

VYSOKÉ UČENÍ FAKULTA
TECHNICKÉ INFORMAČNÍCH
V BRNĚ TECHNOLOGIÍ



6

Směrování v Internetu

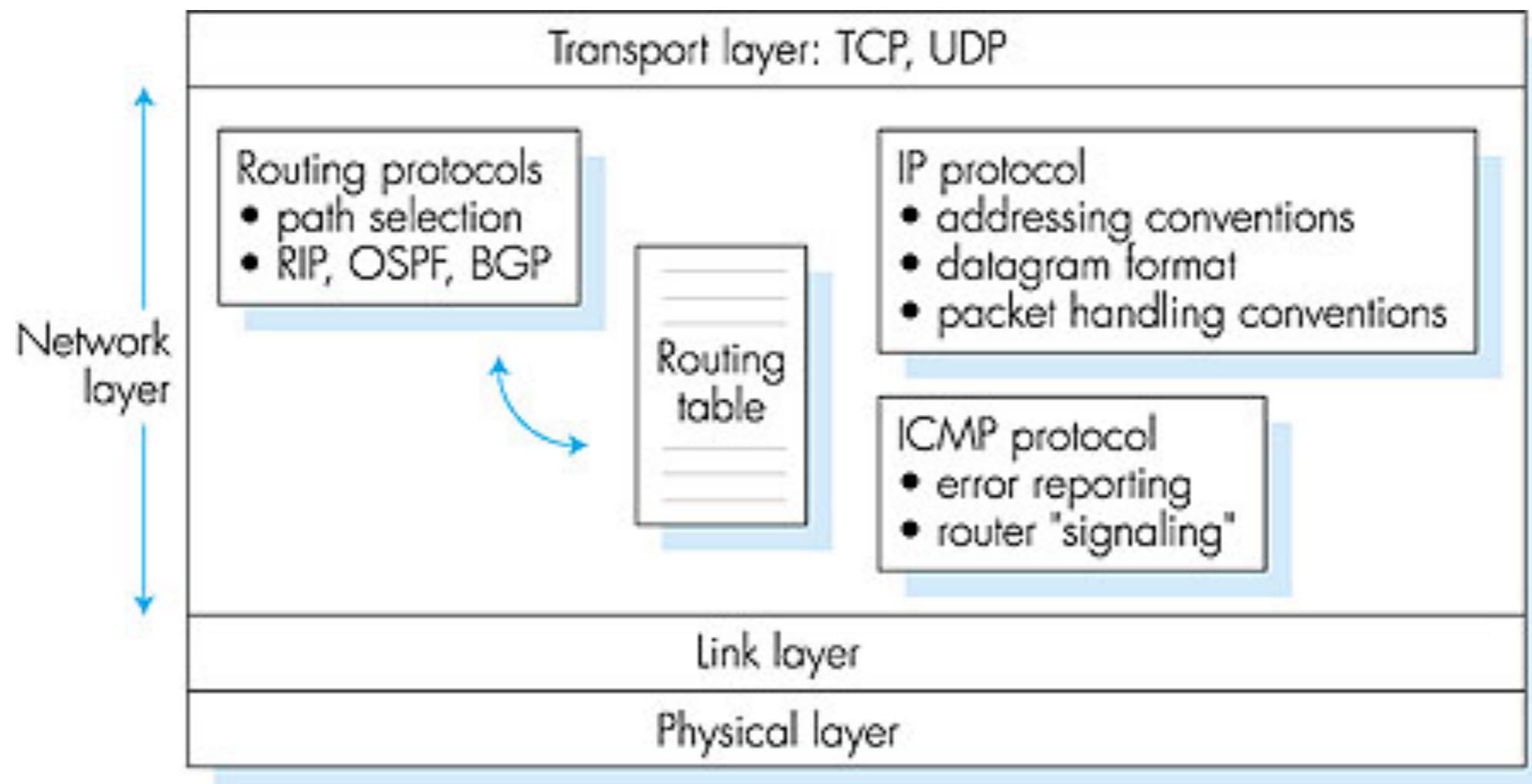
IPK2020L
~~**IPK2018L**~~

Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Bellman-Ford
 - Algoritmus Dijkstra
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

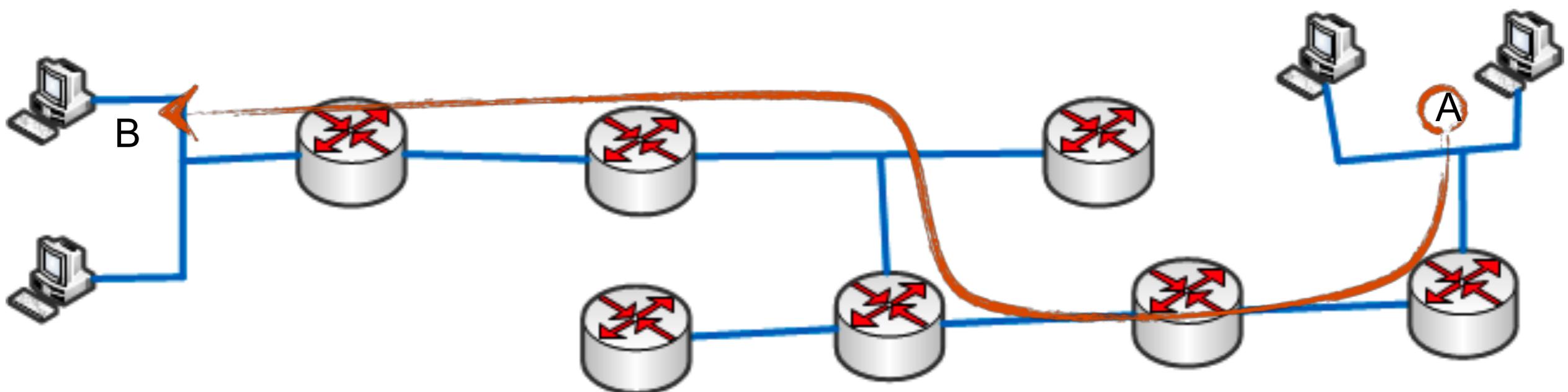
Směrování a síťová vrstva

- Síťová vrstva doručuje pakety koncovým zařízením
- Síťová vrstva zajišťuje šíření směrovacích informací v Internetu pomocí směrovacích protokolů



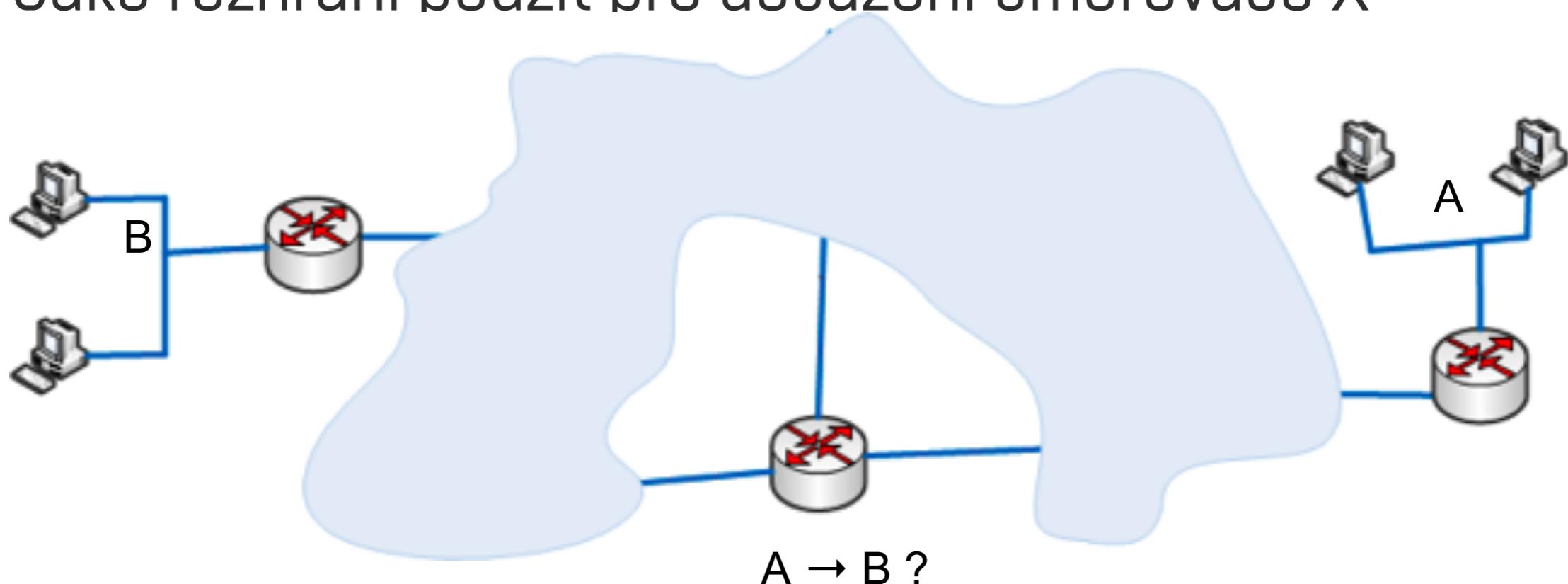
Co je směrování?

↔ Najít vhodnou cestu (tj. sekvenci směrovačů) od zdroje A k cíli B.



Co je směrování?

- Směrovače v síti potřebují vědět:
 - Jaký směrovač X použít pro dosažení cíle B
 - Jaké rozhraní použít pro dosažení směrovače X



Přepínání a směrování

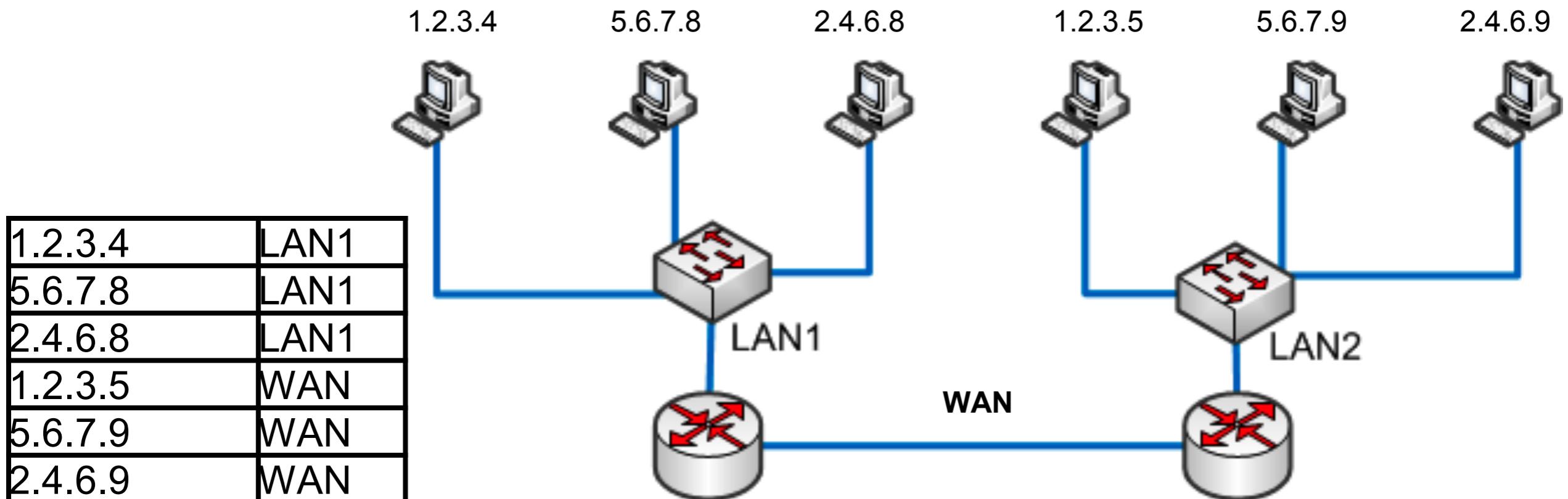
- **Přepínání (forwarding)**
 - směrovač přesune paket, který přichází na vstupní linku na odpovídající výstup
 - lokální akce prováděná směrovačem
- **Směrování (routing)**
 - síťová vrstva musí nalézt cestu pro paket
 - směrovací algoritmus
 - síťový proces, forma distribuovaného výpočtu, na kterém participují směrovače v síti
- Směrovač používá **směrovací tabulku** (forwarding/routing table)

Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Dijkstra
 - Algoritmus Bellman-Ford
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

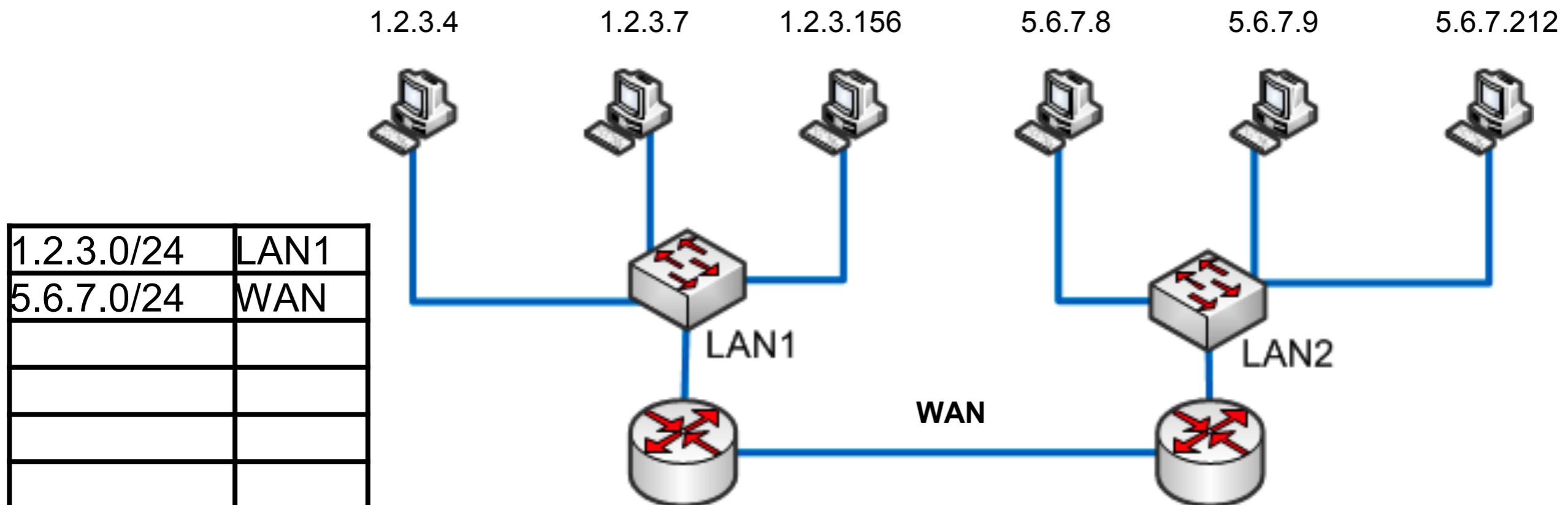
Přepínání paketů

- Směrovač má v tabulce informaci pro každou IP adresu
 - testování shody pro cílovou adresu v paketu
 - jednoznačné určení výstupního rozhraní
- *Velikost směrovacích tabulek!*



Přepínání paketů

- Směrovač má v tabulce informaci pro 24-bitový prefix
 - testování shody pro prefix cílové adresy v paketu
 - vyžaduje rozdělení IP adresy na síťovou část a uzlu
- *Zmenšení velikosti tabulek:) fixní délka prefixu:(*



Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Bellman-Ford
 - Algoritmus Dijkstra
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP

Třídní směrování

- angl. Classful routing
- Položky RT jsou JEN síťové adresy
 - třída adresy definuje prefix (8, 16, 24)
 - třídu lze poznat podle nejvyšších bitů
 - Classful routing protokol (např. RIPv1 či IGRP neposílá masku sítě)

Address Class	Bit Pattern of First Byte	First Byte Decimal Range	Host Assignment Range in Dotted Decimal
A	0xxxxxxxx	1 to 127	1.0.0.1 to 126.255.255.254
B	10xxxxxx	128 to 191	128.0.0.1 to 191.255.255.254
C	110xxxxx	192 to 223	192.0.0.1 to 223.255.255.254
D	1110xxxx	224 to 239	224.0.0.1 to 239.255.255.254
E	11110xxx	240 to 255	240.0.0.1 to 255.255.255.255

Beztrídní směrování (1)

- angl. Classless routing
- Položky RT jsou síťové adresy + maska sítě
- Význam masky sítě (subnet mask):
 - 1 = bit je součástí NetID
 - 0 = bit je součástí HostID

158	193	138	40
10011110	11000001	10001010	00101000
11111111	11111111	11111111	00000000

Beztrídní směrování (2)

- Hranice mezi NetId a HostId nemusí být na bytech
- Router porovnává dstIP s každým záznamem v RT (součástí čehož je i vymaskování dstIP SM)
- $158.193.138.40 \ \& \ 255.255.255.224 = 158.193.138.32$

10011110	11000001	10001010	00101000
AND			
11111111	11111111	11111111	11100000
=			
10011110	11000001	10001010	00100000

Příklad

- Na vstupu:

Packet
address:

1.2.3.4 0000 0001.0000 0010.0000 0011.0000 0100

Routing
Table:

1.0.0.0/8 0000 0001.xxxx xxxx.xxxx xxxx.xxxx xxxx

2.0.0.0/8 0000 0010.xxxx xxxx.xxxx xxxx.xxxx xxxx

3.0.0.0/8 0000 0011.xxxx xxxx.xxxx xxxx.xxxx xxxx

216.58.201.0/24 1101 1000.0011 1010.1100 1001.xxxx xxxx

CIDR

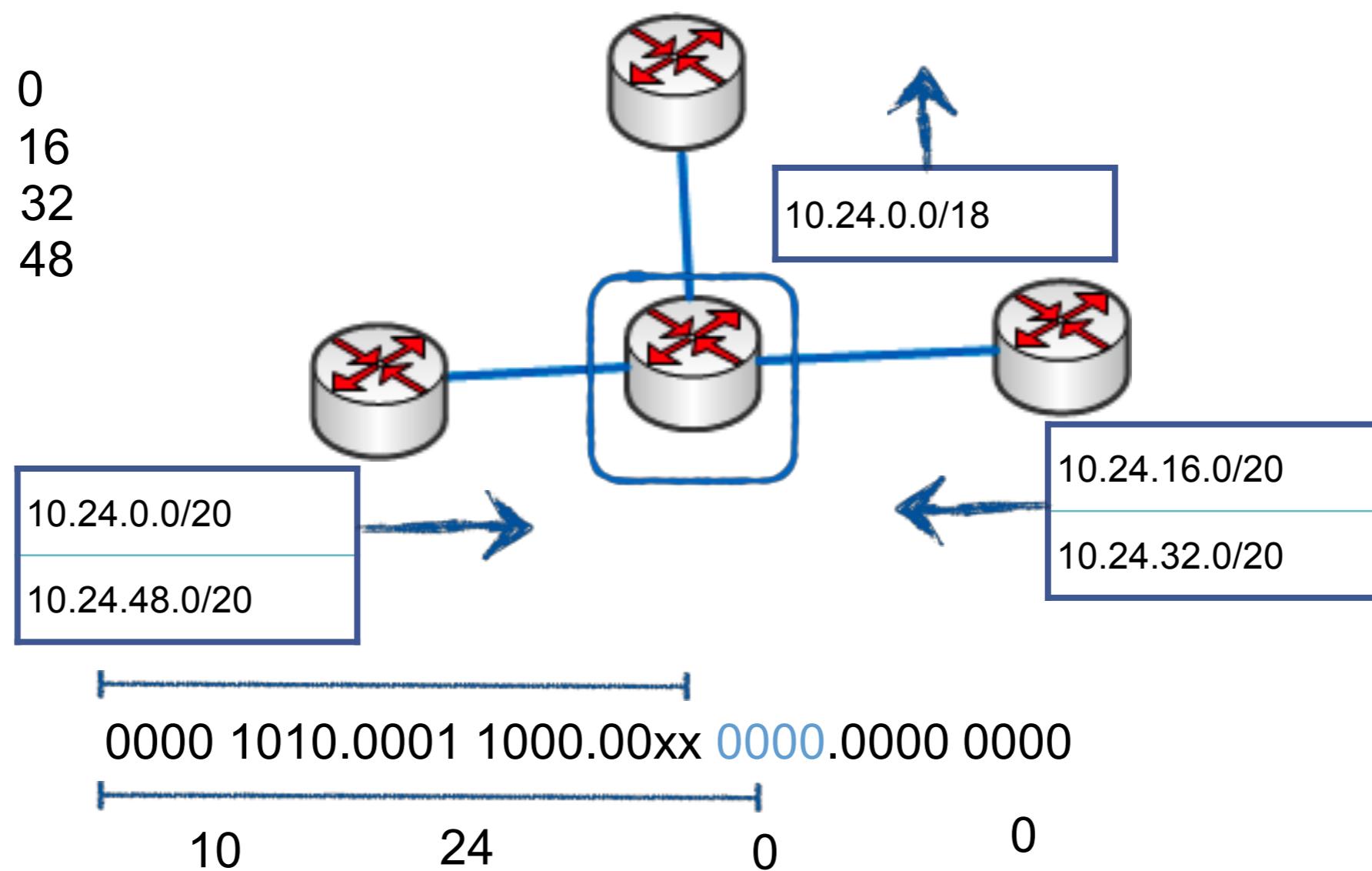
- Počet možných sítí jednotlivých tříd:
 - Třída A: 127
 - Třída B: 16 384
 - Třída C: 2 097 152
- Celkem možných síťových prefixů
 - 2 113 664 (ne všechny prefixy jsou použitelné v Internetu)
 - Kompletní (*a velká*) směrovací tabulka Internetu!

CLASS	IP RANGE (1ST OCTET)	DEFAULT SUBNET MASK	NETWORK/NODE PORTIONS	TOTAL NUMBER OF NETWORKS	TOTAL NUMBER OF USABLE ADDRESSES
A	0–127	255.0.0.0	Net.Node.Node.Node	2^7 or 128	$2^{24} - 2$ or 16,777,214
B	128–191	255.255.0.0	Net.Net.Node.Node	2^{14} or 16,384	$2^{16} - 2$ or 65,534
C	192–223	255.255.255.0	Net.Net.Net.Node	2^{21} or 2,097,151	$2^8 - 2$ or 254
D	224–239	N/A	N/A	N/A	N/A
E	240–255	N/A	N/A	N/A	N/A

CIDR

- CIDR (classless interdomain routing) ↳ agregace adres pro směrování (někdy supernetting)
- Směrovače šíří informaci o prefixu směrování

0000 0000	0
0001 0000	16
0010 0000	32
0011 0000	48



Longest Prefix Match (1)

- Jak se určí, která cesta v RT se vybere?
- Směrovače hledají nejdelší shodu
 - řeší problém možné vícenásobné shody ve směrovací tabulce
 - existují efektivní algoritmy

Cílová adresa: 201.10.6.17

1100 1001.0000 1010.0000 0110.0001 0001

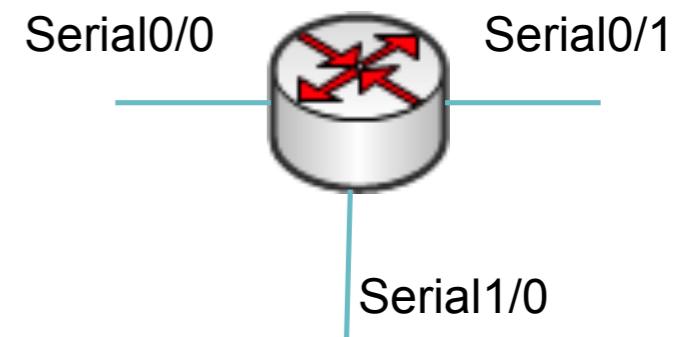
0000 0100.xxxxx xxxx.xxxxx xxxx.xxxxx xxxx /8

0000 0100.0101 0011.1xxx xxxx.xxxxx xxxx /17

1100 1001.0000 1010.0000 0xxx.xxxxx xxxx /21

1100 1001.0000 1010.0000 011x.xxxxx xxxx /23

0111 1110.1111 1111.0110 0111.xxxxx xxxx /24



4.0.0.0/8	Serial0/0
4.83.128.0/17	Serial0/0
201.10.0.0/21	Serial1/0
201.10.6.0/23	Serial0/1
126.255.103.0/24	Serial0/0

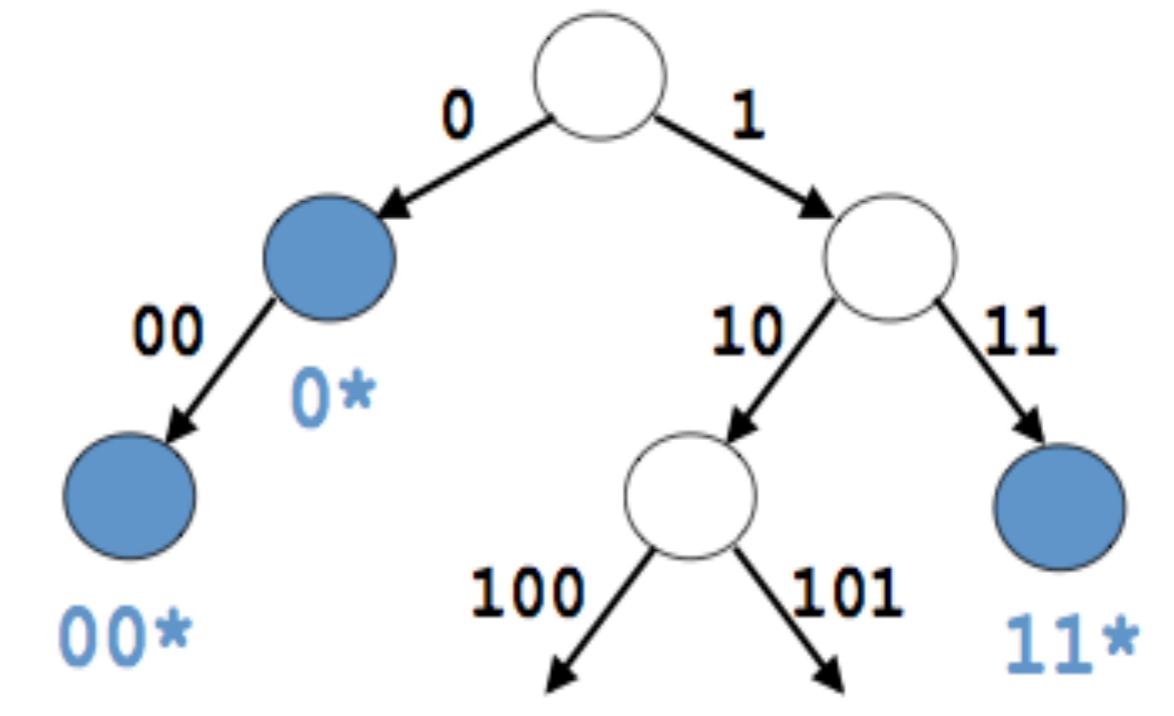
Naivní přístup

- Sekvenční průchod směrovací tabulkou
- Složitost je lineární, nicméně je potřeba lepší algoritmus
- Problém:
 - některé směrovače mohou mít až 350.000 záznamů
 - Čas na zpracování, 10 Gbps směrovač, pakety 64B:

$$\frac{10^{10}}{8 \cdot 64} = 19531250 \text{ p/s} \sim 51 \text{ ns/p}$$

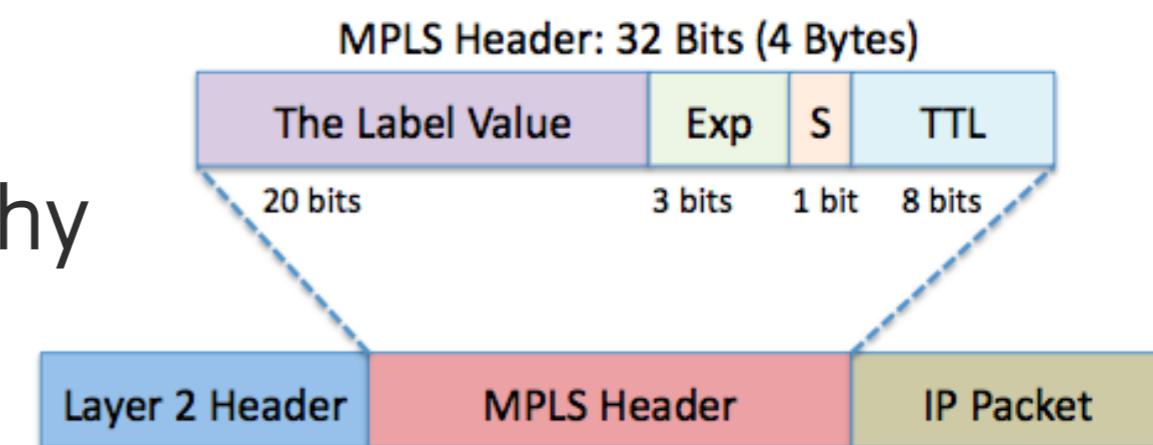
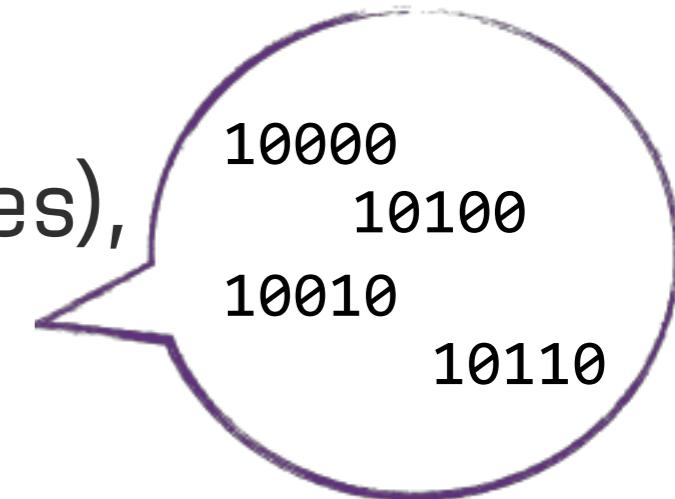
Dobrý přístup

- Prefixové stromy (Patricia Tree, 1968)
 - každá úroveň stromu je jeden bit v adrese
 - některé uzly mají přiřazeny záznamy v tabulce
- Když přijde paket hledá se nejdelší prefix ve stromě pro jeho cílovou adresu
- listy jsou routy RT



„Nejchytrější“ přístup

- Použiji speciální HW
 - Použití CAM (Content Addressable Memories), které implementuje asociativní pole
 - Dnes nejčastěji jako TCAM (hodnoty 0,1,X)
- *Lze to ještě urychlit?*
 - Použiji jiný koncept směrování než na základě dst IP
- například MPLS
 - umožňují vytvářet virtuální okruhy na kterých se přepínají data identifikována pomocí labelů



Kde se berou záznamy v RT?

- Položky mohou být **staticky** definovány
 - administrátor je musí vložit
 - nejsou aktuální v případě selhání zařízení
 - nejsou dynamické, nereflektují aktuální stav sítě
- Nebo zjištěny pomocí **dynamický směrovacích protokolů**
 - IGP protokoly pro směrování uvnitř organizace (RIP, OSPF)
 - EGP protokol pro výměnu informací mezi organizacemi a ISP (BGP)

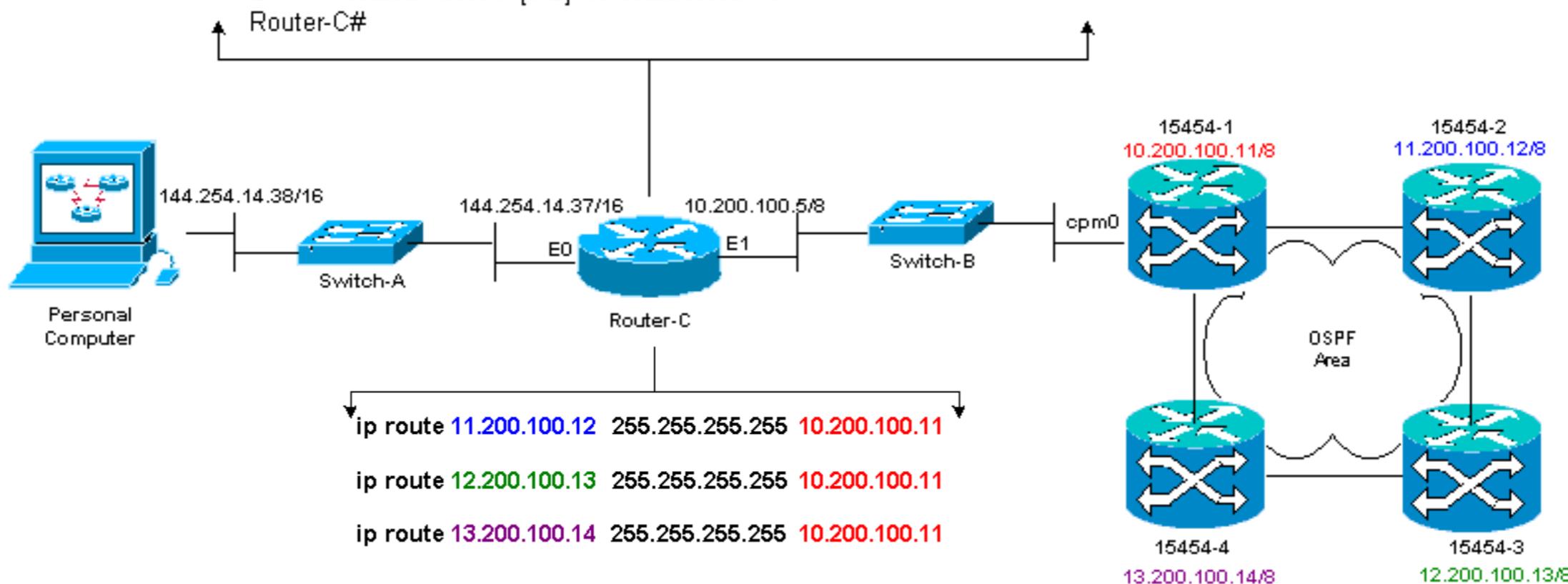
Statické směrování

Router-C# **show ip route**

Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, * - candidate default
U - per-user static route, o - ODR

Gateway of last resort is not set

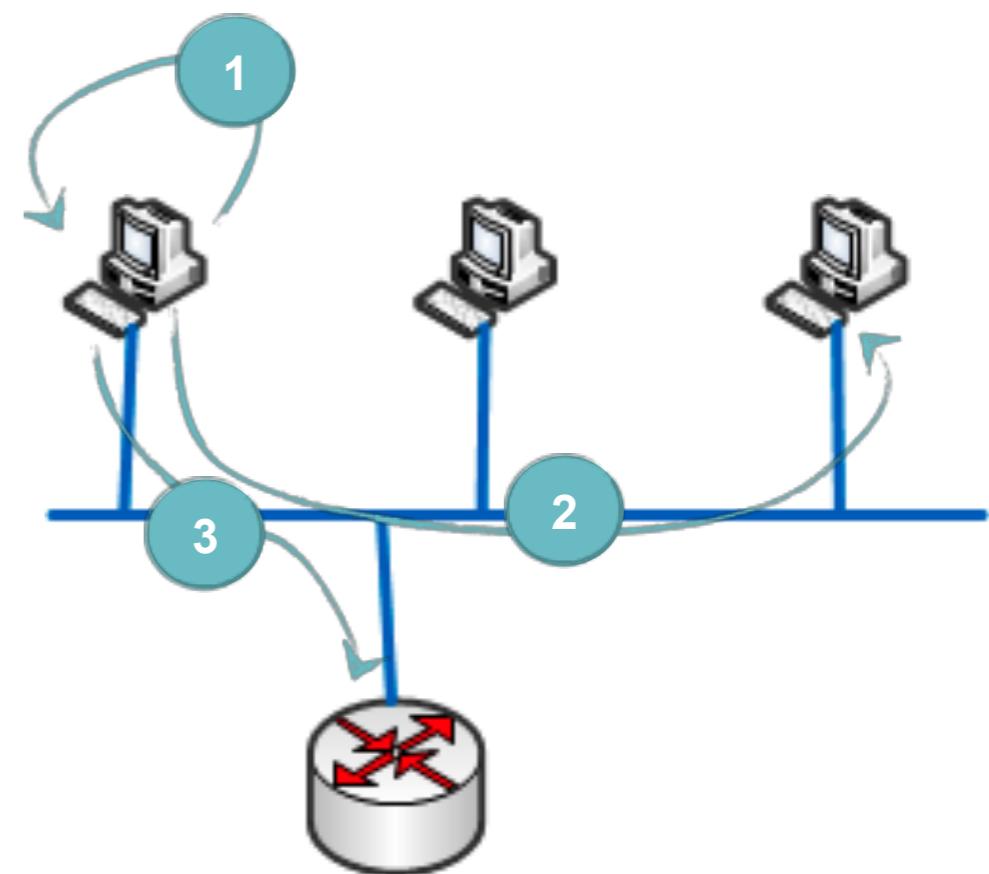
C 10.0.0.0/8 is directly connected, Ethernet0
C 144.254.0.0/16 is directly connected, Ethernet1
S **11.200.100.12** [1/0] via **10.200.100.11**
S **12.200.100.13** [1/0] via **10.200.100.11**
S **13.200.100.14** [1/0] via **10.200.100.11**



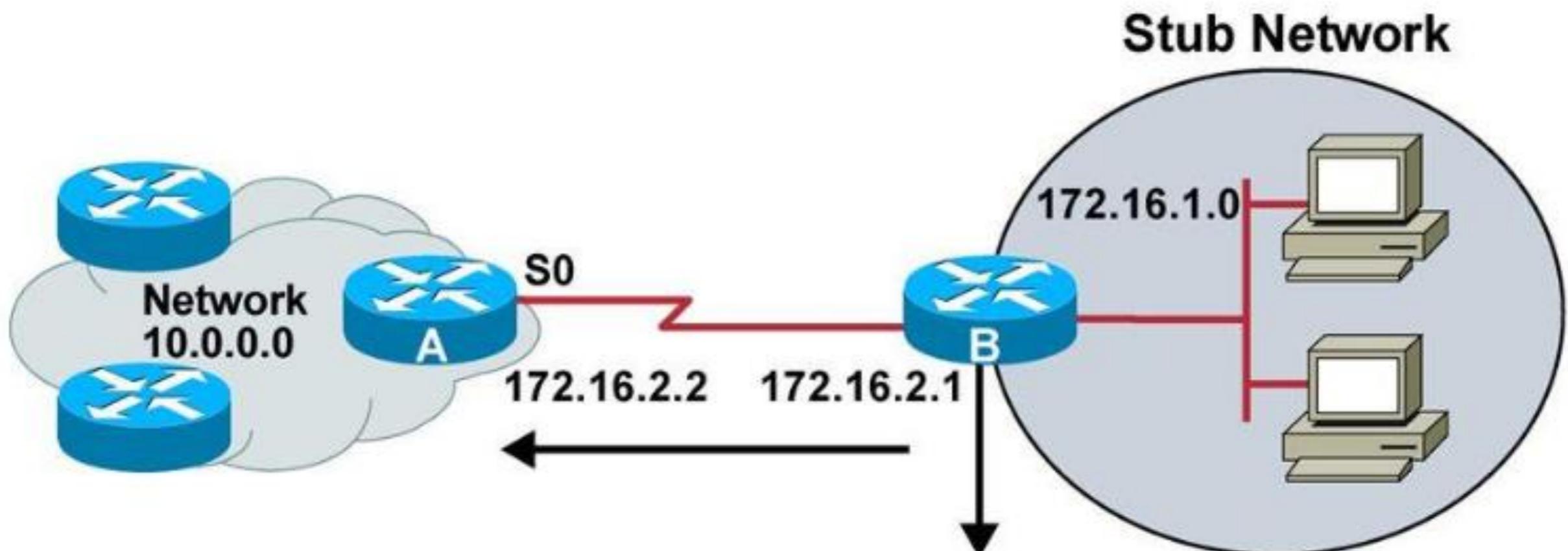
Směrování z pohledu koncové stanice

- V případě, že má koncová stanice pouze jedno rozhraní, nepotřebuje směrování:
 1. Paket na vlastní adresu je doručen lokálně
 2. Paket ostatním uzelům v síti je poslán v Ethernetovém rámci s konkrétní adresou příjemce
 3. Paket vně sítě je poslán na lokální výchozí bránu

```
C:\WINDOWS\system32\cmd.exe
IPv4 Route Table
=====
Active Routes:
Network Destination      Netmask      Gateway      Interface      Metric
          0.0.0.0        0.0.0.0    10.20.40.1    10.20.40.47      25
         10.7.0.0    255.255.0.0        On-link       10.7.0.1      291
         10.7.0.0    255.255.0.0  192.168.255.1  192.168.255.4      259
         10.7.0.1    255.255.255.255        On-link       10.7.0.1      291
        10.7.255.255  255.255.255.255        On-link       10.7.0.1      291
         10.20.40.0    255.255.255.0        On-link    10.20.40.47      281
         10.20.40.47    255.255.255.255        On-link    10.20.40.47      281
        10.20.40.255  255.255.255.255        On-link    10.20.40.47      281
         10.100.0.0   255.255.0.0    192.168.255.1  192.168.255.4      259
         10.200.0.0   255.255.0.0    192.168.255.1  192.168.255.4      259
         127.0.0.0     255.0.0.0        On-link     127.0.0.1      331
         127.0.0.1     255.255.255.255        On-link     127.0.0.1      331
        127.255.255.255  255.255.255.255        On-link     127.0.0.1      331
         169.254.0.0   255.255.0.0        On-link  169.254.86.8      291
         169.254.0.0   255.255.0.0        On-link  169.254.125.121      291
         169.254.0.0   255.255.0.0        On-link  169.254.217.213      291
         169.254.0.0   255.255.0.0        On-link  169.254.150.30      291
         169.254.0.0   255.255.0.0        On-link  169.254.83.108      291
         169.254.0.0   255.255.0.0        On-link  169.254.71.147      291
        169.254.71.147  255.255.255.255        On-link  169.254.71.147      291
        169.254.83.108  255.255.255.255        On-link  169.254.83.108      291
         169.254.86.8   255.255.255.255        On-link  169.254.86.8      291
        169.254.125.121  255.255.255.255        On-link  169.254.125.121      291
        169.254.150.30  255.255.255.255        On-link  169.254.150.30      291
       169.254.217.213  255.255.255.255        On-link  169.254.217.213      291
```



Defaultní cesta

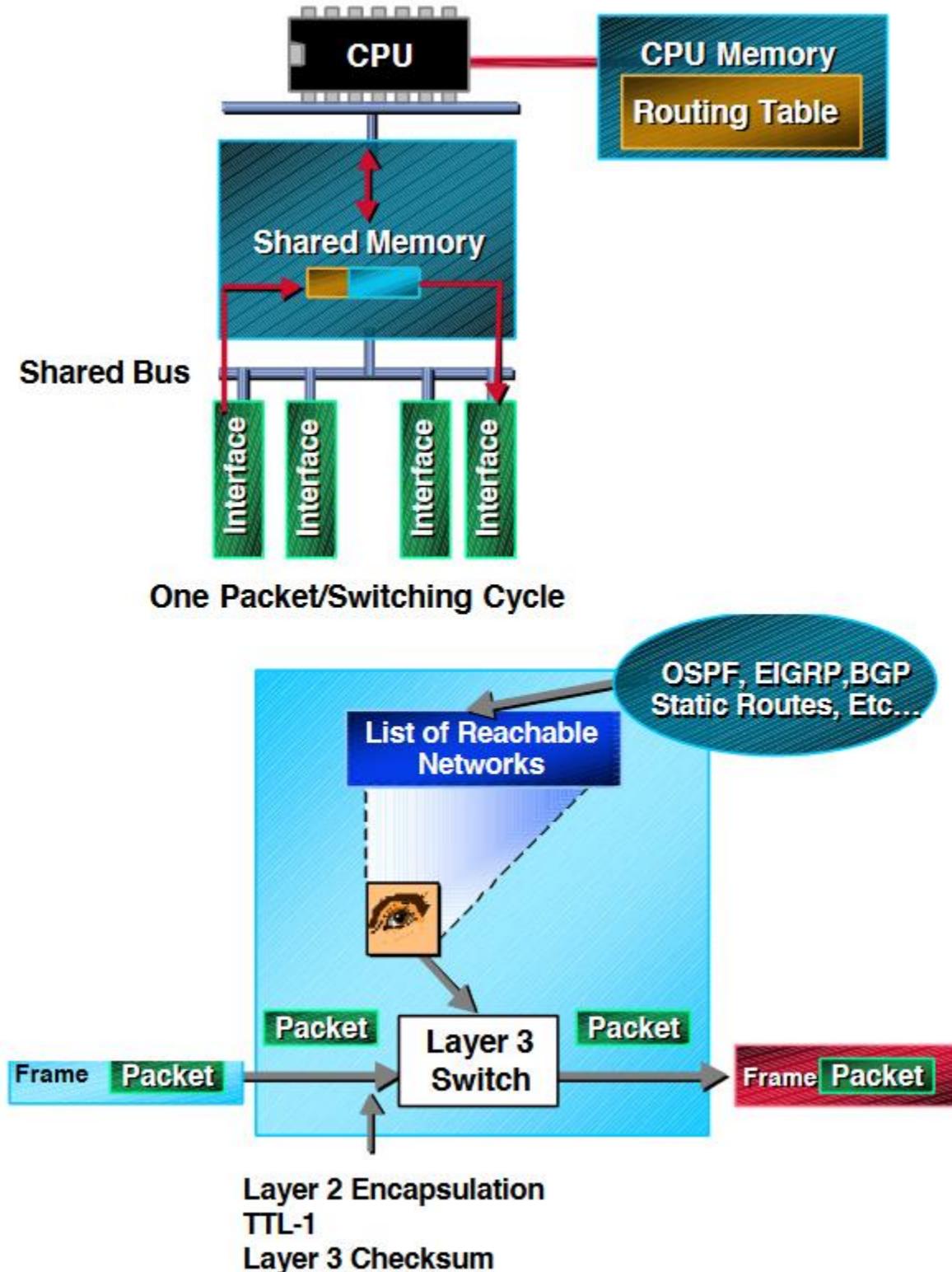


```
Router(config)# ip route 0.0.0.0 0.0.0.0 172.16.2.2
```

Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Bellman-Ford
 - Algoritmus Dijkstra
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

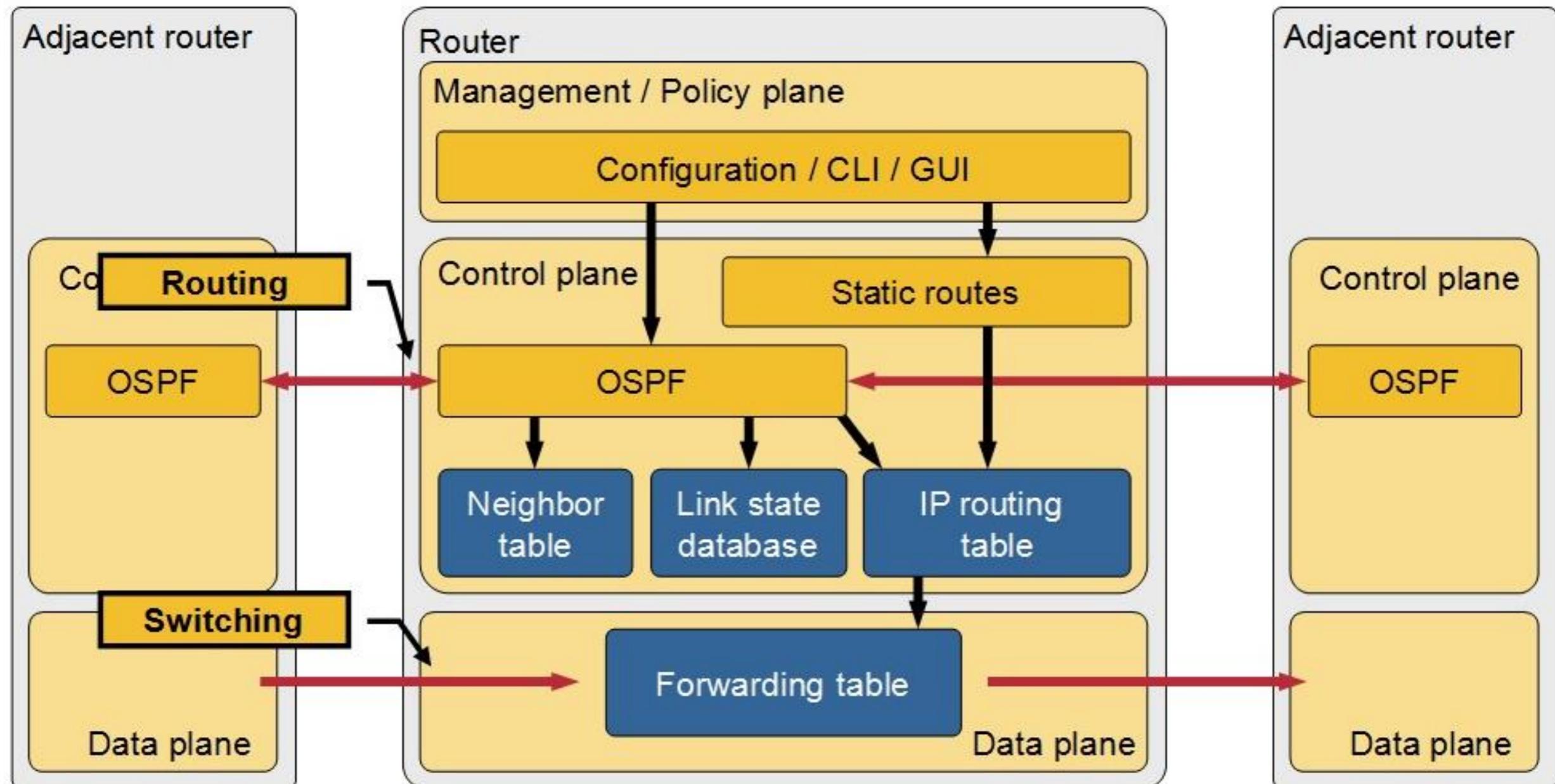
Co je směrovač?



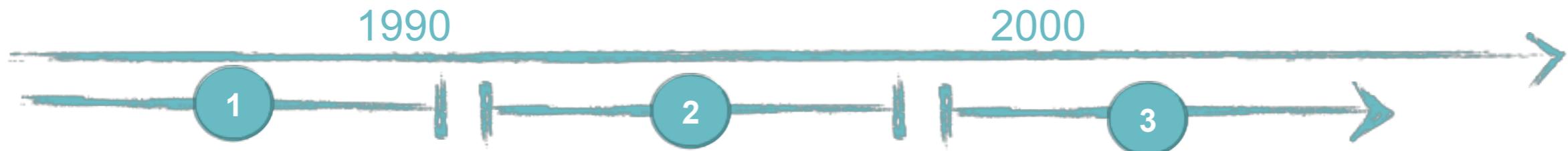
Funkce směrovače

- Zpracování směrovacích informací
 - výpočet nejlepší cesty
 - správa směrovací tabulky
 - provoz směrovacích protokolů
- Přepínaní paketů
 - zjištění cílové adresy paketu
 - nalezení výstupního portu
 - kontrola stáří paketu (TTL)
 - výpočet kontrolního součtu
- Speciální funkce
 - transformace paketů
 - zabalování (tunelování)
 - klasifikace paketů, prioritizace
 - autentizace
 - filtrování paketů
 - správa, účtování, statistika

Struktura směrovače



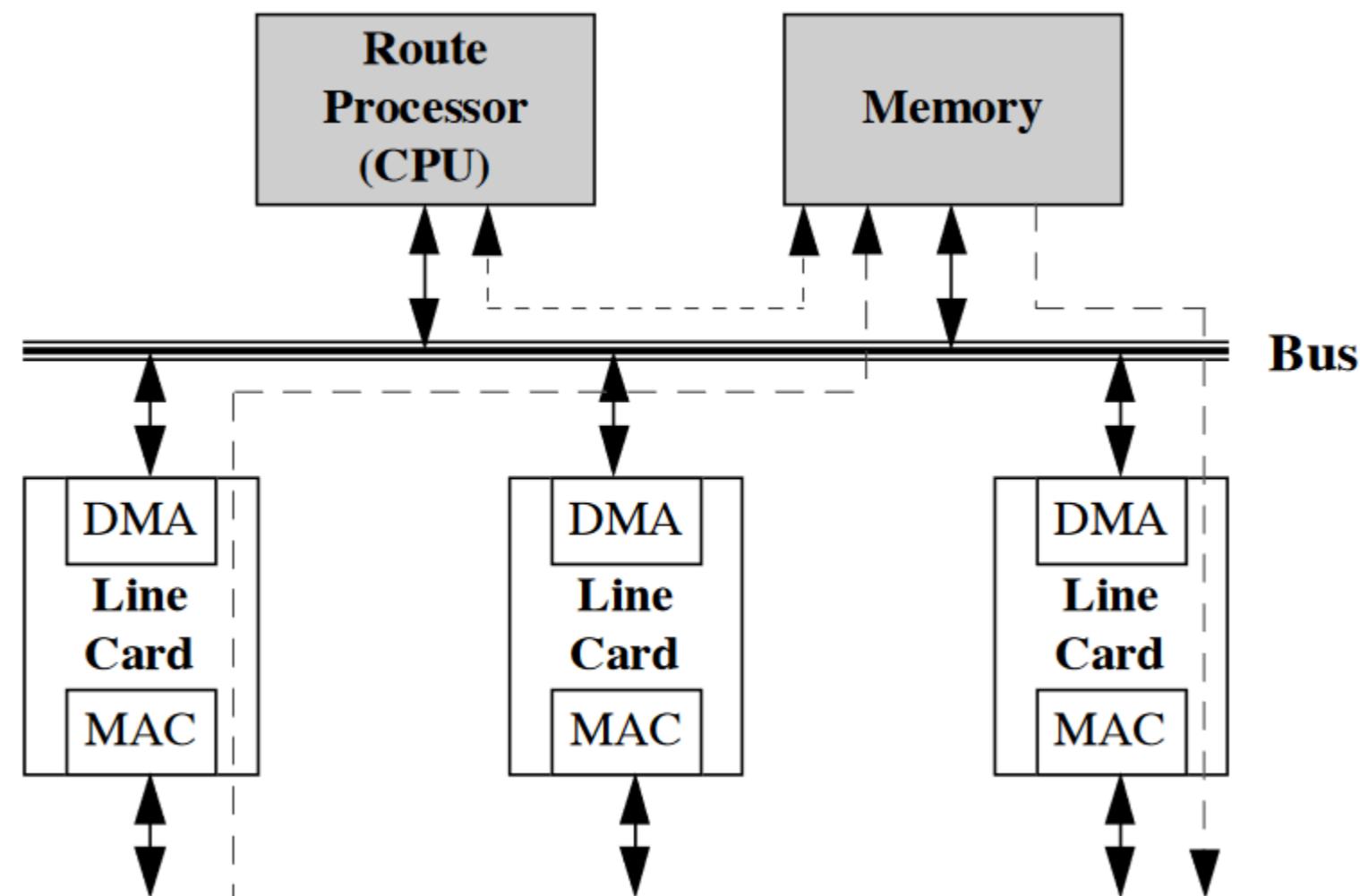
Architektury směrovačů: vývoj



- 1.generace
 - softwarové směrovače
 - standardní PC
- 2.generace
 - sběrnice pro vnitřní komunikaci
 - paralelní zpracování na rozhraní
- 3.generace
 - přepínač pro vnitřní komunikaci
 - distribuovaná architektura

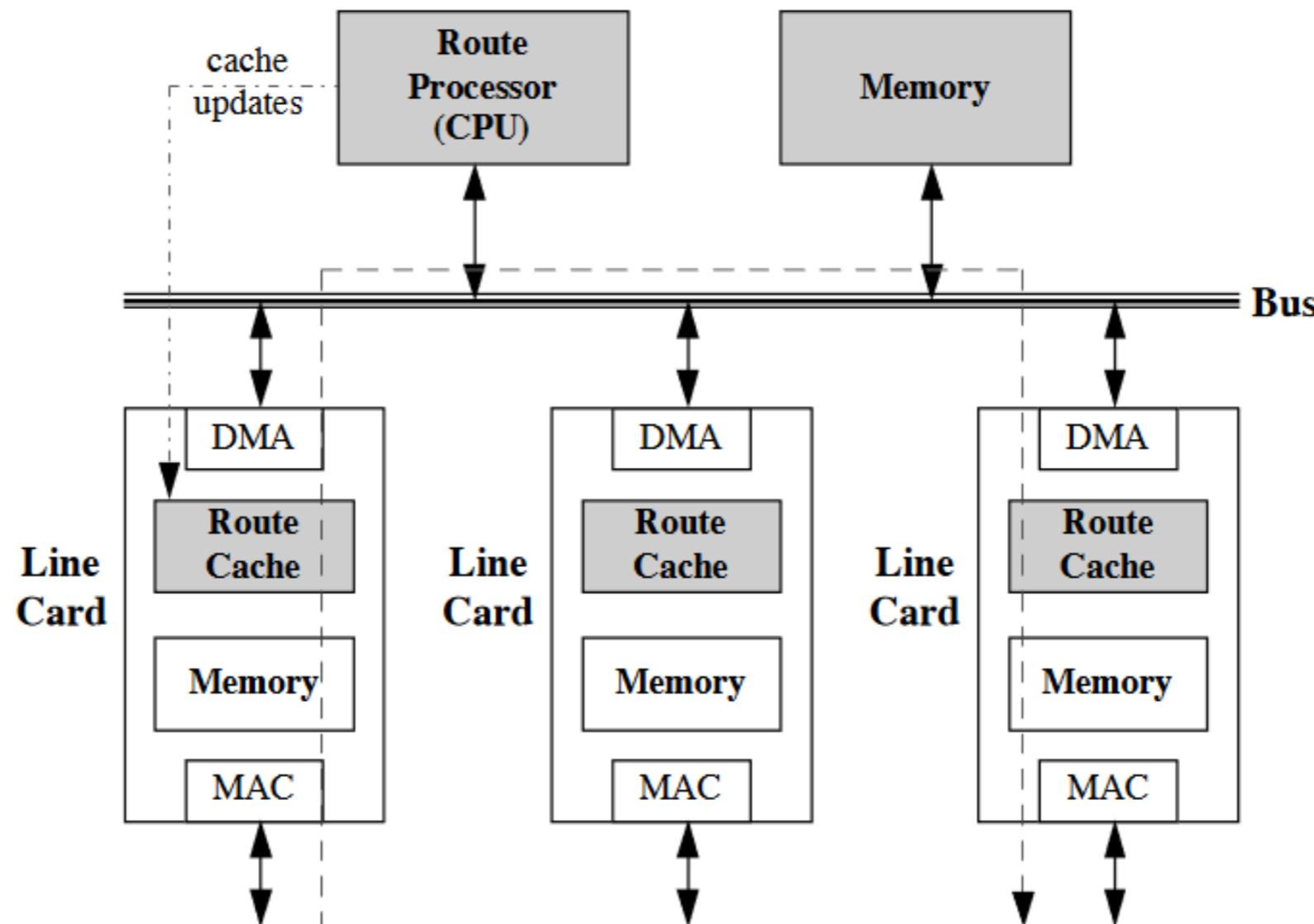
Sběrnice s centrálním CPU

- všechny pakety jsou posílány do CPU, které je analyzuje
- CPU musí dělat i další věci
- malá výkonnost, není škálovatelné



Sběrnice s lokální pamětí cache

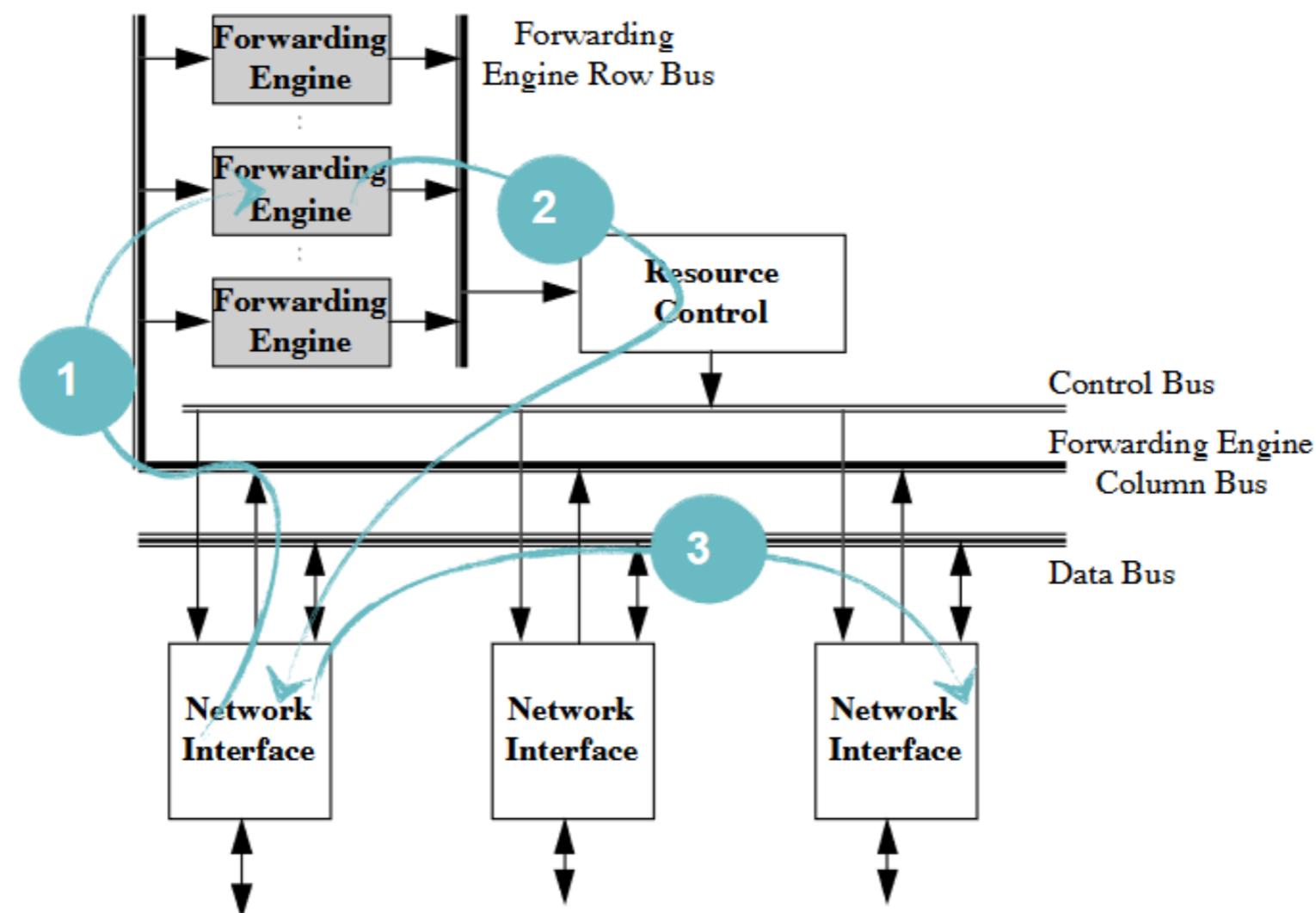
- lokální CPU na kartách obsahují částečné kopie centrální směrovací tabulky
- sběrnice přenáší data mezi kartami
- pakety, které není možné zpracovat na kartě se posílají do centrálního CPU



Sběrnice s paralelním zpracováním

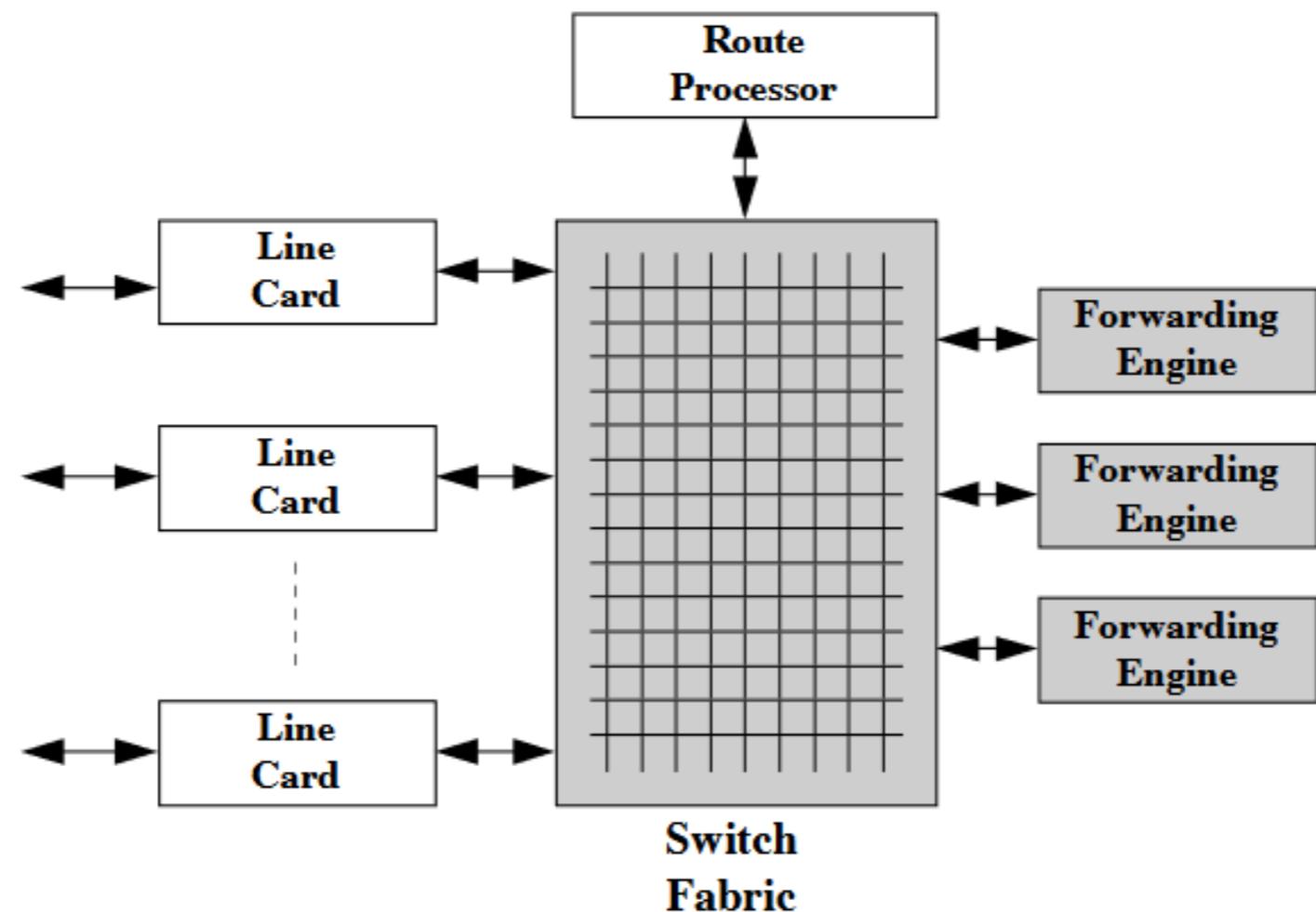
1. síťová karta oddělí hlavičku a pošle ji do FE
2. FE určí z hlavičky výstupní rozhraní pro paket
3. paket uložený v buffer vstupního rozhraní je pak přesunut na výstupní rozhraní

Předpokládá se, že ne všechny porty jsou vždy maximálně zatíženy

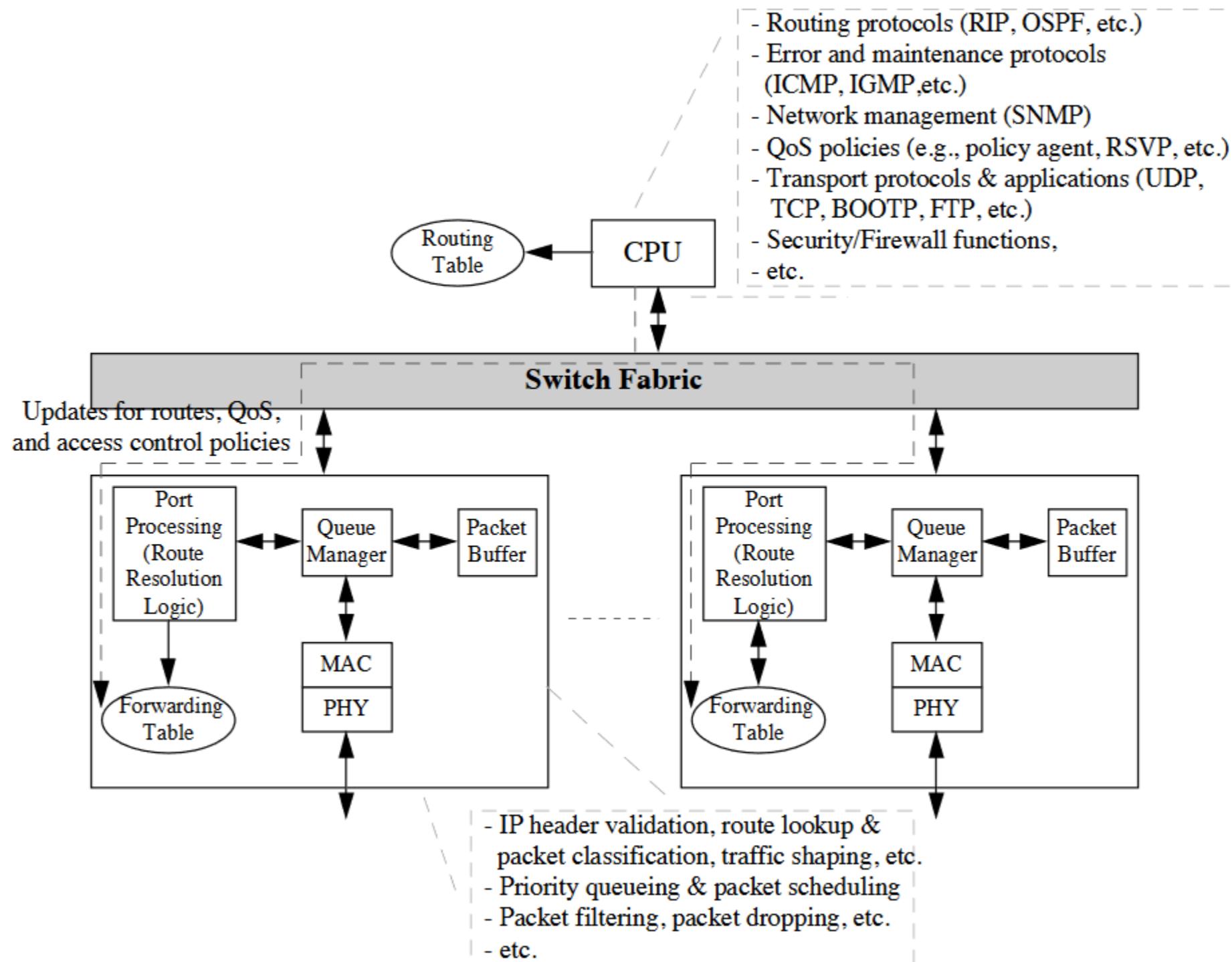


Přepínač s více procesory

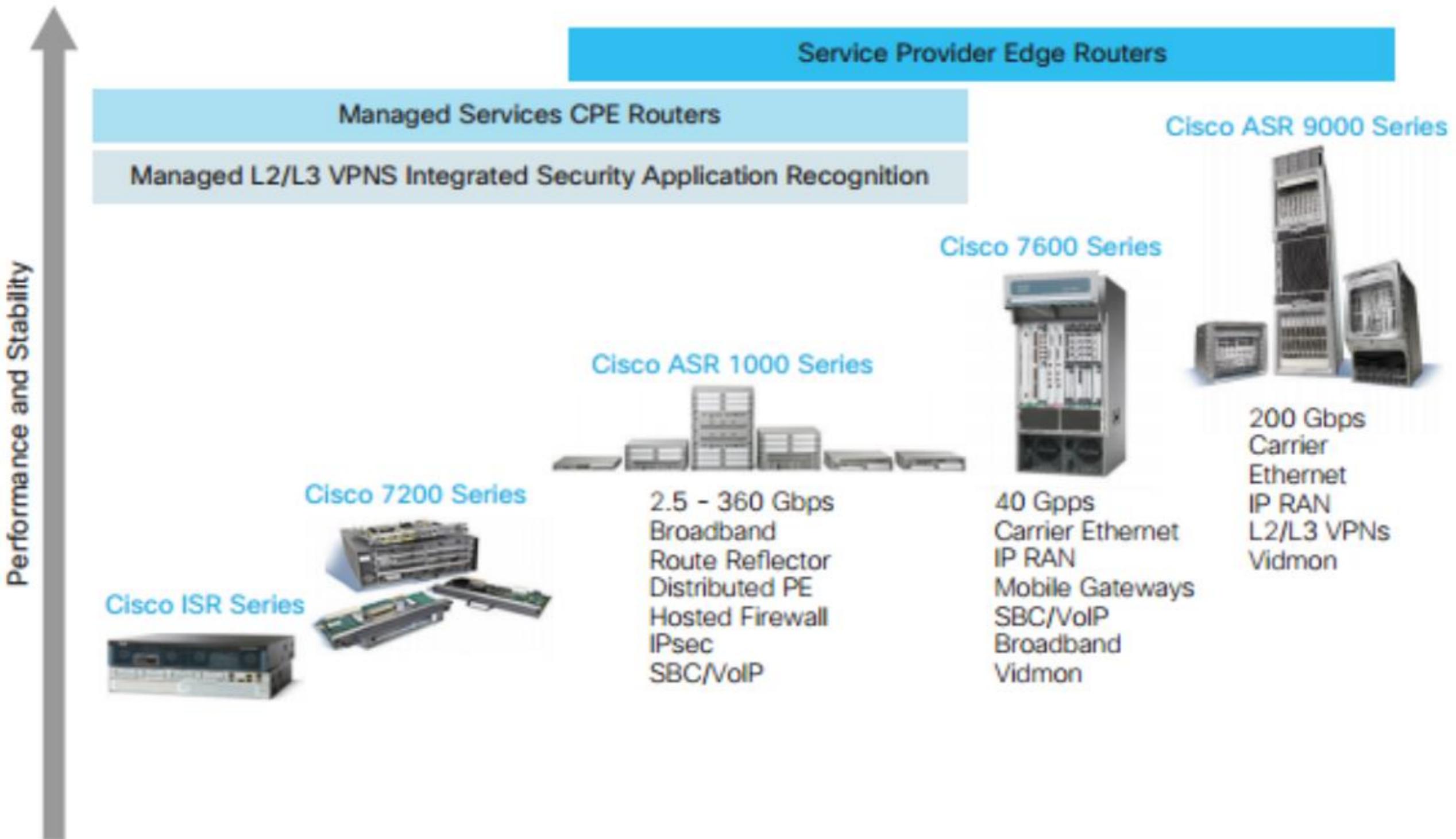
- podobně jak u sběrnice s paralelním zpracováním, nicméně větší rychlosť přenosu mezi vnitřními komponentami
- fast-path: informace je nalezena v cache FE
- slow-path: při výpadku v cache je nutné najít informaci v hlavní směrovací tabulce
- velikost cache je limitujícím faktorem, enterprise směrovače mohou "obsluhovat" statisíce aktivních toků



Distribuovaná architektura



Jaký router potřebuji?

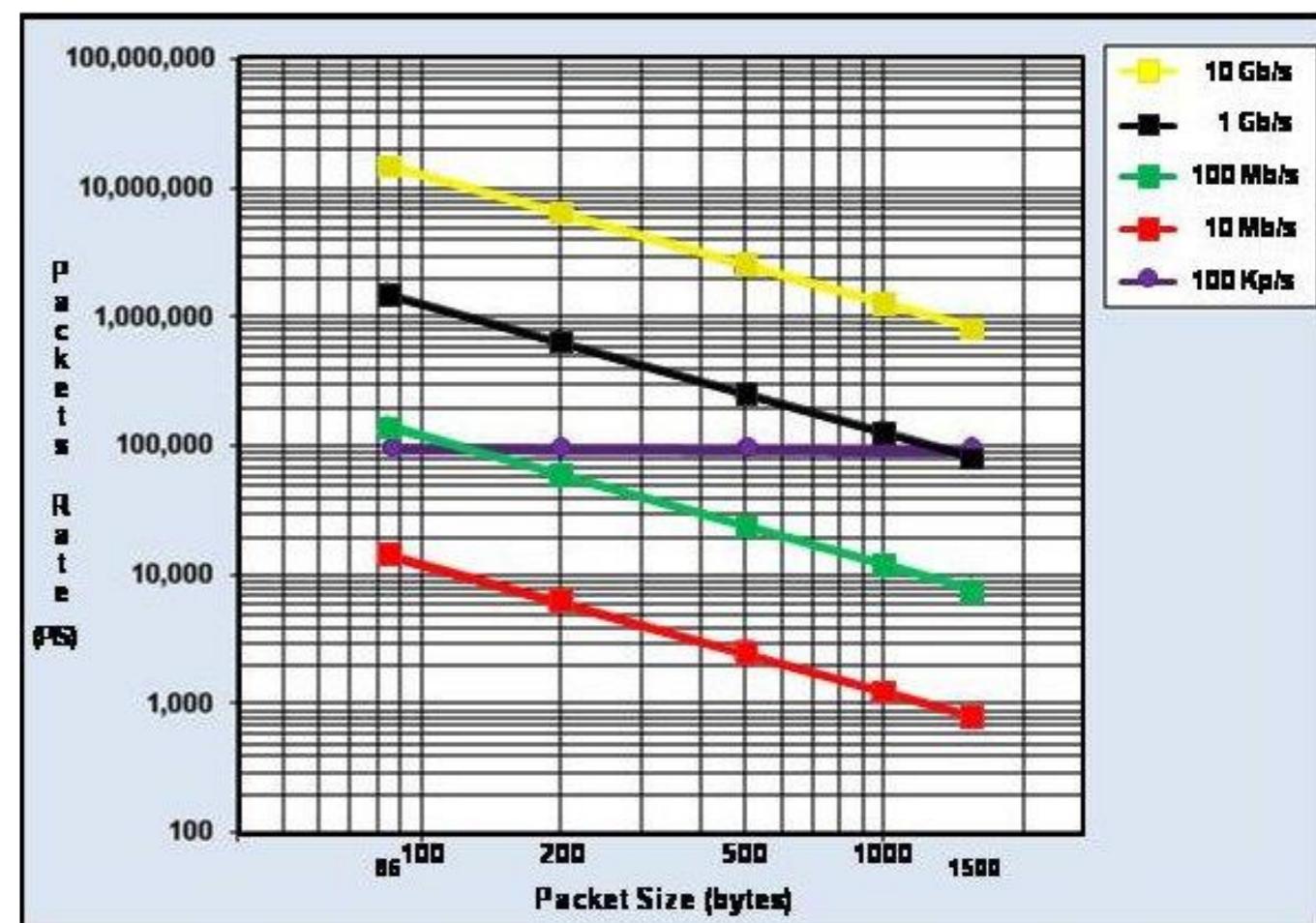


Performance Metrics

- Někdo nám tvrdí:
 - „Cisco ASR 1000 Series Router, is capable of forwarding packets at up to 16 Mp/s with services enabled, it can support the processing of the equivalent of 10 Gb/s of traffic at line rate?“
 - Je to hodně nebo málo?
- Maximum Frame Rate (Minimum Frame Size)
- Maximum Throughput (Maximum Frame Size)

[$1,000,000,000 \text{ b/s} / (84 \text{ B} * 8 \text{ b/B})$] == **1,488,096 f/s (maximum rate)**

[$1,000,000,000 \text{ b/s} / (1,538 \text{ B} * 8 \text{ b/B})$] == **81,274 f/s (minimum rate)**



A kolik mě to bude stát?



Cisco Linksys E4200 v2 Maximum Performance Dual-Band N900 router

Part Number: E4200V2

\$188.49

MSRP: \$199.99

(-6%)



Cisco ISR 4431 - router - rack-mountable

Part Number: ISR4431/K9

\$5,164.11 to \$5,374.99



Cisco 7604 - router - desktop, rack-mountable - with Cisco 7600 Series Route Switch Processor 720 with 10 Gigabit Ethernet (RSP720-3CXL-10GE)

Part Number: 7604-RSP7XL-10G-P

1 Related Model

MSRP: \$54,000.00

Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Bellman-Ford
 - Algoritmus Dijkstra
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP

Bellman-Fordův algoritmus

- Používá se u distance-vector protokolů
- Každý uzel komunikuje se svými sousedy (*routing by rumor*)
- Každý uzel počítá na základě dostupných informací vlastní nejkratší cestu k cíli
- Vlastnosti
 - Iterativní - Běží pokud jsou nové informace vyměňovány mezi sousedy
 - Asynchronní - Není potřeba synchronizace mezi uzly pro zajištění správnosti výpočtu
 - Reaktivní - změna cen lokální linky a zpráva od sousedního uzlu s aktualizací
 - Distribuovaný
 - Každý uzel oznamuje změnu (hodnotu Distance Vector)
 - Změny mohou způsobit změnu na sousedním uzlu, který ji potom šíří dále

Distance-Vector (DV) Algorithm

$$D_x(y) = \min_v \{c(x,v) + D_v(y)\}$$

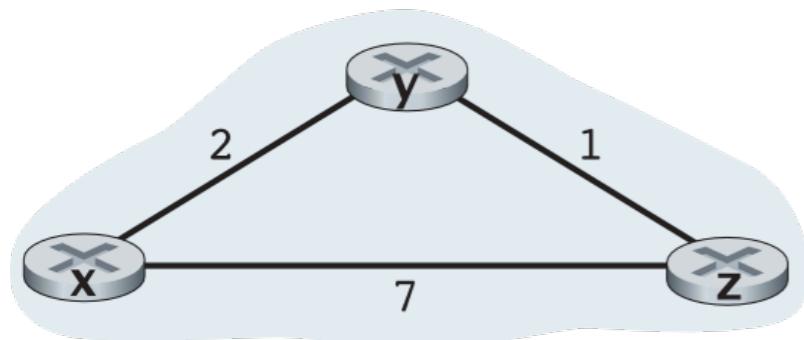
At each node, x :

```
1  Initialization:  
2      for all destinations y in N:  
3          Dx(y) = c(x,y) /* if y is not a neighbor then c(x,y) = ∞ */  
4      for each neighbor w  
5          Dw(y) = ? for all destinations y in N  
6      for each neighbor w  
7          send distance vector Dx = [Dx(y): y in N] to w  
8  
9  loop  
10     wait (until I see a link cost change to some neighbor w or  
11         until I receive a distance vector from some neighbor w)  
12  
13     for each y in N:  
14         Dx(y) = minv{c(x,v) + Dv(y)}  
15  
16     if Dx(y) changed for any destination y  
17         send distance vector Dx = [Dx(y): y in N] to all neighbors  
18  
19 forever
```

Node x table

		cost to		
		x	y	z
from	x			
	y			
	z			

Node y table



		cost to		
		x	y	z
from	x			
	y			
	z			

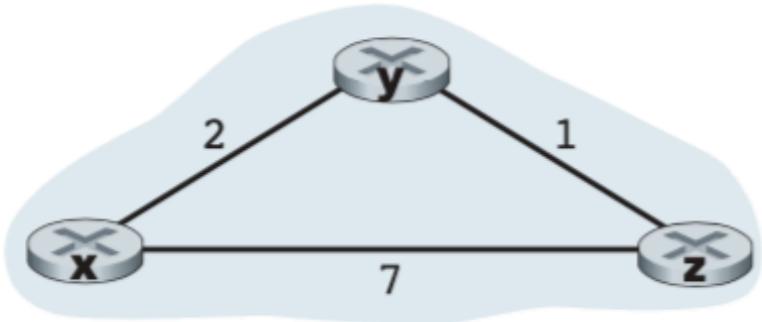
Node z table

		cost to		
		x	y	z
from	x			
	y			
	z			

Time

Node x table

		cost to		
		x	y	z
from	x	0	2	7
	y	∞	∞	∞
	z	∞	∞	∞



Node y table

		cost to		
		x	y	z
from	x	∞	∞	∞
	y	2	0	1
	z	∞	∞	∞

Node z table

		cost to		
		x	y	z
from	x	∞	∞	∞
	y	∞	∞	∞
	z	7	1	0

Time

Obsah

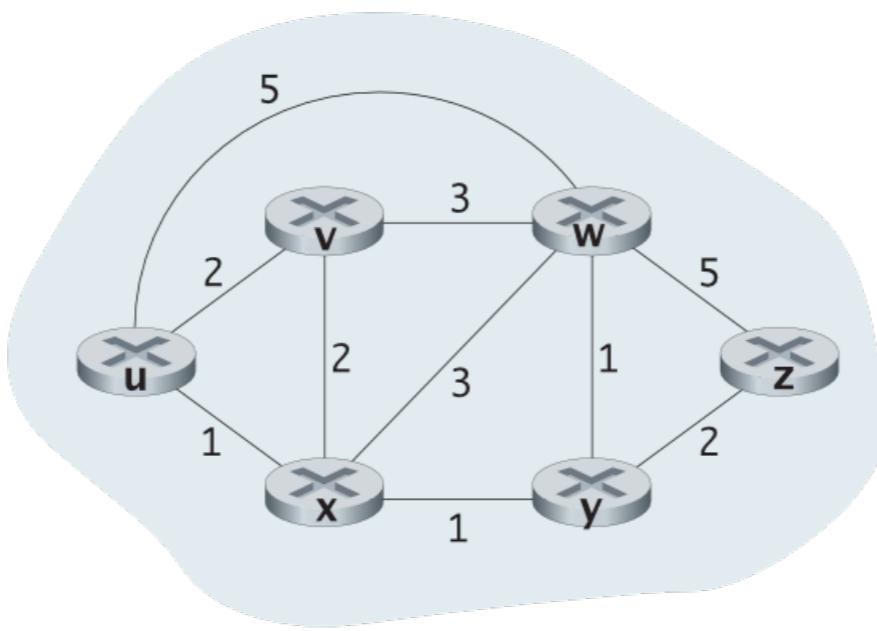
- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Bellman-Ford
 - Algoritmus Dijkstra
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

Dijkstrův algoritmus

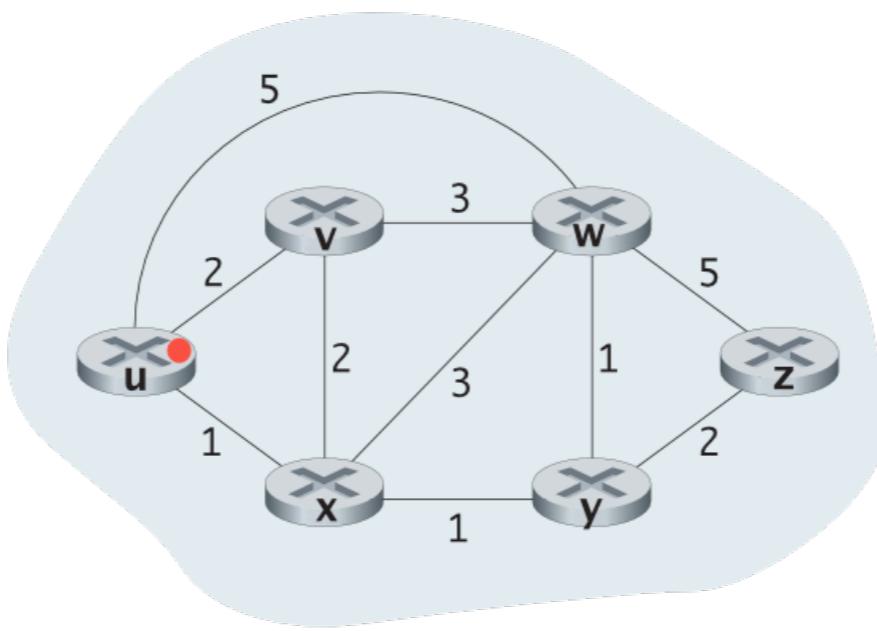
- Používá se u link-state protokolů
- Předpoklady
 - Topologie sítě včetně ceny všech linek je předem známa každému uzlu
 - Všechny uzly mají stejnou informaci o stavu sítě
- Každý směrovač vypočítá cestu s nejmenší cenou pro každou cílovou síť
 - Začátek cesty je vždy aktuální směrovač
 - Iterativní výpočet

Link-State (LS) Algorithm for Source Node u

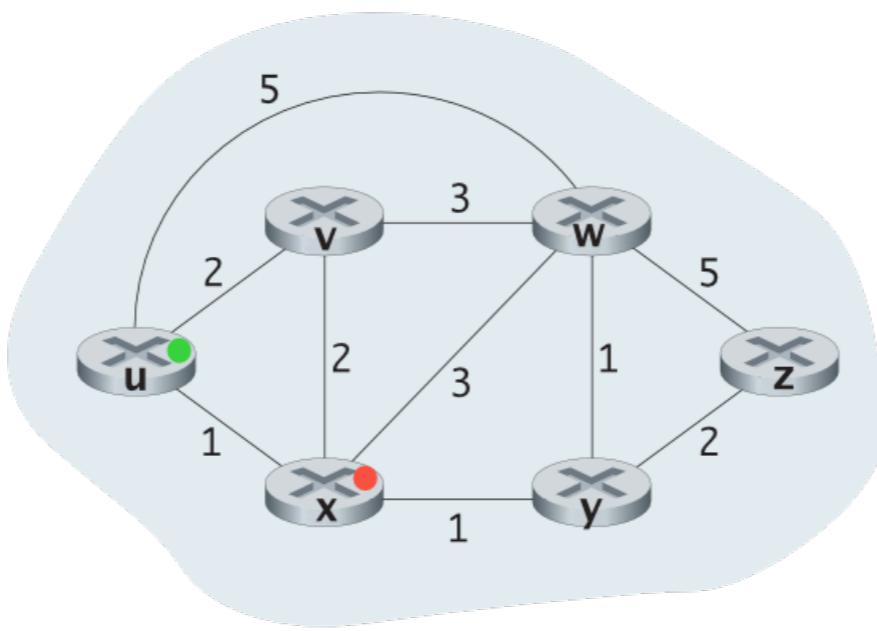
```
1 Initialization:
2    $N' = \{u\}$ 
3   for all nodes  $v$ 
4     if  $v$  is a neighbor of  $u$ 
5       then  $D(v) = c(u,v)$ 
6     else  $D(v) = \infty$ 
7
8 Loop
9   find  $w$  not in  $N'$  such that  $D(w)$  is a minimum
10  add  $w$  to  $N'$ 
11  update  $D(v)$  for each neighbor  $v$  of  $w$  and not in  $N'$ :
12     $D(v) = \min(D(v), D(w) + c(w,v))$ 
13  /* new cost to  $v$  is either old cost to  $v$  or known
14  least path cost to  $w$  plus cost from  $w$  to  $v$  */
15 until  $N' = N$ 
```



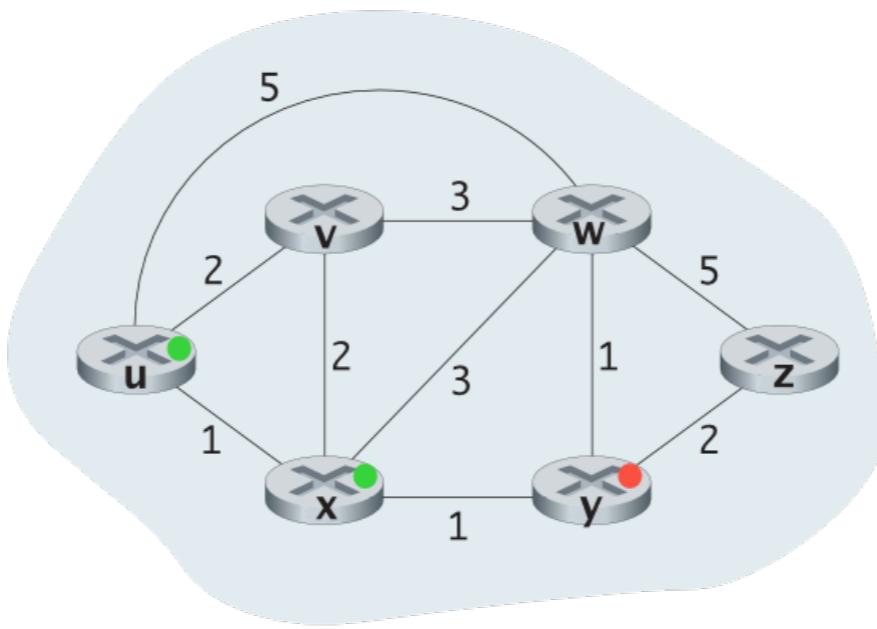
step	N'	$D(v), p(v)$	$D(w), p(w)$	$D(x), p(x)$	$D(y), p(y)$	$D(z), p(z)$



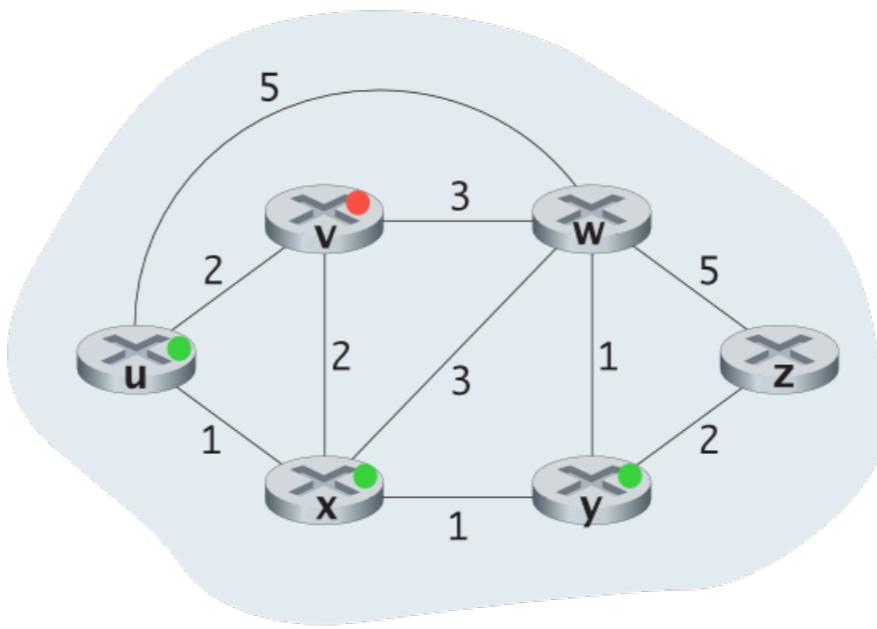
step	N'	$D(v), p(v)$	$D(w), p(w)$	$D(x), p(x)$	$D(y), p(y)$	$D(z), p(z)$
0	u	2, u	5, u	1, u	∞	∞



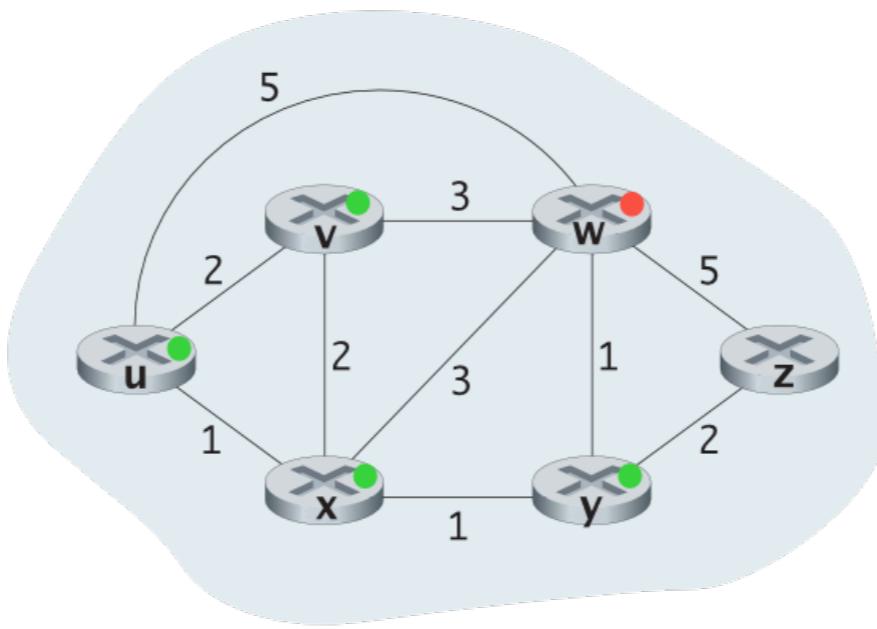
step	N'	$D(v), p(v)$	$D(w), p(w)$	$D(x), p(x)$	$D(y), p(y)$	$D(z), p(z)$
0	u	2,u	5,u	1,u	∞	∞
1	ux	2,u	4,x		2,x	∞



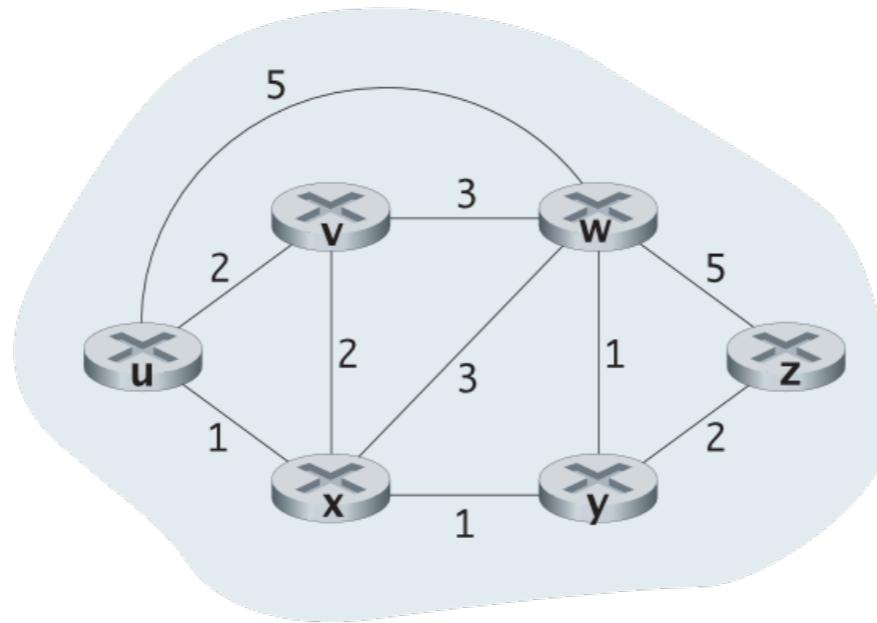
step	N'	$D(v), p(v)$	$D(w), p(w)$	$D(x), p(x)$	$D(y), p(y)$	$D(z), p(z)$
0	u	2,u	5,u	1,u	∞	∞
1	ux	2,u	4,x		2,x	∞
2	uxy	2,u	3,y			4,y



step	N'	$D(v), p(v)$	$D(w), p(w)$	$D(x), p(x)$	$D(y), p(y)$	$D(z), p(z)$
0	u	2,u	5,u	1,u	∞	∞
1	ux	2,u	4,x		2,x	∞
2	uxy	2,u	3,y			4,y
3	uxyw		3,y			4,y



step	N'	$D(v), p(v)$	$D(w), p(w)$	$D(x), p(x)$	$D(y), p(y)$	$D(z), p(z)$
0	u	2,u	5,u	1,u	∞	∞
1	ux	2,u	4,x		2,x	∞
2	uxy	2,u	3,y			4,y
3	uxyw		3,y			4,y
4	uxywv					4,y



step	N'	$D(v), p(v)$	$D(w), p(w)$	$D(x), p(x)$	$D(y), p(y)$	$D(z), p(z)$
0	u	2,u	5,u	1,u	∞	∞
1	ux	2,u	4,x		2,x	∞
2	uxy	2,u	3,y			4,y
3	uxyw		3,y			4,y
4	uxyw					4,y
5	uxywz					

Obsah

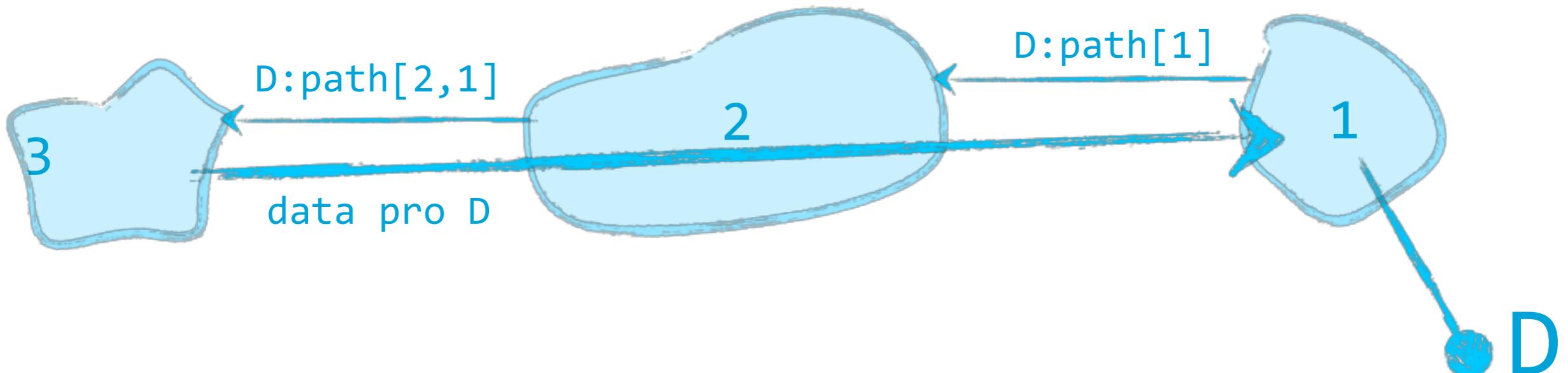
- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Bellman-Ford
 - Algoritmus Dijkstra
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP

Path-vector směrování

- Používá se pro směrování mezi autonomními systémy
 - LS je nevhodné (nelze udělat graf „Internetu“ a zkonzervovat ho)
 - DV je nevhodné (pouze jedna hodnota metriky není schopná uspokojit use-case jako „chci směrovat ne nejkratší, ale nejlevnější cestou“)
- Path-Vector nicméně používá a rozšiřuje myšlenku DV směrování
 - podporuje navíc flexibilní směrovací politiky

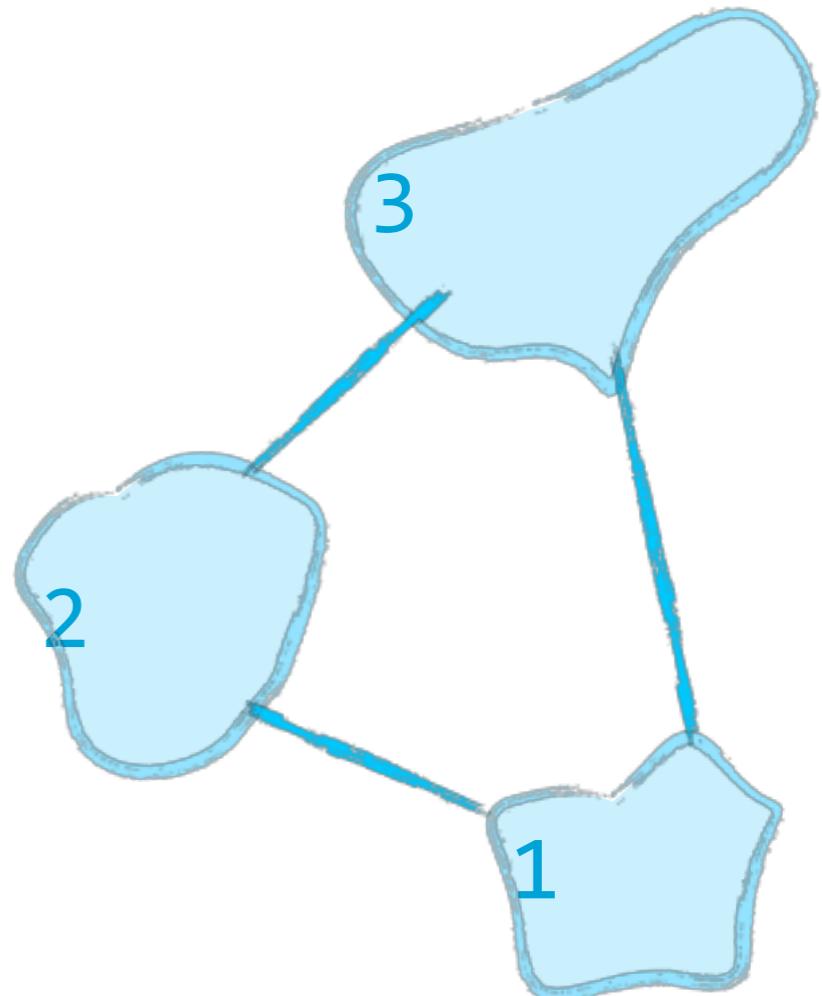
PV směrování

- Základní myšlenka \leftrightarrow oznamovat celou cestu ke koncové síti (místo vzdálenosti)
 - DV: má pro každý cíl D jeho vzdálenost
 - PV: má pro každý cíl D jeho celou cestu
 - je jednoduché detekovat a eliminovat smyčky

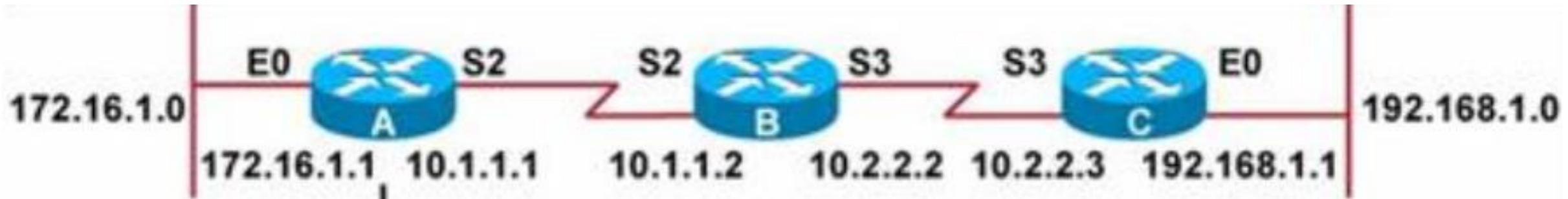


Flexibilní politiky

- Každý uzel může aplikovat lokální politiku
 - Výběr cesty
 - Oznamování cesty
- Například:
 - AS 2 preferuje [2,3,1] místo [2,1]
 - AS 1 nechce říct AS 3 o cestě [1,2]
- Použito v BGP
 - Path odpovídá AS
 - Politiky odpovídají smlouvám mezi ISP



Co je tedy v RT?



```
RouterA#show ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
        D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
        N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
        E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
        i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, * - candidate
        default
        U - per-user static route, o - ODR
        T - traffic engineered route

Gateway of last resort is not set

    172.16.0.0/24 is subnetted, 1 subnets
C      172.16.1.0 is directly connected, Ethernet0
        10.0.0.0/24 is subnetted, 2 subnets
R      10.2.2.0 [120/1] via 10.1.1.2, 00:00:07, Serial2
C      10.1.1.0 is directly connected, Serial2
R      192.168.1.0/24 [120/2] via 10.1.1.2, 00:00:07, Serial2
```

Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Dijkstra
 - Algoritmus Bellman-Ford
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

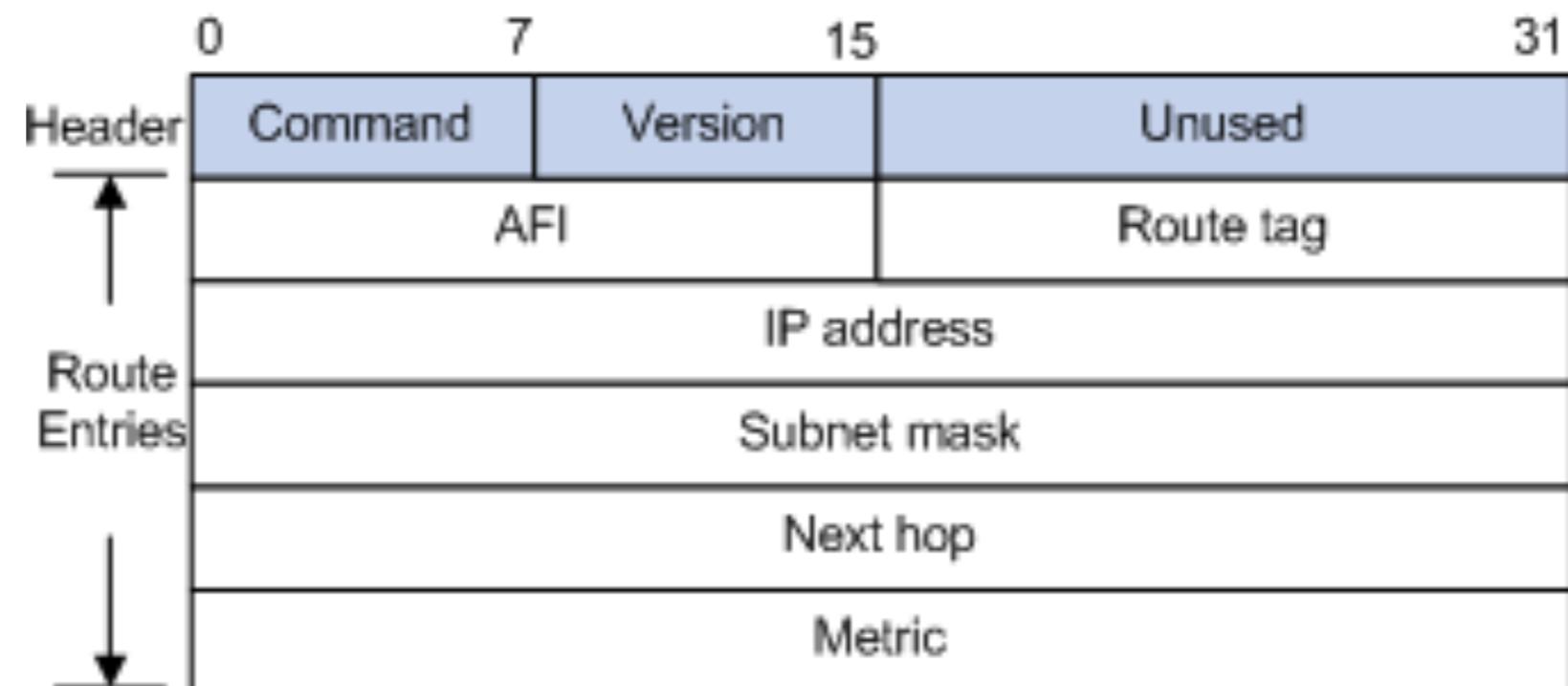
Historie RIP

- Principy definovány v roce 1969 pro ARPANET and CYCLADES
- V 70-tých letech použit v Xerox PUP sítích a posléze v Xerox Network System jako XNS RIP
- XNS RIP se poté uplatnil v IPX RIP, AppleTalk RTMP and IP RIP
- 1982 byl RIP implementován v BSD UNIXu jako routed daemon
- 1988 byl vydan standard RFC 1058 (Charles Hedrick)

Charakteristika

↔ RIP je DV, který používá Bellman-Fordův algoritmus

- Metrikou je počet skoků (hop-count)
- Hop-count > 15 označuje nedosažitelnou cestu
- Periodické aktualizace každých 30s
- RIP zprávy jsou neseny v UDP datagramu na port 520
- Verze
 - v1 classful
 - v2 classless
 - ng IPv6



RIP zprávy

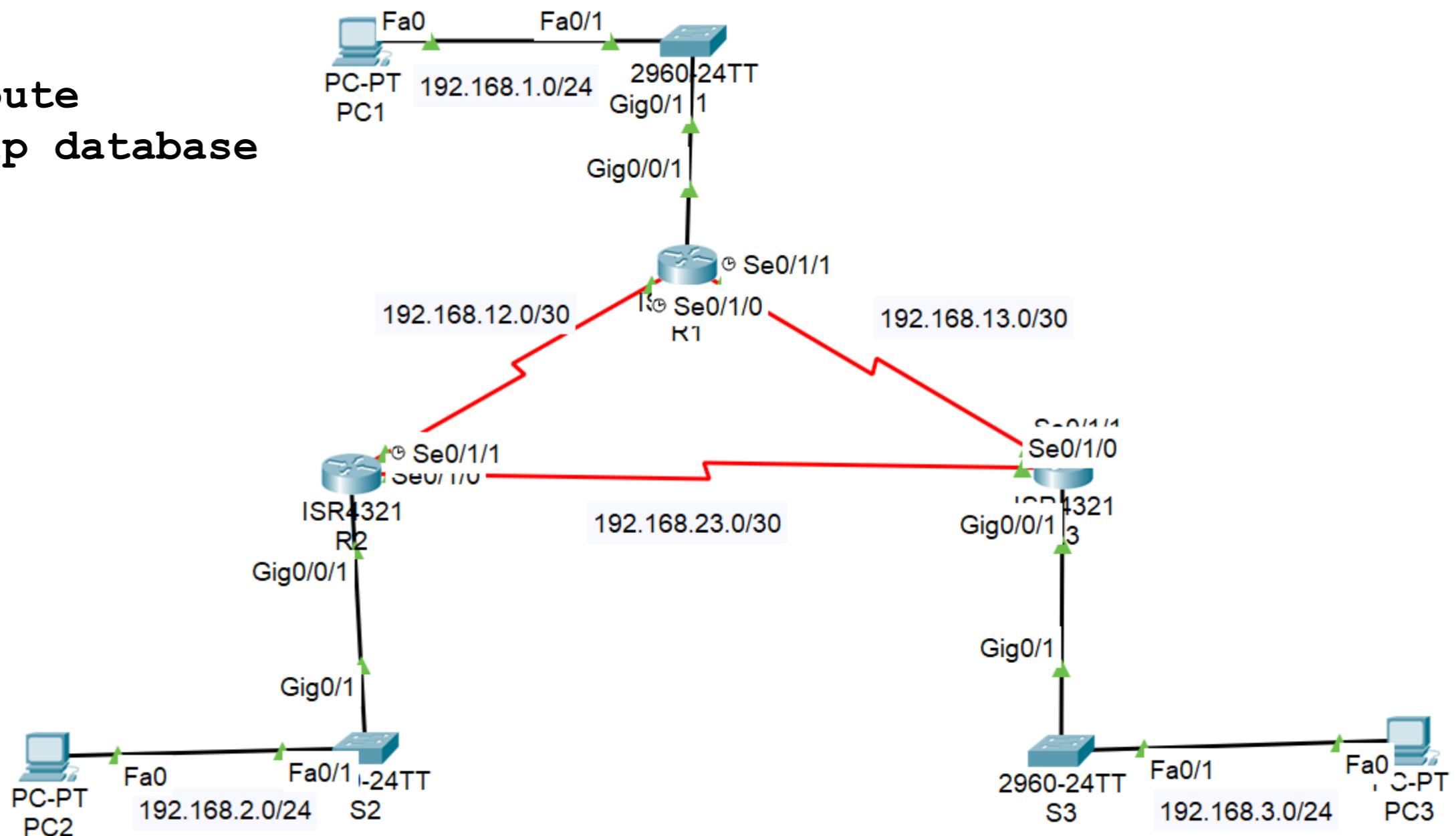
- **Response message**
 - Až 25 směrovacích záznamů
 - Periodické aktualizace
 - Spuštěné aktualizace
 - Odpověď na Request message
- **Request message**
 - Zasíláno při spuštění směrovače
 - Požadavek na zaslání kompletní či částečné směrovací tabulky

Aktualizace stavu linky

- Jestliže nepřijde po dobu 180s aktualizace od souseda je označen za neplatný
 - Cesty vědoucí přes tohoto souseda jsou zneplatněny
 - Jsou informováni sousedé
 - Sousedé aktualizují svá data a pokud došlo ke změně ve směrovací tabulce posílají oznámení
- Selhání linky/směrovače je takto šířeno celou sítí
- **Poison Reverse** je použit pro zabránění vytváření směrovacích smyček

RIP Demo

```
debug ip rip  
router rip  
network  
show ip route  
show ip rip database
```



RIP Packets

- RIPv1 Updates

<https://www.cloudshark.org/captures/00d58e1f4dd5>

- RIPv2 Updates

<https://www.cloudshark.org/captures/00bdca4b449a>

- RIPv2 Unreachable Update

<https://www.cloudshark.org/captures/016c88c0e465>

Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Dijkstra
 - Algoritmus Bellman-Ford
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP

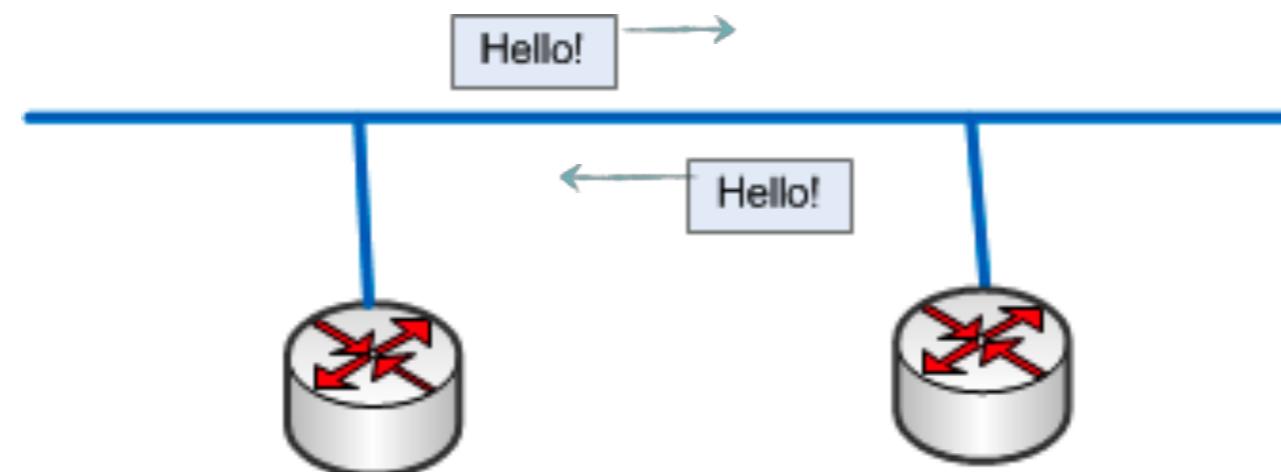
Open Shortest Path First

↳ OSPF je LS využívající Dijkstrův algoritmus

- Šíří informaci o změně v celé oblasti (LSA flooding)
- Ceny linek odráží jejich rychlosti (bandwidth, administrátor může redefinovat)
- Informace jsou posílány okamžitě v případě změny, či alespoň jednou za 30 minut
- OSPF testuje stav linky a objevuje sousedy pomocí HELLO paketů
- Dvě verze
 - v2 jen pro IPv4
 - v3 původně jen pro IPv6, nyní IPv4 + IPv5

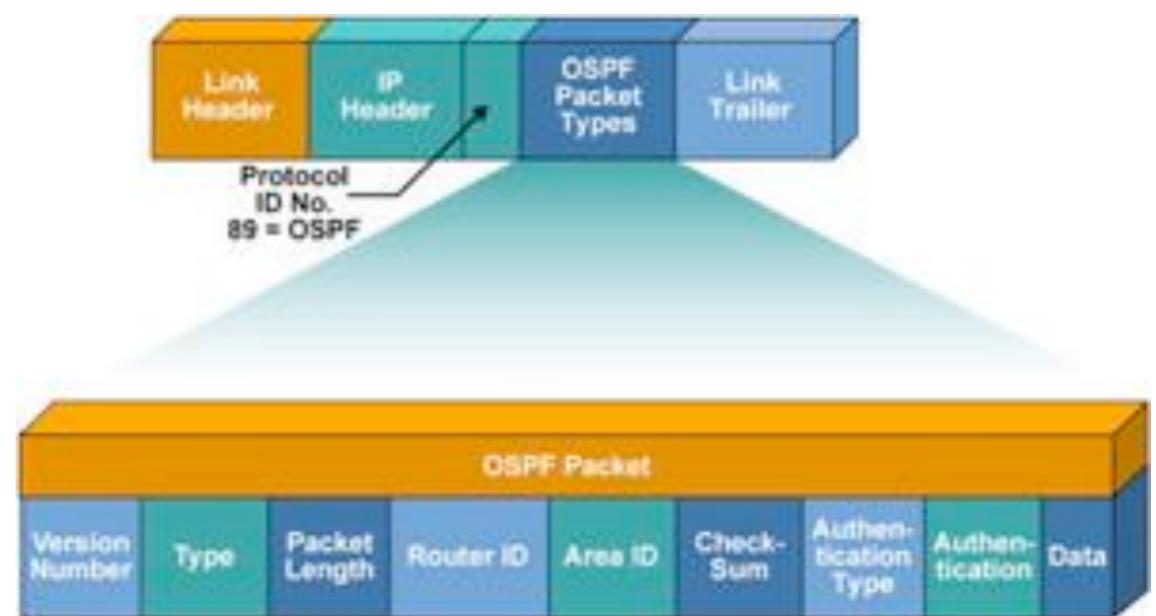
Detekce změny topologie

- **Keepalive / Beaconing**
 - pravidelné posílání krátkých zpráv oběma směry
 - detekce selhání při ztrátě několika těchto zpráv
- Není ideální
 - rychlosť detekce odpovídá intervalu zpráv
 - režie posílání zpráv bez datového obsahu
 - možnost mylné detekce



Šíření informací

- OSPF protokol je nesen IP protokolem (má číslo 89)
- Používá záplavové šíření (flooding)
 - směrovač posílá LS informace všem sousedům
 - sousedé šíří tyto informace dál
- Vyžaduje spolehlivé šíření informací
 - každý směrovač musí informaci dostat
 - všichni musí mít stejnou informaci
- Používá přímo IP paket pro přenos dat, musí řešit:
 - ztrátu paketů
 - přijetí v jiném pořadí
- Řešení
 - ACK a znovaodesílání
 - sekvenční čísla

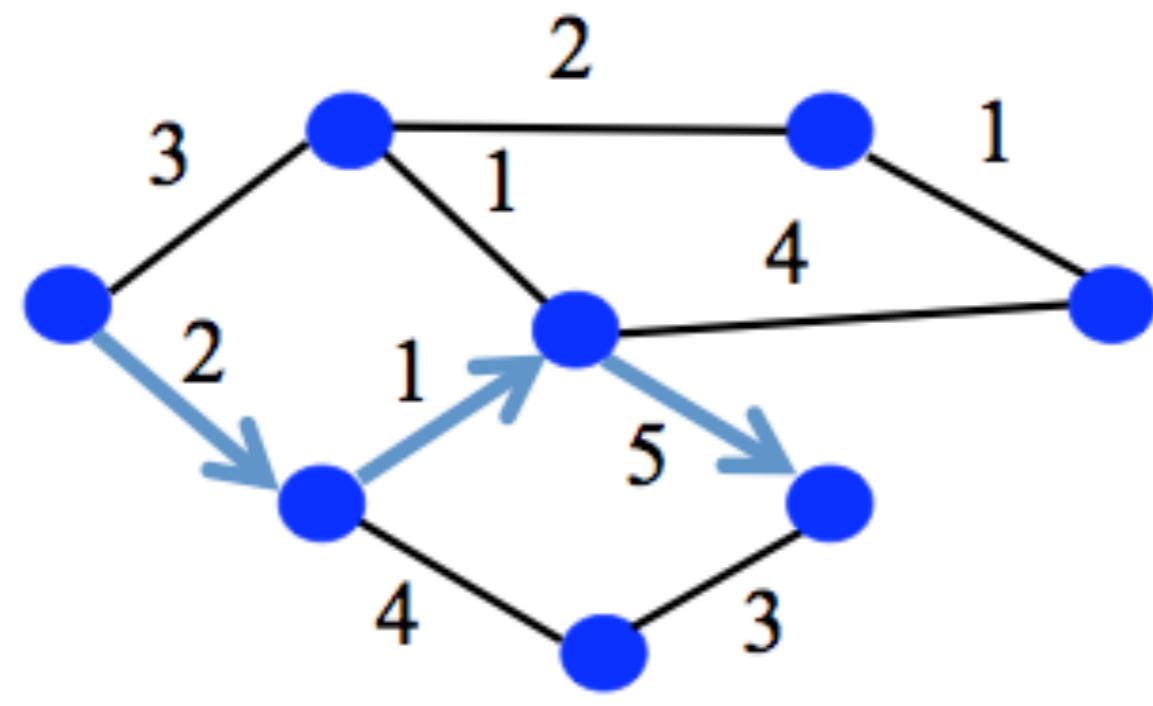


Kdy se posílají aktualizace

- Změna topologie
 - selhání linky/sousedů
 - nalezení linky/sousedů
- Změna konfigurace
 - změna nastavené ceny linky
- Periodicky
 - Většinou každých 30 minut
 - Korekce stavu
 - Pro jistotu, že všichni mají stejné informace

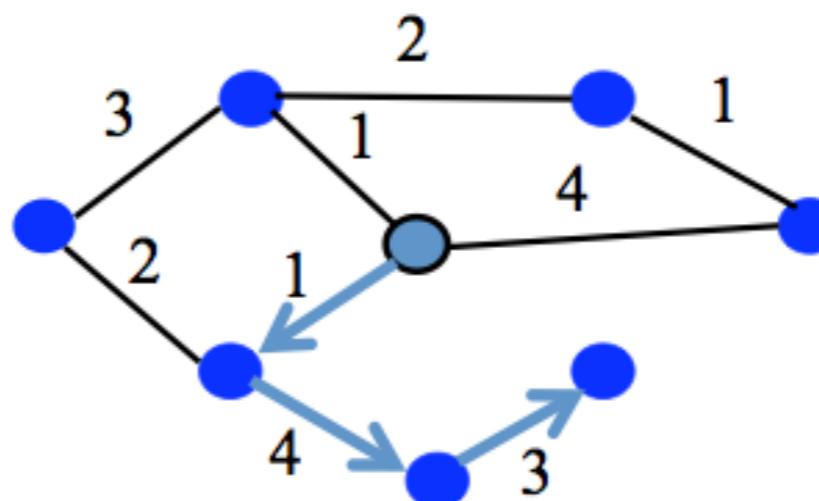
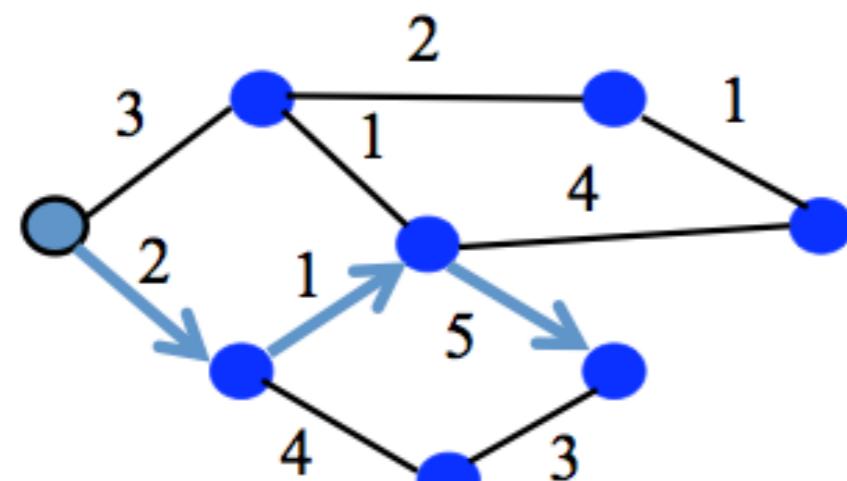
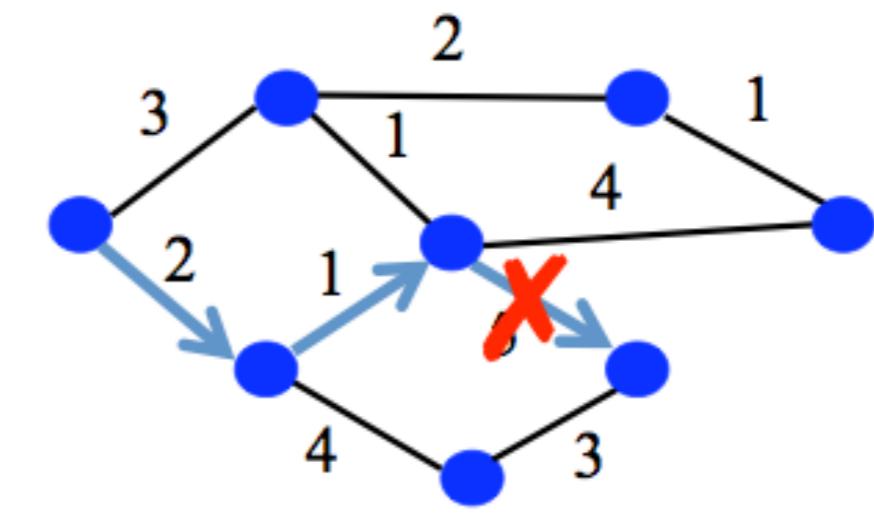
Konvergencie

- Konzistentní informace ve všech uzlech
 - všechny uzly mají stejnou LS databázi
- Směrování je konzistentní v konvergovaném stavu
 - všechny uzly mají stejnou LS databázi - stejná topologie
 - pakety jdou nejkratší cestou



Problémy

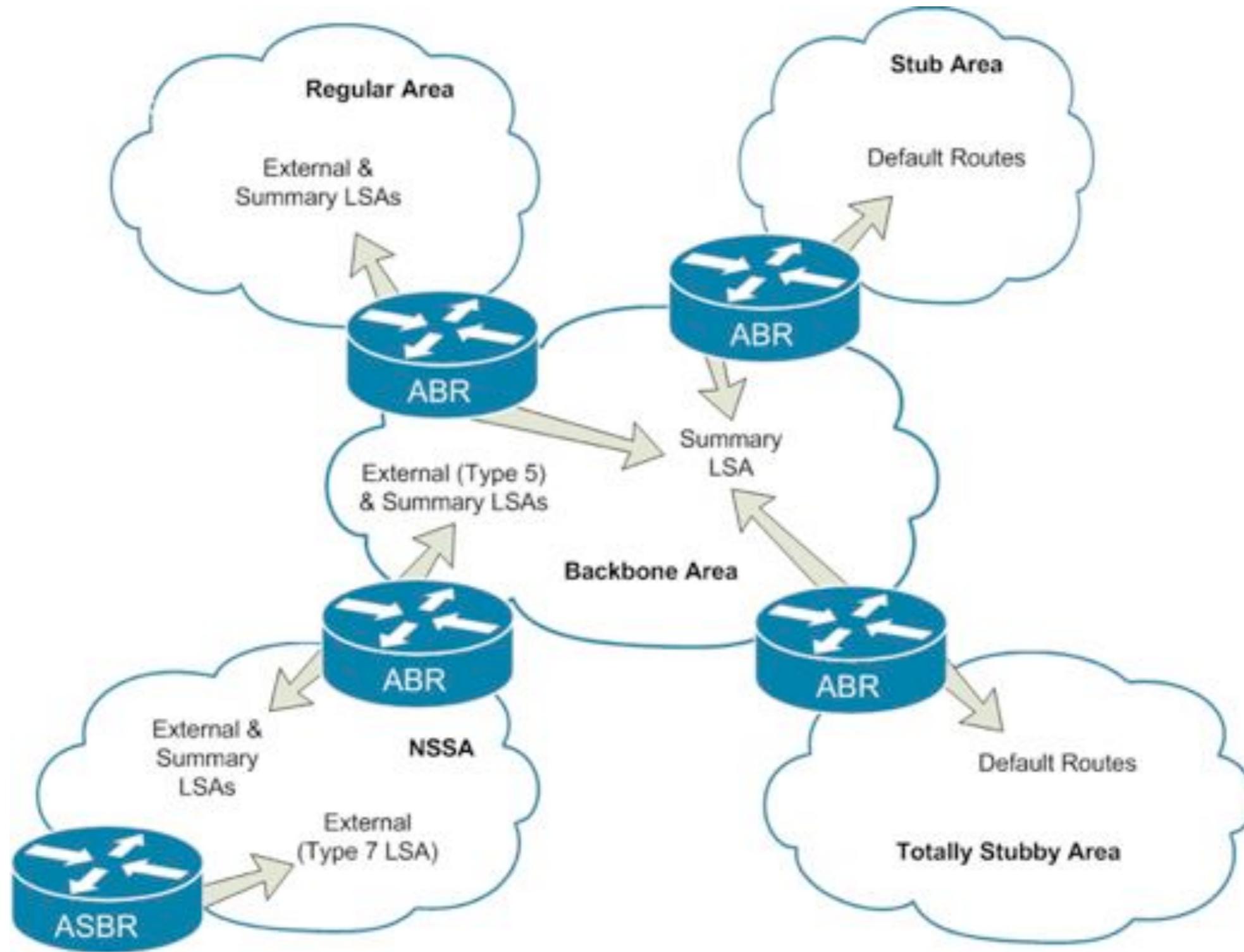
- Nekonsistentní databáze
 - některé směrovače mají jinou informaci
 - mohou vědět o selhání zatímco jiné ještě ne
 - můžezpůsobit
 - zvýšenou ztrátovost
 - nemožnost komunikovat
 - dočasnou směrovací smyčku



Doba konvergence

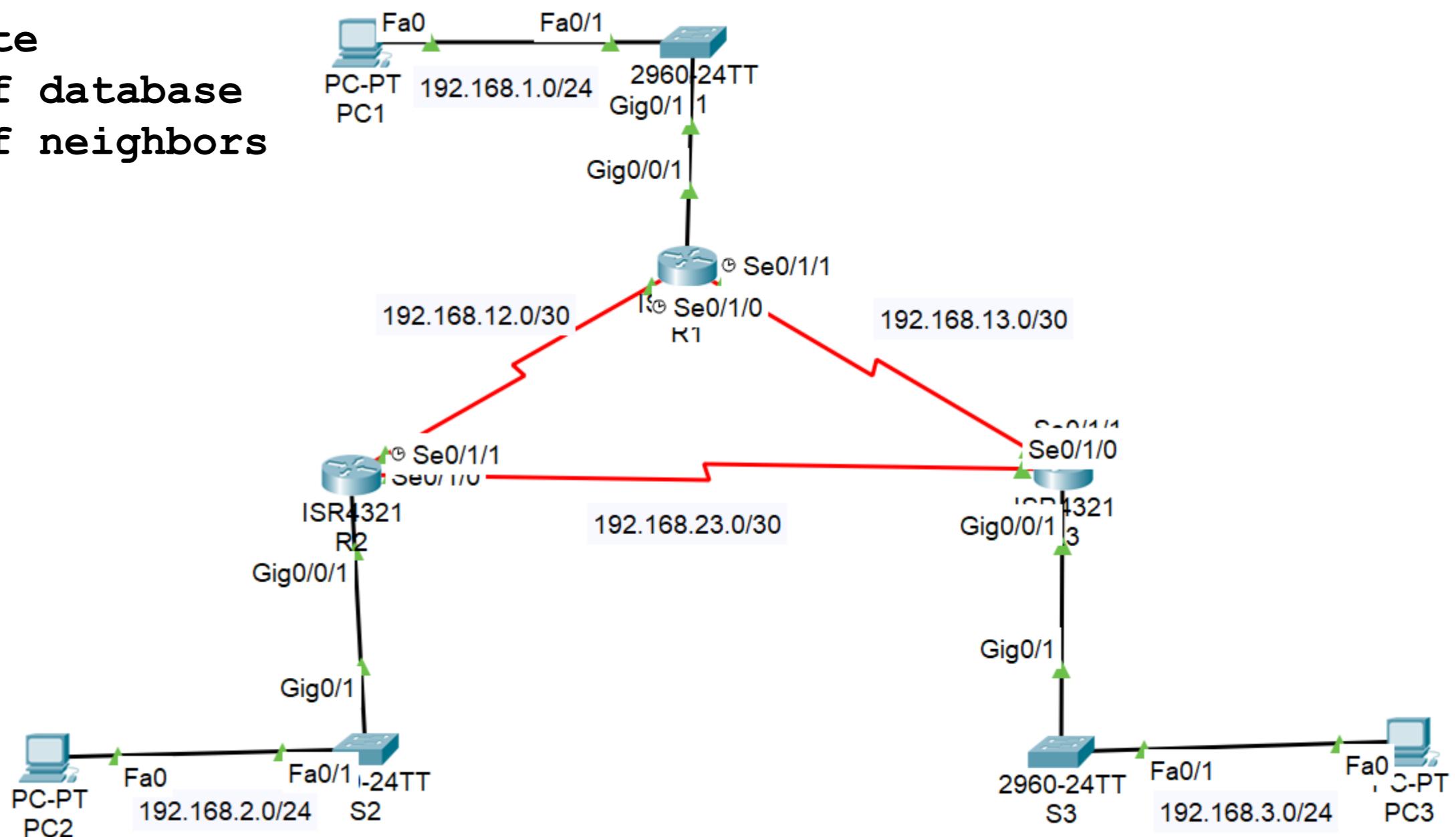
- Zdroje zpoždění
 - detekce
 - záplava LS informací
 - výpočet SP
 - naplnění FT
- Důsledky
 - ztracené pakety
 - nedoručitelné pakety zabírají zdroje
 - pakety mimo pořadí
 - citlivé aplikace (VoIP, video)
- Opatření
 - Rychlejší detekce
 - kratší HELLO interval
 - detekce na L2
 - Rychlejší šíření LS
 - okamžité informování
 - priorita pro LS pakety
 - Rychlejší výpočet
 - rychlejší HW
 - inkrementální DA
 - inkrementální změna FT

Škálovatelnost díky oblastem



OSPF Demo

```
debug ip ospf  
router ospf  
network  
show ip route  
show ip ospf database  
show ip ospf neighbors
```



OSPF Messages

- OSPF LSA Types

<https://www.cloudshark.org/captures/0062204357ab>

- OSPF With Authentication

<https://www.cloudshark.org/captures/007bba156585>

Obsah

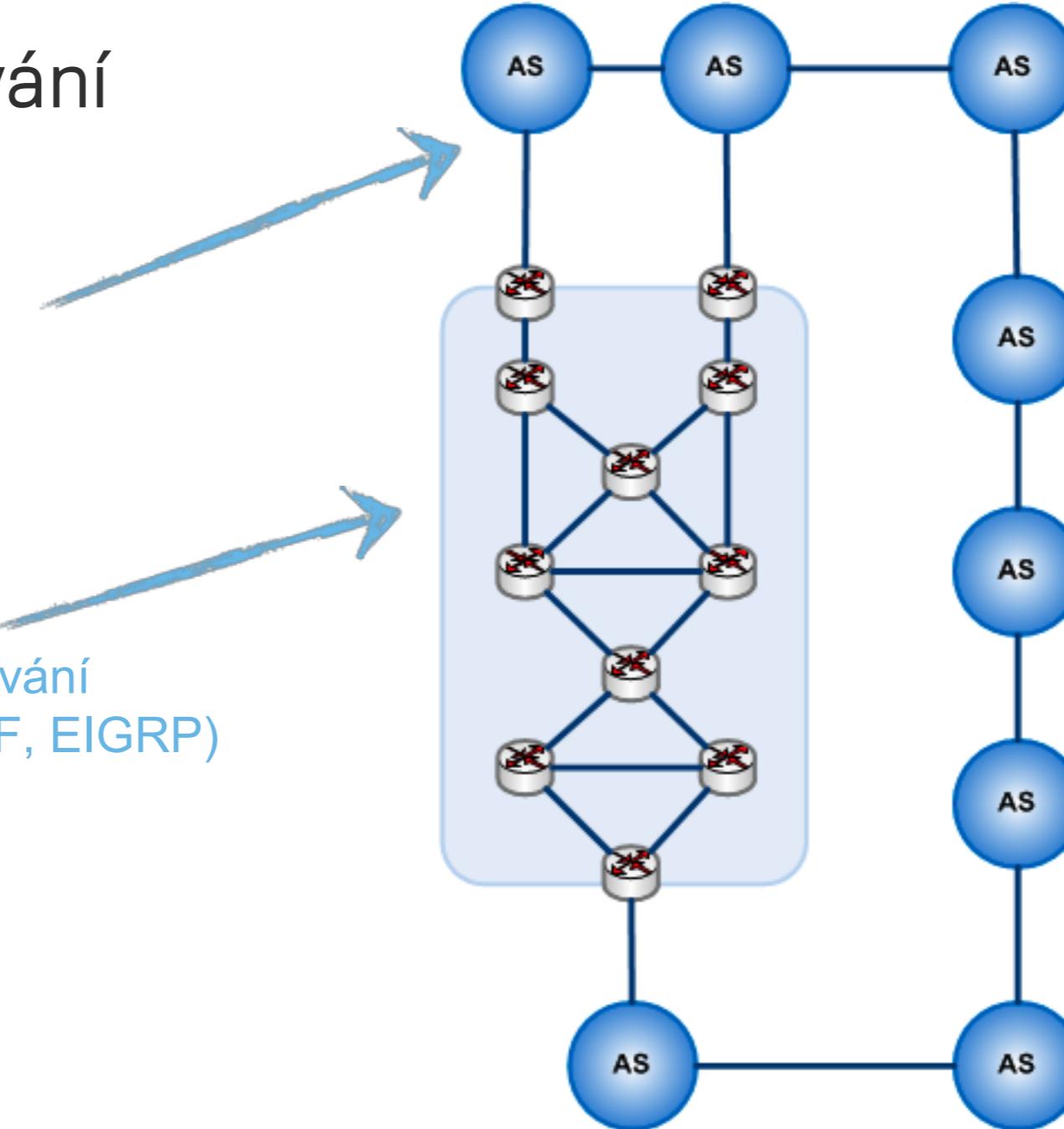
- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Dijkstra
 - Algoritmus Bellman-Ford
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

Směrování v Internetu

- Dvě úrovně směrování

Interdomain směrování
mezi domény
(BGPv4)

Doména
IGP směrování
(RIP, OSPF, EIGRP)



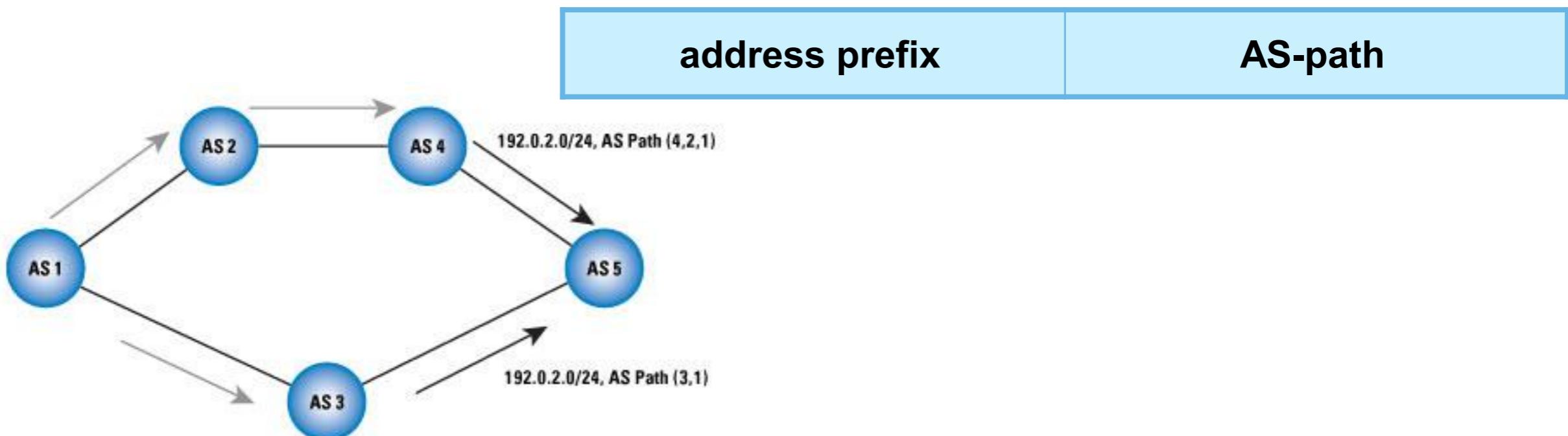
Internet je soubor domén - AS

- Je rozdělen do domén zvaných **autonomní systémy**
 - samostatné oblasti (z pohledu Internetu)
 - síť spravované jednou institucí
 - poskytovatelé služeb, firmy, univerzity, ...

An AS is a group of IP networks operated by one or more network operator(s) that has a single and clearly defined external routing policy. Exterior routing protocols are used to exchange routing information between ASes.
[RFC1930]

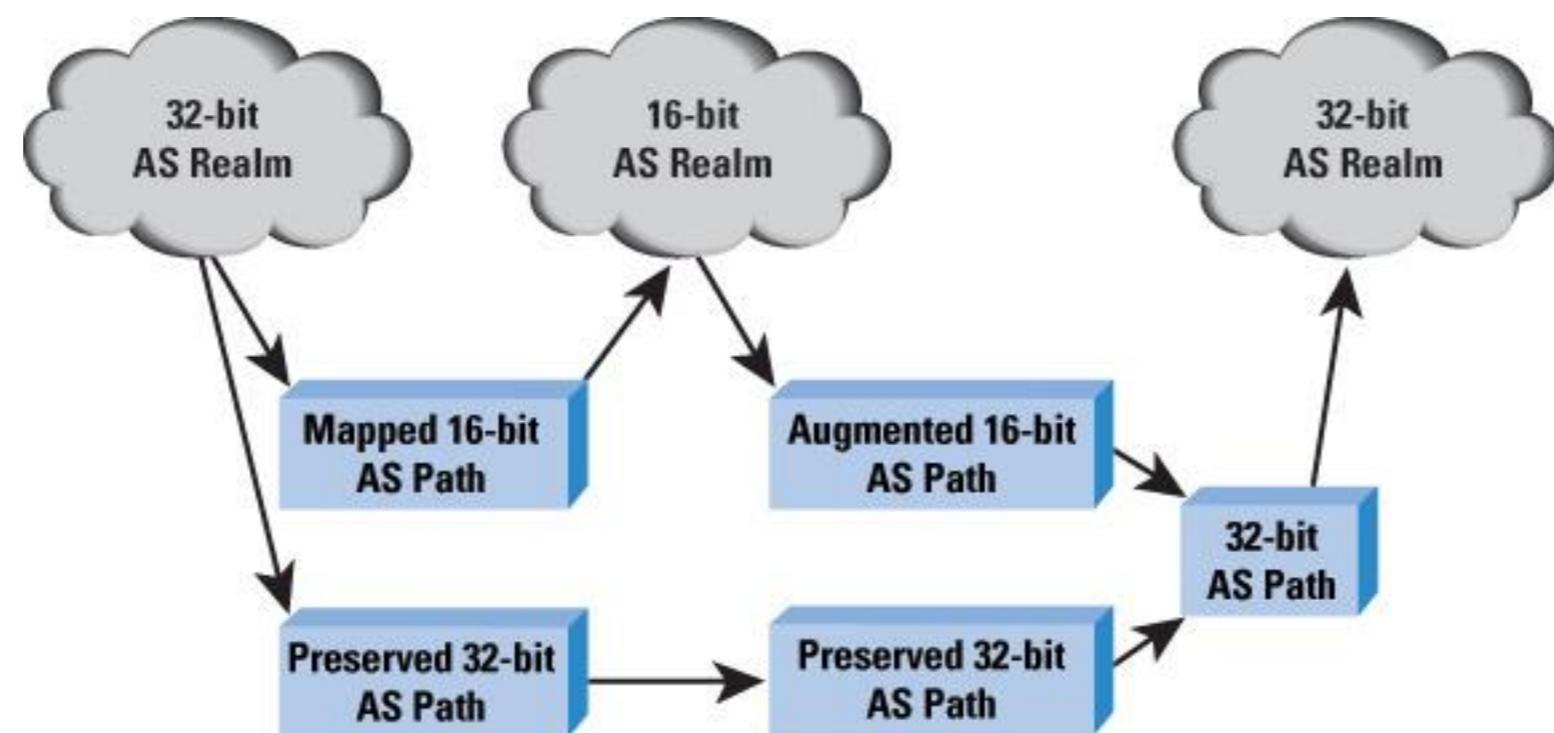
ASN

- Identifikovány pomocí autonomous system number (ASN)
 - <http://bgp.potaroo.net/cidr/autnums.html>
- ASN je buď 16 bitové číslo, a nebo 32b
 - 1-64511: veřejné AS
 - 64512-65534: privátní AS
- Použití v interdomain směrování:



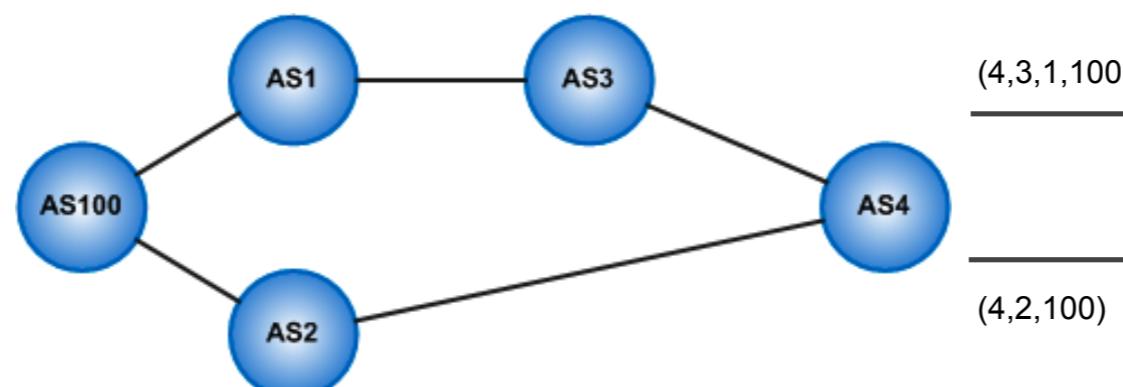
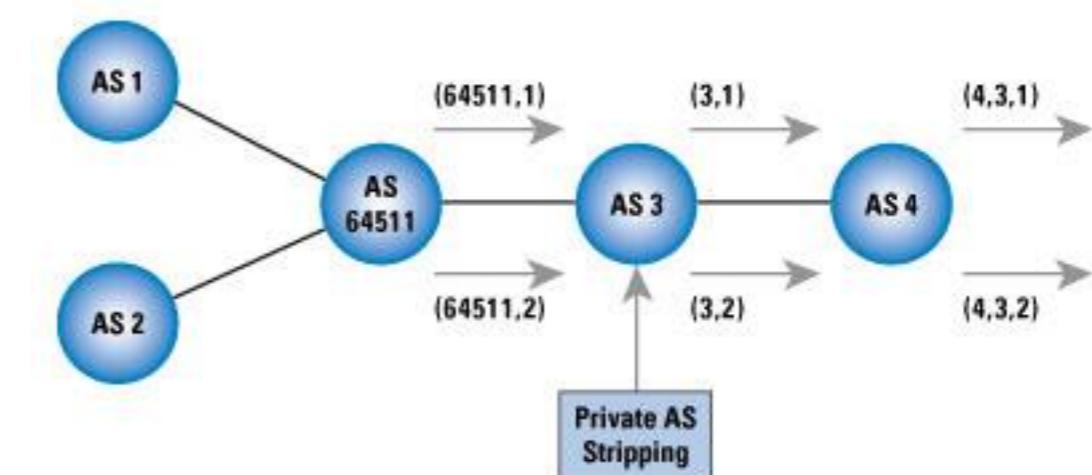
Potřeba většího počtu ASN

- podle předpokladů budou 16-ti bitové ASN vyčerpány
- změna AS z 16-ti bitů na 32 bitů: X.Y [RFC4893, 2007]
- vyžaduje změnu v BGP protokolu
- v Internetu musí spolukomunikovat BGP pro 16-bitů AS a BGP pro 32-bitů AS



Kdo potřebuje ASN?

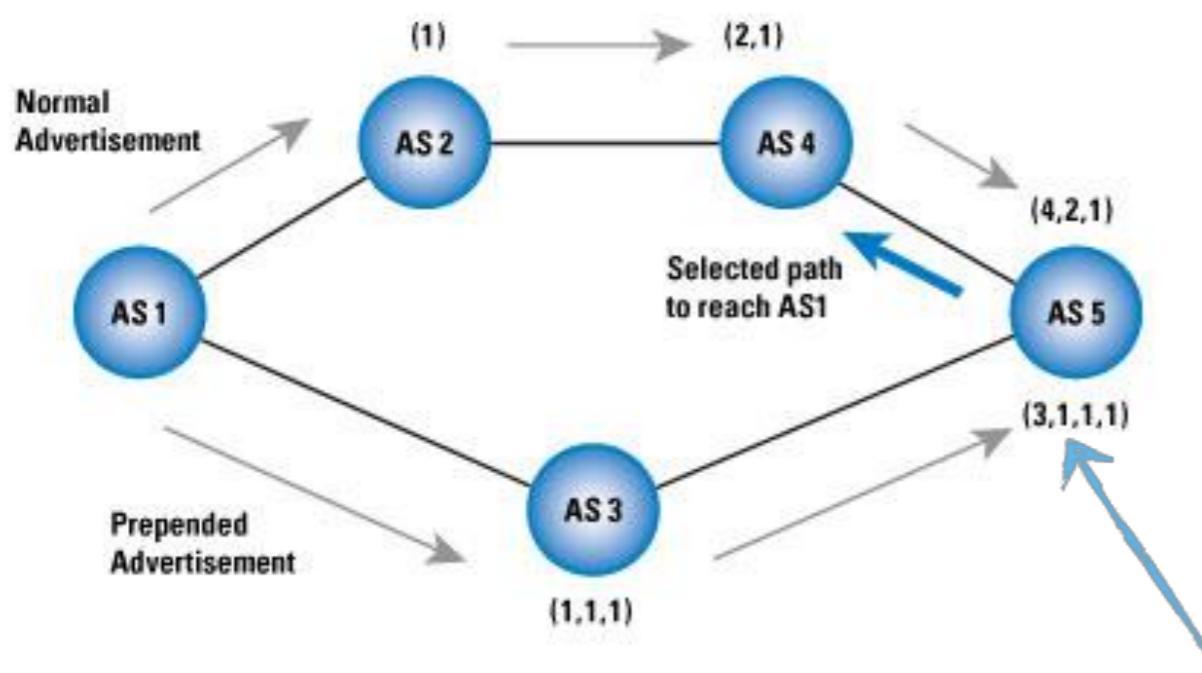
- ASN identifikují sítě s rozdílními směrovacími politikami
- běžní zákazníci mají ASN providerů
 - v případě použití BGP lze přiřadit privátní ASN
 - Private ASN Stripping
- veřejné ASN zákazníka
 - multihoming
 - nutnost rozlišit ve směrování více možných cest do sítě



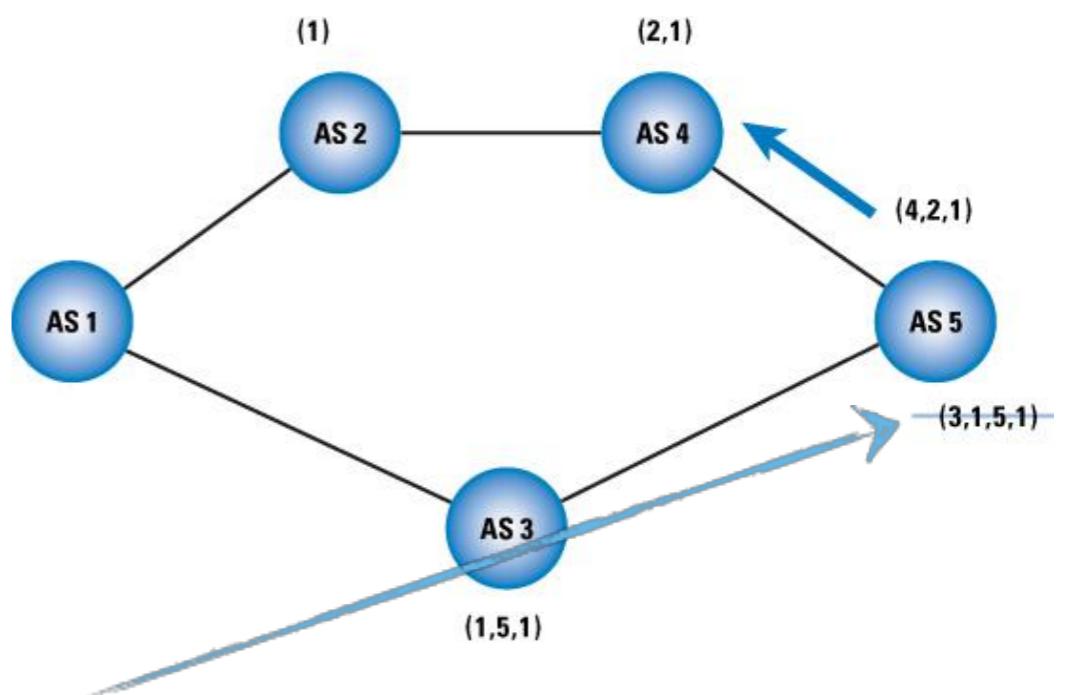
Výběr nejlepší cesty podle ASN

- BGP implicitně vybírá cestu i podle nejkratší AS-Path
- Modifikace tohoto chování je možná:
 - jednoduché, ale může způsobit komplikace
 - lepší řešení BGP community

AS path prepending



AS path poisoning



jaký je zde rozdíl?

Využití AS

- nárast ASN cca 3500 každý rok
 - především způsoben multihomingem
 - použitím v MPLS VPN
 - ISP mají více ASN pro vyjádření různých politik směrování
- ASN spravuje IANA, která alokuje bloky po 1024 jednotlivým RIR
 - RIR alokuji ASN pro ISP a koncové sítě

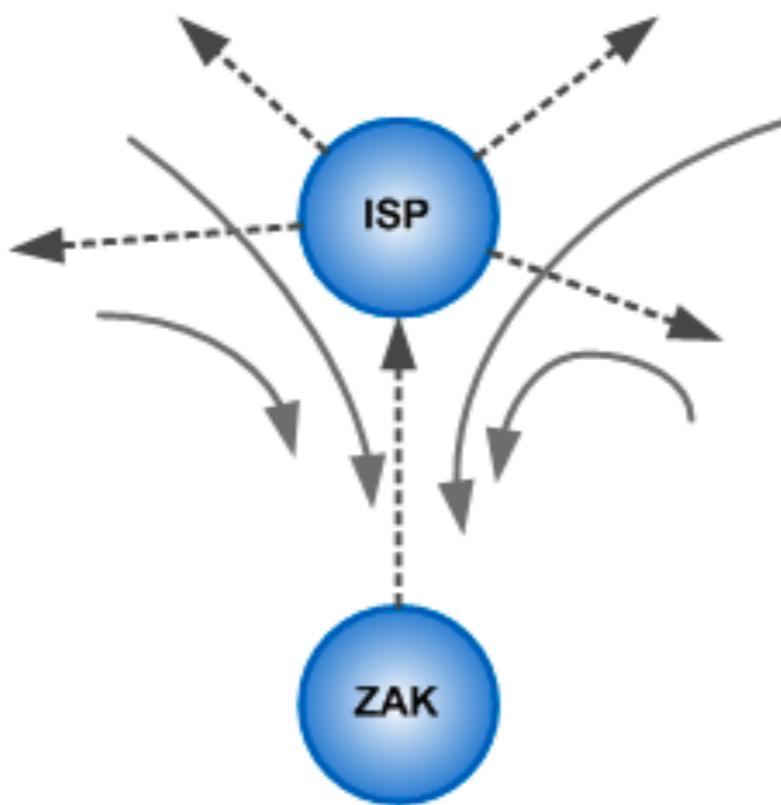
AS Reports (plots)	Data Sets(txt)	Additional Reports (plots)	Data Sets(txt)
Unique ASes	64575	ASes visible in only one AS path	38386
Origin only ASes	54254	Origin ASs announced via a single AS path	37561
Transit only ASes	345	Originating AS ATOM count	64230
Mixed ASes	9976	Originating AS ATOM compression	0.0800
Multi-Origin Prefixes	8670		
ASes originating a single prefix	23342		
Average entries per origin AS	12.5033	Maximum entries for an origin AS (AS61317)	771
Average address range span for an origin AS	44276.4318	Maximum address range for an origin AS (AS4134)	117678336

Vztah mezi AS

- Sousední AS mají mezi sebou smluvní vztah ([service level agreement](#))
 - kolik dat budou přenášet
 - do jakých cílových cílů budou doručovat data
 - kolik se za to bude platit
- Typické vztahy
 - zákazník-poskytovatel
 - VUT je zákazník CESNETU
 - ČD-Telematika je národní ISP pro lokální ISP
 - peer-peer
 - CESNET je peer Telefonica O2 CZ ([NIX.CZ](#))
 - AT&T je peer Sprintu

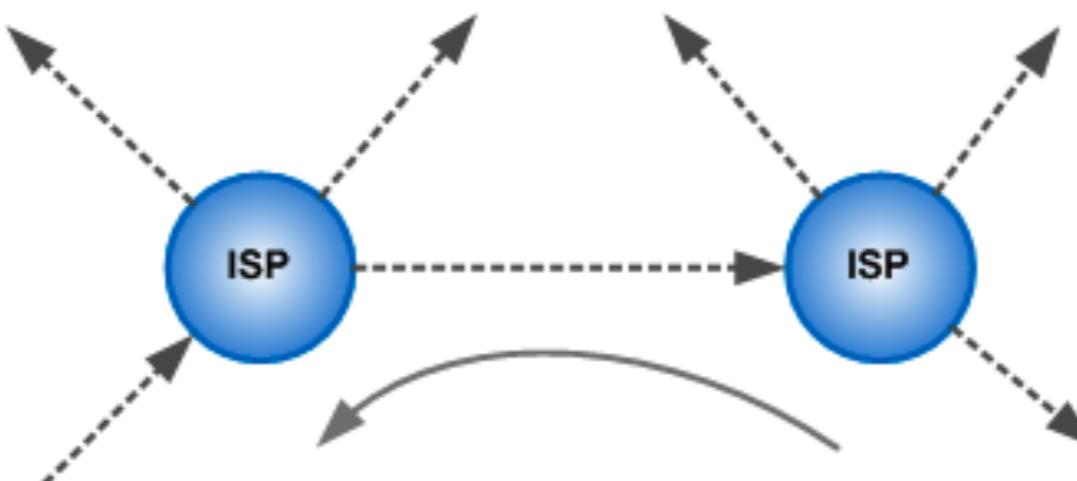
Zákazník-poskytovatel

- Zákazník vyžaduje dostupnost
 - poskytovatel oznamuje sousedům informace o zákazníkovi
- Zákazník nechce dostávat data, která mu nepatří
 - transit traffic



Peer-peer vztah

- Vyměňují si data svých zákazníků
 - pouze sítě zákazníků jsou šířeny ostatním
 - sítě ostatních peerů jsou oznamovány zákazníkům
 - často bez finančního vyrovnání

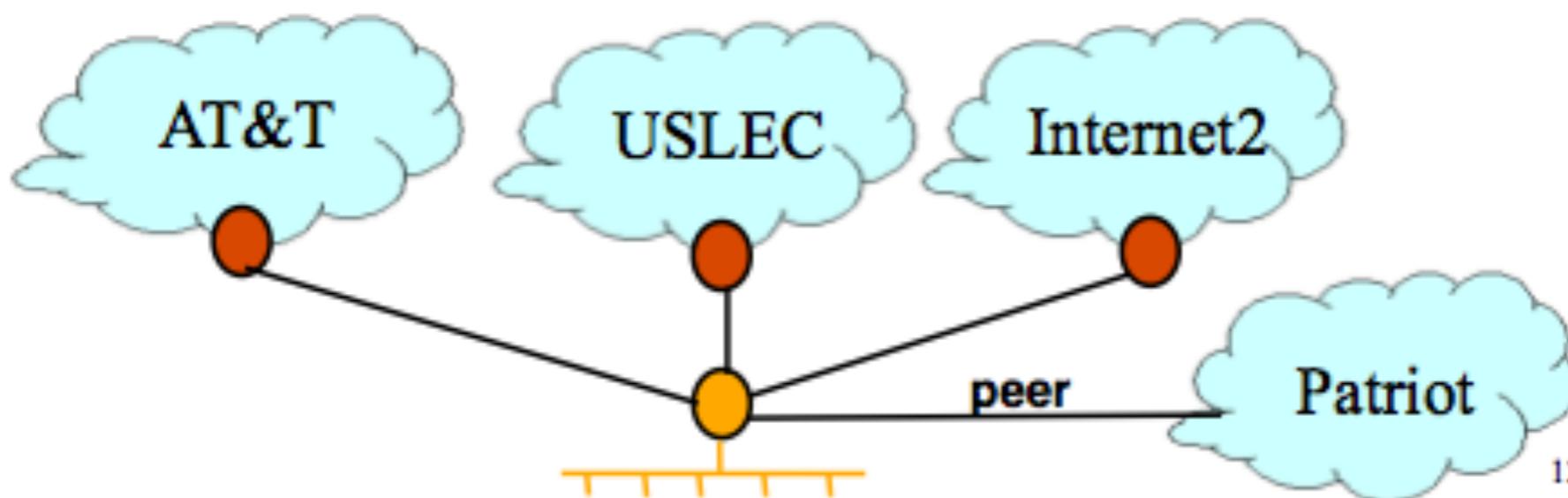


Vícenásobné vztahy

Princeton Example



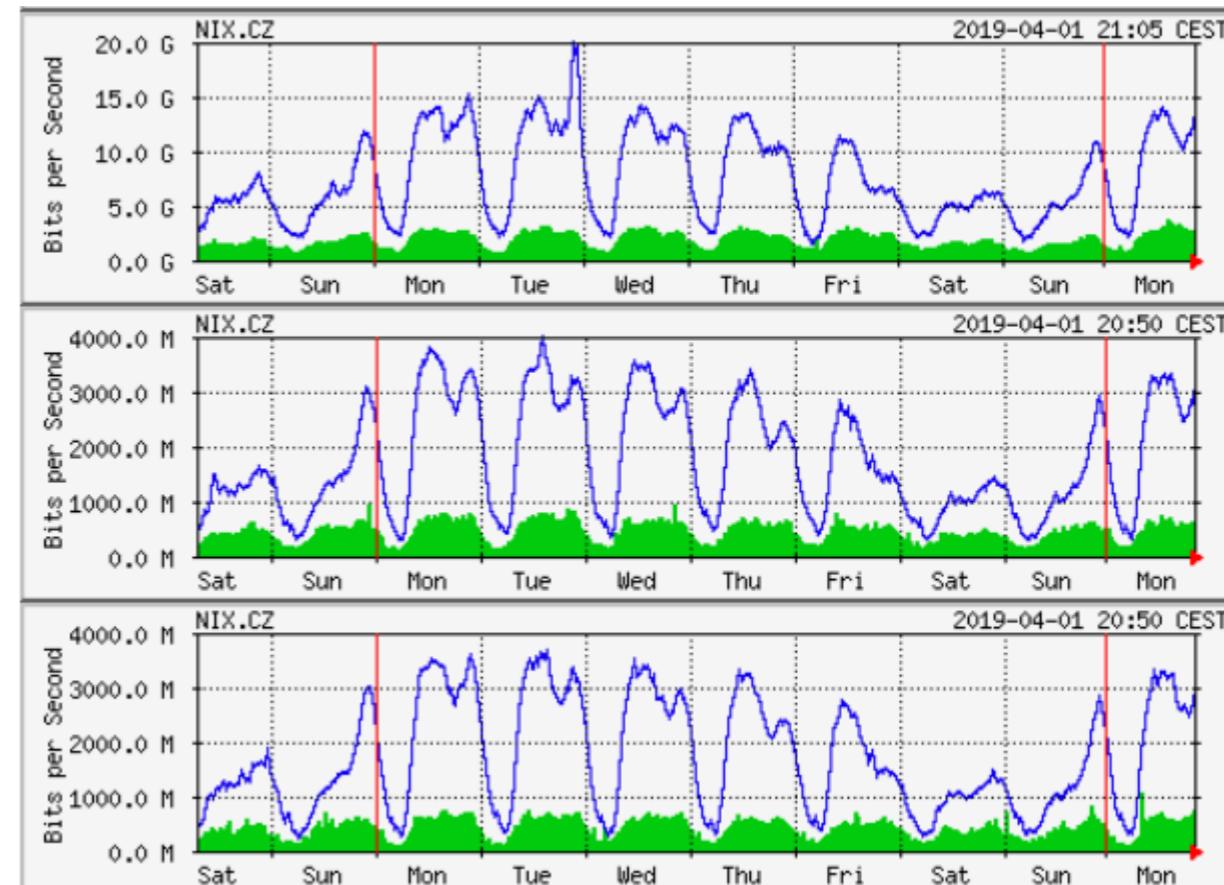
- Internet: customer of AT&T and USLEC
- Research universities/labs: customer of Internet2
- Local residences: peer with Patriot Media
- Local non-profits: provider for several non-profits



12

Peeringová centra

- Propojují různé ISP
 - NIX.CZ (www.nix.cz)
CESNET
NIX1-acc3
Ethernet3/2
100 Gb up
 - BRIX (<http://br-ix.cz>)
- Podmínky připojení:
 - vlastní ASN
 - konektivita do NIX
 - POP
- Cena
 - dle typu připojení + náklady za konektivitu do centra

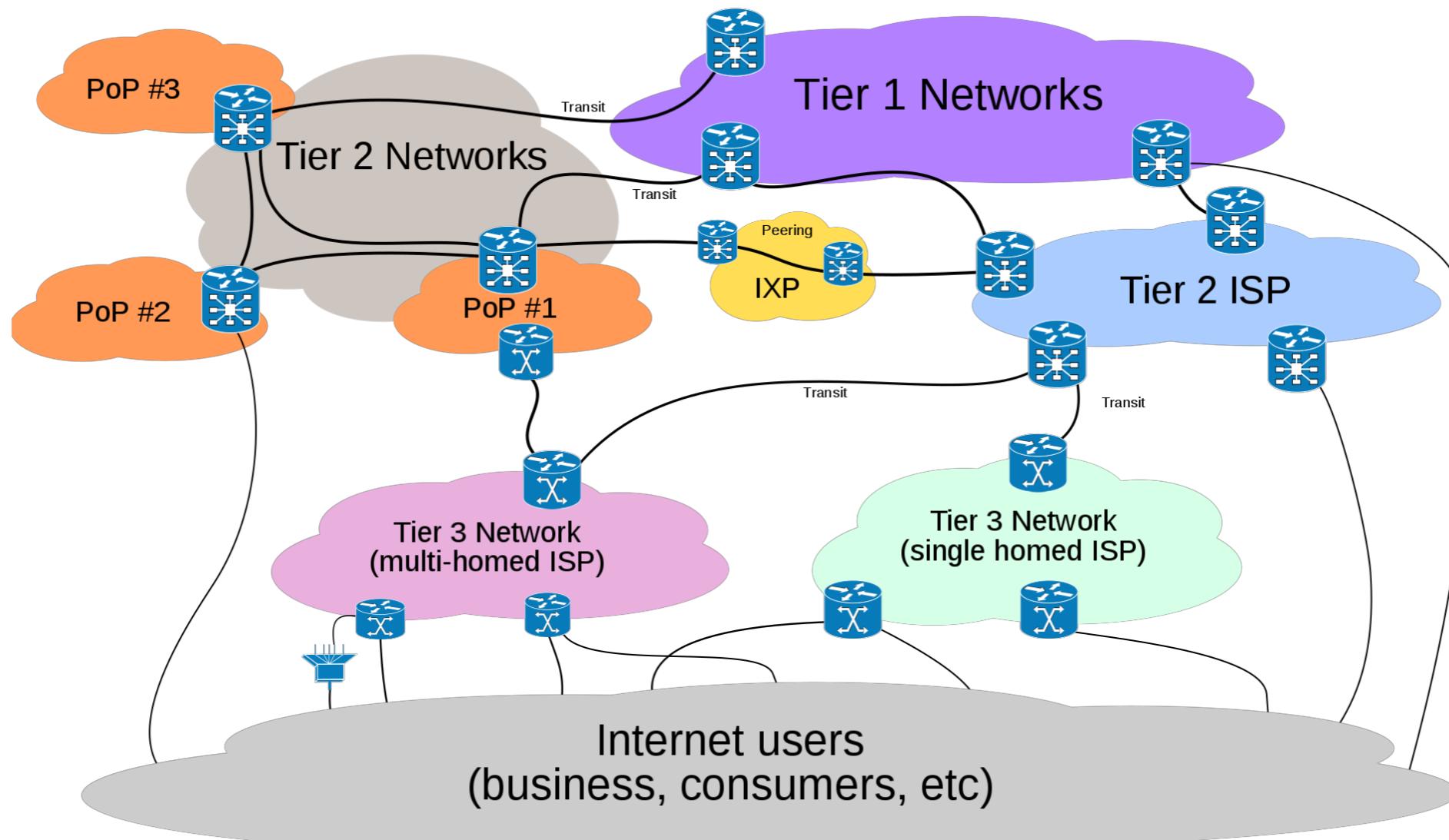


Příklad CESNET

- Podmínky pro peering
 - Partner musí mít vlastní NOC (Network Operating Centre) v nepřetržitém provozu.
 - Partner musí provozovat vlastní internetovou síť s homogenní směrovací politikou.
 - Partner musí být LIR (Local Internet Authority), musí mít vlastní adresní prostor a autonomní systém.
 - Směrovací tabulky musí být maximálně agregovány (mechanismem CIDR), CESNET2 nepřijímá síťové prefixy delší než /24.
 - Partner nesmí šířit defaultní cestu (route of last resort) přes propojení se sítí CESNET2.
 - Všechnen provoz mezi sítěmi CESNET2 a partnera musí být směrován přímým propojením mezi sítěmi.
 - Sdružení umožňuje propojení infrastrukturou NIX.CZ. Připojení k infrastruktuře NIX.CZ realizuje každá strana samostatně.

Struktura Internetu: Tier-1

- Není ve vztahu zákazník s žádným jiným ISP
 - Má vlastní páteřní síť
 - Plný peering mezi Tier-1 ISP



Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Dijkstra
 - Algoritmus Bellman-Ford
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

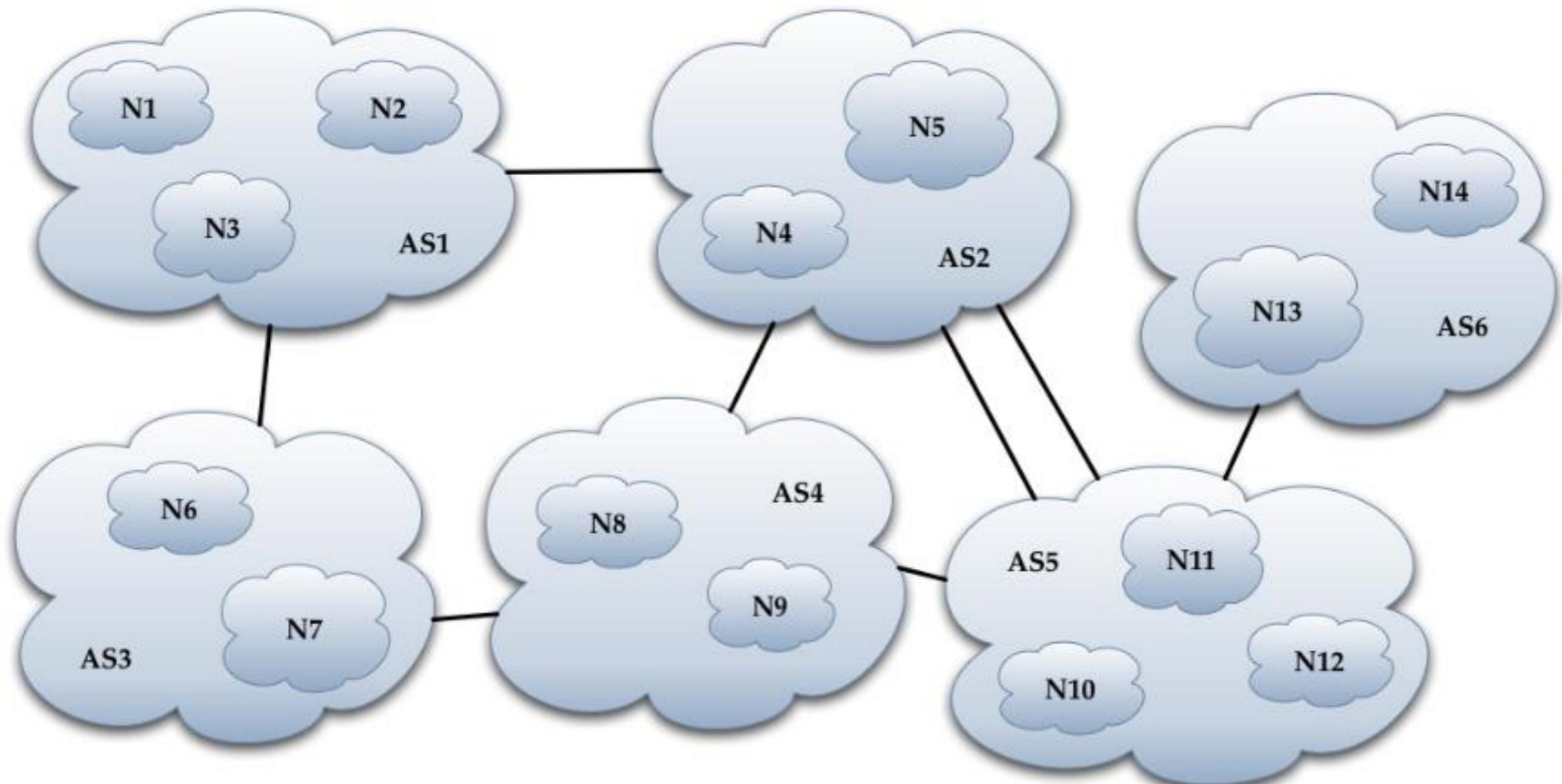
Border Gateway Protocol

- vytvořen koncem 1980/začátkem 1990
- nyní ve verzi BGP-4
- definováno v RFC 4271
- informace jsou neseny protokolem TCP, defaultní port 179
- Příklad BGP tabulky:
 - <http://bgp.potaroo.net/as2.0/bgptable.txt>

BGP

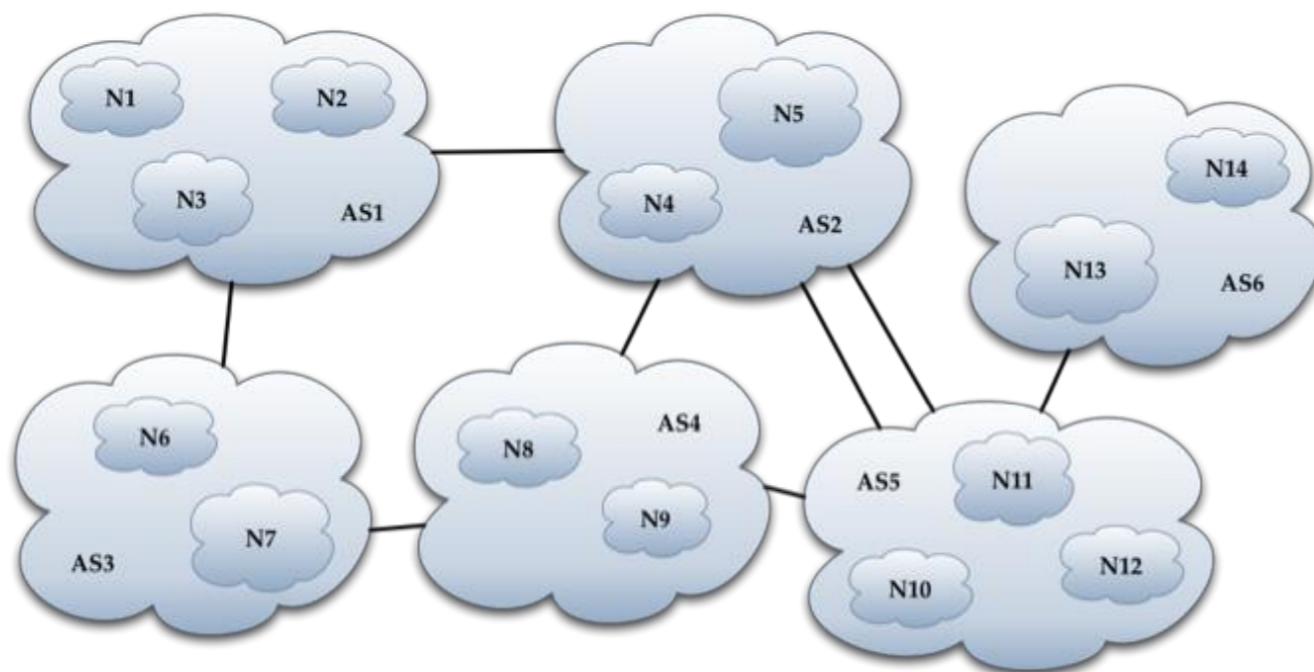
- Co je BGP?
 - ↔ směrovací protokol, který spojuje celý Internet dohromady
 - vyměňuje si informace o sítích spravovaných v rámci AS
 - síť je zde myšlen blok IP prefixů
 - BGP pracuje v rámci propojených AS
 - AS představují superuzel pro BGP
 - BGP agent pro tento superuzel komunikuje s agentem pro sousední superuzel
 - agent = BGP speaker
 - spojení mezi agenty = BGP session
 - cena spojení = AS-hop cost

BGP struktura



Komunikace

- AS6 posílá AS5, že vlastní dva adresové bloky
 $(AS6) \rightarrow \{N13, N14\}$
- Každý AS přidá svoje ID a informuje sousedy



$(AS2, AS5, AS6) \rightarrow \{N13, N14\}$

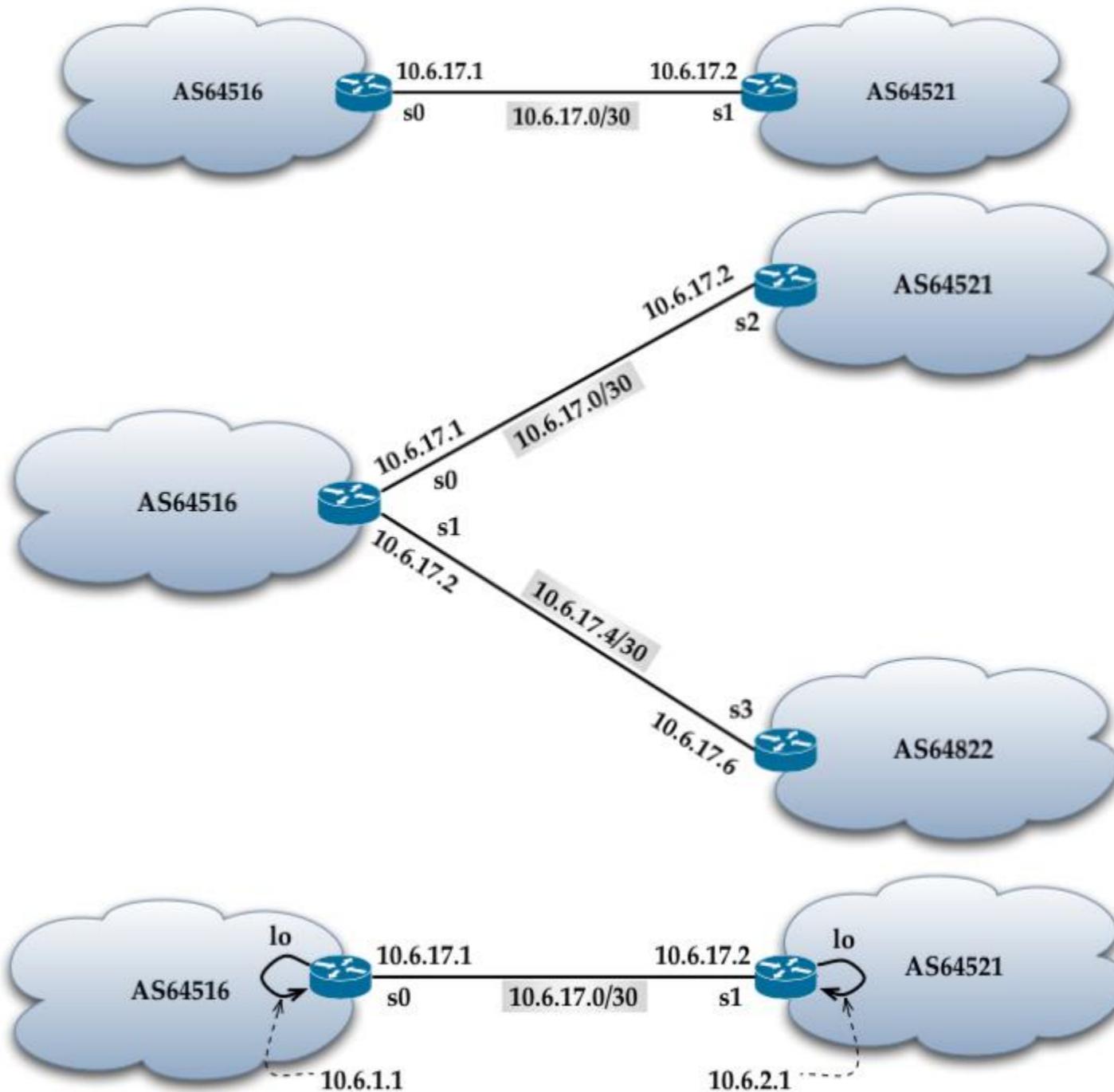
$(AS3, AS4, AS5, AS6) \rightarrow \{N13, N14\}$

lepší dle AS-path

Zprávy

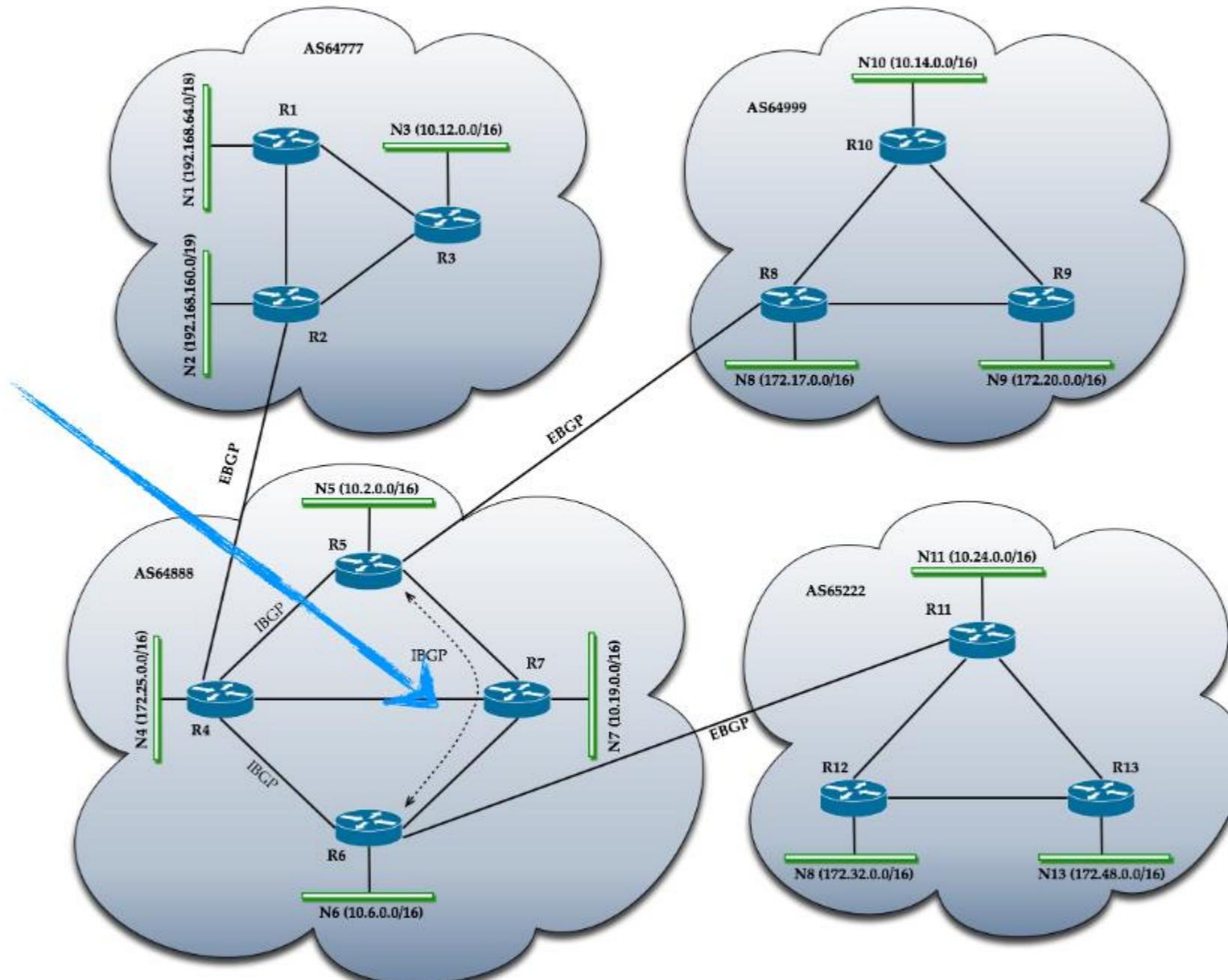
- BGP pracuje nad TCP na portu 179
 - spolehlivé doručování zpráv
- Po ustanovení TCP spojení mezi dvěma BGP uzly:
 - **OPEN**: zahajení komunikace, definuje hold time
 - **UPDATE**: výměna informací o IP prefixech
 - **KEEPALIVE**: periodické udržování spojení
 - **NOTIFICATION**: korektní ukončení relace
 - **ROUTE-REFRESH**: aktivní dotazovaní informace (novinka BGP-4)

Sousedství



I-BGP

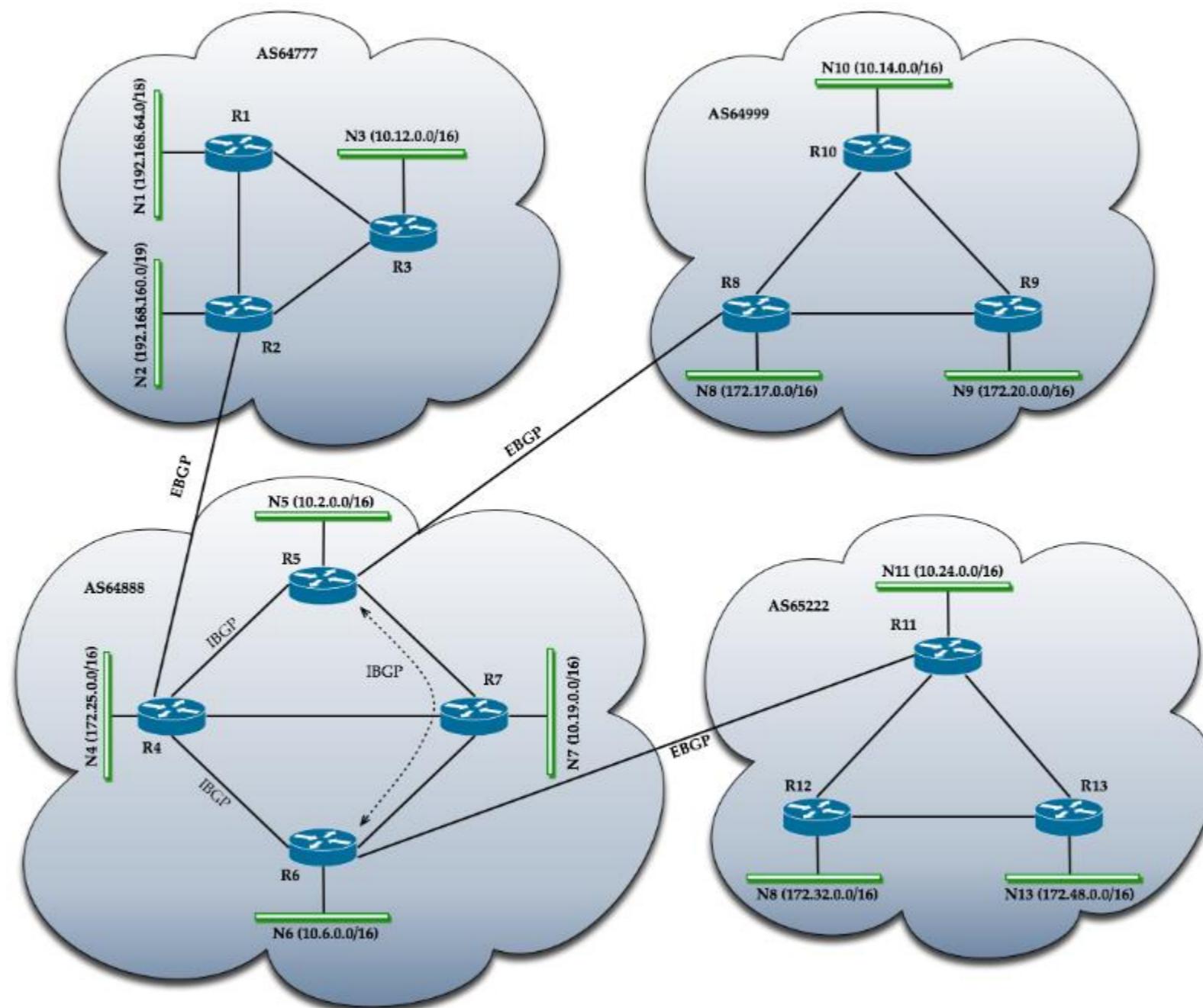
- Směrovače potřebují znát cestu k sousedním AS ze svého AS



I-BGP a E-BGP

- intra-AS ~ I-BGP
- inter-AS ~ E-BGP
- I-BGP
 - uvnitř AS je potřeba vyměňovat si informace mezi různými BGP uzly
 - Pravidla
 - BGP uzel může oznamovat prefix, který se naučil od E-BGP všem I-BGP sousedům
 - BGP může oznamovat prefix, který se naučil od I-BGP všem E-BGP sousedům
 - **I-BGP nemůže oznamovat prefix naučený od I-BGP jiným I-BGP sousedům (mohla by vzniknout smyčka)**

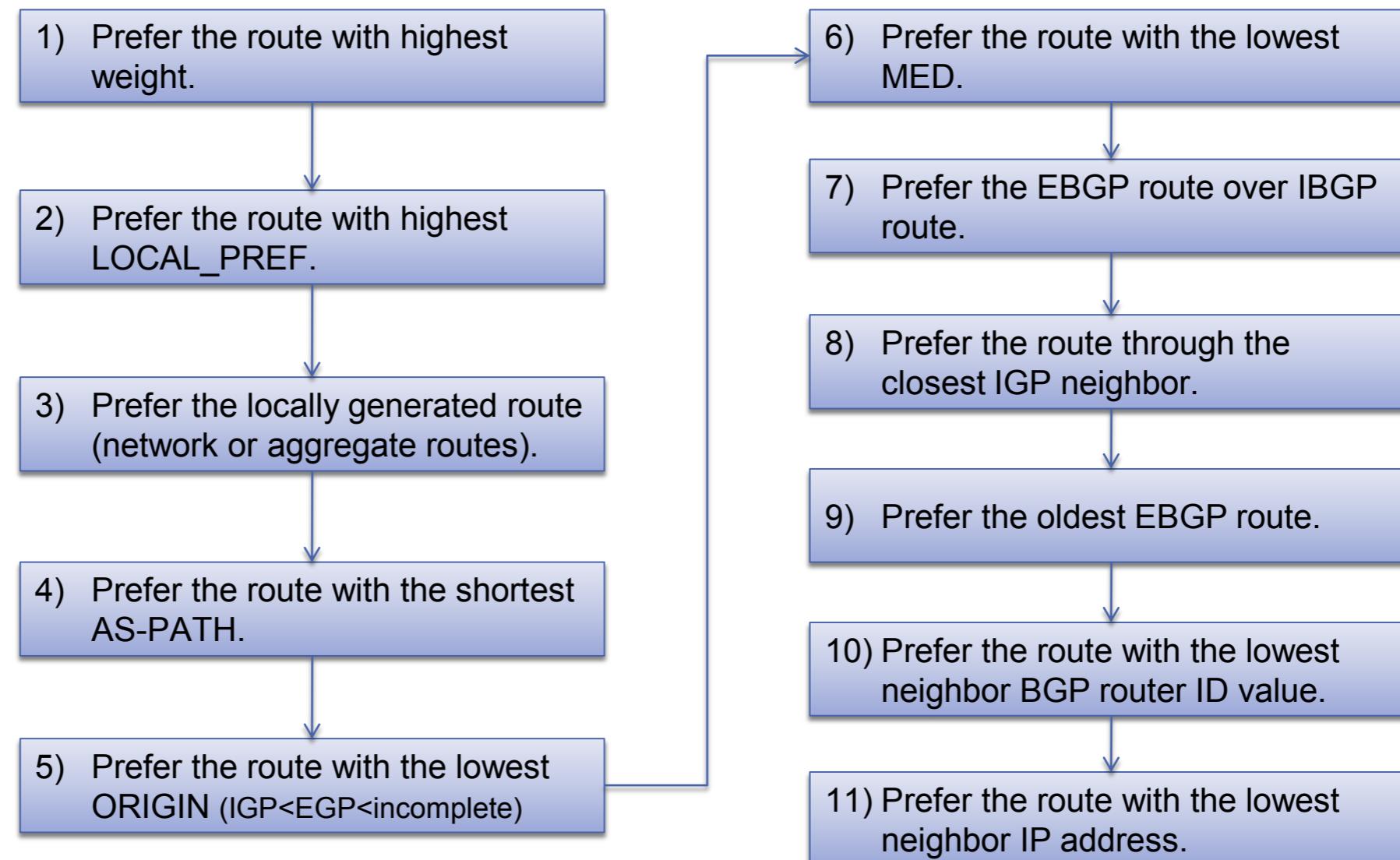
I-BGP a E-BGP



R4 se dozví o N11, N8 a N13 od AS65222 od R6
R4 nemůže informovat R5 o N11, N12 a N13

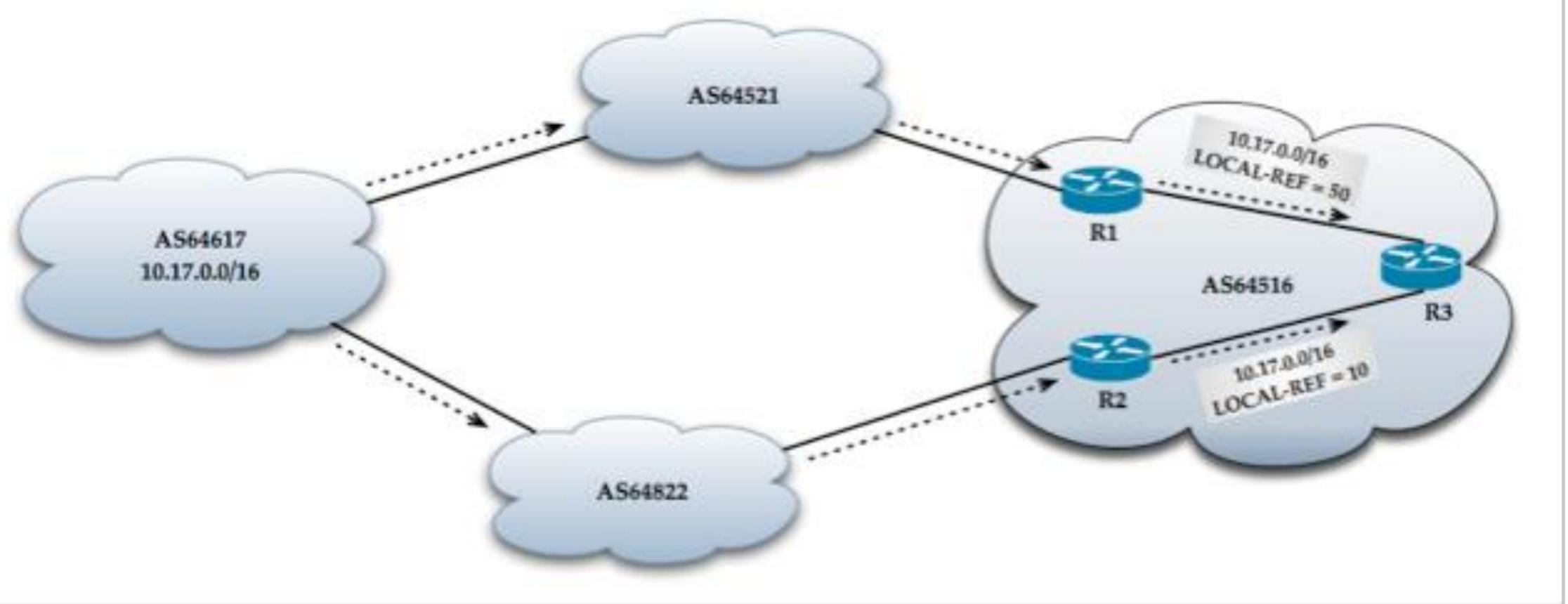
BGP Atributy

- Kritéria pro rozhodnutí o nejlepší cestě
 - well-known mandatory vs. discretionary
 - optional transitive vs. nontransitive



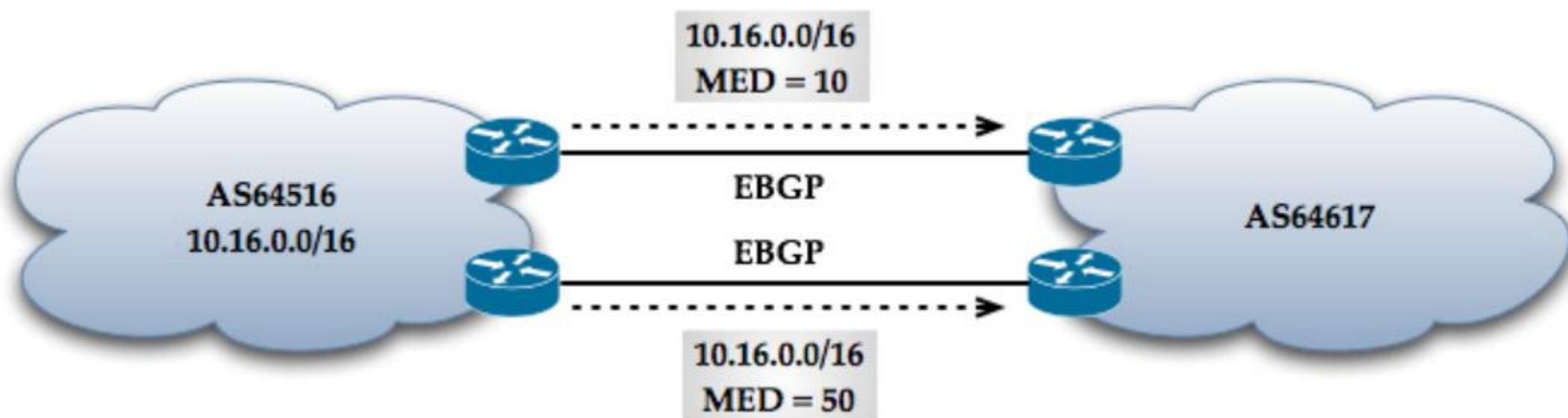
Local Preference

- Uvnitř AS specifikujeme preferované „výstupy“
- Větší je preferovanější



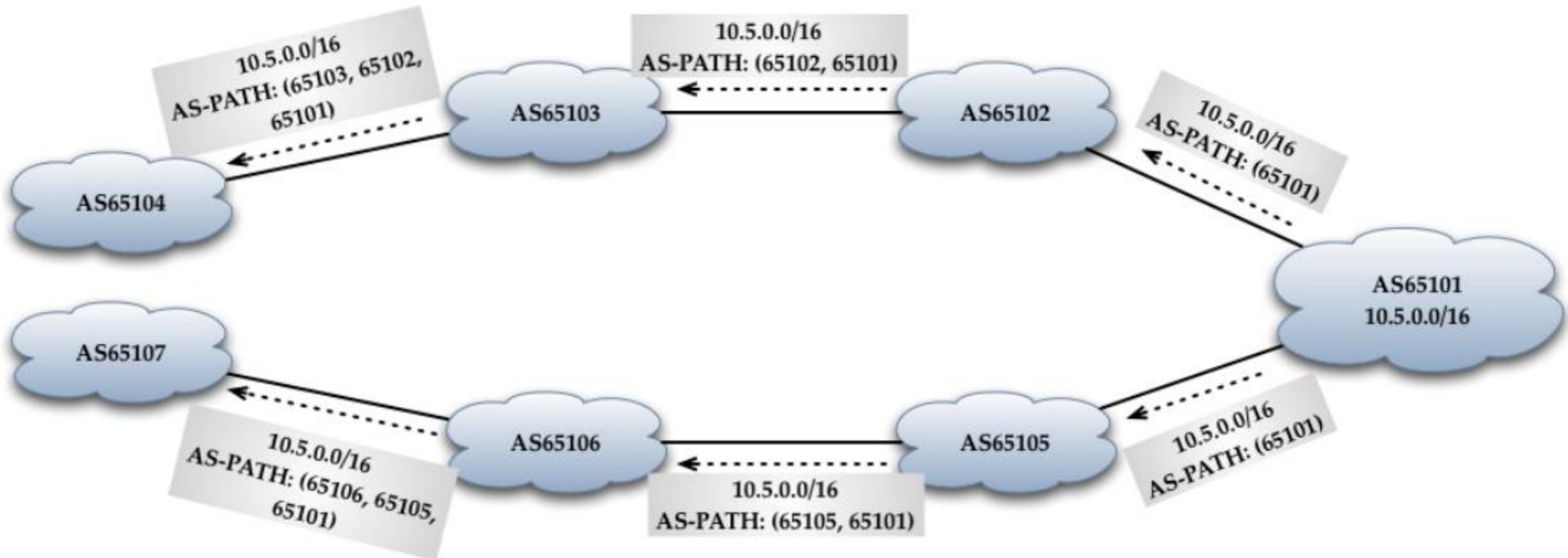
MED

- Multi Exit Discriminator
- Vně AS specifikujeme preferované „vstupy“
- Menší je preferovanější

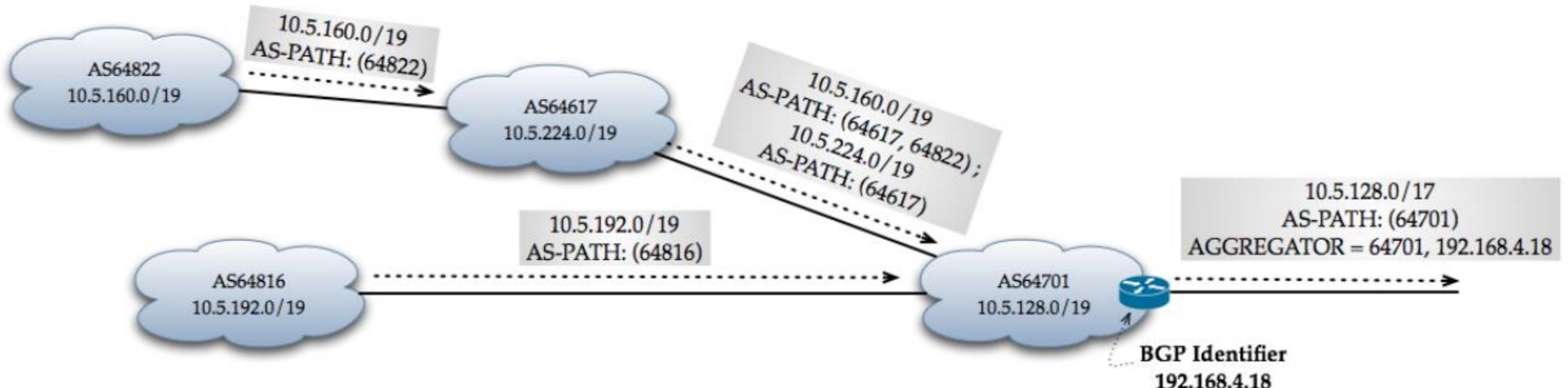


AS-PATH

- Kratší je preferovanější



Route Aggregation



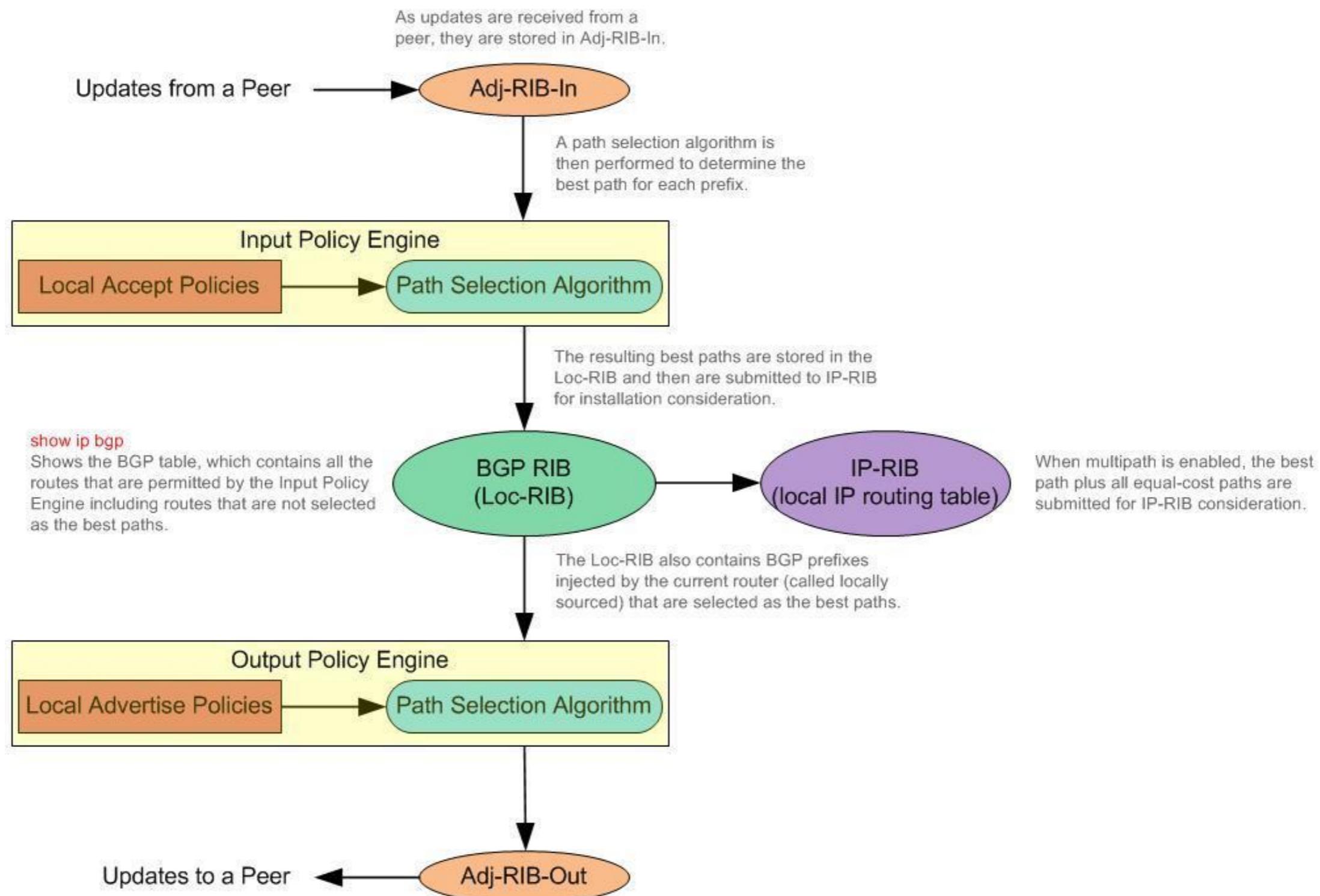
Politiky (1)

- Pravidla, která se použijí při zpracování BGP informace na směrovači

TABLE 8.1 Examples of import and export policies at a BGP speaker.

Import Policy	Export Policy
<ul style="list-style-type: none">– Do not accept default 0.0.0.0/0 from AS64617.– Assign 192.168.1.0/24 coming from AS64617 preference to receiving it from AS64816.– Accept all other IP prefixes.	<ul style="list-style-type: none">– Do not propagate default route 0.0.0.0/0 except to internal peers.– Do not advertise 192.168.1.0/24 to AS64999.– Assign 172.22.8.0/24 a MED metric of 10 when sent to AS64999.

Politiky (2)



BGP Message

- Cloud Shark Links

BGP Update

<https://www.cloudshark.org/captures/0224f4ab8f63>

Forming BGP Adjacency

<https://www.cloudshark.org/captures/00249be4441f>

Obsah

- 1) Úvod do směrování
- 2) Směrování paketů
 - Směrovací tabulky
 - Algoritmus výběru
 - Architektura směrovače
- 3) Směrování podle nejkratší cesty
 - Algoritmus Dijkstra
 - Algoritmus Bellman-Ford
 - Path-Vector směrování
- 4) IGP směrovací protokoly
 - RIP
 - OSPF
- 5) EGP směrování
 - BGP
- 6) Shrnutí

Shrnutí

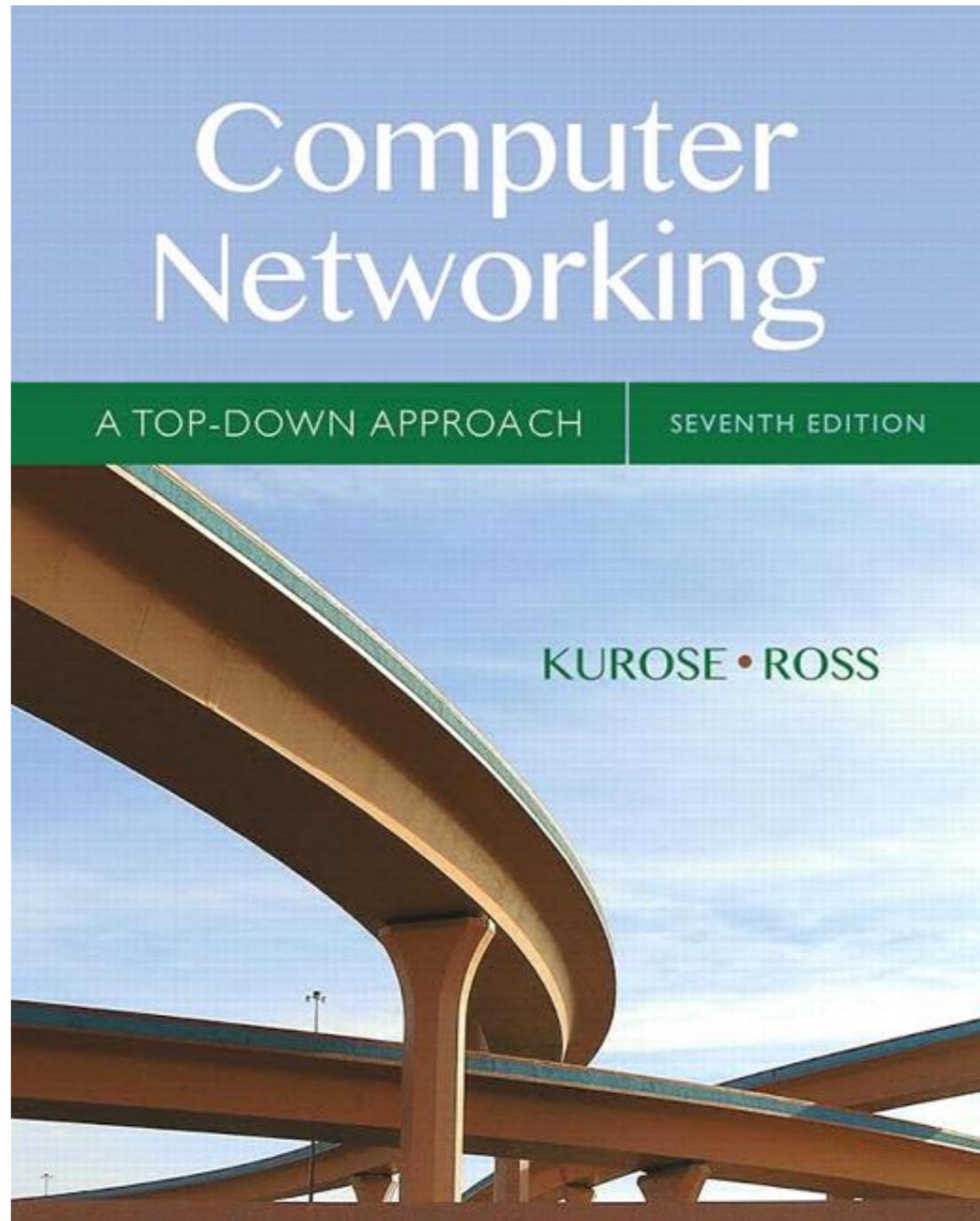
- Internet je složen z autonomních systémů
 - každý systém má své unikátní ASN
- Směrování je na dvou úrovních:
 - EGP - mezi autonomními systémy
 - IGP - uvnitř autonomních systémů
- BGP je protokol pro EGP
 - Path-vector protokol
 - jiné požadavky a tudíž atributy než IGP
 - implementuje směrovací politiky

Literatura

- Kurose J.F., Ross K.W.: Computer Networking, A Top-Down Approach Featuring the Internet. Addison-Wesley, 2003.
- Rita Pužmanová: Routing and Switching: Time of convergence? Addison-Wesley, 2002.
- A.Zinin: Cisco IP Routing: Packet Forwarding and Intra-domain Routing Protocols. 2001.
- Příslušná RFC...

Domácí úkol

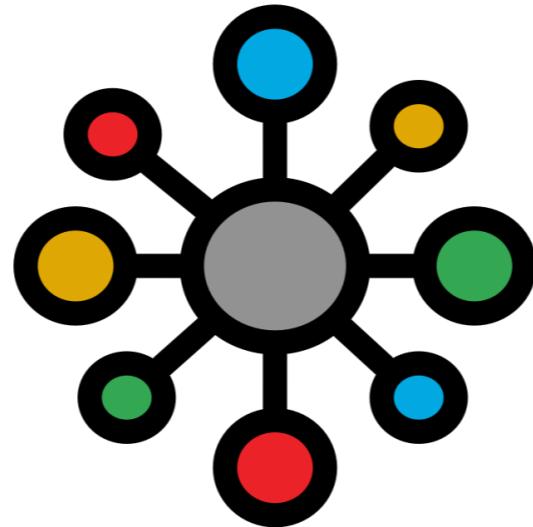
- Kapitola o směrování, sekce problémy a otázky.



Kontakty

Ondřej Ryšavý rysavy@fit.vutbr.cz

Vladimír Veselý veselyv@fit.vutbr.cz



<https://nesfit.github.io>