BGP

Matěj Grégr
igregr@fit.vutbr.cz

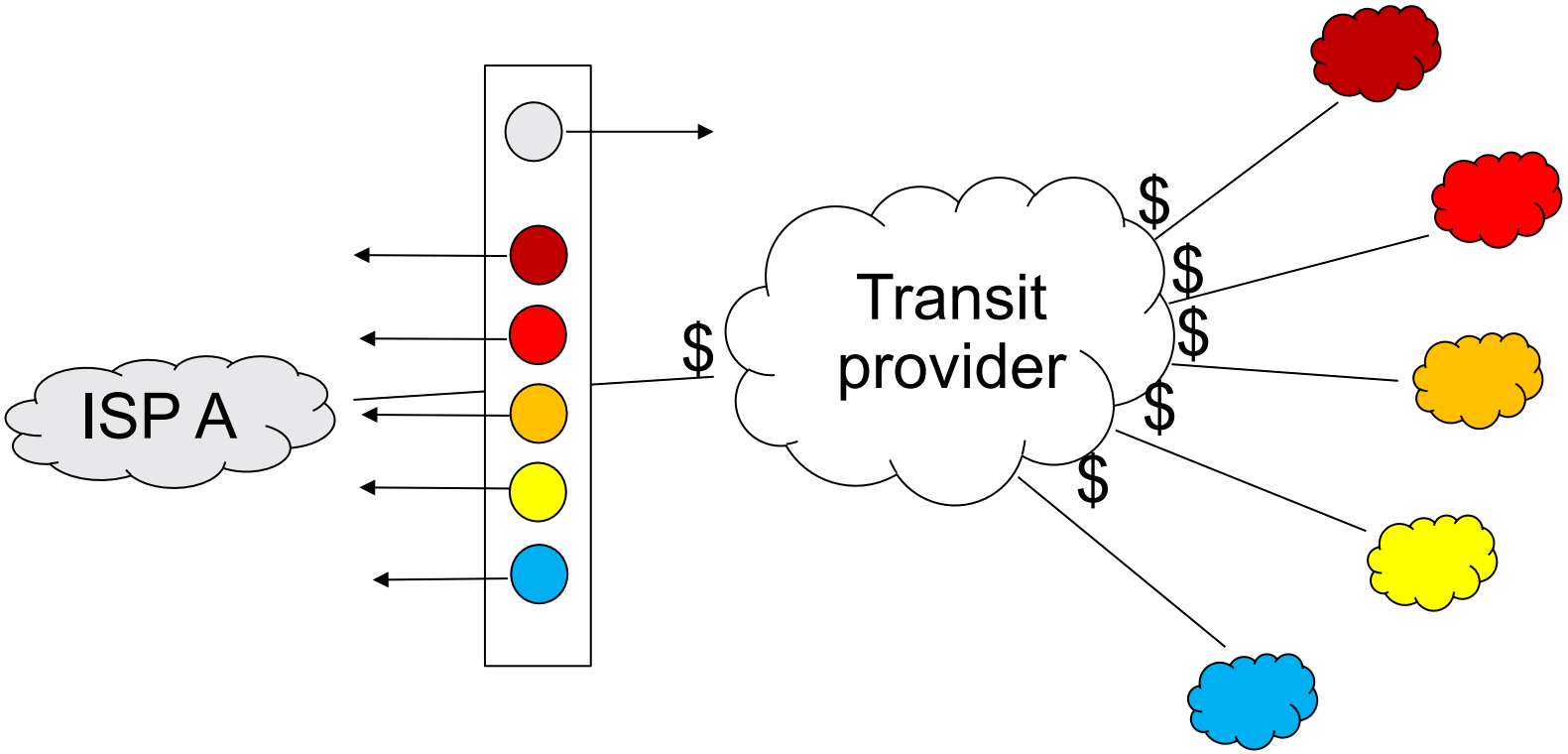
Routing – outgoing vs incomming

```
ping 8.8.8.8
```

How to obtain addresses?

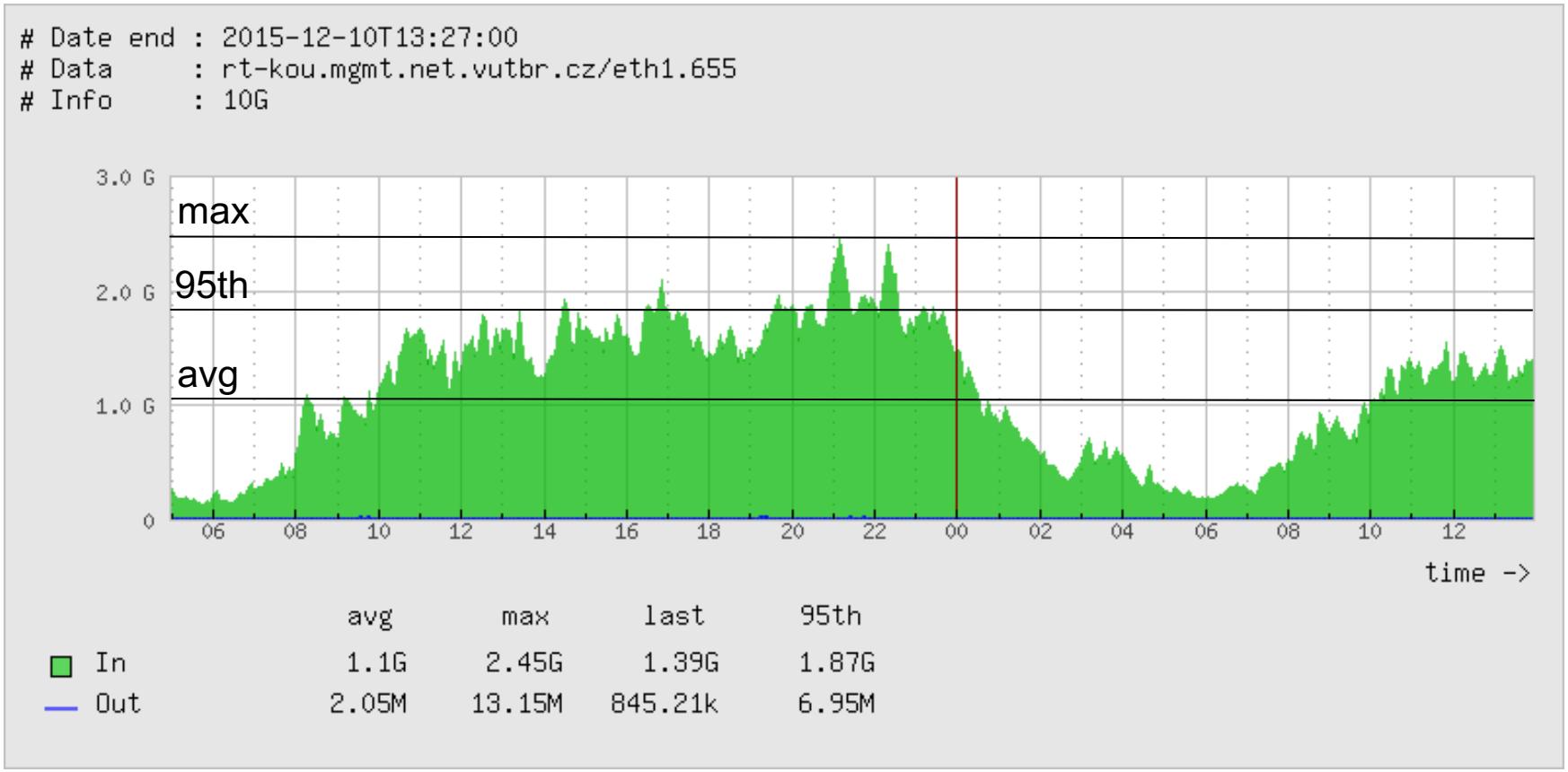
How to connect to the Internet?

- Connect to another network that is already connected
- **Buy** from the network Internet Transit



Costs?

- Standard practice is 95/5 percentil



Costs?

Internet Transit Prices (1998-2015) U.S. Internet Region		
Year	Prices (in Mbps, min commit)	% Decline
1998	$1200 per Mbps	
1999	$800 per Mbps	33%
2000	$675 per Mbps	16%
2001	$400 per Mbps	40%
2002	$200 per Mbps	50%
2003	$120 per Mbps	40%
2004	$90 per Mbps	25%
2005	$75 per Mbps	17%
2006	$50 per Mbps	33%
2007	$25 per Mbps	50%
2008	$12 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%

Source: DrPeering.net

Every year the ISPs say to me things like,

"*Transit prices can't get any lower*," and "*No one is making any money at these prices.*" And every year, the prices drop again.

-William B. Norton

Costs with commit

Commit	Unit Price per Mbps	Min Spend
10 Mbps	$12	$120
100 Mbps	$5	$500
1 Gbps	$3.5	$3500
10 Gbps	$1.20	$12 000
100 Gbps	$0.70	$70 000

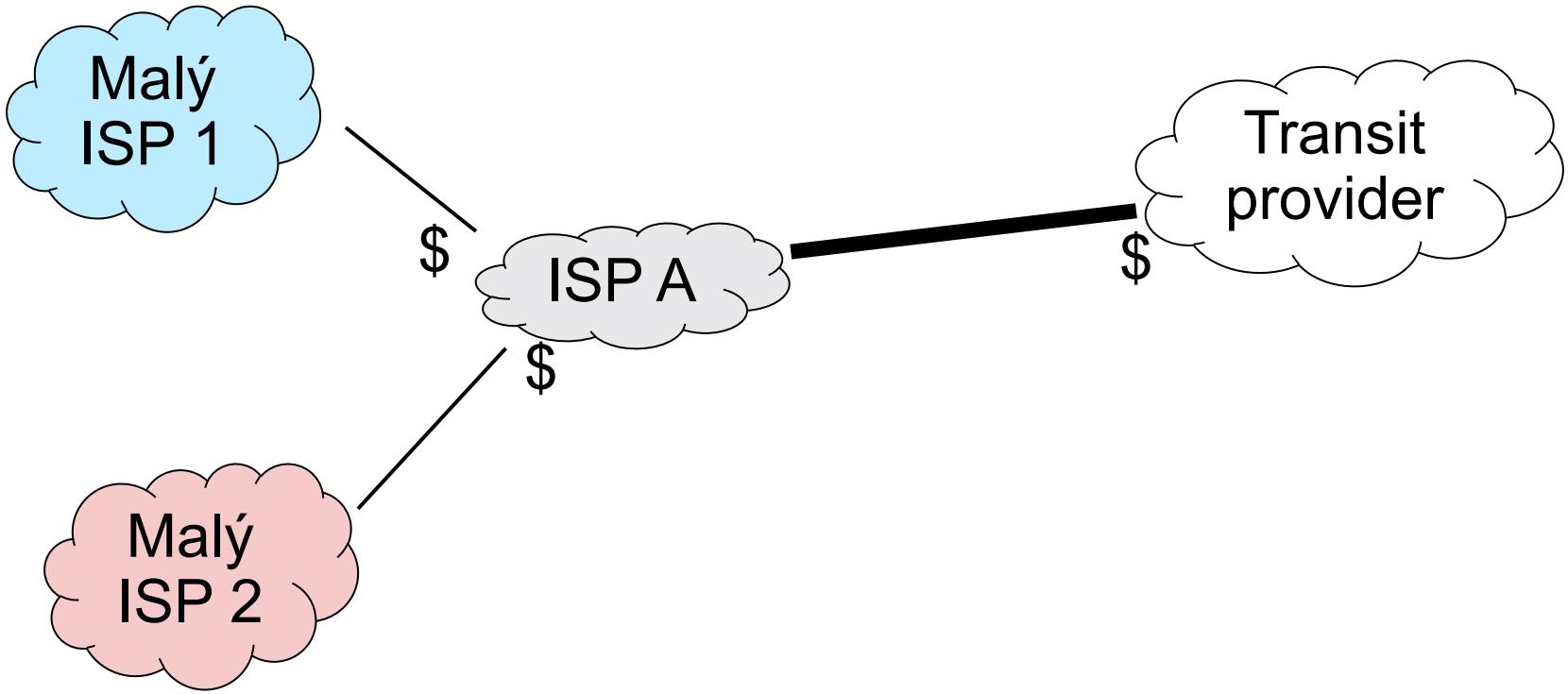
Example

Monthly Bill = max(TransitVolumeAt95th*Price, Commit Volume*Price)

ISP A buys transit from ISP B for $5 per Mbps. ISP A sends 500 Mbps, receives 800 Mbps. 95/5 percentil is used for cost calculation.

If ISP A receives offer 3$ per Mbps with 1Gbps commit. Is the offer budget-wise?

Reselling

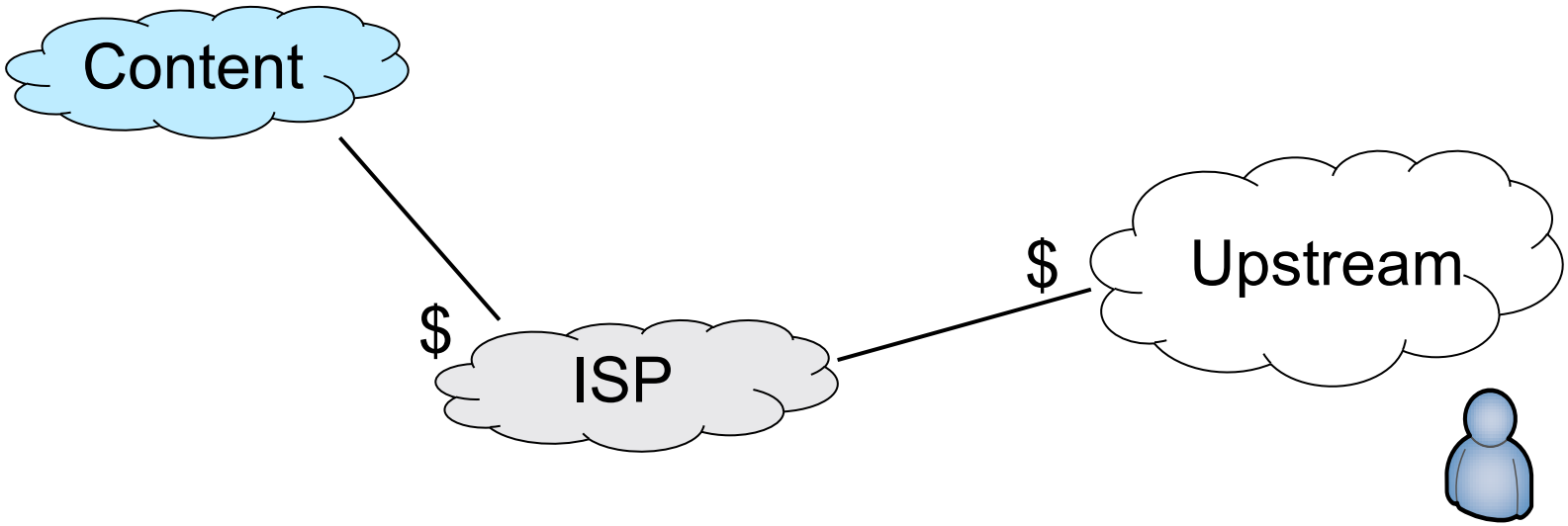


Free band usage

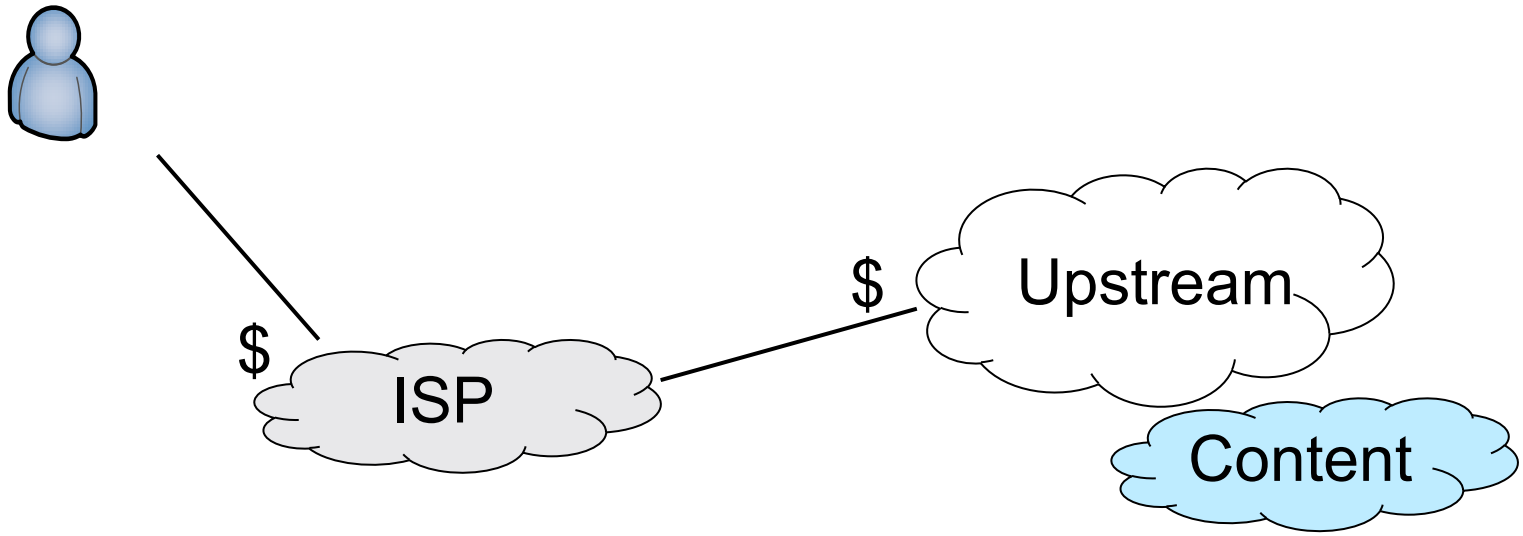
- Traffic is usually predictable
  - Users at work
  - Home users

- If a link is not used, it‘s possible to use it for different purpose
  - Preload data from CDN cache
  - Mix business and home users

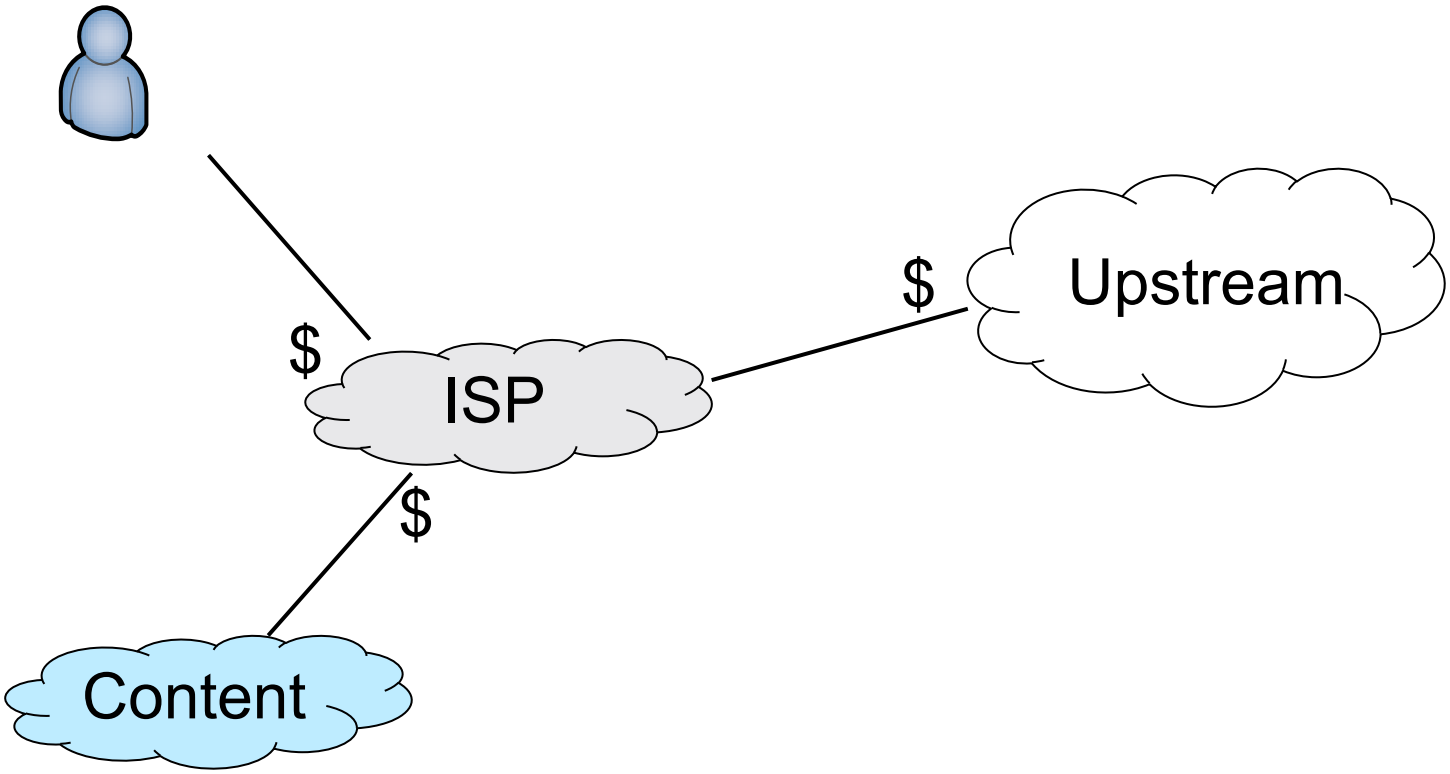
Capture access + content



Capture access + content



Capture access + content



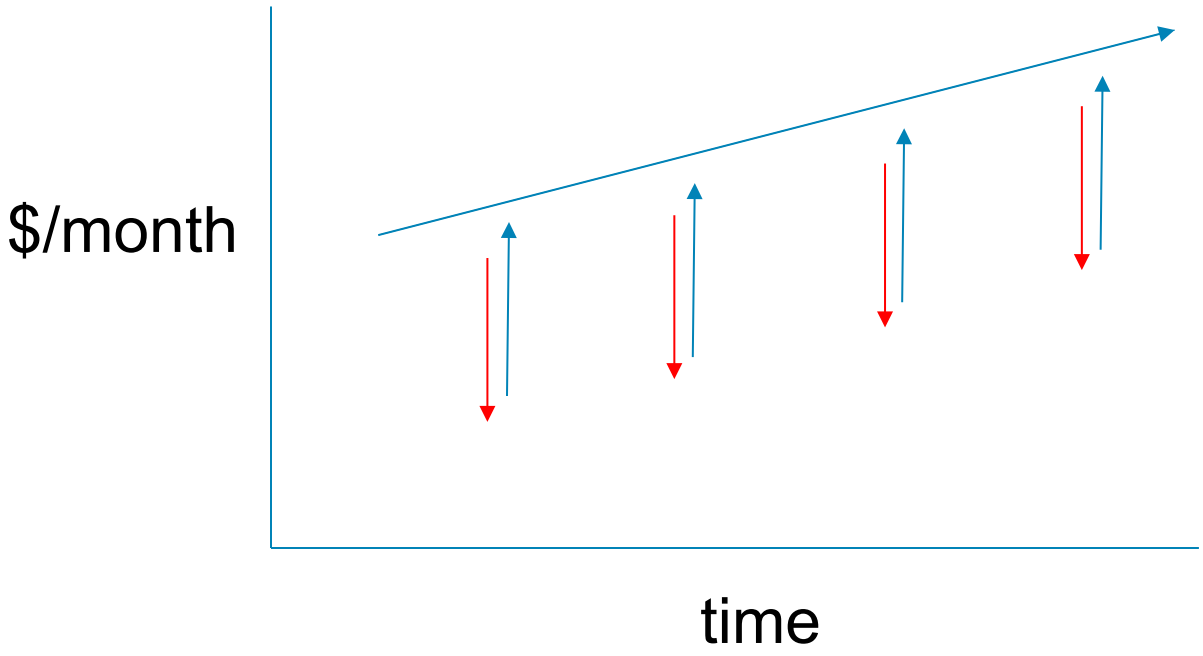
Other transit games

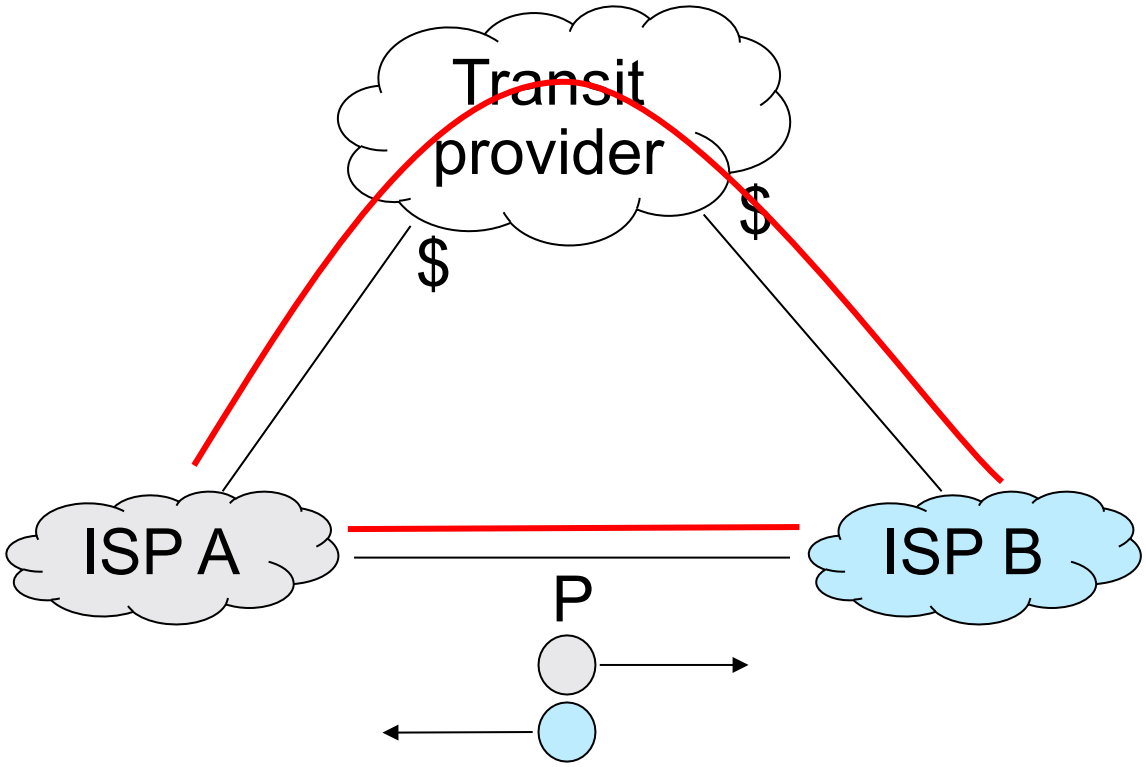
- 95/5 percentil usage
- Multi-Homing
- Routing „Magic“
- Link to cheaper place

Peering

Peering

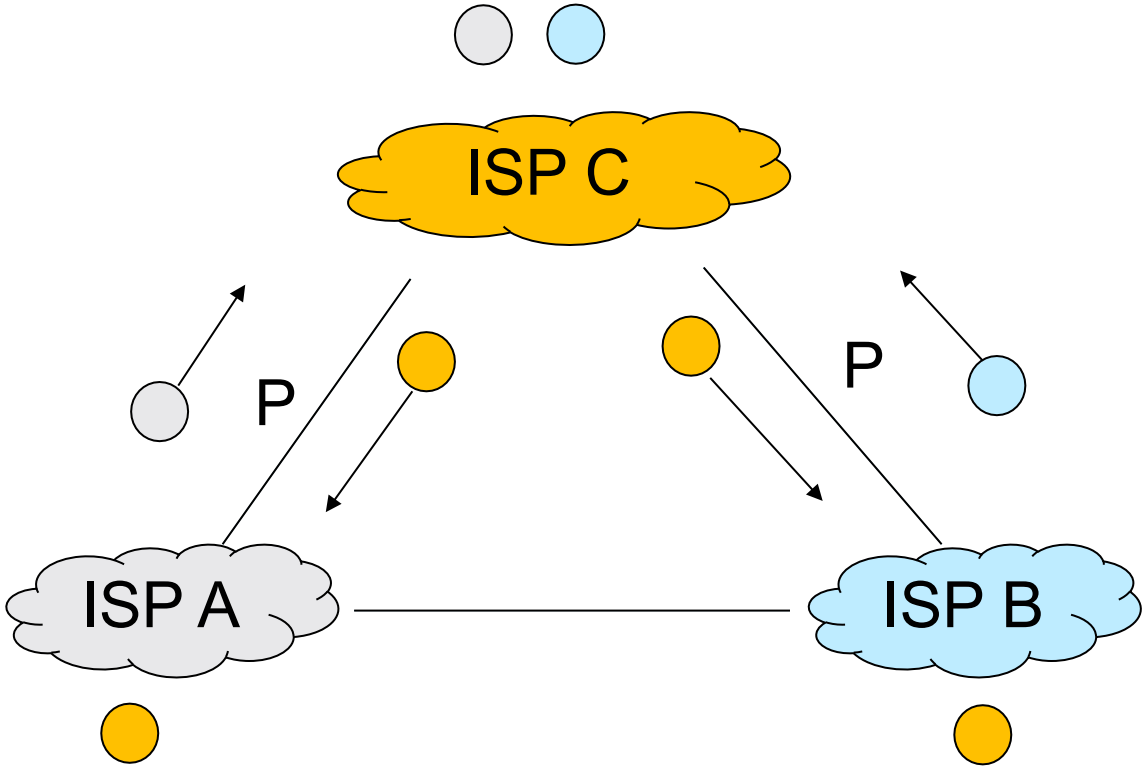
- A *voluntary* interconnection of networks for the purpose of exchanging traffic between the users of each network
- Why it is neccessary if transit price goes down?
  - VOLUME!




Transit
provider
$
$
ISP A
ISP B
P

Peering – properties

- Peering is not transitive!
  - If A peers with B and C peers with B, it does not mean that A can reach C via B!
- Peering is not a substitution for transit connection!
- Peering is usually without direct fees (*free*)
  - Pay peering exists as well
  - Indirect cost can apply

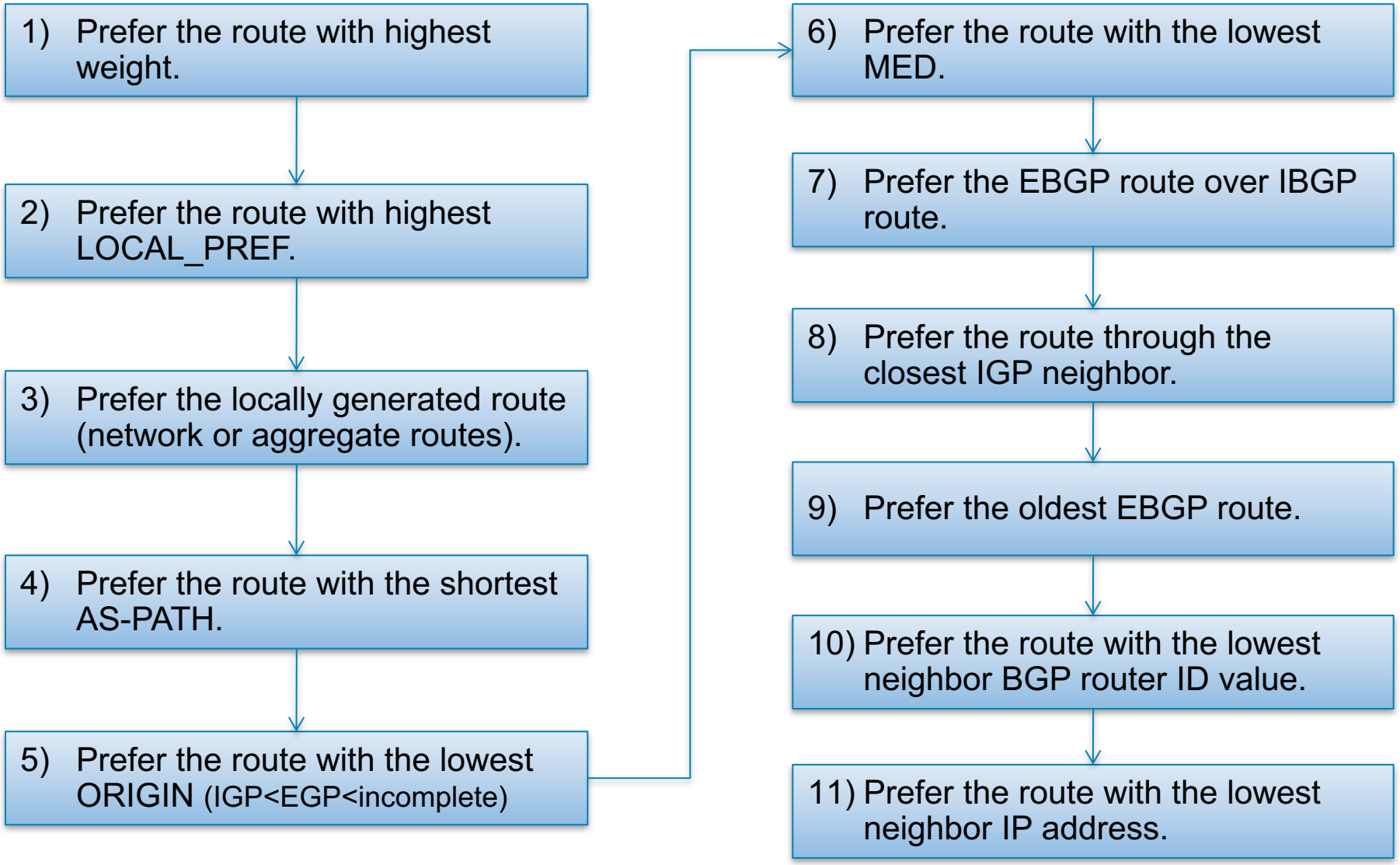

ISP C
P
P
ISP A
ISP B

Peering – motivation

- Lower transit costs
- Better *user-experience*
- Lower latency and packet loss → more user traffic
  - If user is accounted based on number of transferred data, the user will transfer more data = higher revenue
- Marketing advantage

BGP Path manipulation

Best Route Selection Process



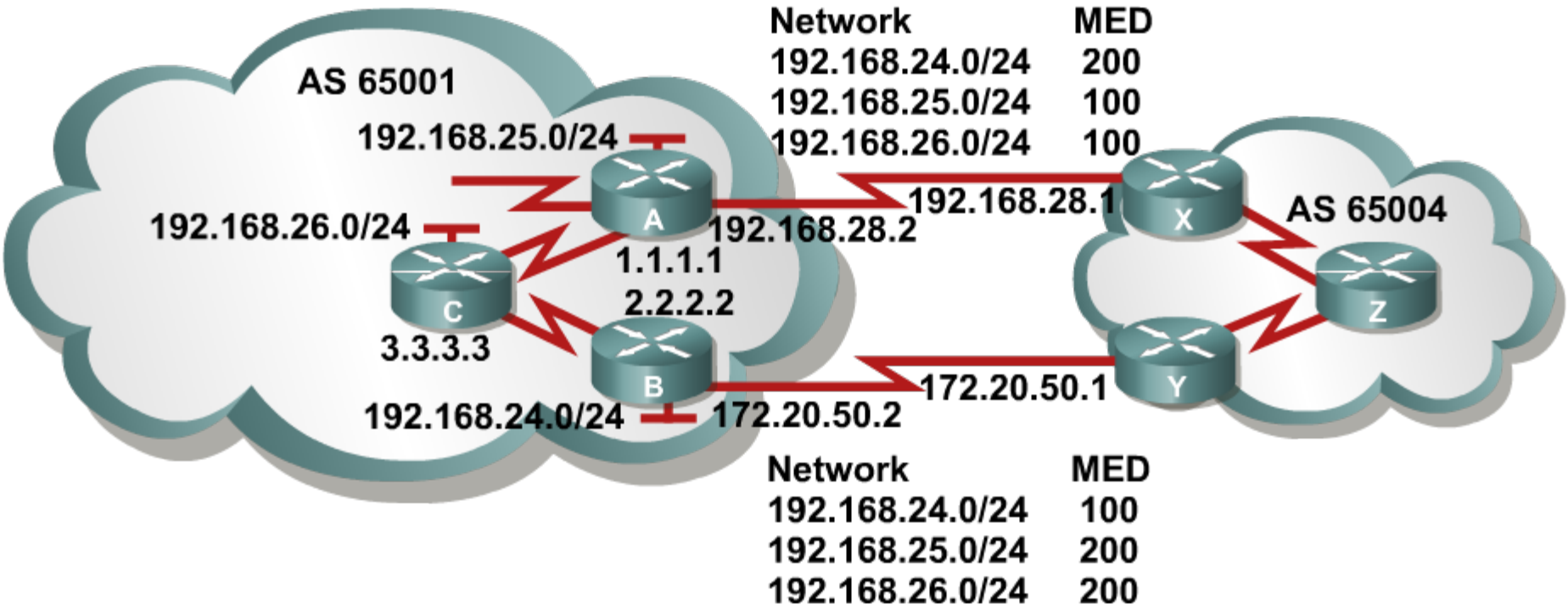
How to influence the traffic flow?

- Attribute **LOCAL_PREF** describes route preference
  - This attribute is exchanged only through iBGP neighborship because routers in different AS could prefer other routes
  - *Higher means better!*
  - By default routes have LOCAL_PREF set on 100
    - Depends on the vendor

How to influence the traffic flow?

- **MED** indicates to other AS which route to use when entering our AS
  - MED is initially send from our AS via eBGP to neighbor AS. Overthere, MED value is employed but does not cross borders of this AS (it traverses to other ASs with value set on 0)
  - MED is called also metric and it behaves just like metric of any IGP – *Lower means better*
  - Default MED is 0

MED Use-Case


Network MED
192.168.24.0/24 200
192.168.25.0/24 100
192.168.26.0/24 100
AS 65001
192.168.25.0/24
192.168.26.0/24
A
C
B
192.168.28.2
1.1.1.1
2.2.2.2
3.3.3.3
192.168.24.0/24
172.20.50.2
192.168.28.1
172.20.50.1
X
Y
Z
AS 65004
Network MED
192.168.24.0/24 100
192.168.25.0/24 200
192.168.26.0/24 200

How to influence the traffic flow?

- **AS-Prepending**
- manipulates the AS Path that its advertised to BGP neighbors
- Prepend number of own ASN to make the path longer
- Influence the incoming traffic flow

```
R(config)# route-map PREPEND permit 10
R(config-route-map)# set as-path prepend  65000 65000
R(config-route-map)# router bgp 65000
R(config-router)# neighbor 192.168.1.1 route-map PREPEND out
```

BGP communities

What are BGP communities?

- Defined by RFC1997 (August 1996)
- a 32-bit integer which is attached to a BGP route as an optional transitive attribute
  - AKA: Not required, and exportable between neighbors
  - Multiple communities can be attached to one route
- Well-known (hard-coded) communities exist
  - No-Export, No-Advertise, etc.
- But mostly, the communities and how they are interpreted are defined by each individual network.

How are communities used

- To add additional information to a BGP route
  - Any data you can encode into a 32 bit integer
  - From you to others (providing information)
  - From others to you (requesting actions)
- To take action based on that information
  - E.g. Alter route attributes on demand
- Control the import/export of routes

How are communities used

- A 32-bit integer isn't always easy to work with
  - More common convention is to split into two 16-bit values
  - First value is intended to define the scope or "target"
  - Second value is arbitrary data for the targeted network
    - Whatever data you're trying to encode

- For example: 701:1234
  - Intended for AS701
  - Community value is "1234"

Usage of BGP Communities

- **Informational tags**
  - Communities set by and sent from a provider network, to tell their customers (or other interested parties) something about that route
  - Encode simple arbitrary data
    - No standards, each network defines its own mapping
    - Which you must then publish somewhere for others to use
    - Ex: Continent
    - (1 = North America, 2 = Europe, etc)
    - Ex: Relationship (1 = Transit, 2 = Public Peer, etc)

Informational tags example

- For example: 1234:TCRPP
  - T = Type of Relationship
  - C = Continent Code
  - R = Region Code
  - PP = POP Code
- The community 1234:31311 could be parsed as:
  - Private Peer
  - North America
  - Mid-Atlantic Region
  - Ashburn VA POP Code

Usage of BGP Communities

- **Action tags**
  - Communities set by and sent from a customer network, to influence the routing policies of the provider network
- Two main types of actions exists
  - Export control (do or do not announce the prefix to X)
  - BGP Attribute manipulation
- Typical BGP attributes to be influenced include
  - AS-PATH
  - Local-Preference
  - Multi-Exit Discriminator (MED)
  - Next-Hop Address
- Anything else you can set in policy (color/weight/etc)

Action tags examples

- Often the first half specifies the target ASN
  - E.g. 1239:1234 = Apply code “1234” to neighbor ASN 1239

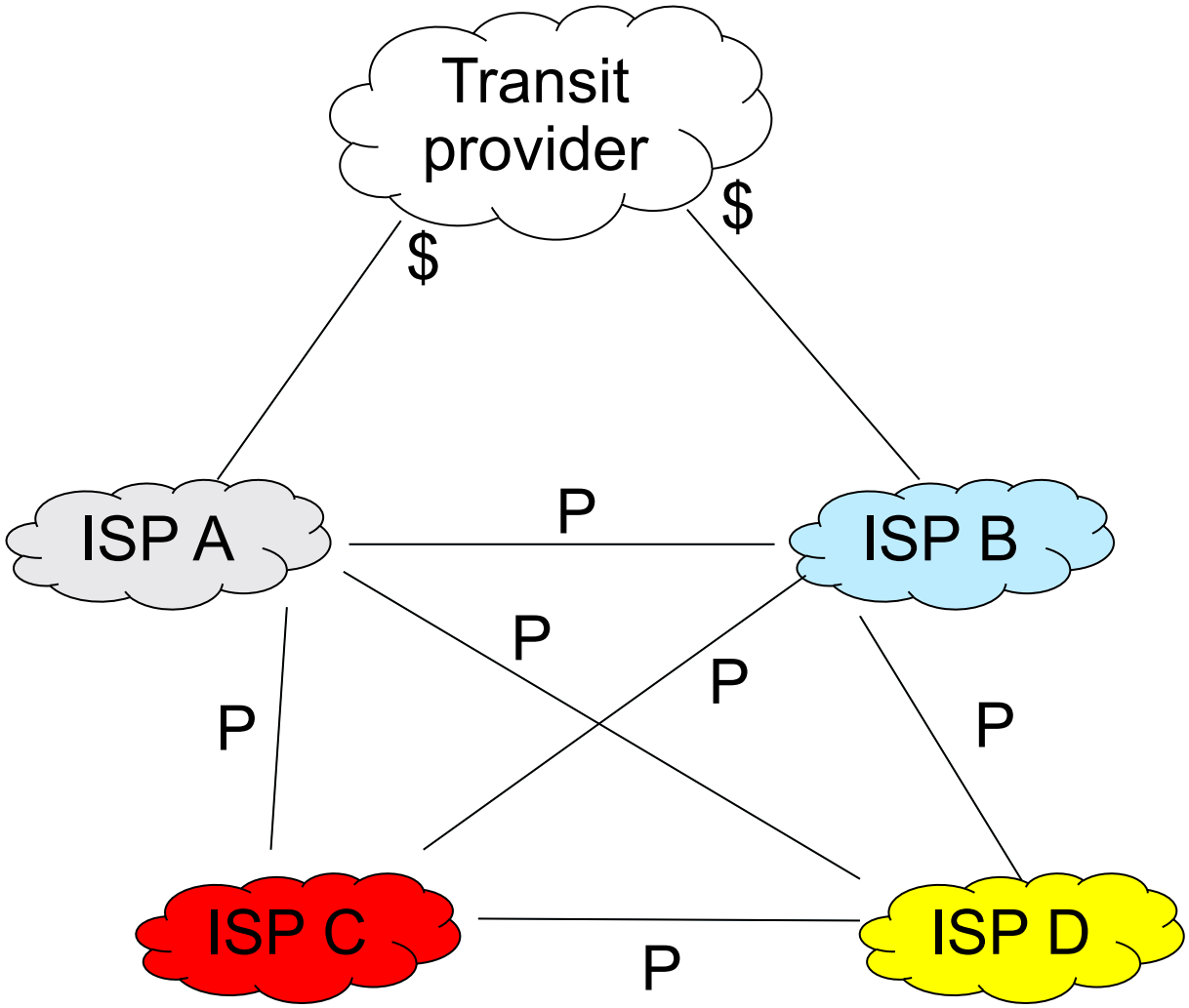
- Example of actions:
  - Apply action only to a specific geographic region
    - E.g. No-export to neighbors in North America, etc.
  - Apply action only to a specific relationship
    - E.g. Prepend to Peers, No-export to Transits, etc.
  - Apply action only to a specific neighbor ASN
    - E.g. Prepend to AS1234
  - The most powerful actions are combinations
    - E.g. No-export to Customers in North America

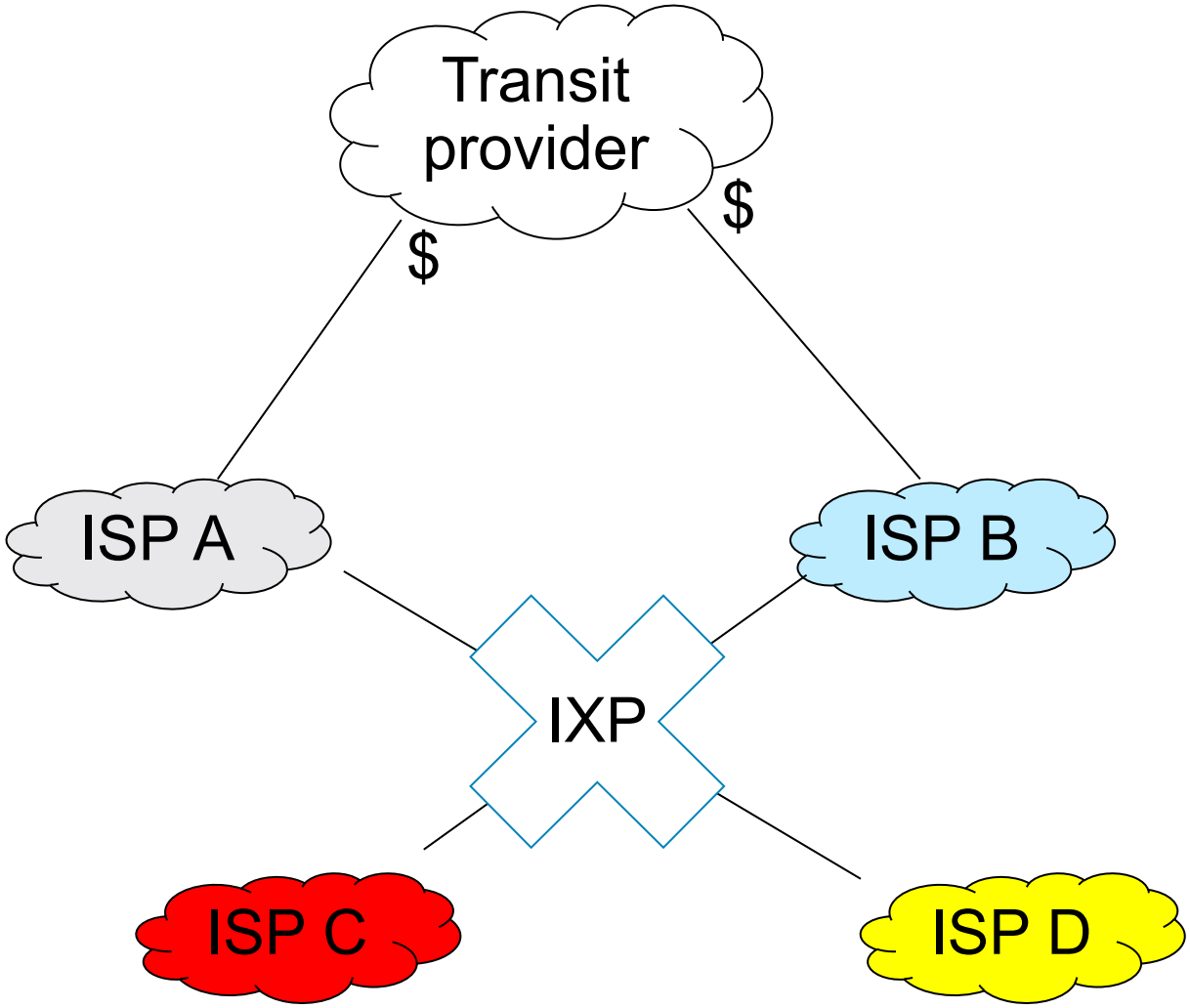
Action tags examples

- Often the first half specifies the target ASN
  - E.g. 1239:1234 = Apply code “1234” to neighbor ASN 1239

- Example of actions:
  - Apply action only to a specific geographic region
    - E.g. No-export to neighbors in North America, etc.
  - Apply action only to a specific relationship
    - E.g. Prepend to Peers, No-export to Transits, etc.
  - Apply action only to a specific neighbor ASN
    - E.g. Prepend to AS1234
  - The most powerful actions are combinations
    - E.g. No-export to Customers in North America

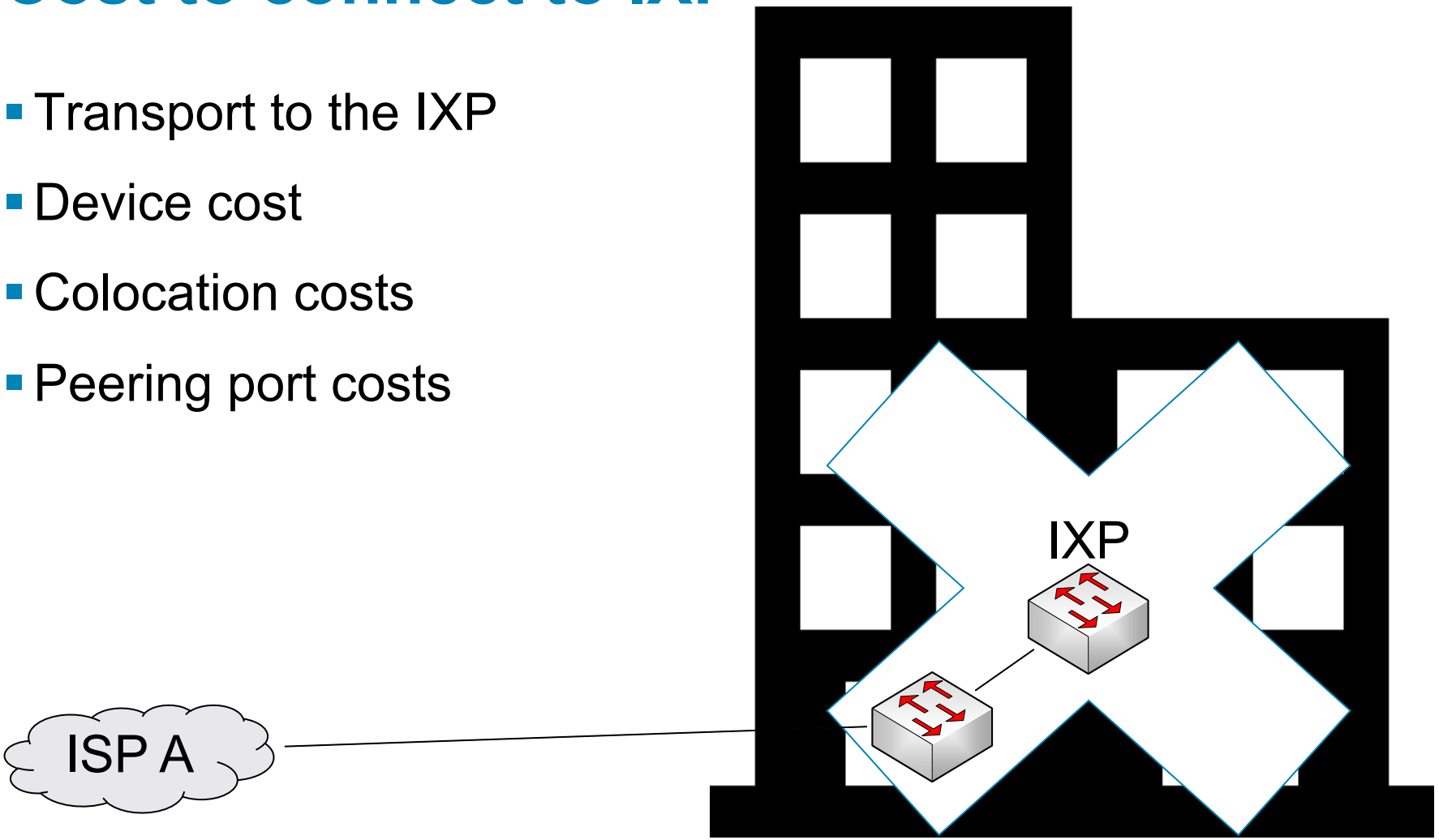
IXP


Transit
provider
$
$
ISP A
P
ISP B
P
P
P
P
ISP C
P
ISP D


Transit
provider
$
$
ISP A
ISP B
IXP
ISP C
ISP D

Cost to connect to IXP

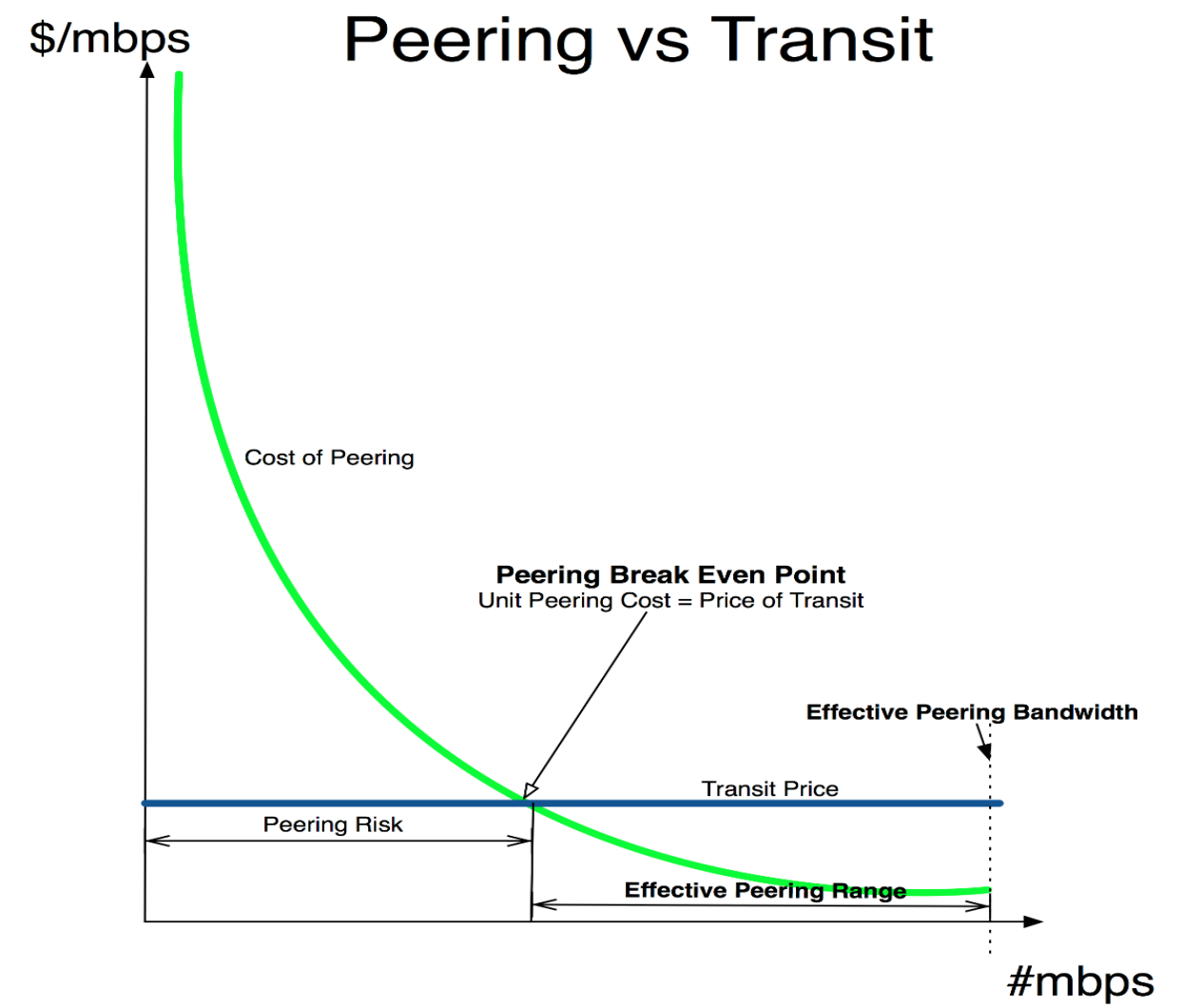
- Transport to the IXP
- Device cost
- Colocation costs
- Peering port costs



Example

- Transport + colocation + peering port cost + device price = $5 000/month
- How much traffic do I have to dispatch in IXP to be cost wise to use it? Let‘s assume 10 Gbps port.

Mbps	Price
100 Mbps	$50 per Mbps
1000 Mbps	$5 per Mbps
1200 Mbps	$4,17
1400 Mbps	$3.57
1500 Mbps	$3.34



Zdroj: drpeering.net

Public peering

- Using IXP switch for peering with everyone in IXP

- Benefits
  - Aggregation – large number of peering session use the same port
  - Easier maintain – rychle otestuji novou peering session
  - Not other cost
  - Can be cheaper in some cases
  - Typically use a link with higher capacity – can eliminates peaks in traffic

Private peerring

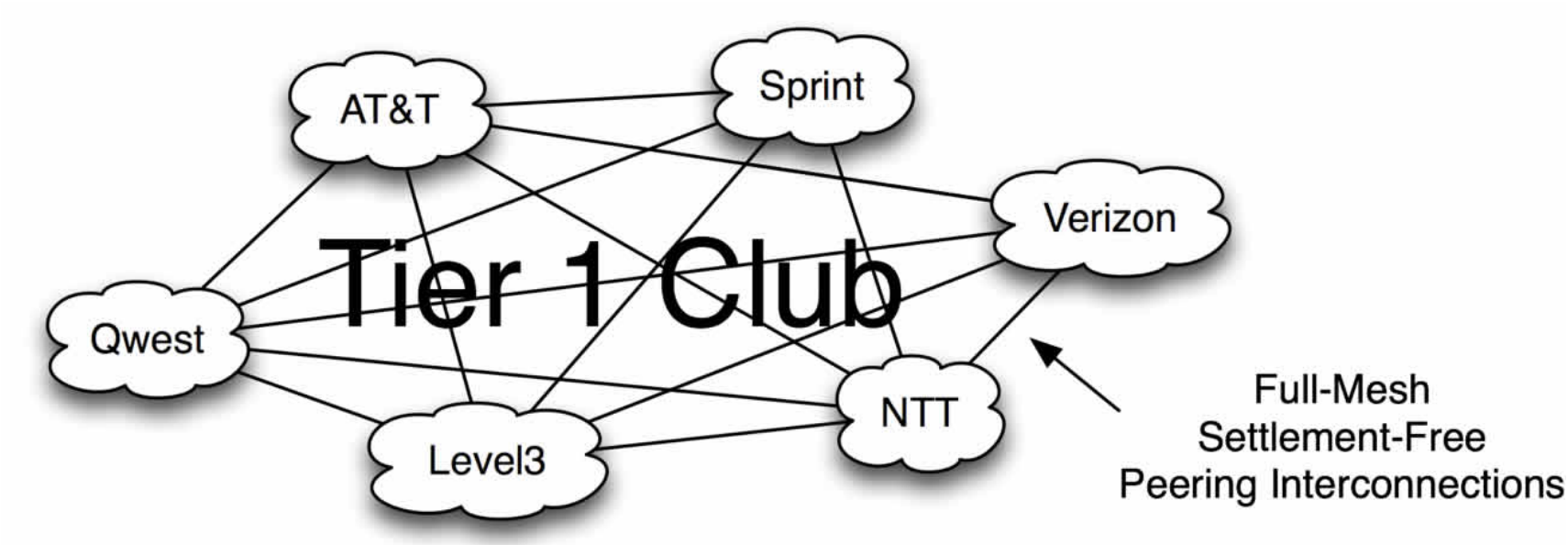
- Use dedicated circuit

- Benefits
  - Easier monitoring – SNMP is enough, NetFlow is not necessary
  - Easier for debuging – less devices, the debuging does not have to be coordinated with IXP operator
  - Security – there‘s no other device between two peers
  - Private peering is sometimes requested/required if traffic cross some threshold

Peering ecosystem

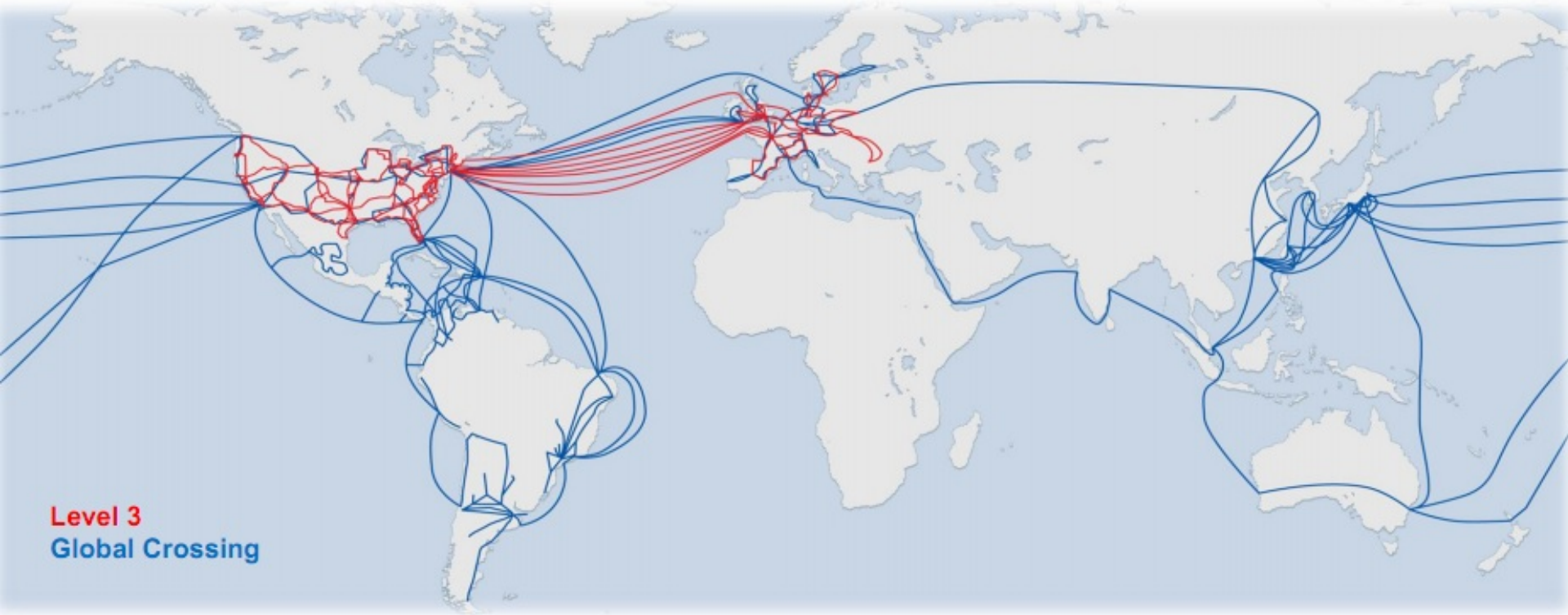
Tier I

- Operator that has access to all networks without paying for transit



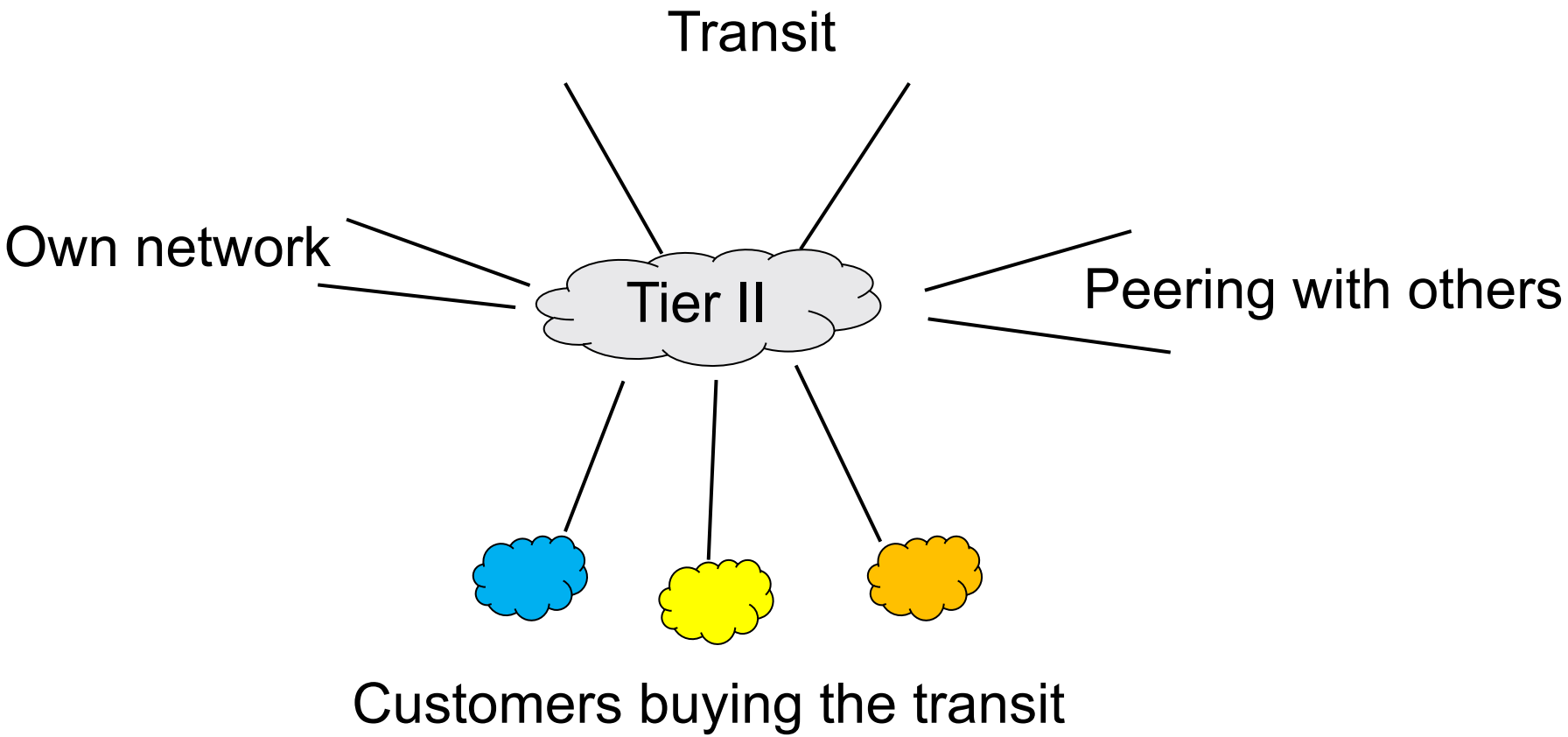
http://drpeering.net/

Tier I – Level3



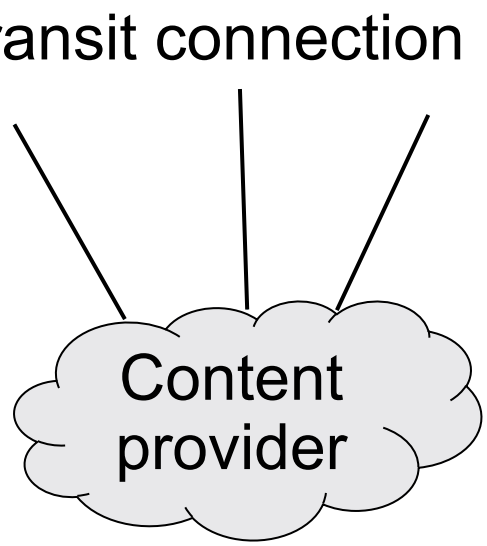
Tier II

- Pays for the transit to some networks in the Internet

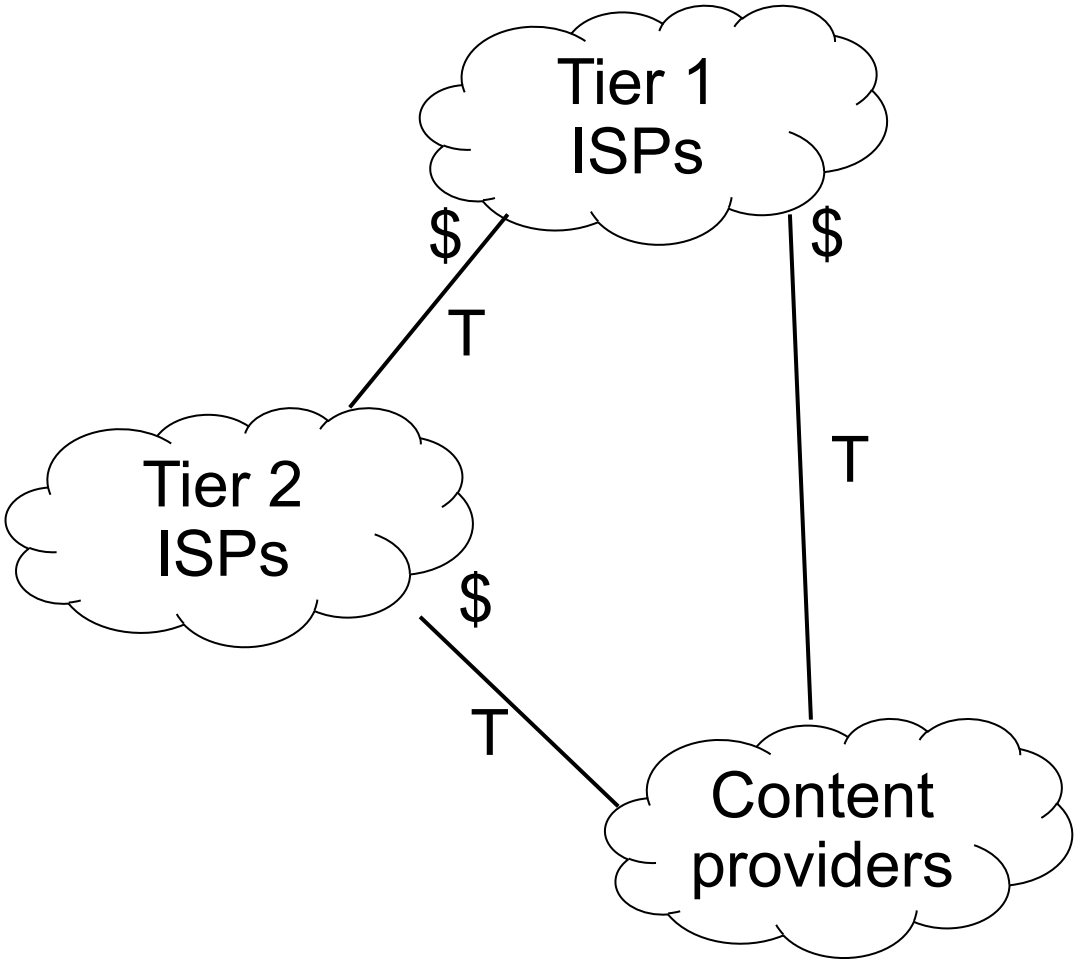


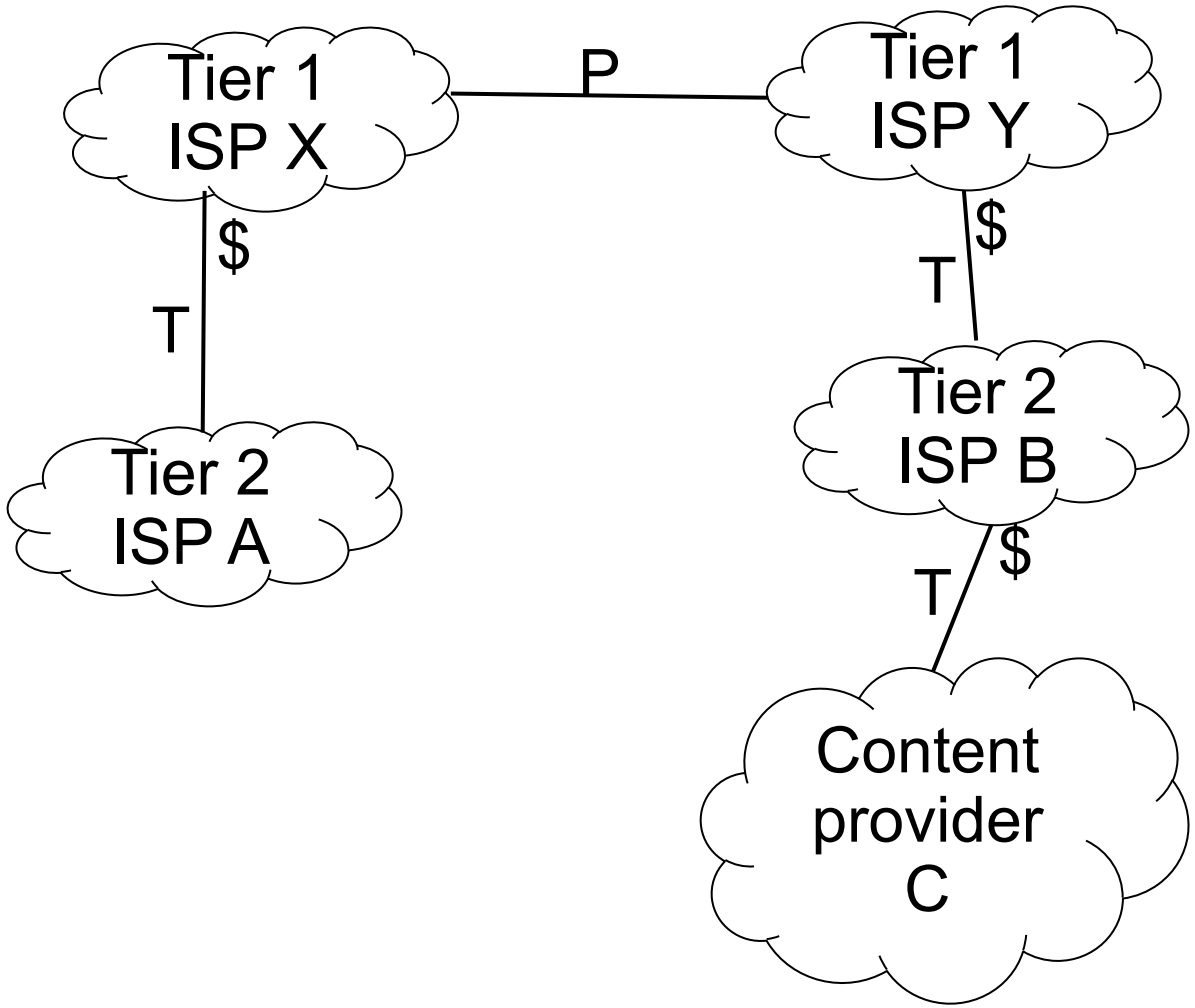
Content provider

- Typically do not peer, only buy transit
  - Exception is big content providers that build their own large content networks (Google, Apple, Microsoft)
- Main purpose is to create a content



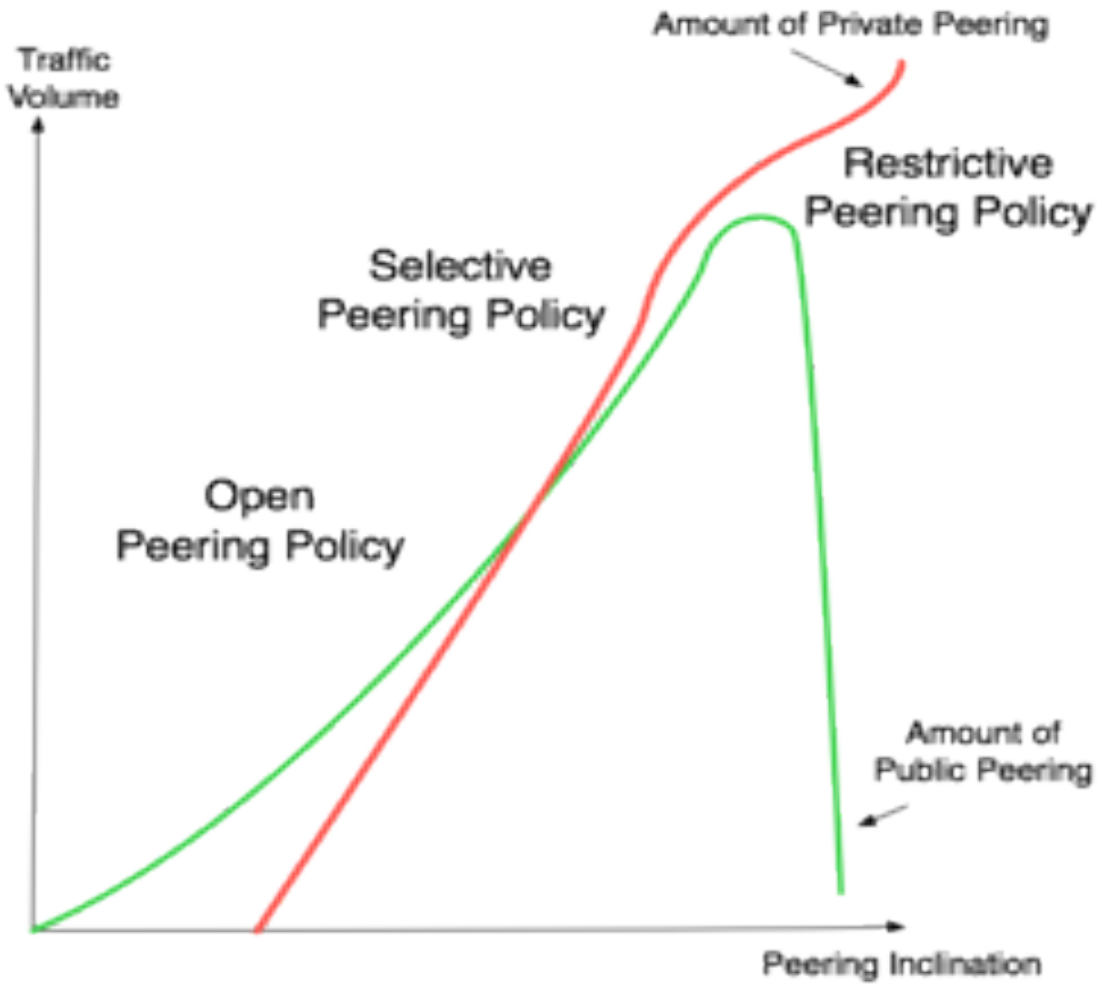
Internet





Will ISP Y accept a peering request from ISB B?
Will ISP Y accept a peering request from CP C?
Will ISP Y accept a peering request from ISP A?

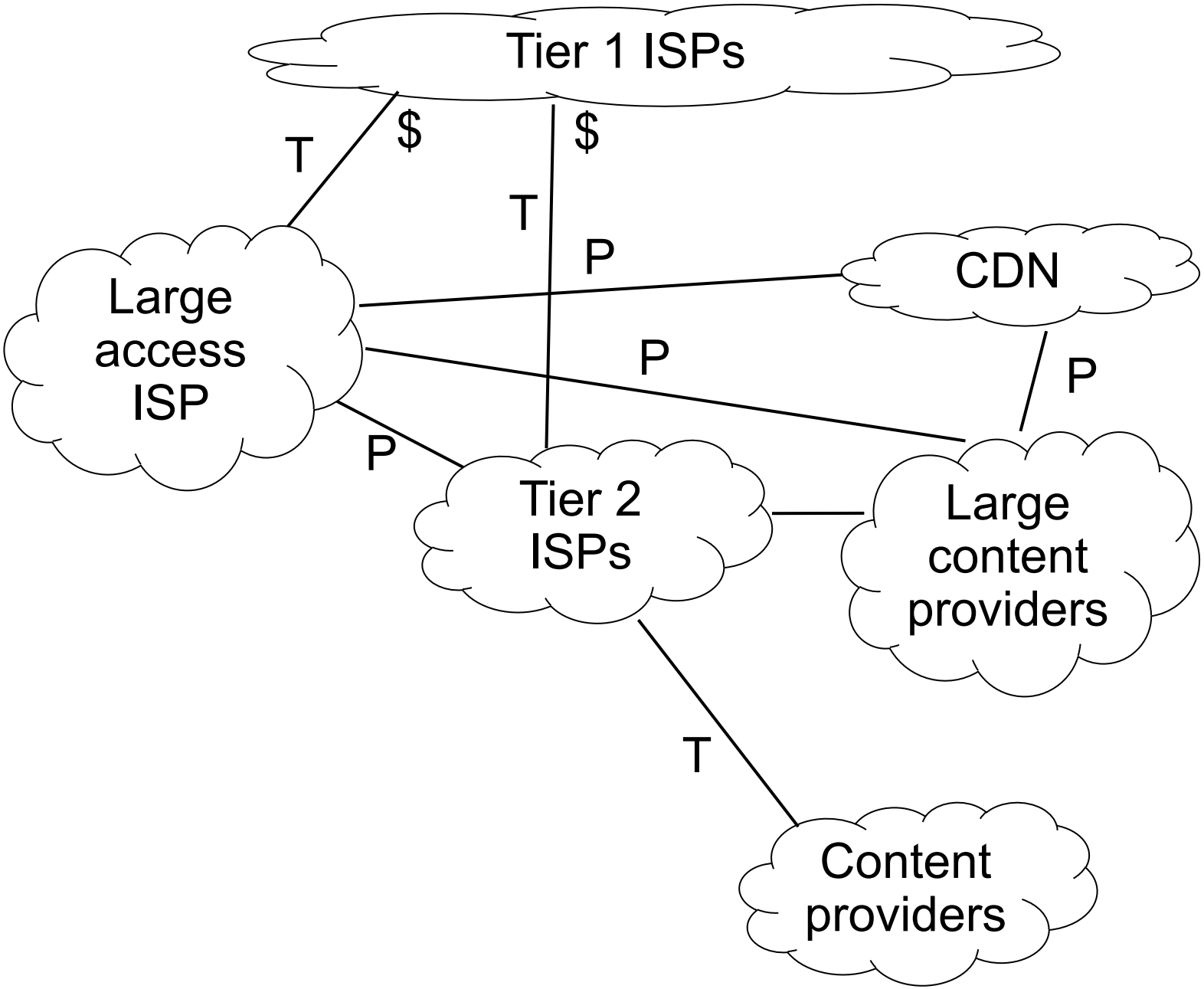
Peering Policy Lifecycle

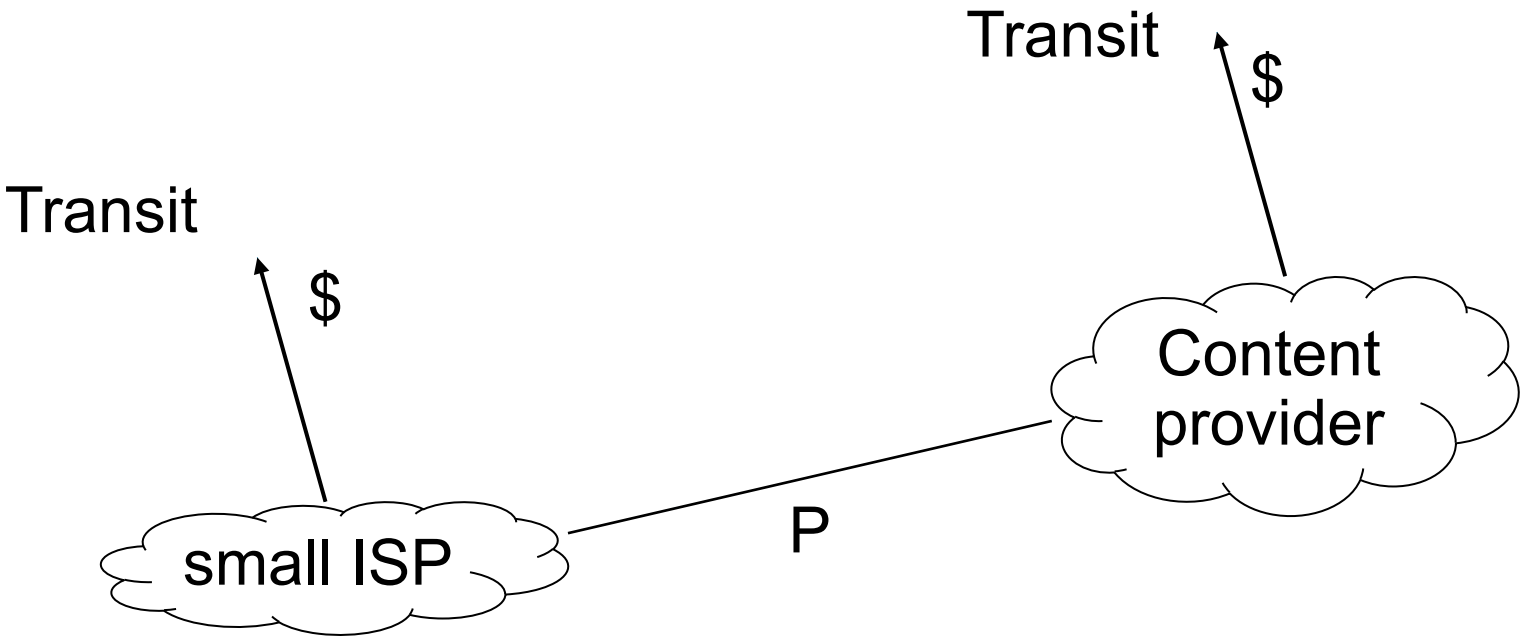


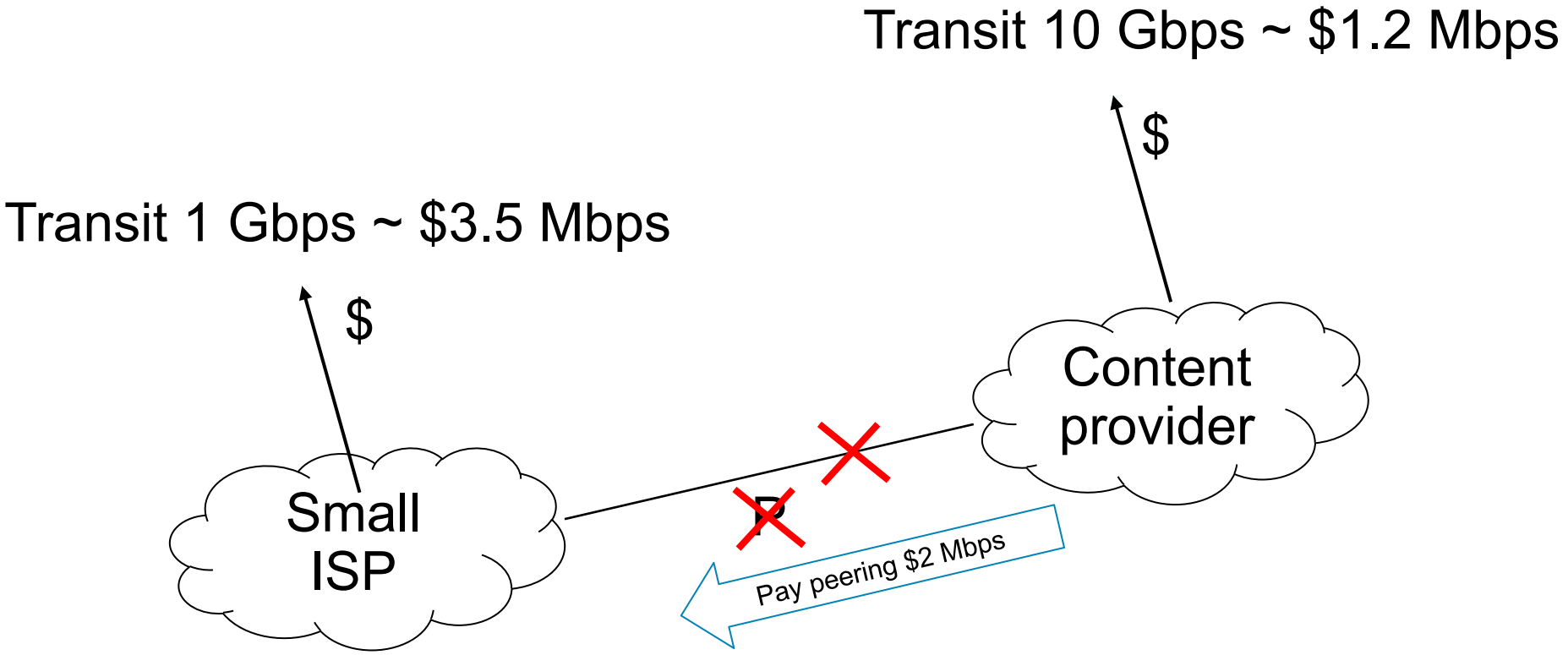
http://drpeering.net/

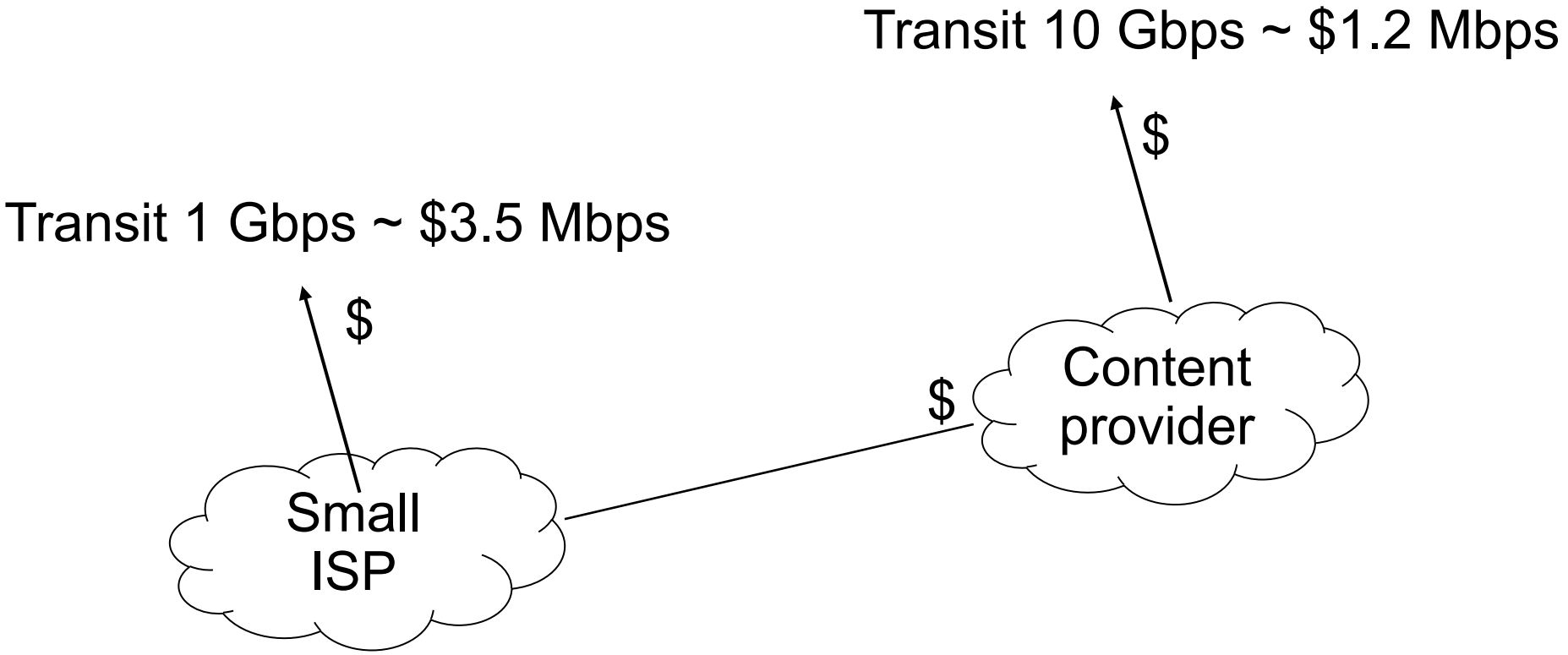
Changes in peering ecosystem

- Emerging of very large ISP with huge amount of end-users (eyball networks)
- Emerging of very large content providers that builts their own networks
  - Large Scale Network Savvy Content Providers
- Emerging of CDN
- Video!
- Emerging of pay private peering


Tier 1 ISPs
T
$
$
T
P
CDN
Large
access
ISP
P
P
P
Tier 2
ISPs
Large
content
providers
T
Content
providers


Transit
$
Transit
$
Content
provider
small ISP
P


Transit 10 Gbps ~ $1.2 Mbps
$
Transit 1 Gbps ~ $3.5 Mbps
$
Content provider
Small ISP
P
Pay peering $2 Mbps


Transit 10 Gbps ~ $1.2 Mbps
$
Transit 1 Gbps ~ $3.5 Mbps
$
$
Content
provider
Small
ISP

SuperProblémy se Super(N

www.root.cz/clanky/superproblemy-se-super-network-hosting-em/

SuperProblémy se Super(Network|Hosting)em

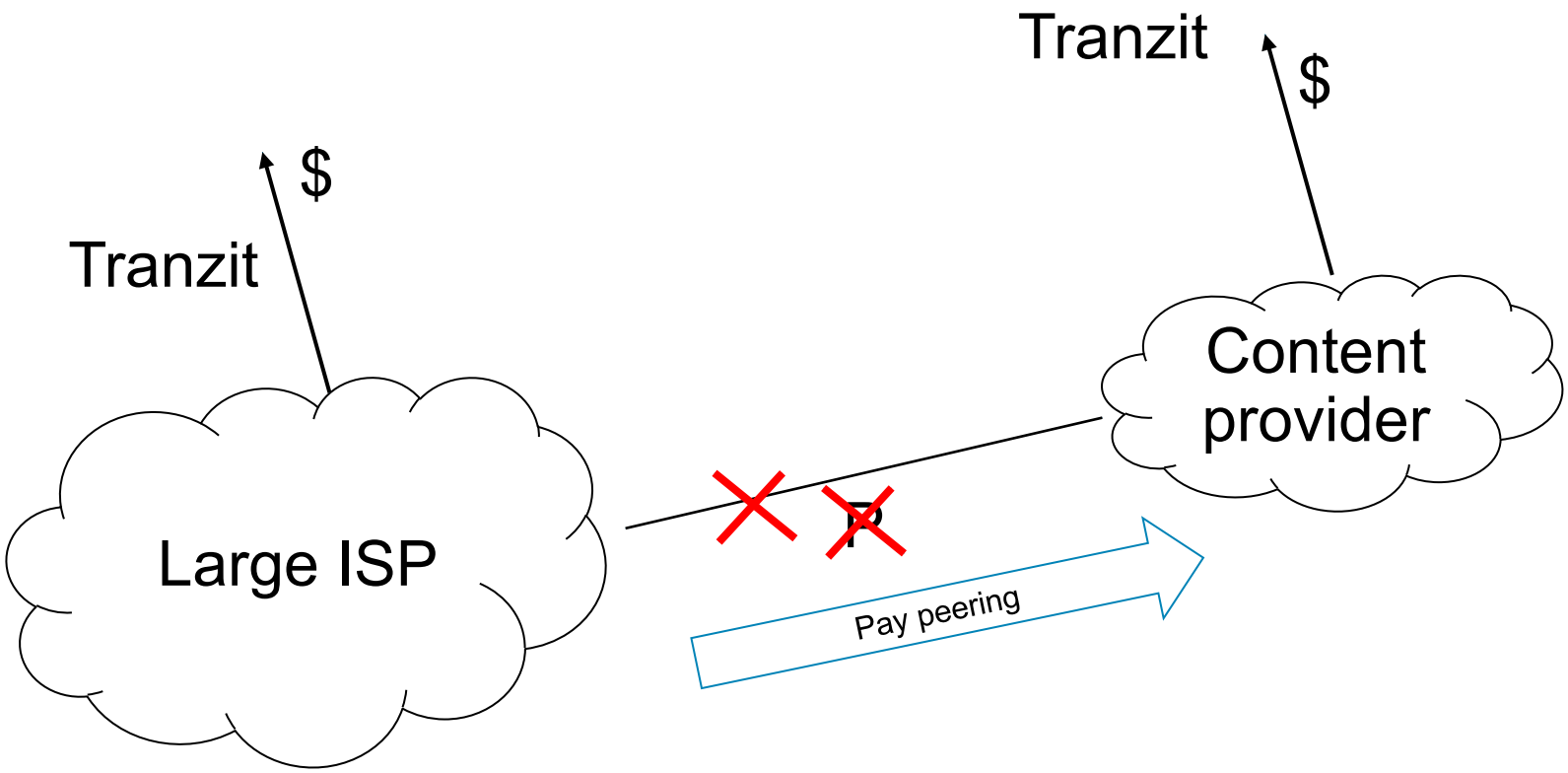
(Aktualizováno) Málokdo z běžných konzumentů internetových služeb tuší, jaké tuhé boje probíhají o jejich pakety ve struktuře sítě internet. Idylické doby vzájemného porozumění mezi správci síťové infrastruktury zaštítěné síťovou neutralitou často berou za své a nastupuje prosazování tvrdých komerčních zájmů.

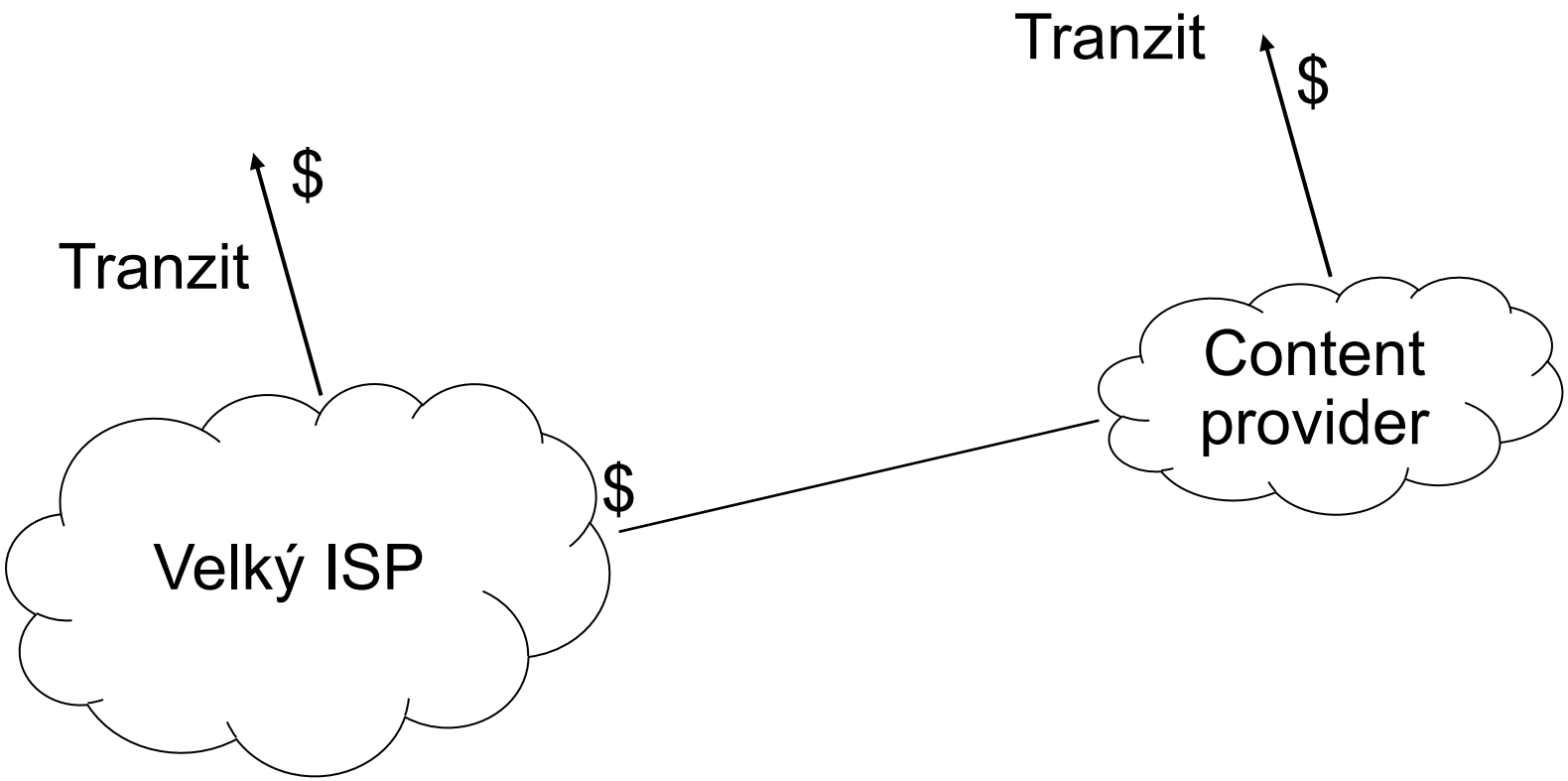
9. 9. 2013 0:00 Adam Covex Přibyl

Sdílet na Facebooku | Odeslat na Twitter | Sdílet na Google+

Nálepky: NIX.CZ

Jedním z hráčů na tomto trhu, který si nebere servítky, je právě SuperNetwork. SuperNetwork je již několik let znám svými aktivitami, kterými se v podstatě snaží obcházet sdružení NIX, kterého je sám členem. Jen krátce – zájmové sdružení právnických osob NIX (Neutral Internet eXchange) sdružuje velké hráče na trhu poskytování internetového připojení a obsahu. Trochu zjednodušeně se dá říci, že si v rámci něj bezplatně vyměňují své datové toky, protože vzájemné účtování by často skončilo nulou


Tranzit
$
Tranzit
$
Content
provider
Large ISP
P
Pay peering


Tranzit
$
$
Tranzit
Content
provider
$
Velký ISP

Proč nebyla Lupa včera do × | Proc se dnes nemusite do ×

www.lupa.cz/clanky/proc-nebyla-lupa-vcera-dostupna-ze-site-upc/

LUPA.cz Server o českém Internetu

Lupa.cz do mailu

Spam „Výše pohledávky na vašem účtu" | Konec True

Články Startupy Měření rychlosti Nástroje Fórum Blogy Práce v IT E-commerce

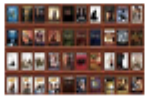
E-government

Nepíšete berňáku elektronicky? Přijde automatická pokuta

SEO

Seznam.cz se probudil a jeho aktualizace hýbe pozicemi

Video na Internetu



Distributoři: diváci pořád chtějí filmy spíš na BluRay a DVD

Telemedicína

Teleme proráže ji zdrav

Proč nebyla Lupa včera dostupná ze sítě UPC?

Výpadek peeringu, který zákazníkům UPC znepřístupnil Lupu a další servery v síti Master Internetu, nebyl technickou závadou. Šlo o důsledek obchodního sporu, který se týká zpoplatnění peeringu.

2. 9. 2011 6:30 Jiří Peterka

BitTorrent S zjednodušuj souborů a tý spolupráci

29. 8. 2014 | Jiří

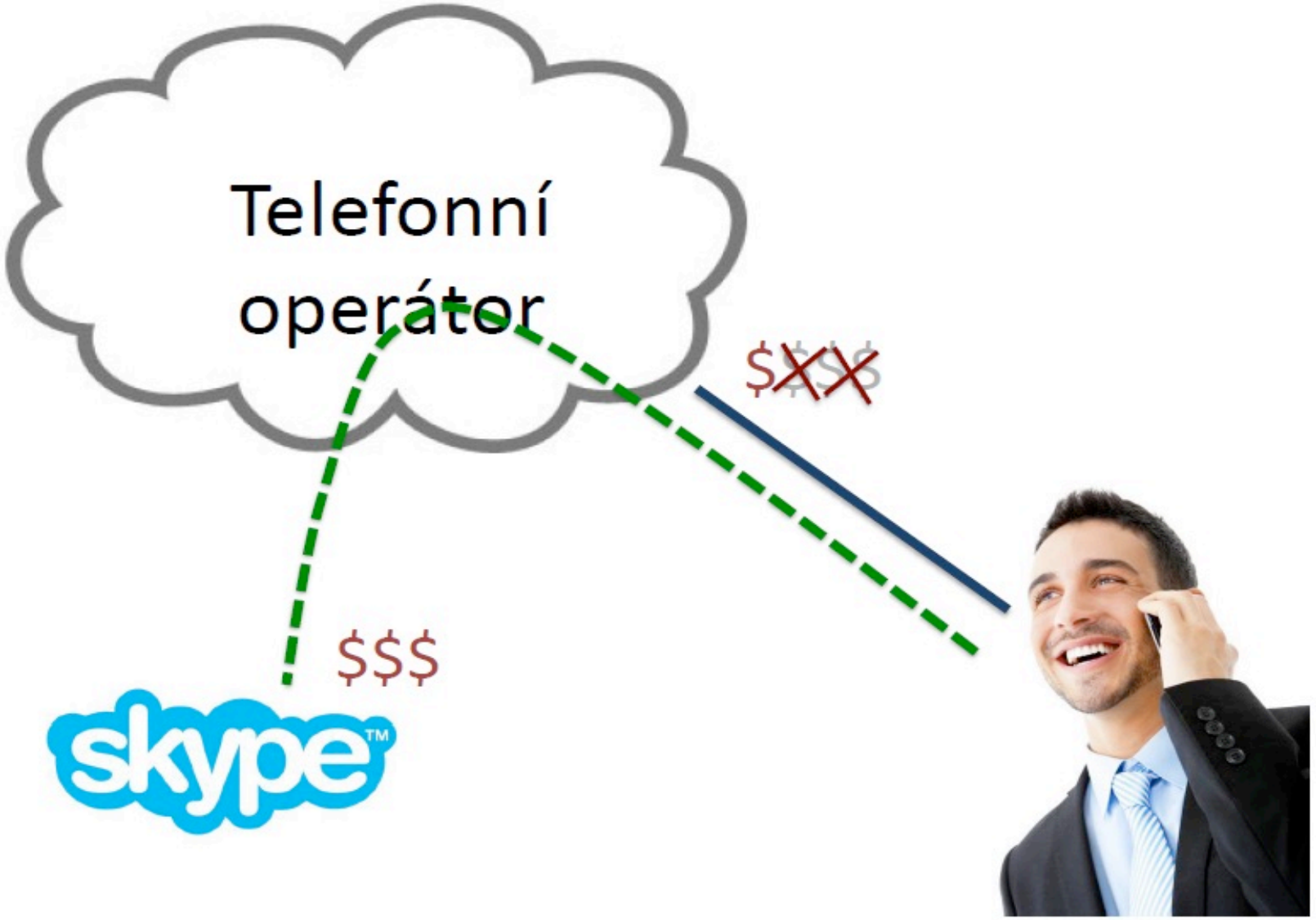
3 nové názory

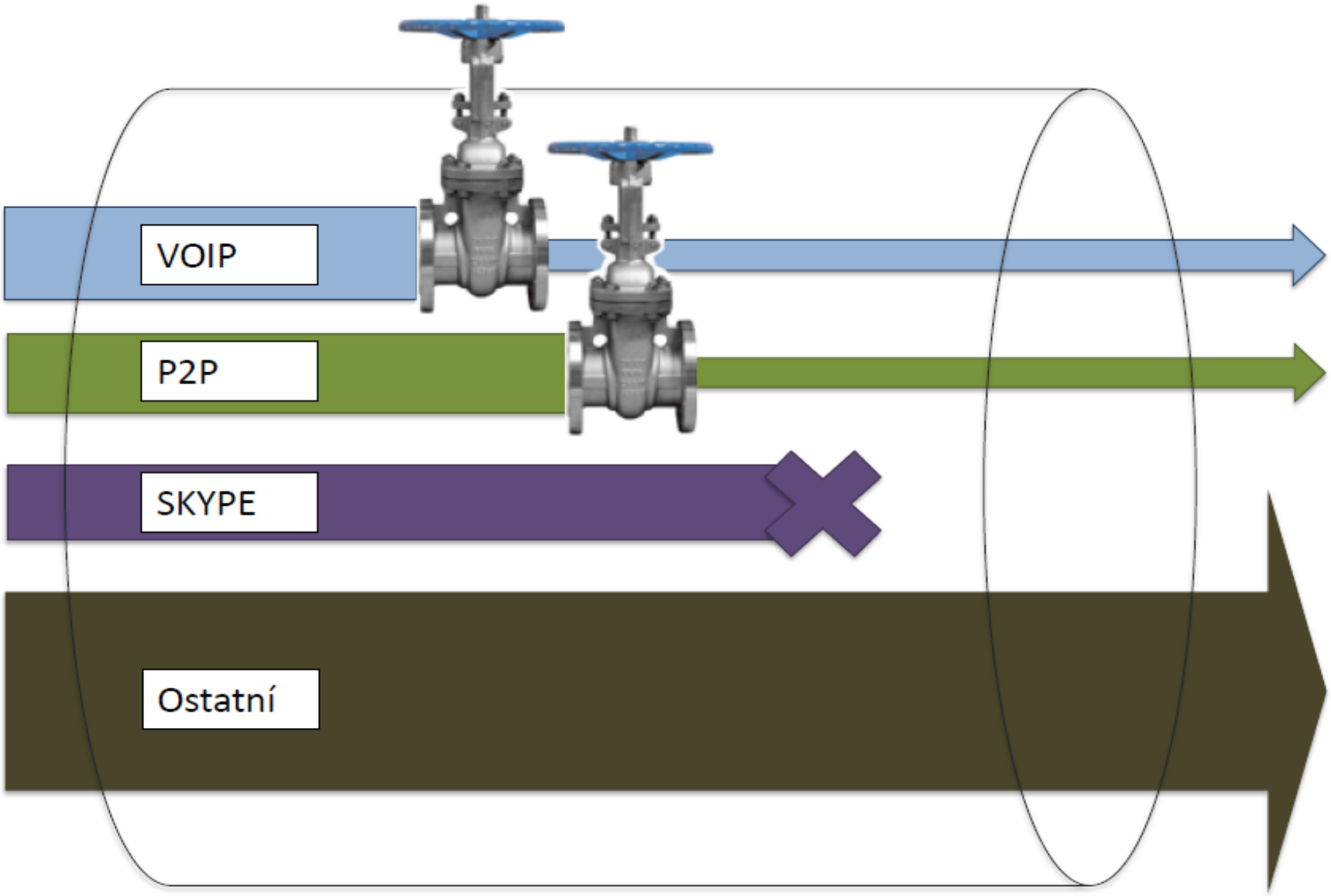
Samsung do telefonů s A

Síťová neutralita

Telefonní operátor

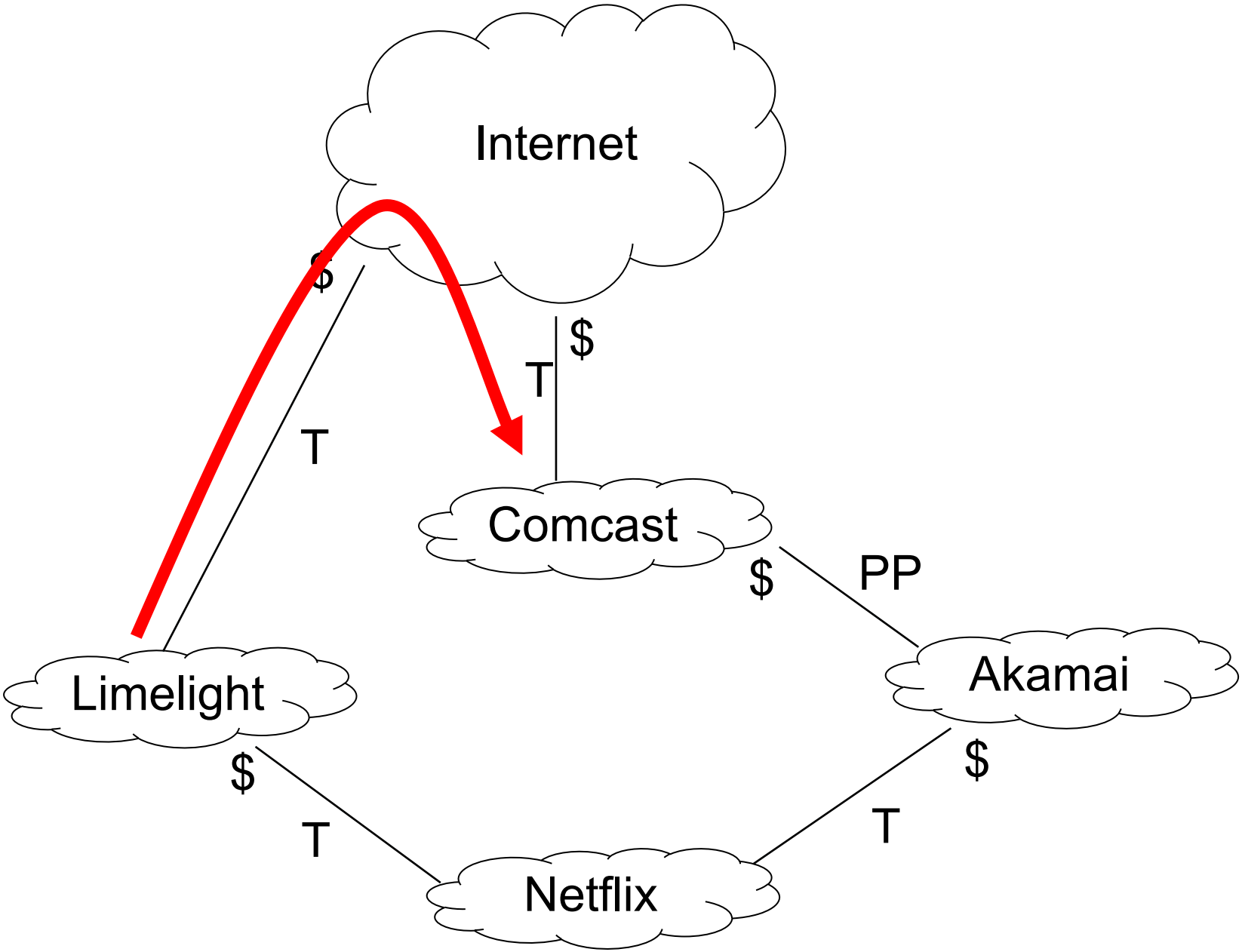
$$$$

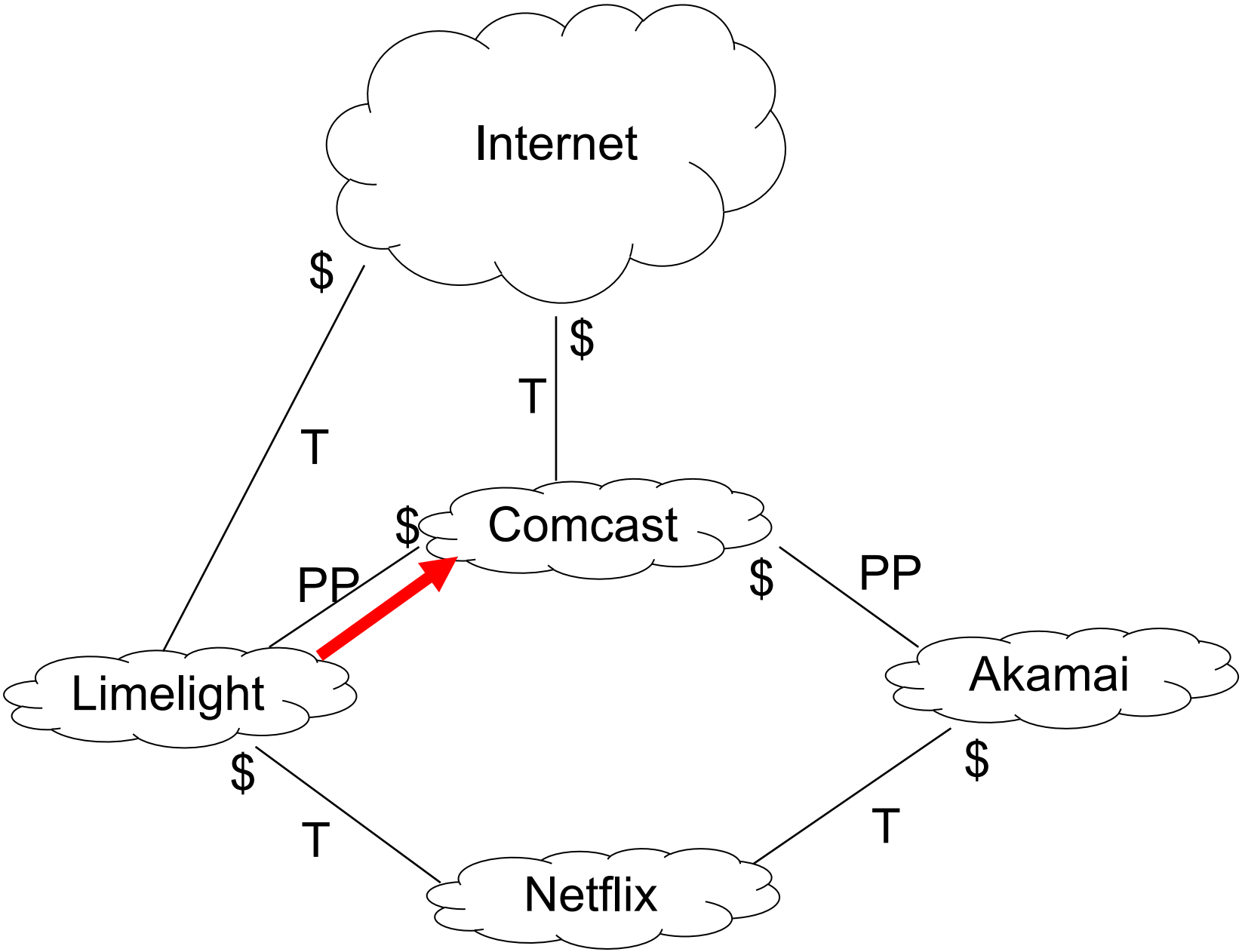

Telefonní operátor
$$$
$$$
skype™

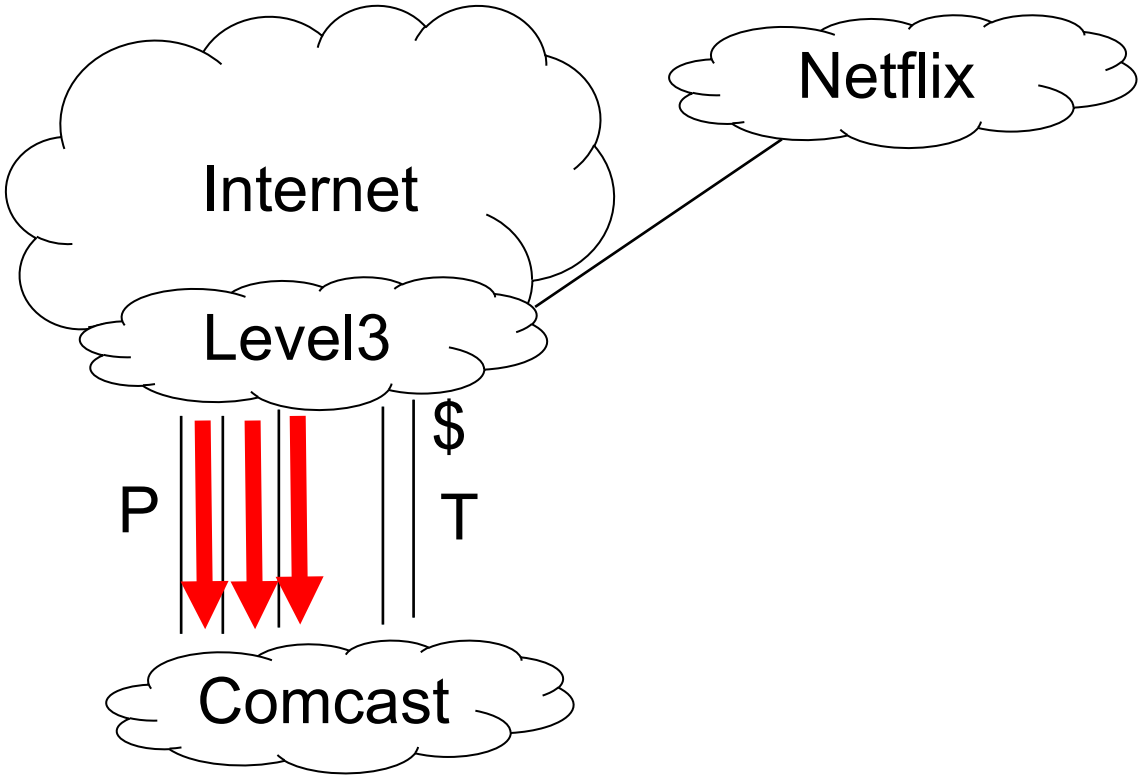

VOIP
P2P
SKYPE
Ostatní

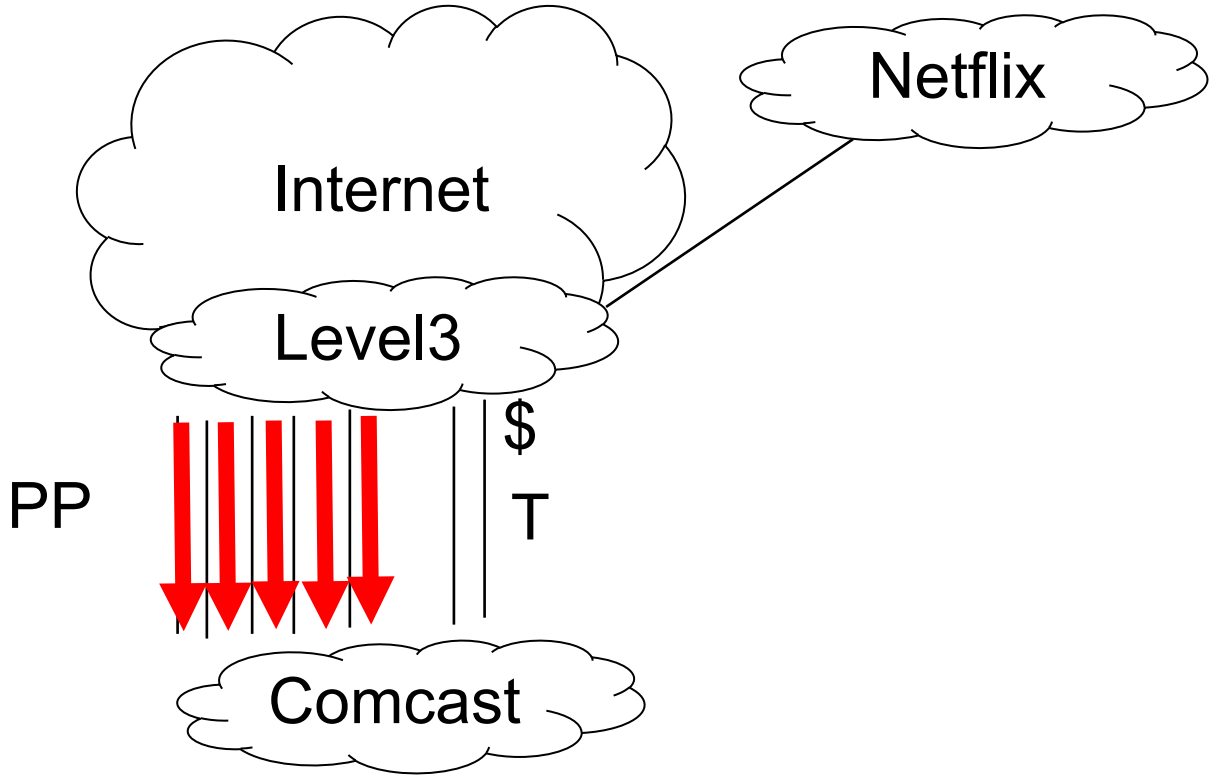

NETWORK
NEUTRALITY

Hráči


Internet
$
$
T
T
Comcast
$
PP
Akamai
Limelight
$
T
$
T
Netflix


Internet
$
$
T
T
$
Comcast
PP
$
PP
Limelight
Akamai
$
$
T
T
Netflix

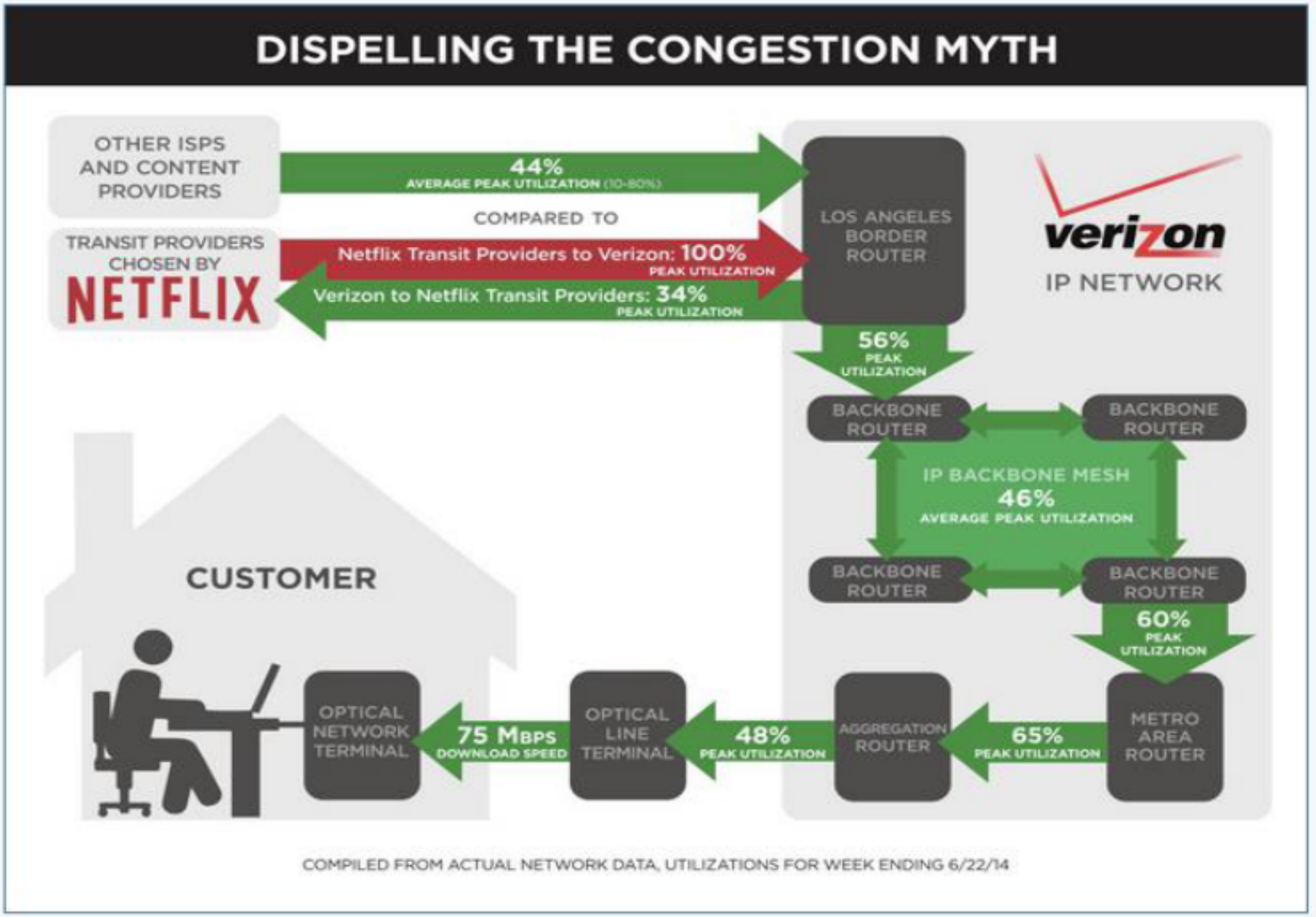

Netflix
Internet
Level3
$
P
T
Comcast

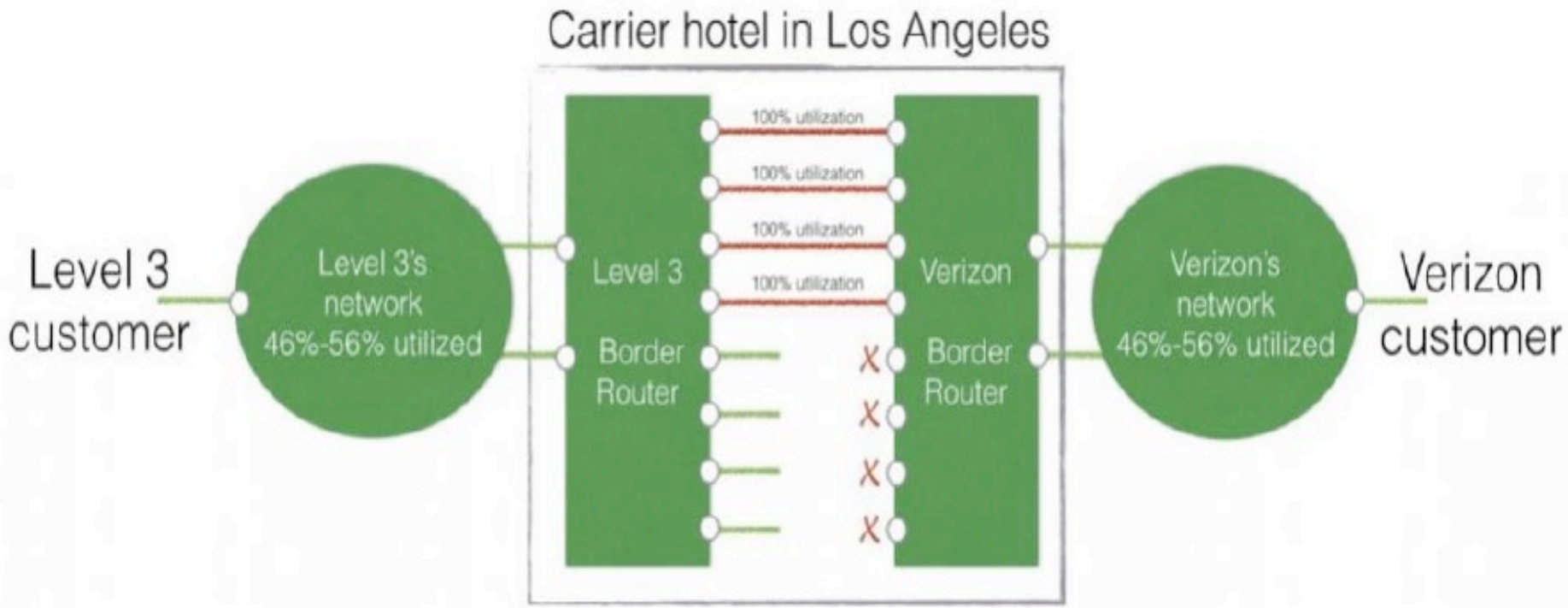

Netflix
Internet
Level3
$
PP
T
Comcast


Netflix
Hulu
Gaming
ISP Content & Services
Email
Facebook
Skype
YouTube
Amazing!
Access for
$9.99/mo
Package Sale!
Access both for just
$29.99/mo
A Steal!
Get Youtube
for only
$9.99/mo


STOP
CABLE
COMPANY
FUCKERY
KILL the
INTERNET
MoveOn.org
CREDO
dc408


KEEP THE
INTERNET
FREE
CODEPINK.ORG


DISPELLING THE CONGESTION MYTH
OTHER ISPS AND CONTENT PROVIDERS
44%
AVERAGE PEAK UTILIZATION (10-80%)
COMPARED TO
TRANSIT PROVIDERS CHOSEN BY
NETFLIX
Netflix Transit Providers to Verizon: 100%
PEAK UTILIZATION
Verizon to Netflix Transit Providers: 34%
PEAK UTILIZATION
LOS ANGELES BORDER ROUTER
verizon
IP NETWORK
56%
PEAK UTILIZATION
BACKBONE ROUTER
BACKBONE ROUTER
IP BACKBONE MESH
46%
AVERAGE PEAK UTILIZATION
BACKBONE ROUTER
BACKBONE ROUTER
60%
PEAK UTILIZATION
CUSTOMER
OPTICAL NETWORK TERMINAL
75 MBPS
DOWNLOAD SPEED
OPTICAL LINE TERMINAL
48%
PEAK UTILIZATION
AGGREGATION ROUTER
65%
PEAK UTILIZATION
METRO AREA ROUTER
COMPILED FROM ACTUAL NETWORK DATA, UTILIZATIONS FOR WEEK ENDING 6/22/14


Carrier hotel in Los Angeles
Level 3 customer
Level 3's network 46%-56% utilized
Level 3
Border Router
100% utilization
100% utilization
100% utilization
100% utilization
Verizon
Border Router
Verizon's network 46%-56% utilized
Verizon customer