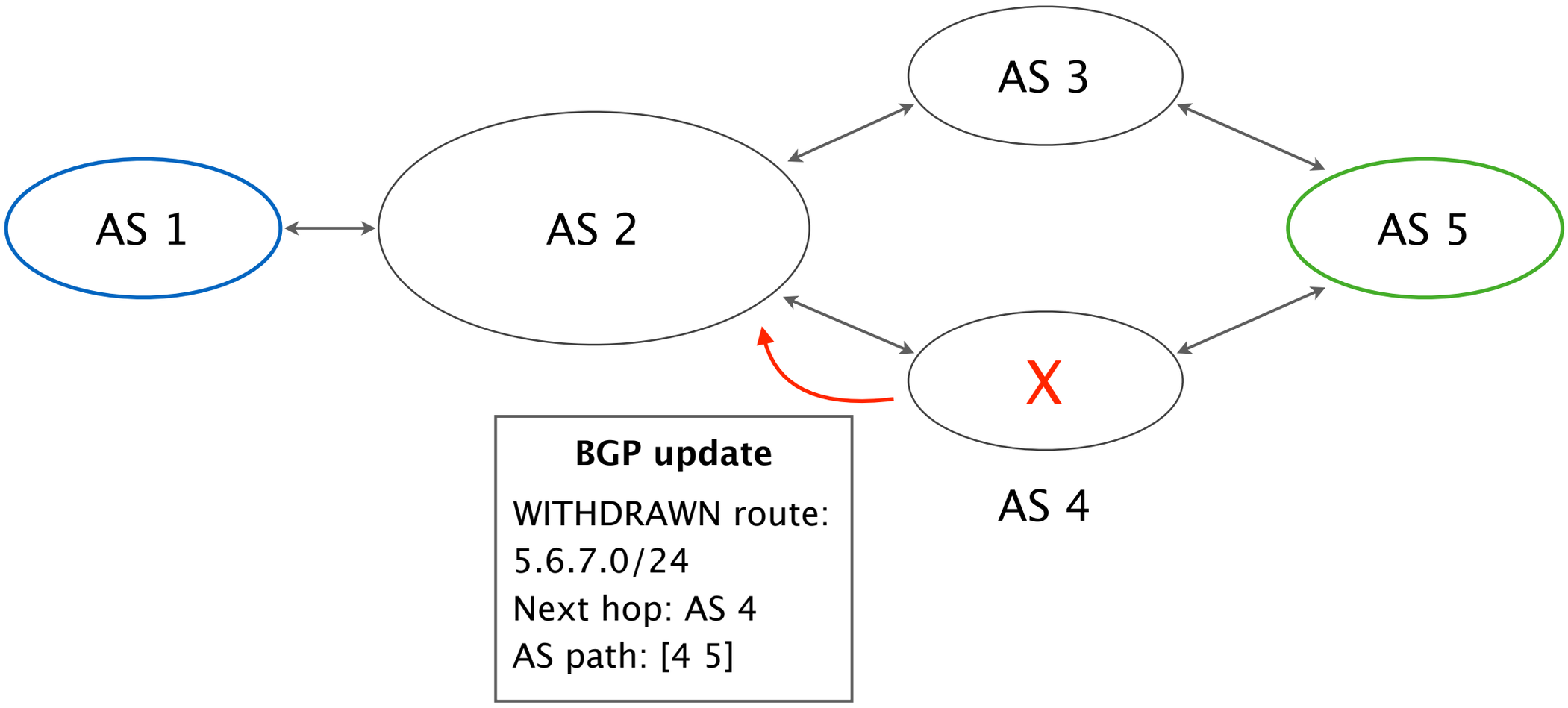
src IP: 5.6.7.8
src port: 554
dst IP: **192.168.1.20**
dst port: **5555**

TCP SYN+ACK

Our host is now able to watch a video on `my.video.com` using the AS path [1 2 4 5]



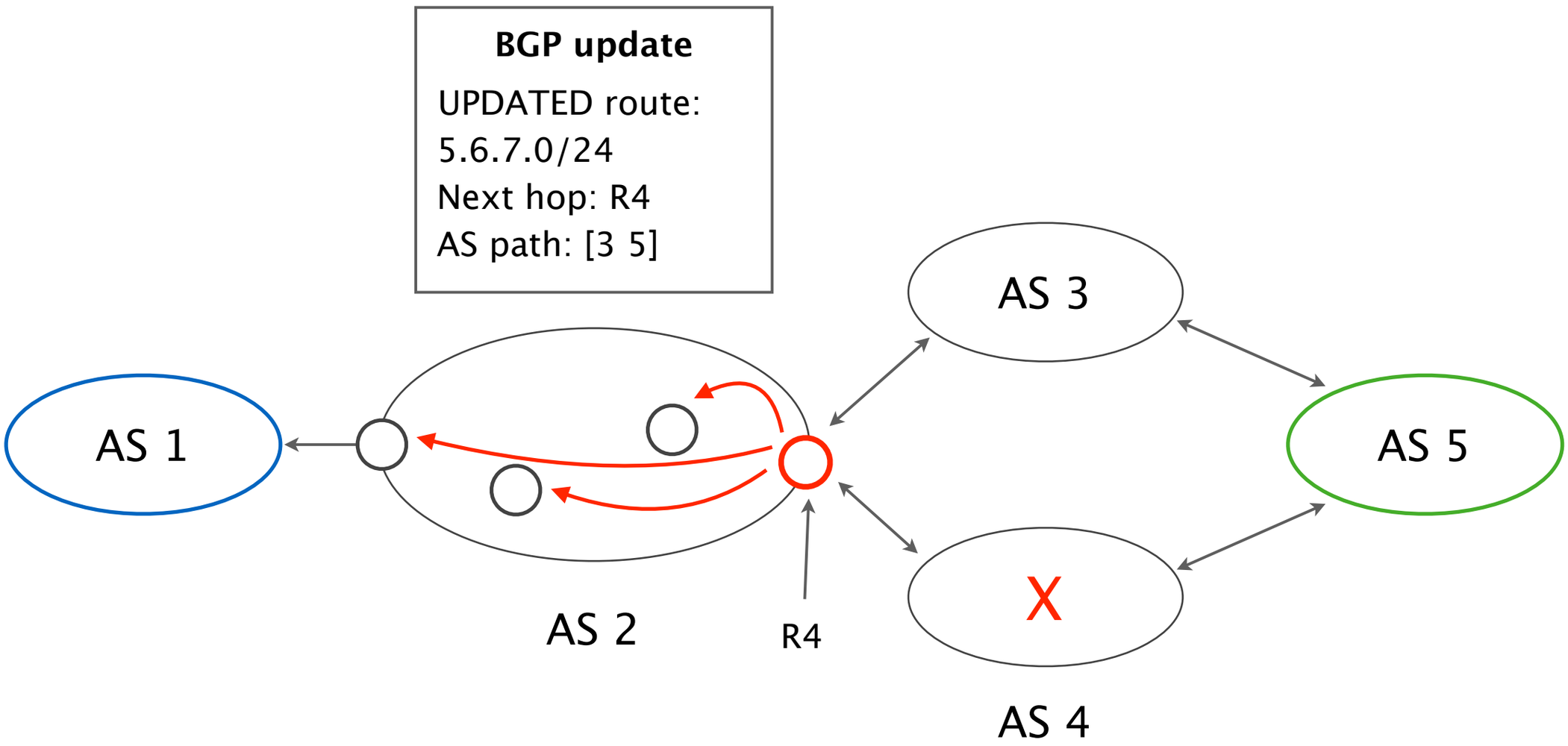
But suddenly AS 4 withdraws the route due to internal link failures



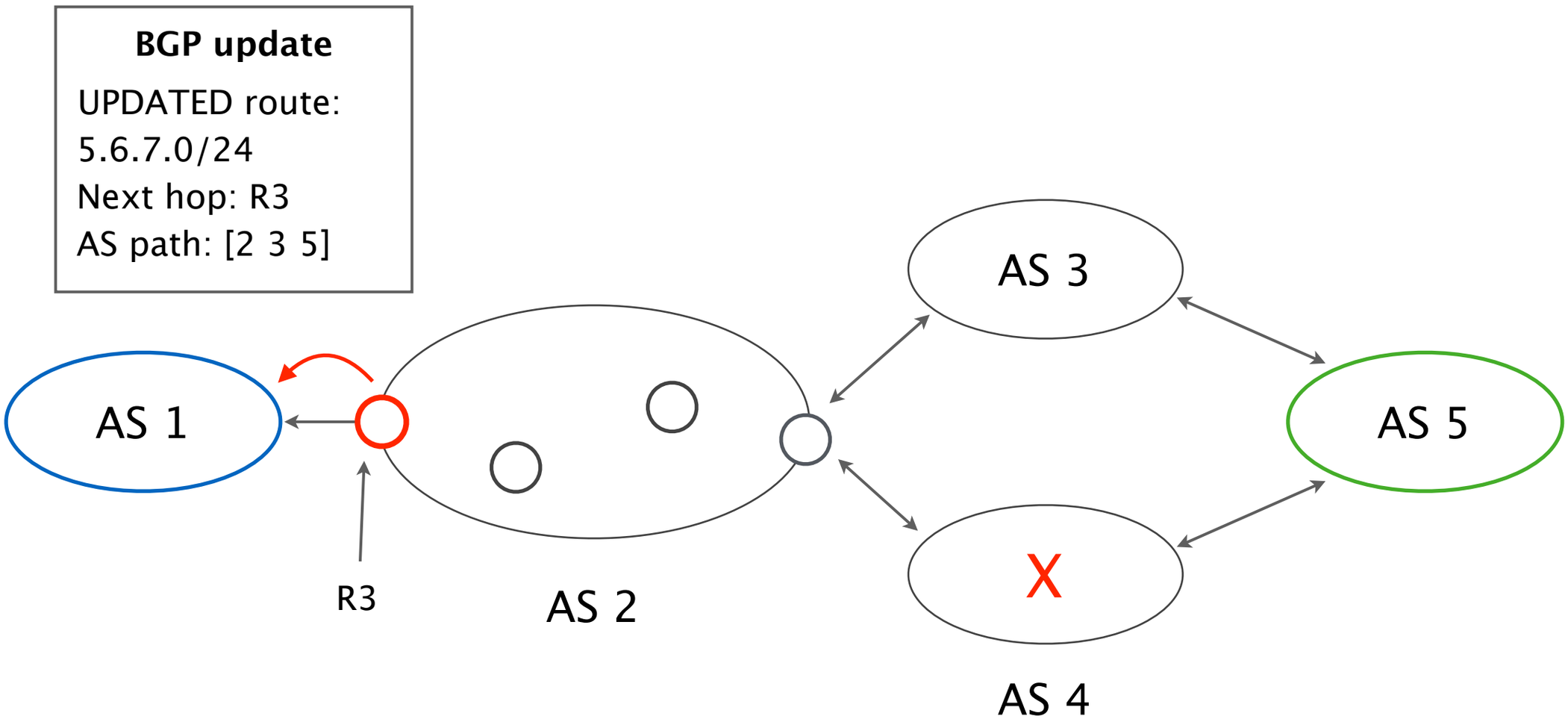
Problem: How to find new BGP routes after failures or BGP attribute changes?

BGP decision algorithm	The BGP decision algorithm finds a new best route
	Based on all currently available routes towards a prefix
Convergence	The new route is distributed over iBGP and eBGP
	Unexpected forwarding behavior during the convergence

Router R4 selects a new best route via AS 3 and distributes it via **iBGP**



Finally, the new route is advertised via **eBGP** to AS 1 which now reaches 5.6.7.0/24 via [1 2 3 5]



What happens to our packets during the convergence?

Some packets are dropped immediately

E.g., on the failed links or in a buffer

Other packets might be part of a forwarding loop

They are eventually dropped once the TTL value reaches 0

Problem: How to handle lost or reordered packets?

Reliable Transport

TCP is the most-used Reliable Transport protocol

UDP is an example for an unreliable protocol

Features

Reliable transport protocols provide:

- correctness, data is delivered in order & unmodified
- timeliness, minimized time until data is transferred
- efficiency, optimal use of bandwidth
- fairness, between concurrent flows

Transport Project

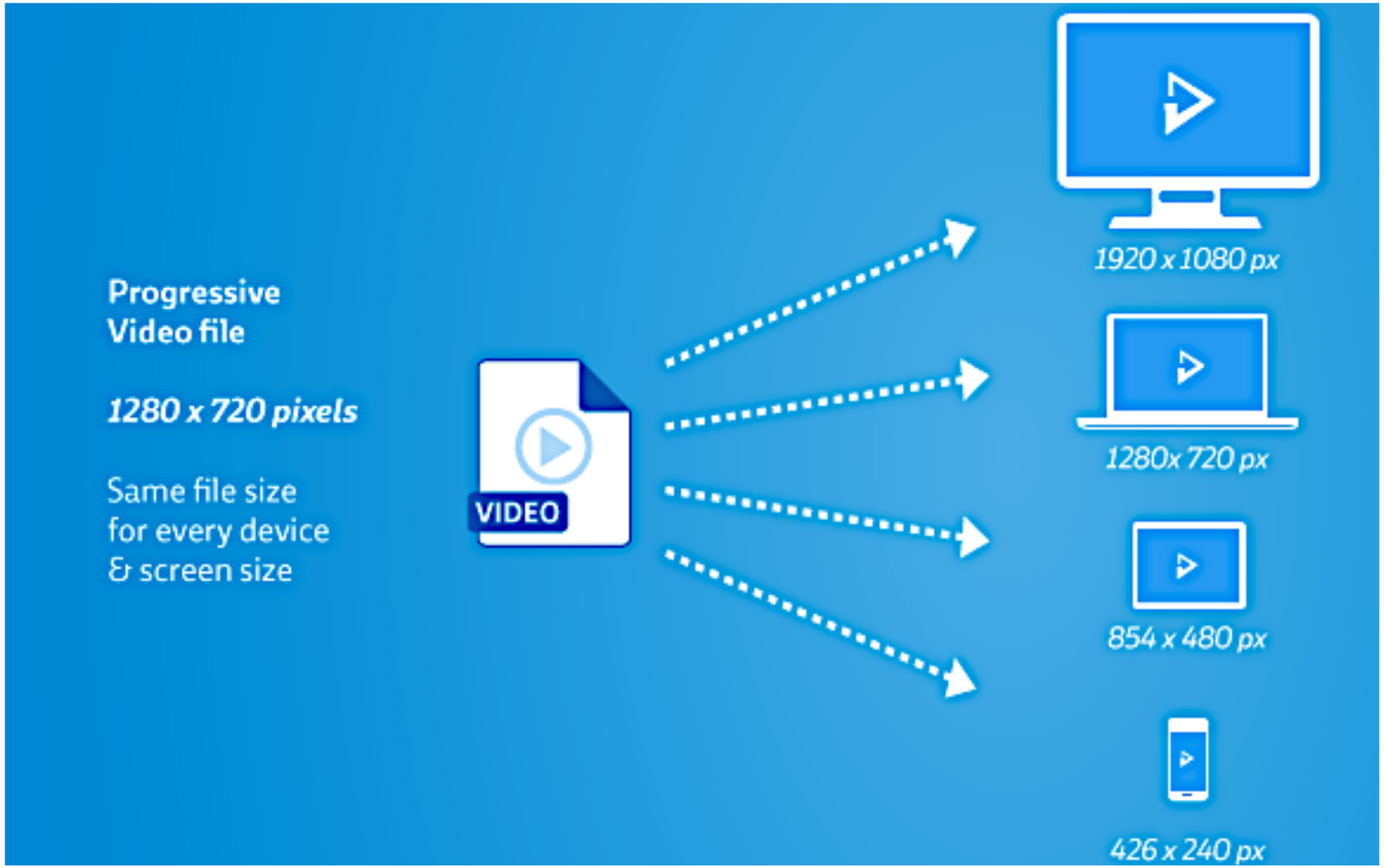
Your GBN sender and receiver provide some of these features

But for example, we do *not* provide fairness

Problem: How to guarantee the highest video quality?

Without seeing this …

A naive approach: one-size-fits-all



[bitmovin.com]

The three steps behind most contemporary solutions

- Encode video in multiple bitrates
- Replicate using a content delivery network
- Video player picks bitrate adaptively
  - Estimate connection's available bandwidth
  - Pick a bitrate ≤ available bandwidth

Encoding

Replication

Adaptation

Encoding

Replication

Adaptation

Video size: 1920 x 1080 px

Screen size: 1920 x 1080 px

Video size: 1280x 720 px

Screen size: 1280x 720 px

Video size: 854 x 480 px

Screen size: 854 x 480 px

Video size: 426 x 240 px

Screen size: 426 x 240 px

Fast Internet

1920 x 1080 px

Screen size: 1920 x 1080 px
With ***fast*** internet.

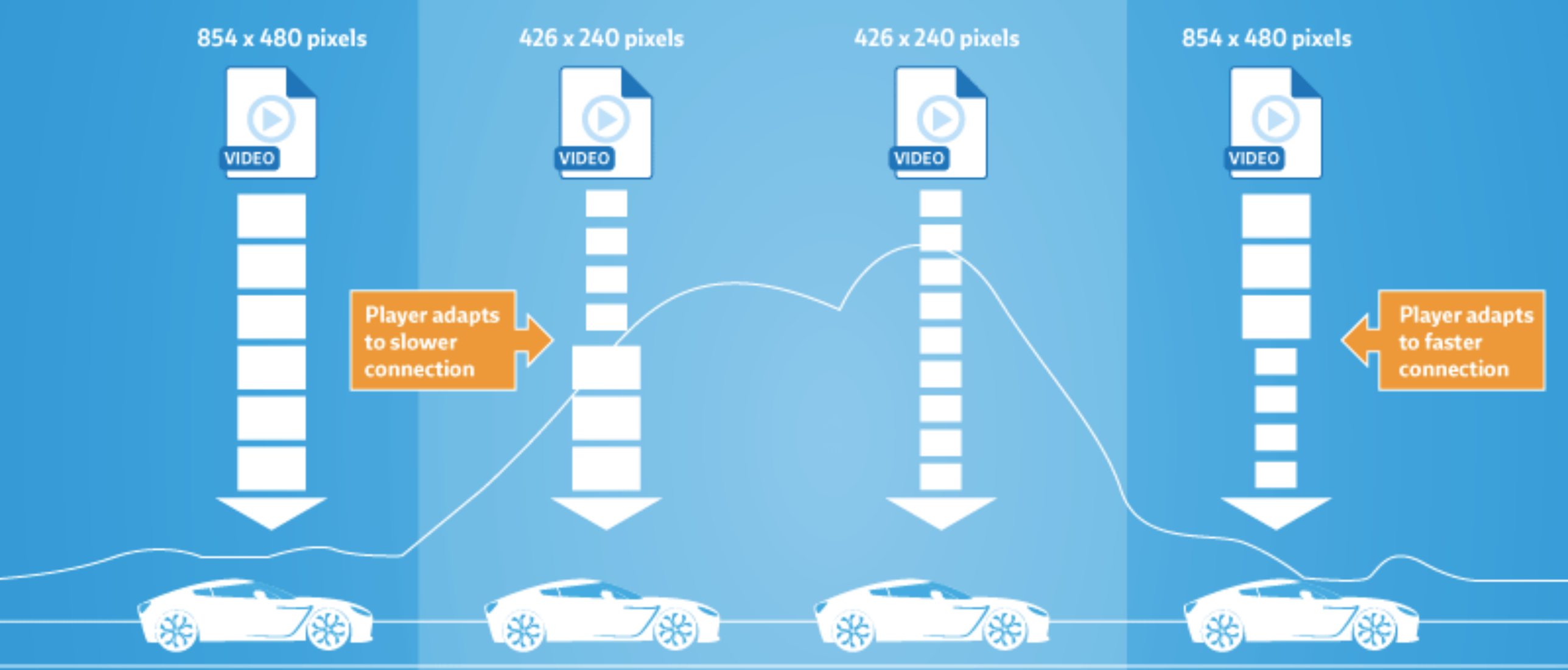
Video plays at **high quality**
1920 x 1080 px with **no buffering**

1280x 720 px

Screen size: 1920 x 1080 px
With ***slower*** internet.

Video plays at **medium quality**
1280x 720 px with **no buffering**

Slow Internet



Normal connection:
The Player downloads the best quality video

Poor connection:
The Player changes to downloading a smaller, faster video file

Normal connection:
The Player returns to the maximum quality video file

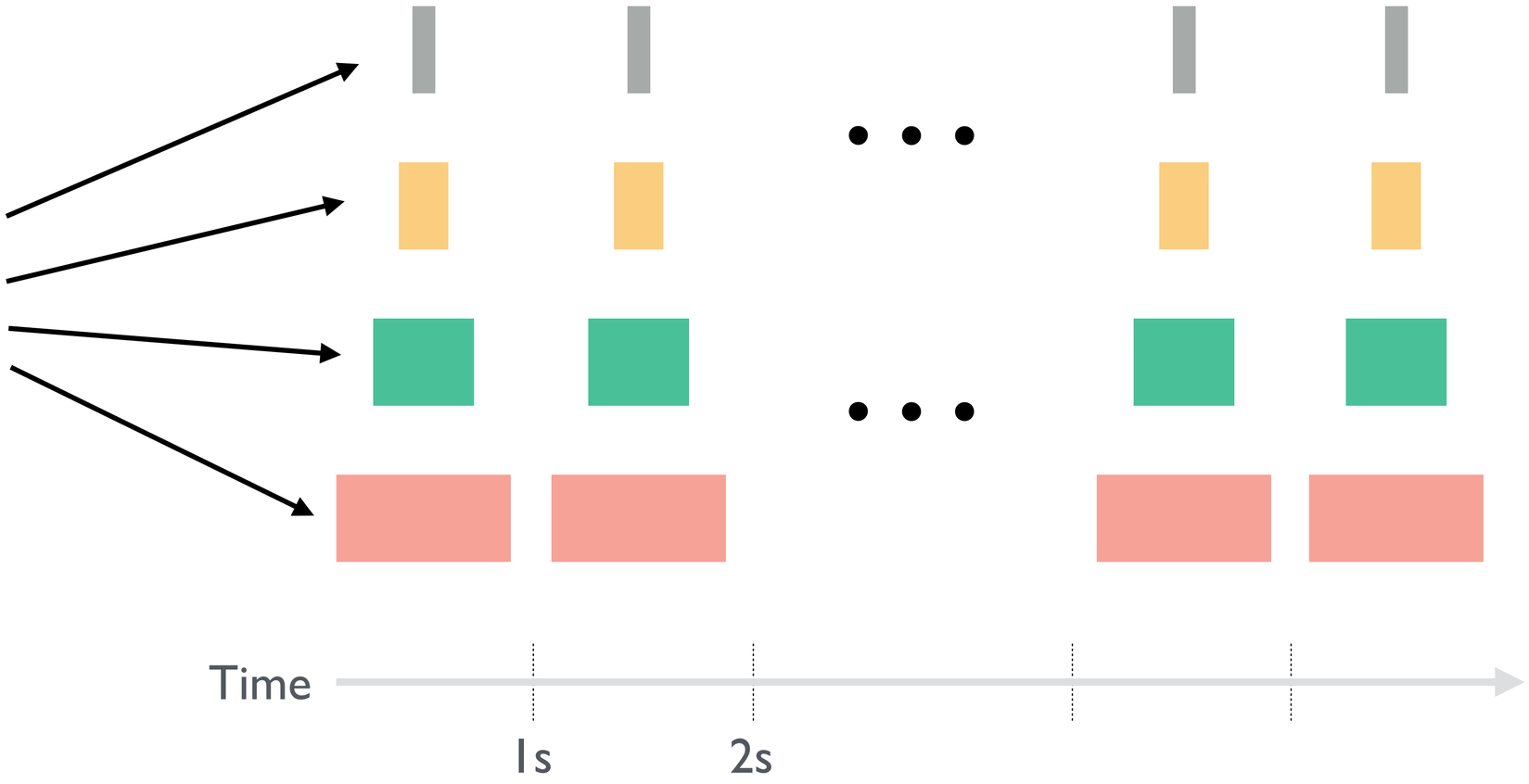
Simple solution for encoding: use a “bitrate ladders”

Bitrate (kbps)	Resolution
235	320x240
375	384x288
560	512x384
750	512x384
1050	640x480
1750	720x480
2350	1280x720
3000	1280x720
4300	1920x1080
5800	1920x1080

[netflix.com]

Your player download “chunks” of video at different bitrates

[netflix.com]



Depending on your network connectivity, your player fetches chunks of different qualities

[netflix.com]

Your player gets metadata about chunks via "Manifest"

```
<?xml version="1.0" encoding="UTF-8"?>
<MPD xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns="urn:mpeg:DASH:schema:MPD:2011"
    xsi:schemaLocation="urn:mpeg:DASH:schema:MPD:2011"
    profiles="urn:mpeg:dash:profile:isoff-main:2011"
    type="static"
    mediaPresentationDuration="PT0H9M56.46S"
    minBufferTime="PT15.0S">
  <BaseURL>http://witestlab.poly.edu/~ffund/video/2s_480p_only/</BaseURL>
  <Period start="PT0S">
        <AdaptationSet bitstreamSwitching="true">
     <Representation id="0" codecs="avc1" mimeType="video/mp4"
        width="480" height="360" startWithSAP="1" bandwidth="101492">
       <SegmentBase>
         <Initialization sourceURL="bunny_2s_100kbit/bunny_100kbit.mp4"/>
       </SegmentBase>
       <SegmentList duration="2">
         <SegmentURL media="bunny_2s_100kbit/bunny_2s1.m4s"/>
         <SegmentURL media="bunny_2s_100kbit/bunny_2s2.m4s"/>
         <SegmentURL media="bunny_2s_100kbit/bunny_2s3.m4s"/>
         <SegmentURL media="bunny_2s_100kbit/bunny_2s4.m4s"/>
         <SegmentURL media="bunny_2s_100kbit/bunny_2s5.m4s"/>
         <SegmentURL media="bunny_2s_100kbit/bunny_2s6.m4s"/>
```

[witestlab.poly.edu]

Encoding

Replication

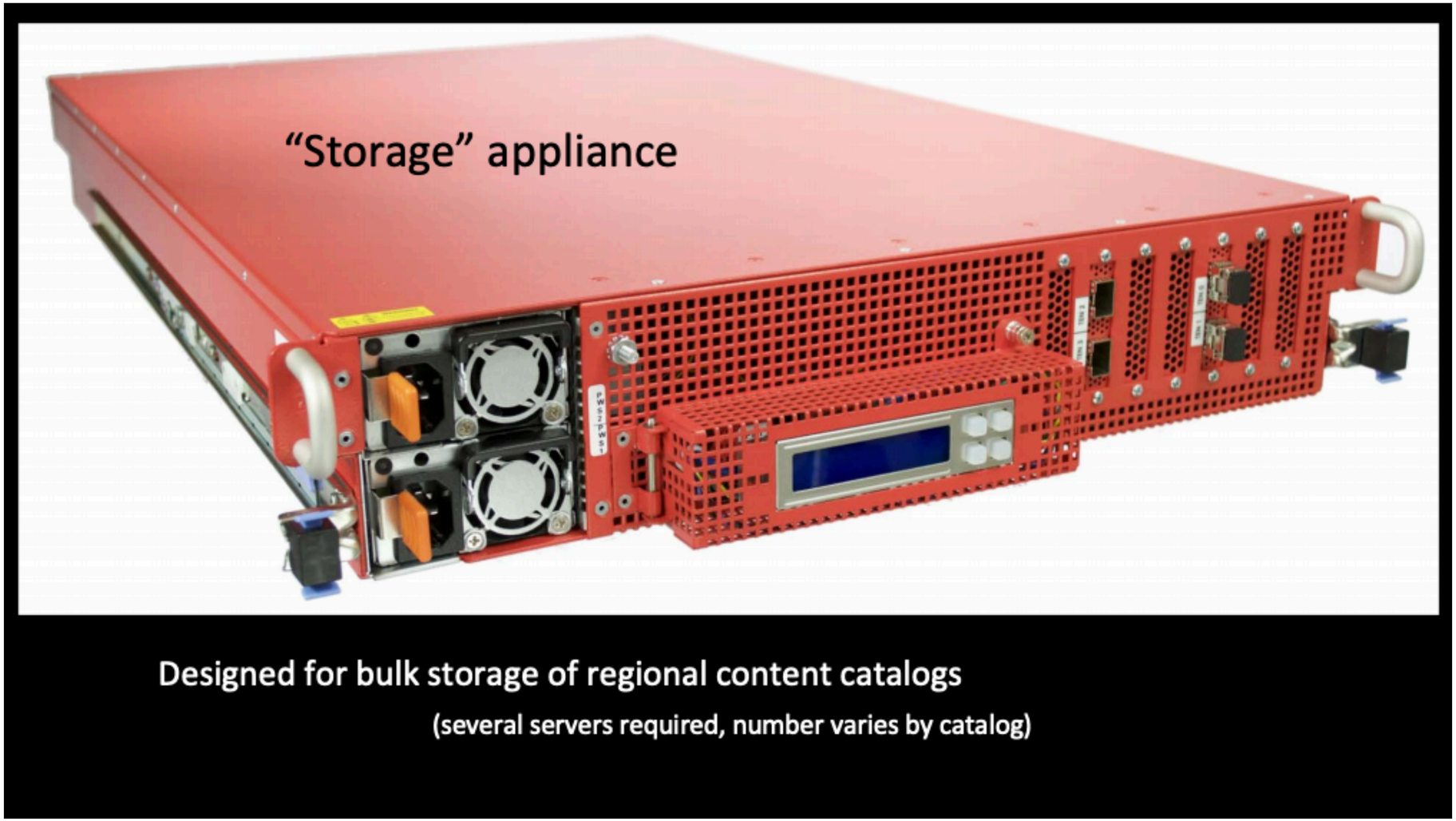
Adaptation

NETFLIX

Open Connect: Starting from a Greenfield (a mostly Layer 0 talk)

Dave Temkin

06/01/2015



[more-ip-event.net]

Storage Appliances

Storage appliances are 2U servers that are focused on reliable dense storage and cost effective throughput. This appliance is used to hold the Netflix catalog in many IX locations around the world and embedded at our larger ISP partner locations.

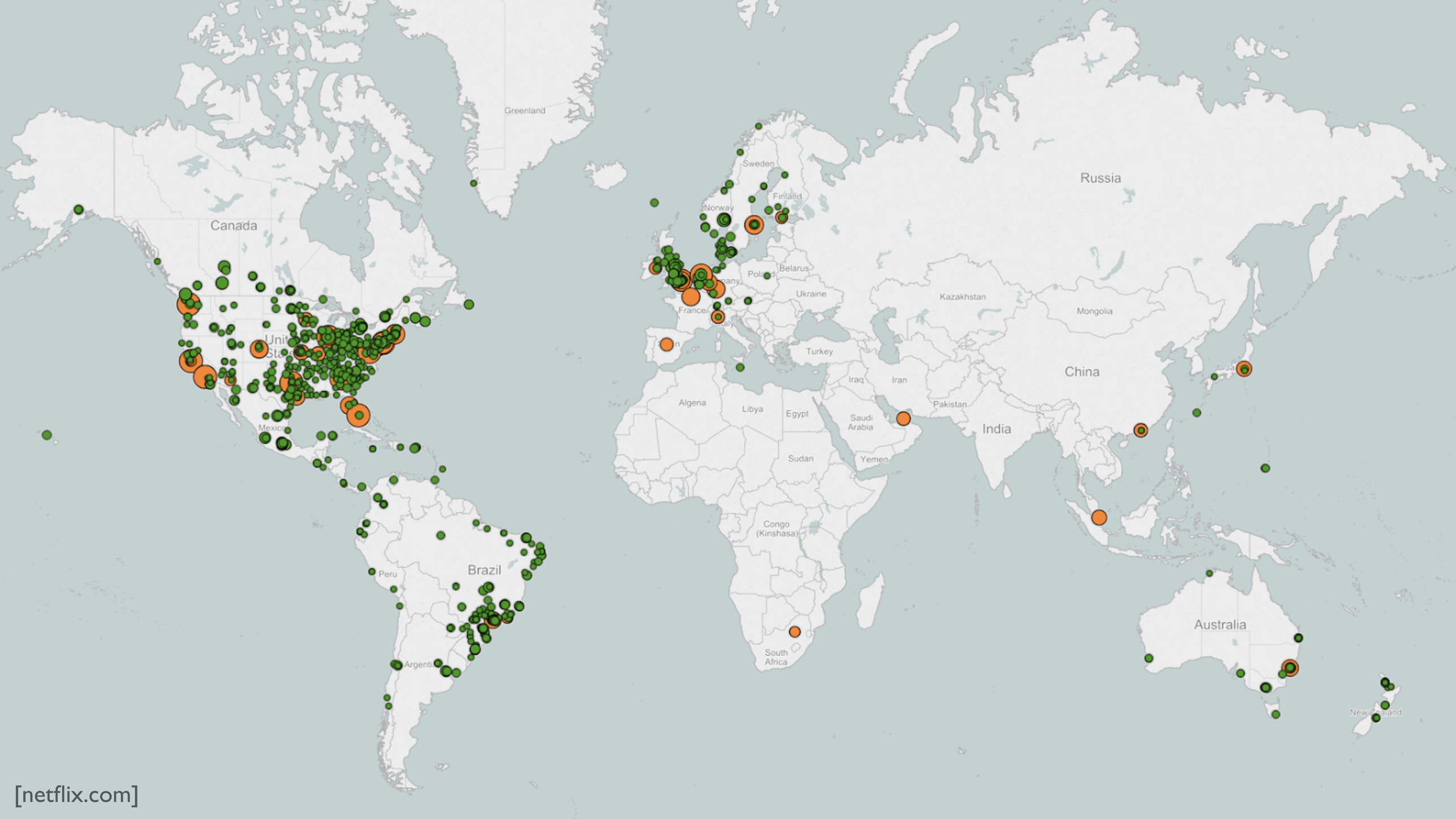
Storage appliance focus areas

- Large storage capacity
- 2U for rack efficiency (no deeper than 29 inches)
- Enough low cost NAND to reach 10GB/s of throughput (<0.3 DWPD)
- Network flexibility to connect at 6x10GE LAG or 1x100GE
- 2 and 4 post racking
- AC or DC power
- Single processor

Storage appliance high-level specifications

Option	Vendors
Chassis	Sanmina
Motherboard	Supermicro
Processor	Intel
Memory	Micron
Hard Drive	HGST
Solid State Drive	Micron, Toshiba
Network Controller	Chelsio
Power draw operational (peak)	~500W
Power Supply Unit	Redundant Hot Swap AC/DC
Operational throughput	~36Gbps
Raw storage capacity	~288 TB

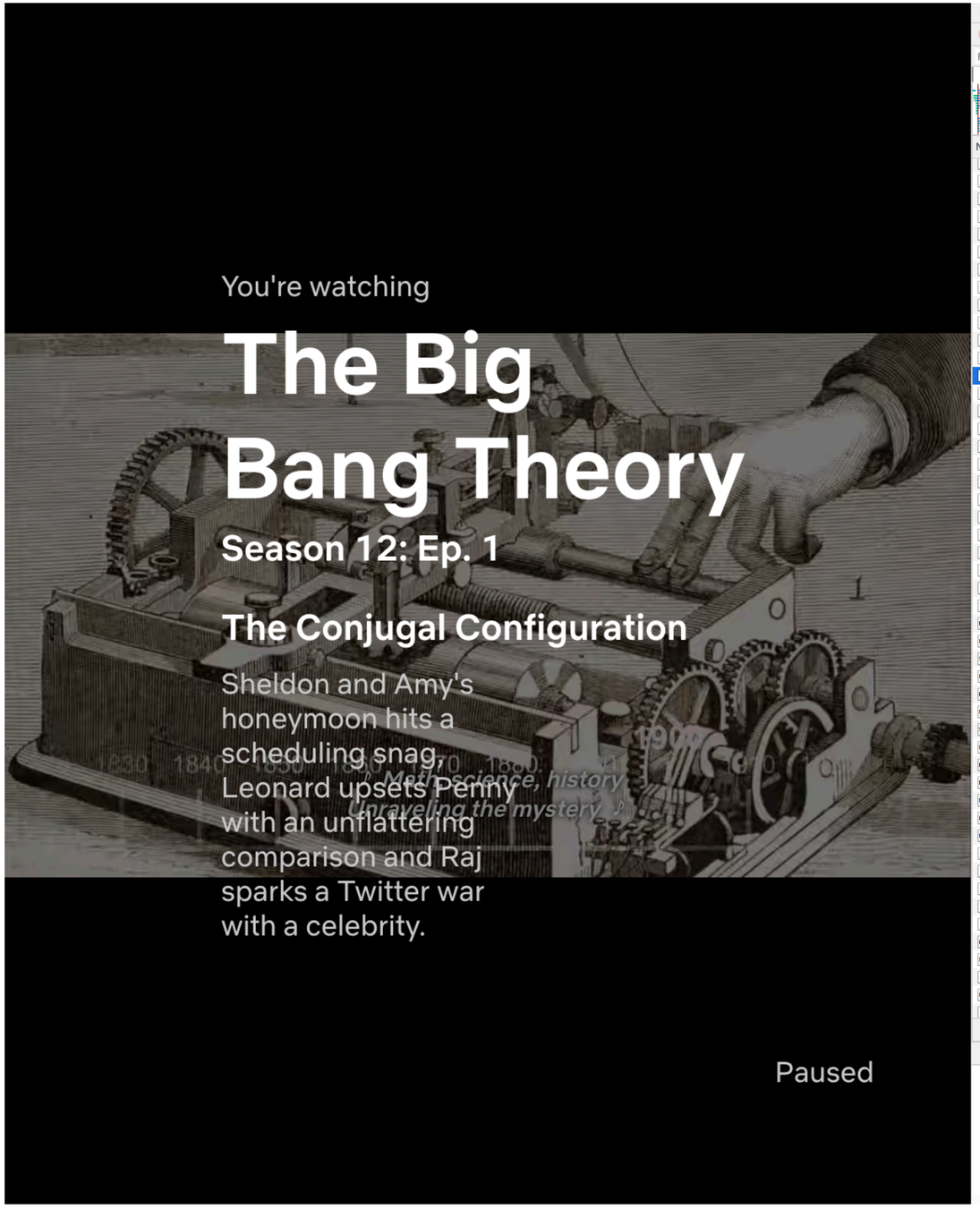
[openconnect.netflix.com]

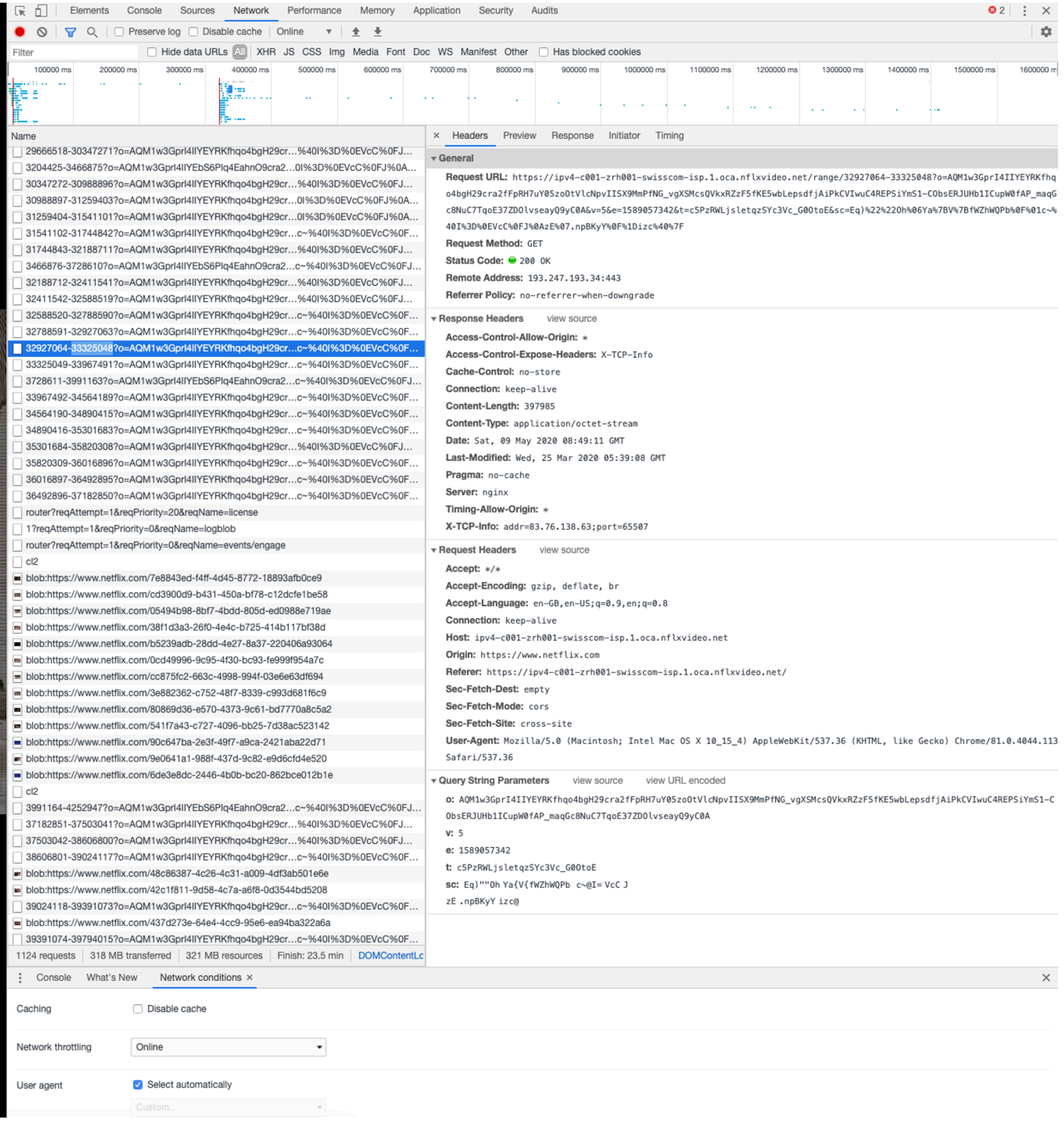

Greenland
Canada
Sweden
Finland
Norway
Russia
Belarus
Ukraine
Kazakhstan
Mongolia
China
Turkey
Iraq
Iran
Pakistan
India
Algeria
Libya
Egypt
Saudi Arabia
Sudan
Yemen
Congo (Kinshasa)
South Africa
Brazil
Peru
Australia

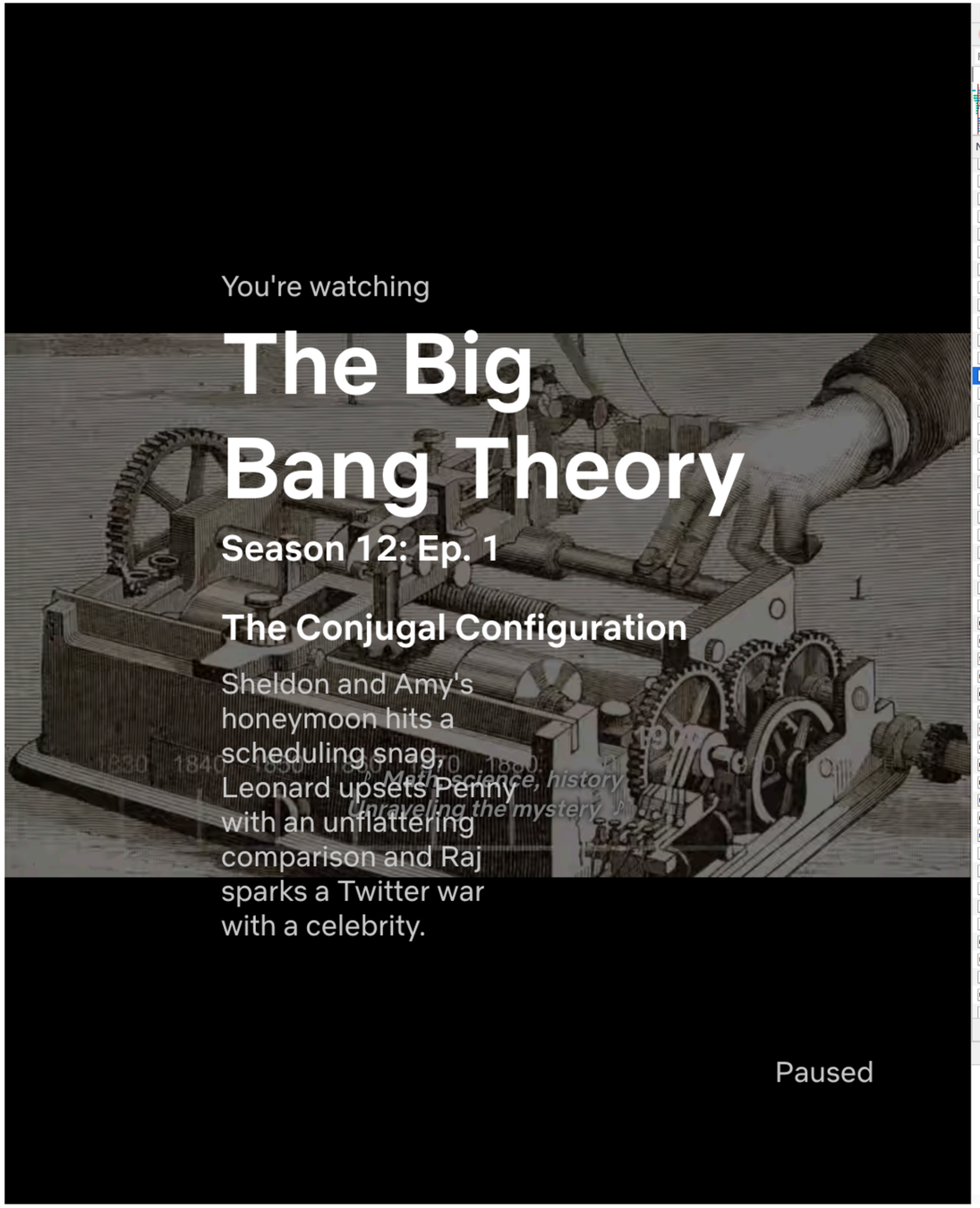
[netflix.com]

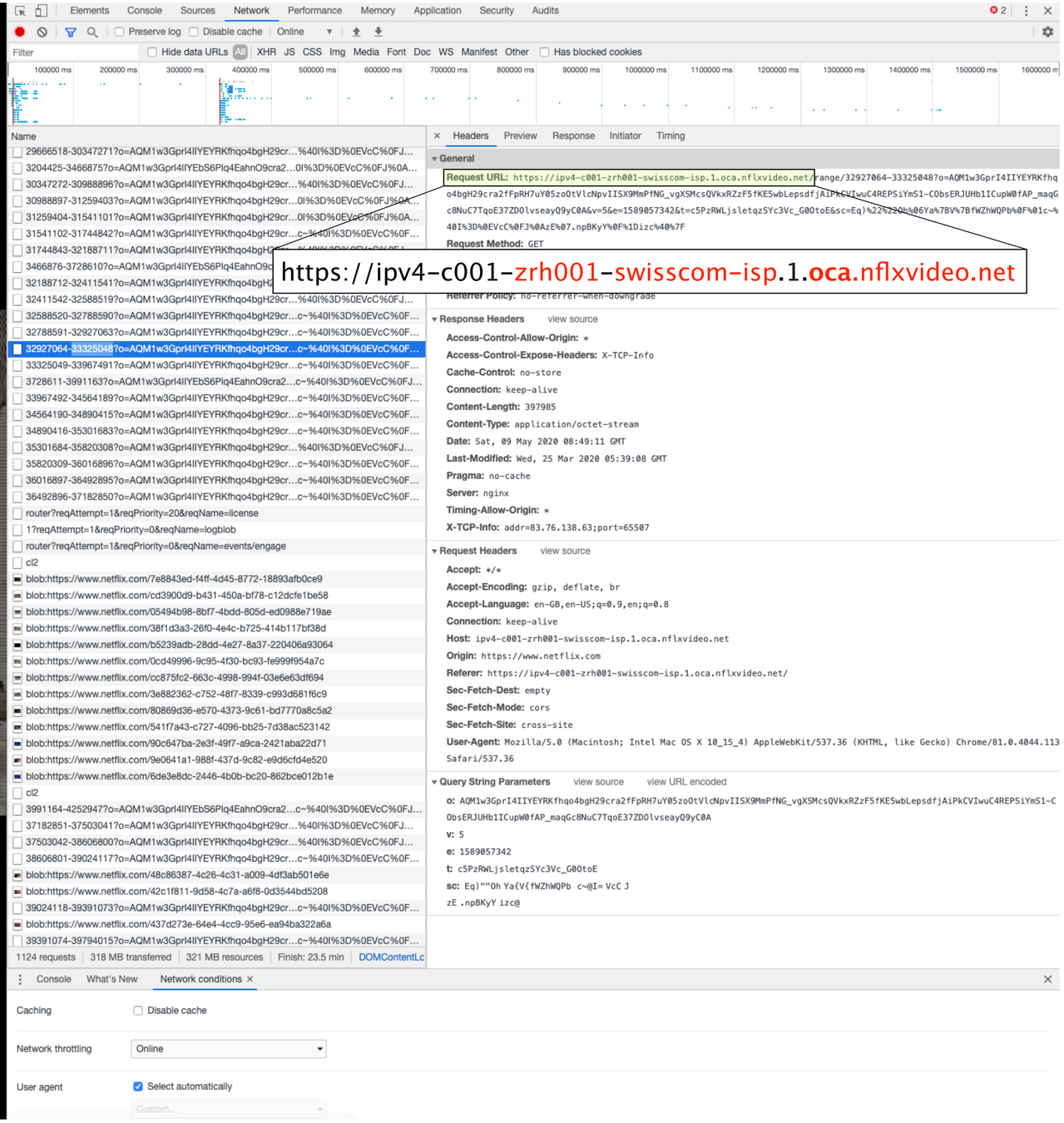

WEST COAST US
DAREDEVIL
DAREDEVIL
DAREDEVIL
AUSTRALIA
DAREDEVIL

[netflix.com]

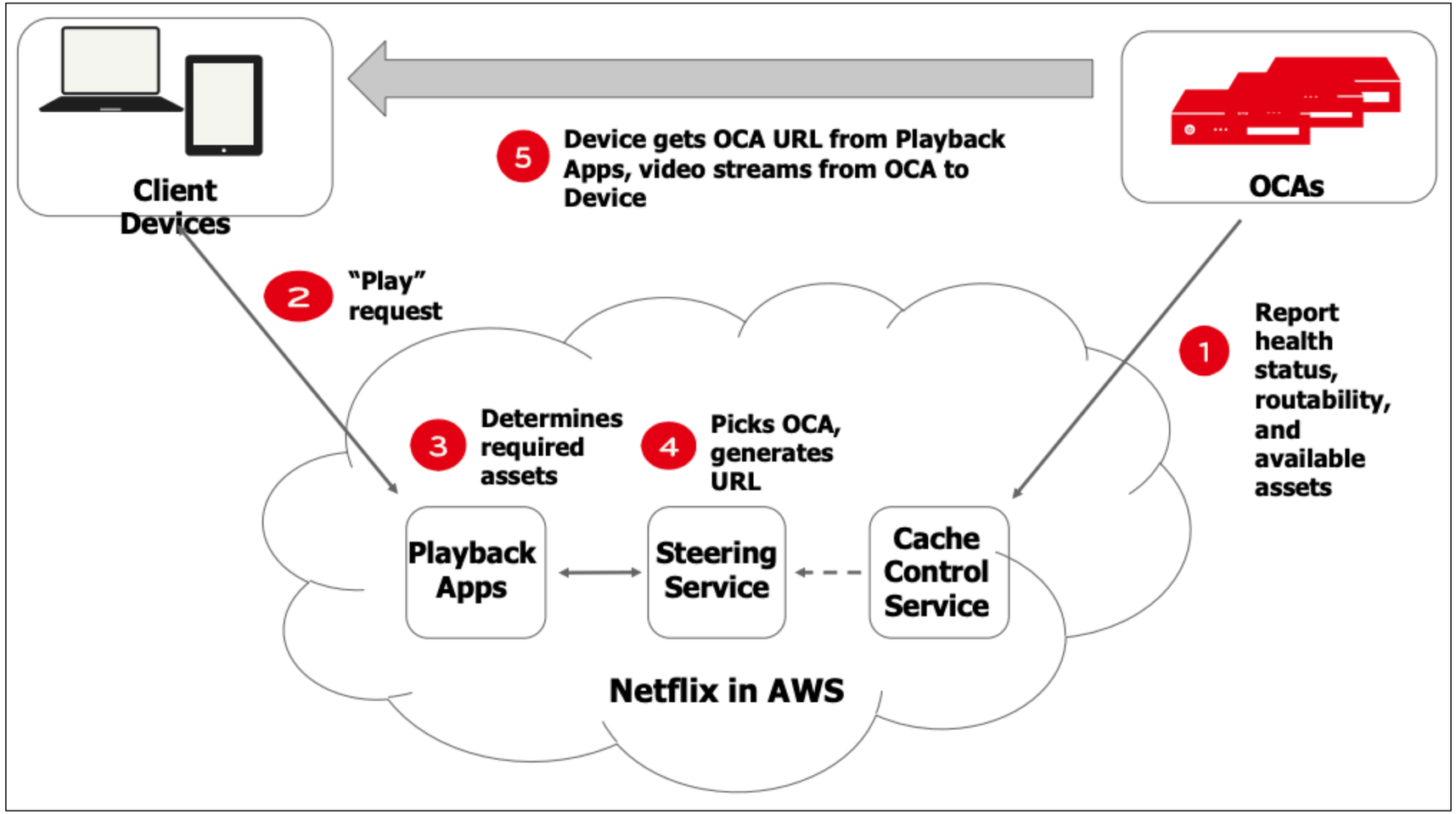

You're watching
The Big Bang Theory
Season 12: Ep. 1
The Conjugal Configuration
Sheldon and Amy's honeymoon hits a scheduling snag, Leonard upsets Penny with an unflattering comparison and Raj sparks a Twitter war with a celebrity.
Paused


Elements Console Sources Network Performance Memory Application Security Audits
Preserve log Disable cache Online
Filter Hide data URLs All XHR JS CSS Img Media Font Doc WS Manifest Other Has blocked cookies
Name
Headers Preview Response Initiator Timing
General
Request URL: https://ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net/range/32927064-33325048?o=AQM1w3GprI4IIYEYRKfhqo4bgH29cra2fFpRH7uY05zoOtVlcNpvIISX9MmPfNG_vgXSMcsQVkxRZzF5fKE5wbLepsdfjAiPkCVIwuC4REPSiYmS1-CObsERJUHb1ICupW0fAP_maqGc8NuC7TqoE37ZDOlvseayQ9yC0A&v=5&e=1589057342&t=c5PzRWLjsletqzSYc3Vc_G0OtoE&sc=Eq)%22%22Oh%06Ya%7BV%7BfWZhWQPb%0F%01c~%40I%3D%0EVcC%0FJ%0AzE%07.npBKyY%0F%1Dizc%40%7F
Request Method: GET
Status Code: 200 OK
Remote Address: 193.247.193.34:443
Referrer Policy: no-referrer-when-downgrade
Response Headers view source
Access-Control-Allow-Origin: *
Access-Control-Expose-Headers: X-TCP-Info
Cache-Control: no-store
Connection: keep-alive
Content-Length: 397985
Content-Type: application/octet-stream
Date: Sat, 09 May 2020 08:49:11 GMT
Last-Modified: Wed, 25 Mar 2020 05:39:08 GMT
Pragma: no-cache
Server: nginx
Timing-Allow-Origin: *
X-TCP-Info: addr=83.76.138.63;port=65507
Request Headers view source
Accept: */*
Accept-Encoding: gzip, deflate, br
Accept-Language: en-GB,en-US;q=0.9,en;q=0.8
Connection: keep-alive
Host: ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net
Origin: https://www.netflix.com
Referer: https://ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net/
Sec-Fetch-Dest: empty
Sec-Fetch-Mode: cors
Sec-Fetch-Site: cross-site
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.113 Safari/537.36
Query String Parameters view source view URL encoded
o: AQM1w3GprI4IIYEYRKfhqo4bgH29cra2fFpRH7uY05zoOtVlcNpvIISX9MmPfNG_vgXSMcsQVkxRZzF5fKE5wbLepsdfjAiPkCVIwuC4REPSiYmS1-CObsERJUHb1ICupW0fAP_maqGc8NuC7TqoE37ZDOlvseayQ9yC0A
v: 5
e: 1589057342
t: c5PzRWLjsletqzSYc3Vc_G0OtoE
sc: Eq)""Oh Ya{V{fWZhWQPb c~@I= VcC J
zE .npBKyY izc@
1124 requests 318 MB transferred 321 MB resources Finish: 23.5 min DOMContentLc
Console What's New Network conditions
Caching Disable cache
Network throttling Online
User agent Select automatically


You're watching
The Big Bang Theory
Season 12: Ep. 1
The Conjugal Configuration
Sheldon and Amy's honeymoon hits a scheduling snag, Leonard upsets Penny with an unflattering comparison and Raj sparks a Twitter war with a celebrity.
Paused


Elements Console Sources Network Performance Memory Application Security Audits
Preserve log Disable cache Online
Filter Hide data URLs All XHR JS CSS Img Media Font Doc WS Manifest Other Has blocked cookies
Name
Headers Preview Response Initiator Timing
General
Request URL: https://ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net/range/32927064-33325048?o=AQM1w3GprI4IIYEYRKfhqo4bgH29cra2fFpRH7uY05zoOtVlcNpvIISX9MmPfNG_vgXSMcsQVkxRZzF5fKE5wbLepsdfjAiPkCVIwuC4REPSiYmS1-CObsERJUHb1ICupW0fAP_maqGc8NuC7TqoE37ZDOlvseayQ9yC0A&v=5&e=1589057342&t=c5PzRWLjsletqzSyc3Vc_G0OtoE&sc=Eq)%22%22Oh%06Ya%7BV%7BfWZhWQPb%0F%01c~%40I%3D%0EVcC%0FJ%0AzE%07.npBKyY%0F%1Dizc%40%7F
Request Method: GET
https://ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net
Response Headers view source
Access-Control-Allow-Origin: *
Access-Control-Expose-Headers: X-TCP-Info
Cache-Control: no-store
Connection: keep-alive
Content-Length: 397985
Content-Type: application/octet-stream
Date: Sat, 09 May 2020 08:49:11 GMT
Last-Modified: Wed, 25 Mar 2020 05:39:08 GMT
Pragma: no-cache
Server: nginx
Timing-Allow-Origin: *
X-TCP-Info: addr=83.76.138.63;port=65507
Request Headers view source
Accept: */*
Accept-Encoding: gzip, deflate, br
Accept-Language: en-GB,en-US;q=0.9,en;q=0.8
Connection: keep-alive
Host: ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net
Origin: https://www.netflix.com
Referer: https://ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net/
Sec-Fetch-Dest: empty
Sec-Fetch-Mode: cors
Sec-Fetch-Site: cross-site
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.113 Safari/537.36
Query String Parameters view source view URL encoded
o: AQM1w3GprI4IIYEYRKfhqo4bgH29cra2fFpRH7uY05zoOtVlcNpvIISX9MmPfNG_vgXSMcsQVkxRZzF5fKE5wbLepsdfjAiPkCVIwuC4REPSiYmS1-CObsERJUHb1ICupW0fAP_maqGc8NuC7TqoE37ZDOlvseayQ9yC0A
v: 5
e: 1589057342
t: c5PzRWLjsletqzSyc3Vc_G0OtoE
sc: Eq)""Oh Ya{V{fWZhWQPb c~@I= VcC J
zE .npBKyY izc@
1124 requests 318 MB transferred 321 MB resources Finish: 23.5 min DOMContentLc
Console What's New Network conditions
Caching Disable cache
Network throttling Online
User agent Select automatically

Complete Playback Workflow @Netflix



[more-ip-event.net]

How many OCA appliances in Swisscom?

I found at least 35 of them

ipv4-c001-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.34
ipv4-c002-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.35
ipv4-c003-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.36
ipv4-c004-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.37
ipv4-c005-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.38
ipv4-c006-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.39
ipv4-c007-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.40
ipv4-c008-zrh001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.41
ipv4-c001-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.98
ipv4-c002-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.99
ipv4-c003-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.100
ipv4-c004-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.101
ipv4-c005-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.102
ipv4-c006-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.103
ipv4-c007-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.104
ipv4-c008-zrh002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.105
ipv4-c001-zrh003-swisscom-isp.1.oca.nflxvideo.net	193.247.193.242
ipv4-c002-zrh003-swisscom-isp.1.oca.nflxvideo.net	193.247.193.243

ipv4-c001-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.2
ipv4-c002-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.3
ipv4-c003-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.4
ipv4-c004-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.5
ipv4-c005-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.6
ipv4-c006-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.7
ipv4-c007-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.8
ipv4-c009-gva001-swisscom-isp.1.oca.nflxvideo.net	193.247.193.9
ipv4-c001-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.72
ipv4-c002-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.73
ipv4-c003-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.74
ipv4-c005-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.67
ipv4-c006-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.68
ipv4-c007-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.69
ipv4-c008-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.70
ipv4-c009-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.71
ipv4-c010-gva002-swisscom-isp.1.oca.nflxvideo.net	193.247.193.66

Assuming all of them are fully loaded → **10 080 TB** of storage!! (288 TB x 35)

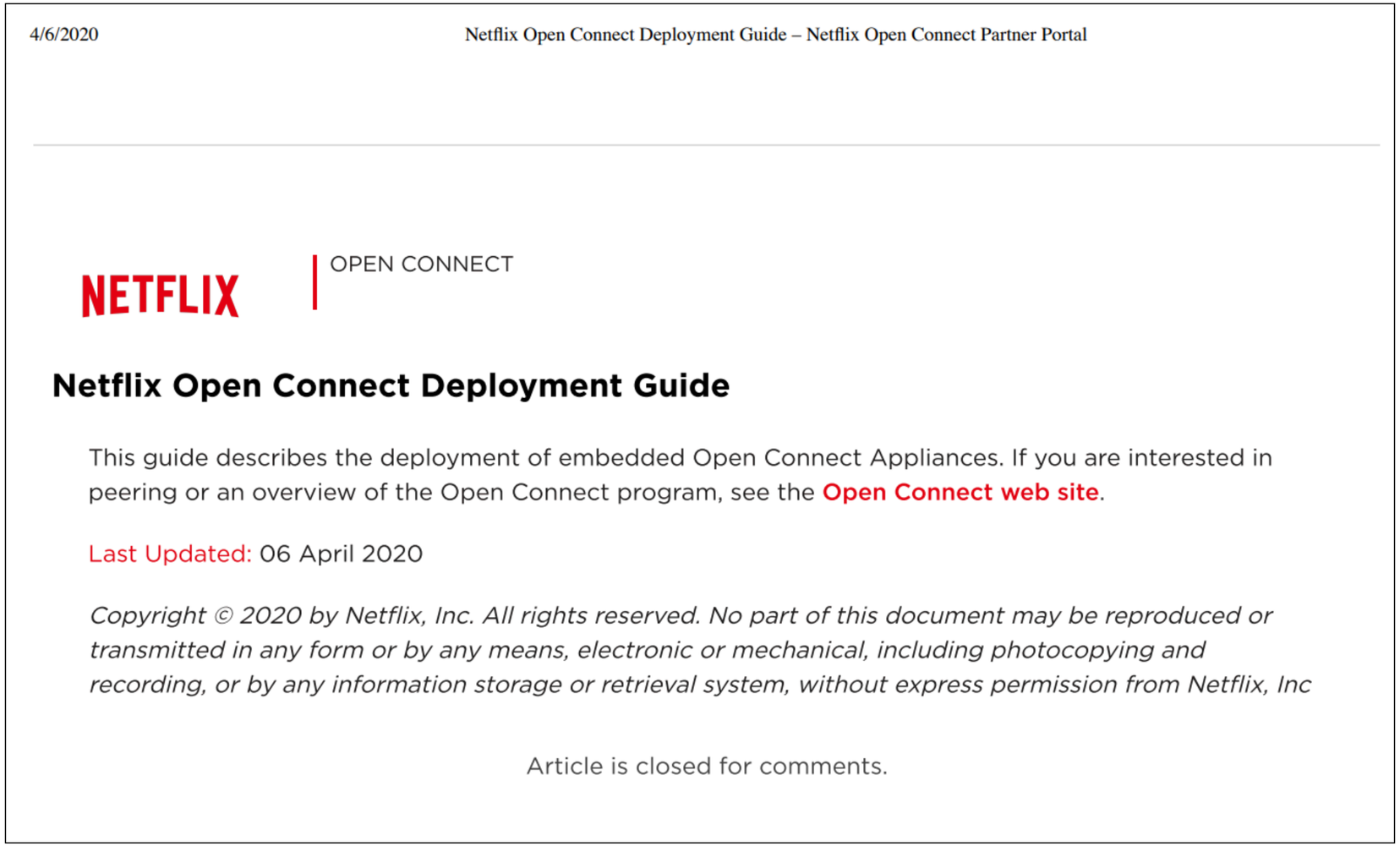
>2 million 1080p movies, assuming 100 min encoded at 5 Mbps

Besides OCAs within ISPs, Netflix also hosts caches at various IXPs and datacenters

ipv4-c001-zrh001-ix.1.oca.nflxvideo.net	45.57.18.130
ipv4-c002-zrh001-ix.1.oca.nflxvideo.net	45.57.18.131
ipv4-c003-zrh001-ix.1.oca.nflxvideo.net	45.57.18.132
ipv4-c004-zrh001-ix.1.oca.nflxvideo.net	45.57.19.130
ipv4-c005-zrh001-ix.1.oca.nflxvideo.net	45.57.19.131
ipv4-c006-zrh001-ix.1.oca.nflxvideo.net	45.57.19.132
ipv4-c007-zrh001-ix.1.oca.nflxvideo.net	45.57.18.133
ipv4-c008-zrh001-ix.1.oca.nflxvideo.net	45.57.18.134
ipv4-c009-zrh001-ix.1.oca.nflxvideo.net	45.57.18.135
ipv4-c010-zrh001-ix.1.oca.nflxvideo.net	45.57.18.136
ipv4-c011-zrh001-ix.1.oca.nflxvideo.net	45.57.19.133
ipv4-c012-zrh001-ix.1.oca.nflxvideo.net	45.57.19.134
ipv4-c013-zrh001-ix.1.oca.nflxvideo.net	45.57.19.135
ipv4-c014-zrh001-ix.1.oca.nflxvideo.net	45.57.19.136
ipv4-c015-zrh001-ix.1.oca.nflxvideo.net	45.57.18.137
ipv4-c016-zrh001-ix.1.oca.nflxvideo.net	45.57.18.138
ipv4-c017-zrh001-ix.1.oca.nflxvideo.net	45.57.19.137
ipv4-c018-zrh001-ix.1.oca.nflxvideo.net	45.57.19.138
ipv4-c019-zrh001-ix.1.oca.nflxvideo.net	45.57.18.139
ipv4-c020-zrh001-ix.1.oca.nflxvideo.net	45.57.18.140
ipv4-c021-zrh001-ix.1.oca.nflxvideo.net	45.57.18.141
ipv4-c022-zrh001-ix.1.oca.nflxvideo.net	45.57.19.139
ipv4-c023-zrh001-ix.1.oca.nflxvideo.net	45.57.19.140
ipv4-c024-zrh001-ix.1.oca.nflxvideo.net	45.57.19.141

At least 24 instances in Zurich Equinix, see https://openconnect.netflix.com/en/peering/#locations

If you are interested in finding out more: check out https://openconnect.netflix.com



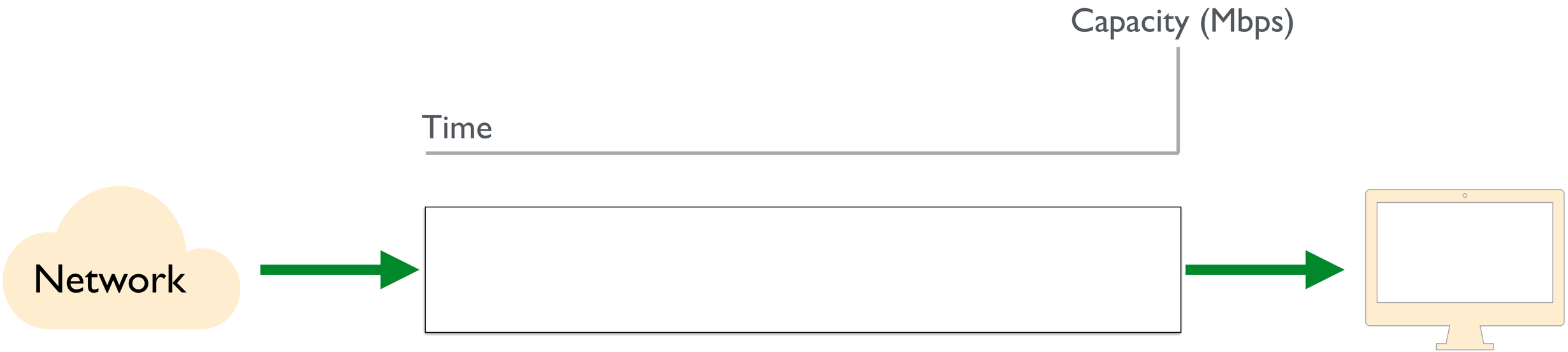
Deployment guide: https://openconnect.netflix.com/deploymentguide.pdf

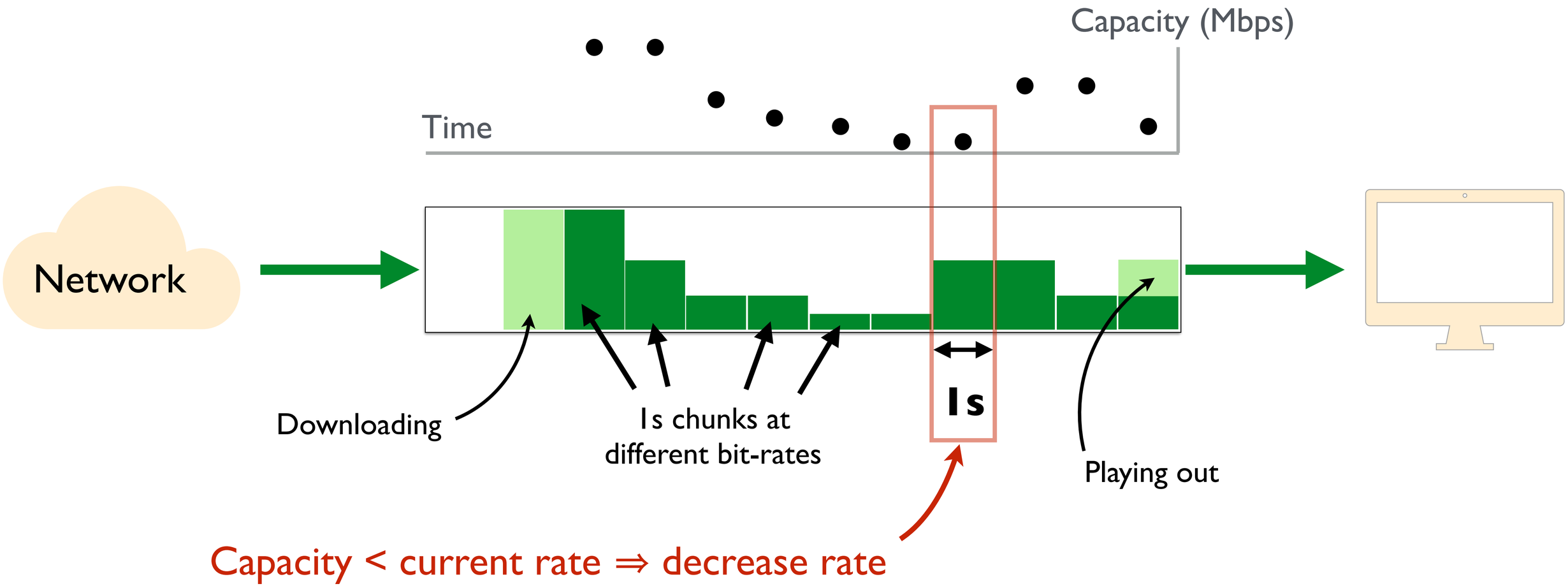
Encoding

Replication

Adaptation


Network

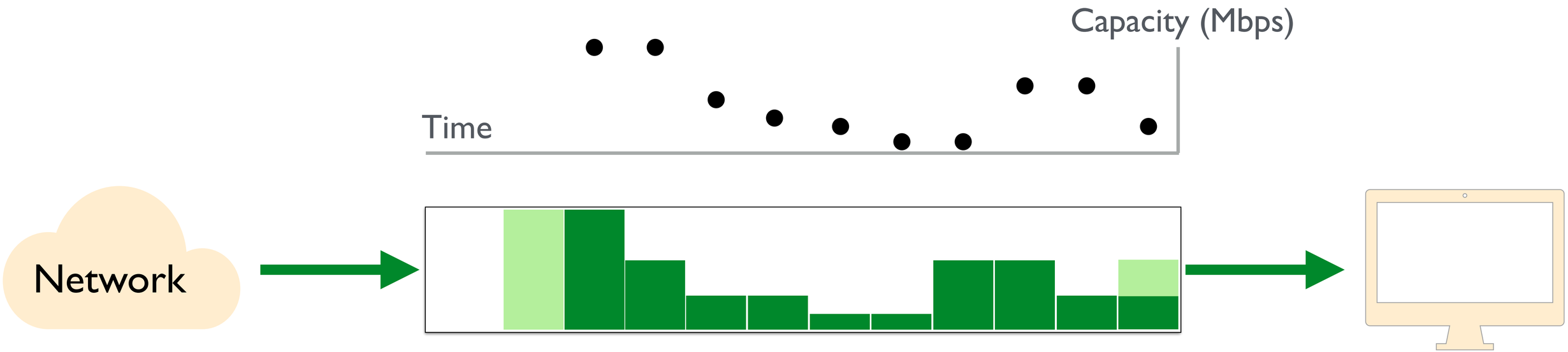

Capacity (Mbps)
Time
Network


Capacity (Mbps)
Time
Network
Downloading
1s chunks at
different bit-rates
1s
Playing out
Capacity < current rate ⇒ decrease rate

Common solution approach

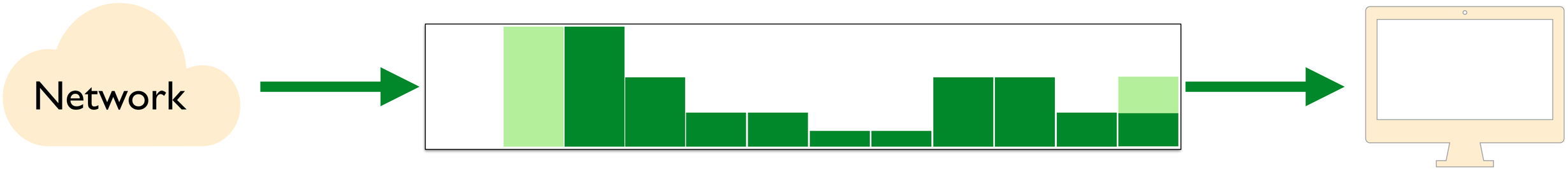
- Encode video in multiple bitrates
- Replicate using a content delivery network
- Video player picks bitrate adaptively
  - Estimate connection's available bandwidth
  - Pick a bitrate ≤ available bandwidth

Buffer-based capacity adaptation



Decide based on the buffer alone?

Buffer-based capacity adaptation

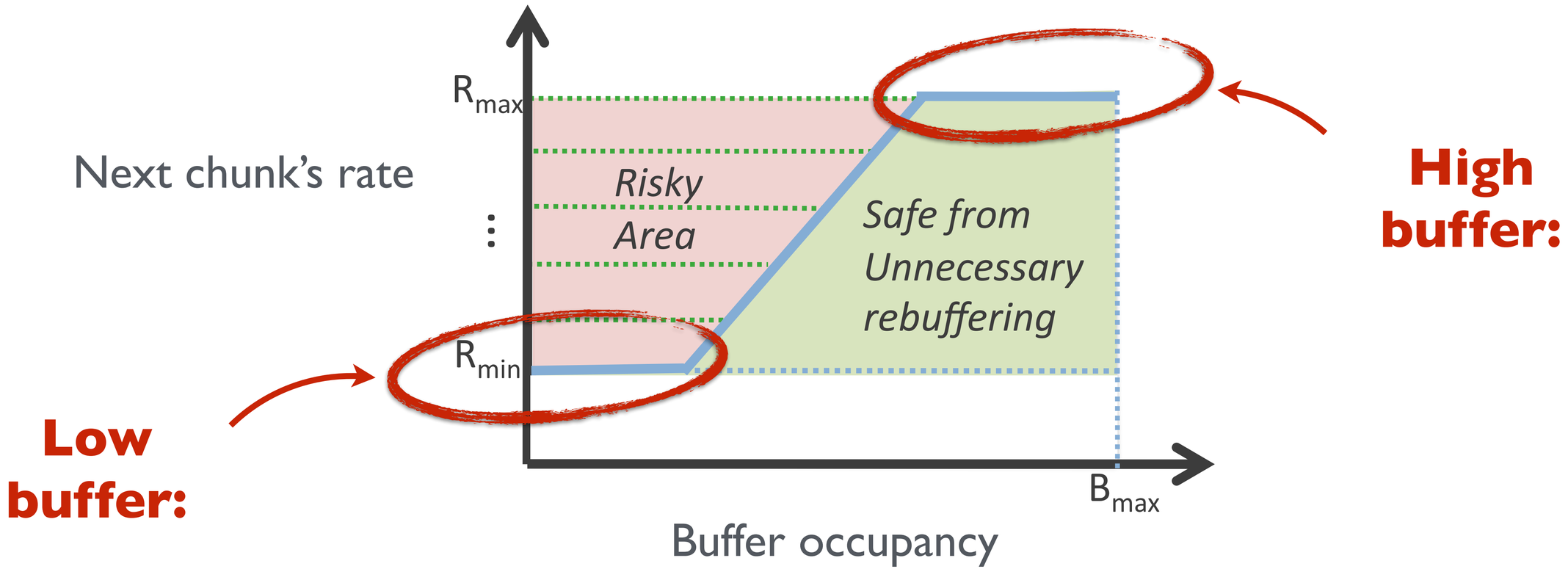


Nearly full buffer ⇒ large rate

Buffer-based capacity adaptation

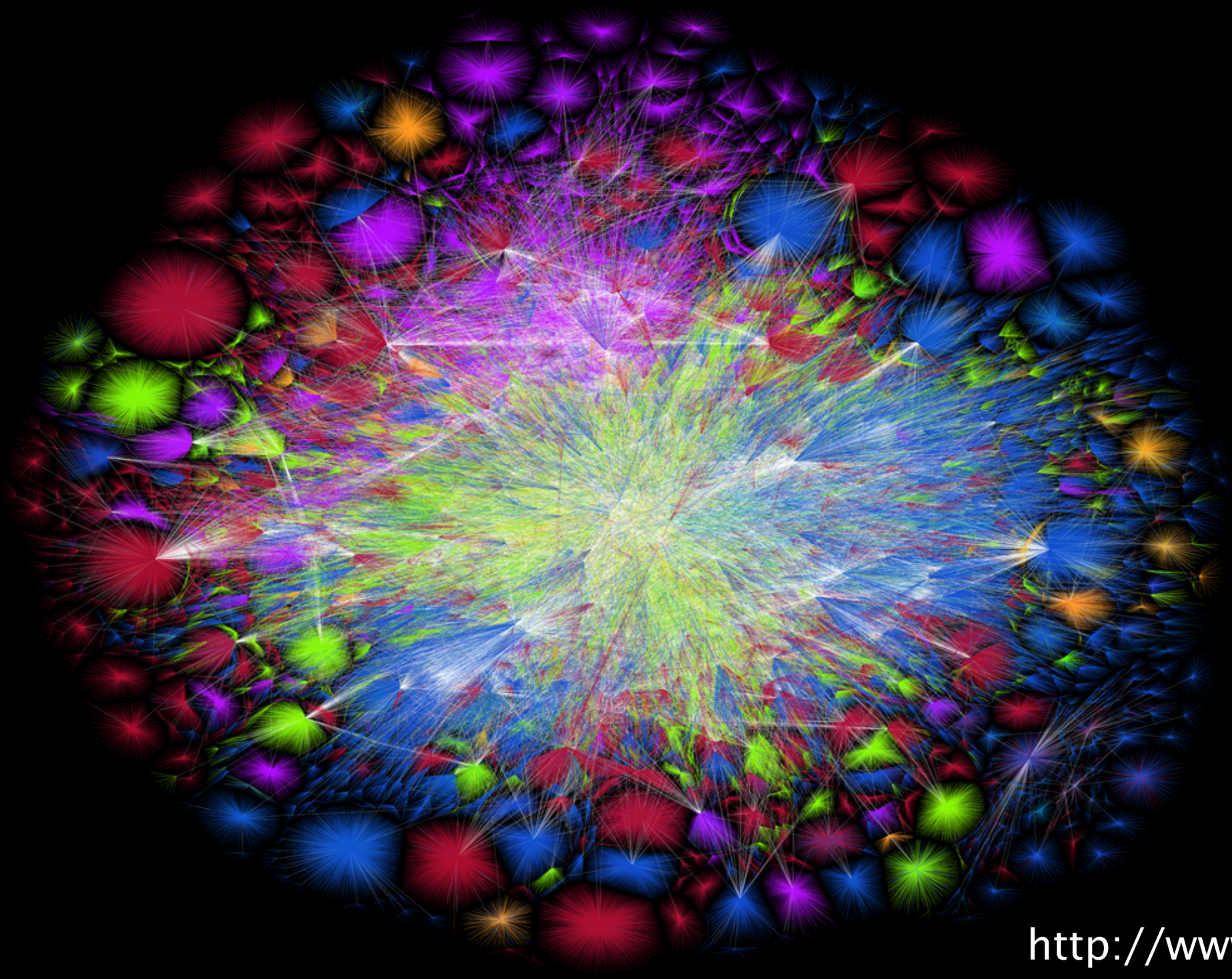
Nearly empty buffer ⇒ small rate

Buffer-based capacity adaptation

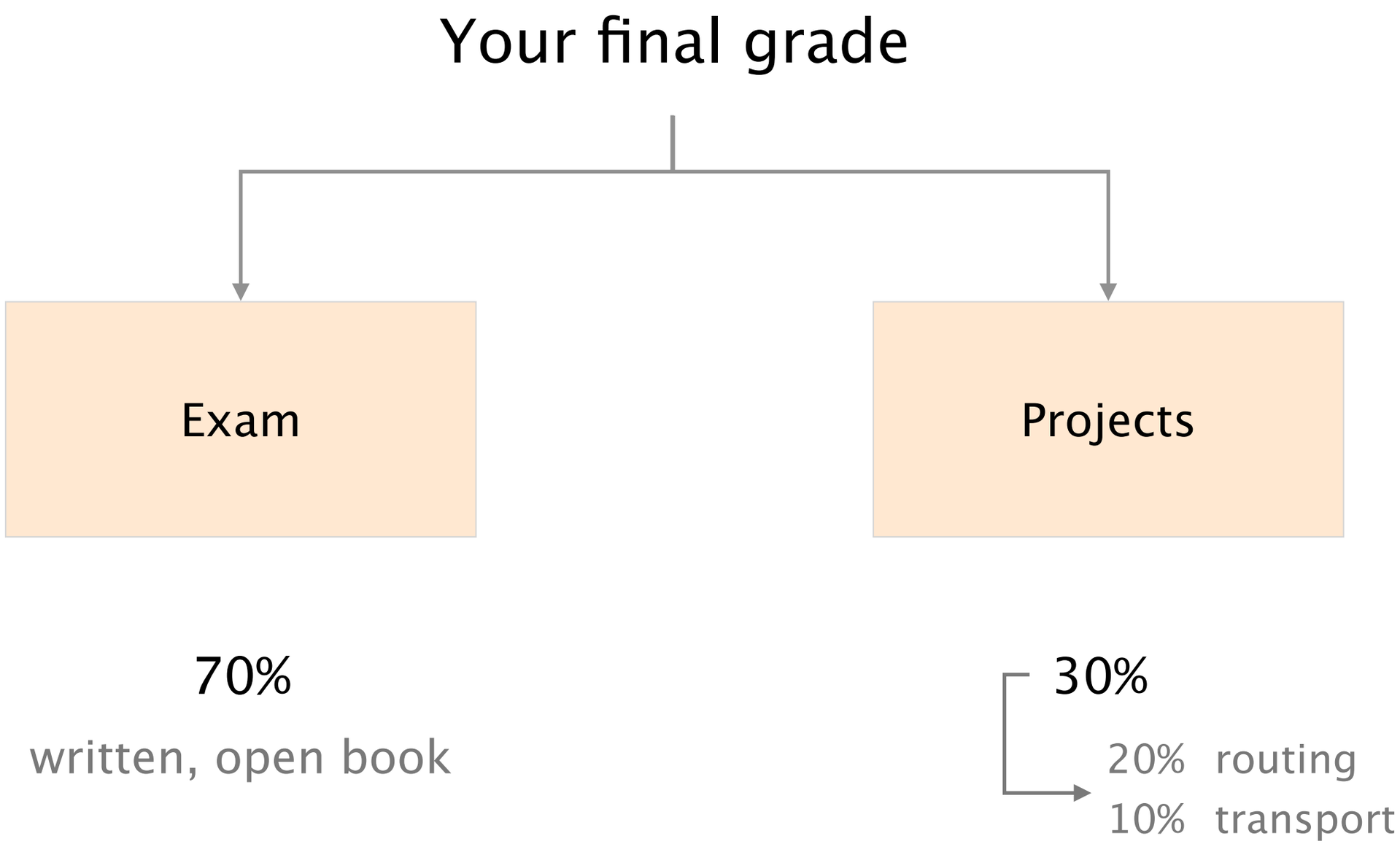


[A Buffer-Based Approach to Rate Adaptation: Evidence from a Large Video Streaming Service, Huang et al., ACM SIGCOMM 2014]

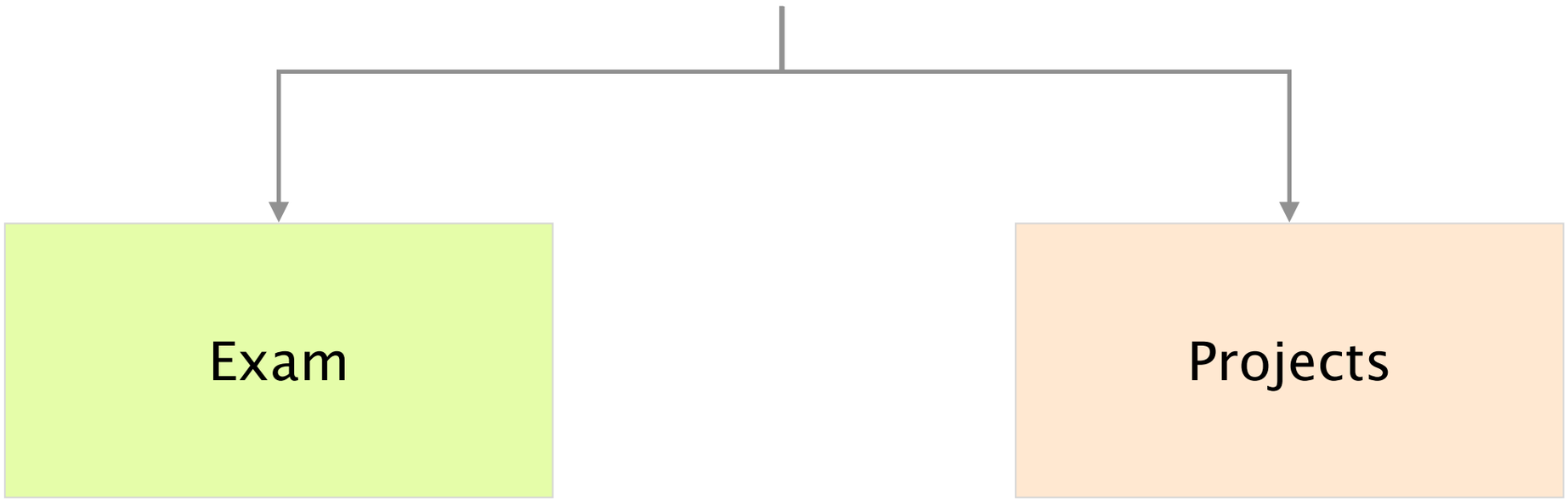
Now you (better) understand this!



http://www.opte.org


Your final grade
Exam
Projects
70%
written, open book
30%
20% routing
10% transport

Your final grade



70%

written, open book

30%

20% routing

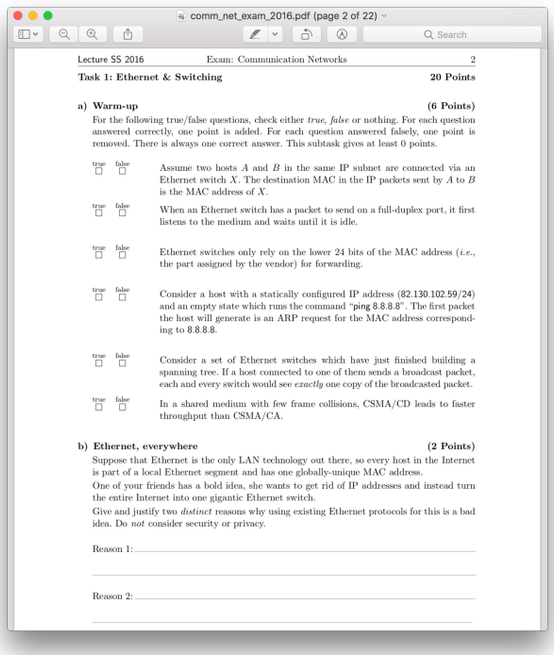
10% transport

The exam will be open book, most of the questions will be open-ended, with some multiple choices

verify your understanding of the material

Make sure you can do *all* the exercises, especially the ones in previous exams

Millesime 2016


comm_net_exam_2016.pdf (page 2 of 22)

Lecture SS 2016 Exam: Communication Networks 2

Task 1: Ethernet & Switching **20 Points**

a) Warm-up **(6 Points)**

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered falsely, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true ☐ false ☐ Assume two hosts A and B in the same IP subnet are connected via an Ethernet switch X. The destination MAC in the IP packets sent by A to B is the MAC address of X.

true ☐ false ☐ When an Ethernet switch has a packet to send on a full-duplex port, it first listens to the medium and waits until it is idle.

true ☐ false ☐ Ethernet switches only rely on the lower 24 bits of the MAC address (*i.e.*, the part assigned by the vendor) for forwarding.

true ☐ false ☐ Consider a host with a statically configured IP address (82.130.102.59/24) and an empty state which runs the command "ping 8.8.8.8". The first packet the host will generate is an ARP request for the MAC address corresponding to 8.8.8.8.

true ☐ false ☐ Consider a set of Ethernet switches which have just finished building a spanning tree. If a host connected to one of them sends a broadcast packet, each and every switch would see *exactly* one copy of the broadcasted packet.

true ☐ false ☐ In a shared medium with few frame collisions, CSMA/CD leads to faster throughput than CSMA/CA.

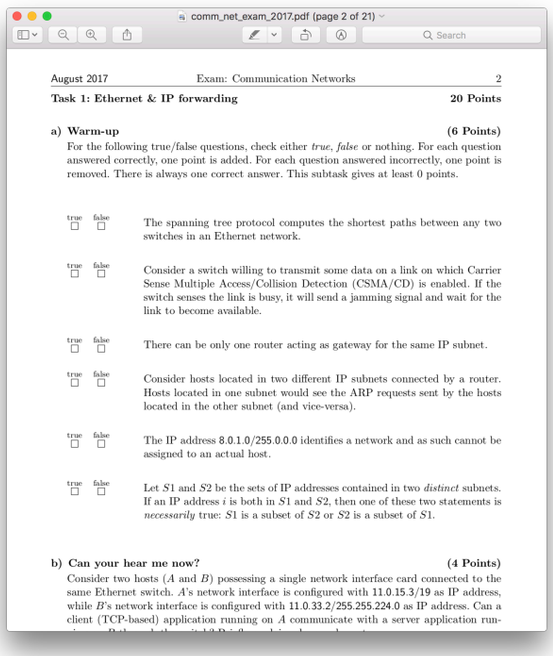
b) Ethernet, everywhere **(2 Points)**

Suppose that Ethernet is the only LAN technology out there, so every host in the Internet is part of a local Ethernet segment and has one globally-unique MAC address.
One of your friends has a bold idea, she wants to get rid of IP addresses and instead turn the entire Internet into one gigantic Ethernet switch.
Give and justify two *distinct* reasons why using existing Ethernet protocols for this is a bad idea. Do *not* consider security or privacy.

Reason 1: _____

Reason 2: _____

Millesime 2017


comm_net_exam_2017.pdf (page 2 of 21)

August 2017 Exam: Communication Networks 2

Task 1: Ethernet & IP forwarding **20 Points**

a) Warm-up **(6 Points)**

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true ☐ false ☐ The spanning tree protocol computes the shortest paths between any two switches in an Ethernet network.

true ☐ false ☐ Consider a switch willing to transmit some data on a link on which Carrier Sense Multiple Access/Collision Detection (CSMA/CD) is enabled. If the switch senses the link is busy, it will send a jamming signal and wait for the link to become available.

true ☐ false ☐ There can be only one router acting as gateway for the same IP subnet.

true ☐ false ☐ Consider hosts located in two different IP subnets connected by a router. Hosts located in one subnet would see the ARP requests sent by the hosts located in the other subnet (and vice-versa).

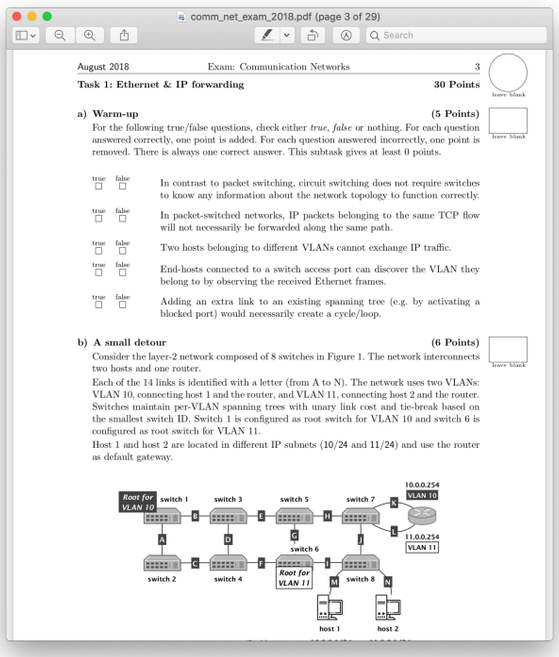
true ☐ false ☐ The IP address 8.0.1.0/255.0.0.0 identifies a network and as such cannot be assigned to an actual host.

true ☐ false ☐ Let $S1$ and $S2$ be the sets of IP addresses contained in two *distinct* subnets. If an IP address i is both in $S1$ and $S2$, then one of these two statements is *necessarily* true: $S1$ is a subset of $S2$ or $S2$ is a subset of $S1$.

b) Can your hear me now? **(4 Points)**

Consider two hosts (A and B) possessing a single network interface card connected to the same Ethernet switch. A's network interface is configured with 11.0.15.3/19 as IP address, while B's network interface is configured with 11.0.33.2/255.255.224.0 as IP address. Can a client (TCP-based) application running on A communicate with a server application run-

Millesime 2018


comm_net_exam_2018.pdf (page 3 of 29)

August 2018 Exam: Communication Networks 3

Task 1: Ethernet & IP forwarding **30 Points**

leave blank

a) Warm-up **(5 Points)**

leave blank

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true ☐ false ☐ In contrast to packet switching, circuit switching does not require switches to know any information about the network topology to function correctly.

true ☐ false ☐ In packet-switched networks, IP packets belonging to the same TCP flow will not necessarily be forwarded along the same path.

true ☐ false ☐ Two hosts belonging to different VLANs cannot exchange IP traffic.

true ☐ false ☐ End-hosts connected to a switch access port can discover the VLAN they belong to by observing the received Ethernet frames.

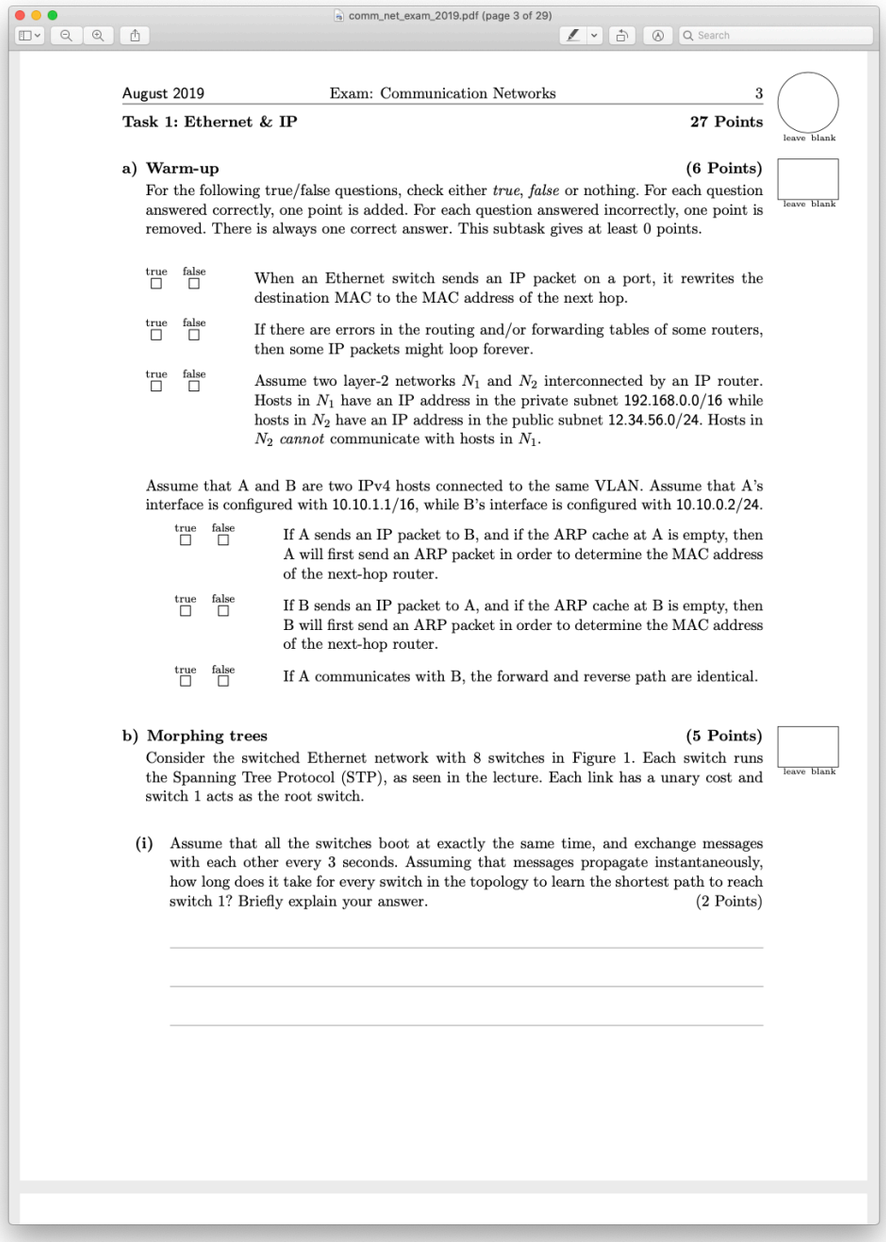
true ☐ false ☐ Adding an extra link to an existing spanning tree (e.g. by activating a blocked port) would necessarily create a cycle/loop.

b) A small detour **(6 Points)**

leave blank

Consider the layer-2 network composed of 8 switches in Figure 1. The network interconnects two hosts and one router.
Each of the 14 links is identified with a letter (from A to N). The network uses two VLANs: VLAN 10, connecting host 1 and the router, and VLAN 11, connecting host 2 and the router. Switches maintain per-VLAN spanning trees with unary link cost and tie-break based on the smallest switch ID. Switch 1 is configured as root switch for VLAN 10 and switch 6 is configured as root switch for VLAN 11.
Host 1 and host 2 are located in different IP subnets (10/24 and 11/24) and use the router as default gateway.

Millesime 2019


comm_net_exam_2019.pdf (page 3 of 29)

August 2019 Exam: Communication Networks 3

Task 1: Ethernet & IP **27 Points**

leave blank

a) Warm-up **(6 Points)**

leave blank

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true ☐ false ☐ When an Ethernet switch sends an IP packet on a port, it rewrites the destination MAC to the MAC address of the next hop.

true ☐ false ☐ If there are errors in the routing and/or forwarding tables of some routers, then some IP packets might loop forever.

true ☐ false ☐ Assume two layer-2 networks N_1 and N_2 interconnected by an IP router. Hosts in N_1 have an IP address in the private subnet 192.168.0.0/16 while hosts in N_2 have an IP address in the public subnet 12.34.56.0/24. Hosts in N_2 *cannot* communicate with hosts in N_1.

Assume that A and B are two IPv4 hosts connected to the same VLAN. Assume that A's interface is configured with 10.10.1.1/16, while B's interface is configured with 10.10.0.2/24.

true ☐ false ☐ If A sends an IP packet to B, and if the ARP cache at A is empty, then A will first send an ARP packet in order to determine the MAC address of the next-hop router.

true ☐ false ☐ If B sends an IP packet to A, and if the ARP cache at B is empty, then B will first send an ARP packet in order to determine the MAC address of the next-hop router.

true ☐ false ☐ If A communicates with B, the forward and reverse path are identical.

b) Morphing trees **(5 Points)**

leave blank

Consider the switched Ethernet network with 8 switches in Figure 1. Each switch runs the Spanning Tree Protocol (STP), as seen in the lecture. Each link has a unary cost and switch 1 acts as the root switch.

(i) Assume that all the switches boot at exactly the same time, and exchange messages with each other every 3 seconds. Assuming that messages propagate instantaneously, how long does it take for every switch in the topology to learn the shortest path to reach switch 1? Briefly explain your answer. (2 Points)

https://comm-net.ethz.ch/#tab-exam

Don't forget the assignments,

they matter

No programming question no Python at the exam

but we could ask you to describe a procedure in English

What would you change in your solution to achieve *X*?

No configuration question no FRRouting at the exam

but we could ask you to describe a configuration in English

How would you enforce policy *X*?

We'll organize another remote Q&A session closer to the exam (details to follow)

Communication Networks

What's next?

Master-level lecture, every Fall semester

Advanced Topics in Communication Networks

Topics
(examples)

Tunneling
Hierarchical routing
Traffic Engineering
Virtual Private Networks
Quality of Service/Scheduling
IP Multicast
Fast Convergence
Network virtualization
Network programmability
Network measurements

+ labs & a project

if you liked the routing project,
you will like this lecture as well

https://adv-net.ethz.ch/

Consider doing one of your theses with our group!

bachelor, semester or master

https://nsg.ee.ethz.ch/theses/

That's all Folks!

Communication Networks

Spring 2022

Laurent Vanbever

nsg.ee.ethz.ch

ETH Zürich (D-ITET)

May 30 2022