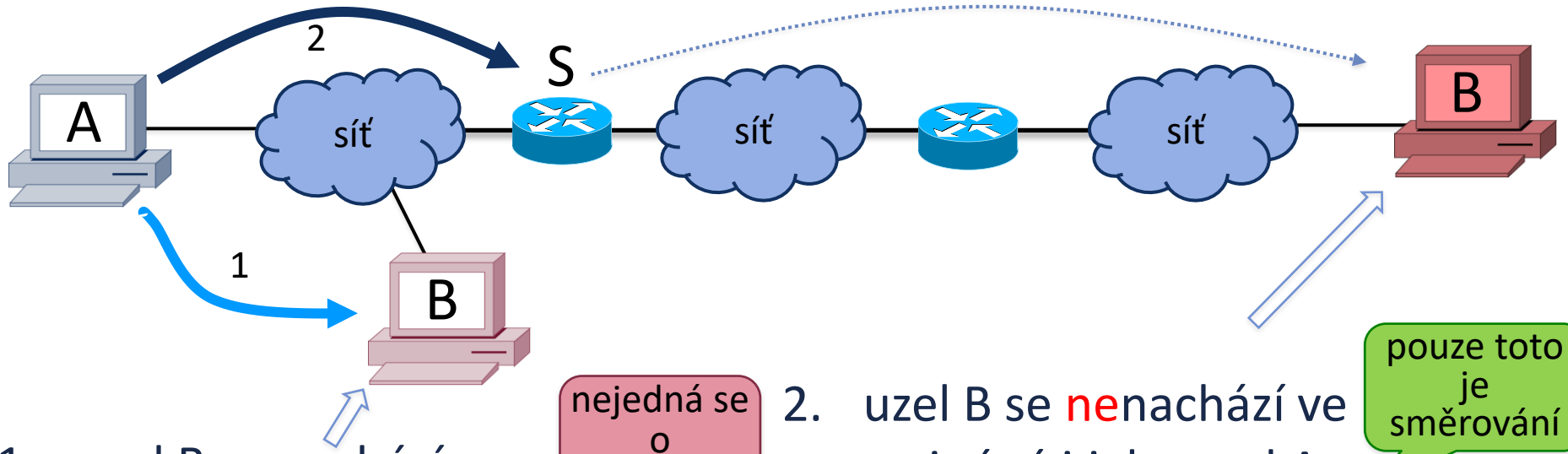
Rodina protokolů TCP/IP
verze 3

Téma 6: Směrování v IP sítích

Jiří Peterka

přímé a nepřímé doručování

- když uzel A odesílá IP datagram uzlu B, mohou nastat 2 různé případy:



1. uzel B se nachází ve stejné síti jako uzel A

nejedná se o směrování

- pak se jedná o přímé doručování (direct delivery)
  - uzel A předá paket své vrstvě síťového rozhraní, k doručení přímo uzlu B
    - ke směrování (volbě směru, na síťové vrstvě) zde nedochází !!!

2. uzel B se nenachází ve stejné síti jako uzel A

pouze toto je směrování

- pak se jedná o nepřímé doručování (indirect delivery)
  - uzel A musí najít „next hop“
    - zjistit adresu směrovače S na cestě k uzlu B
      - dochází zde ke směrování, jako volbě směru dalšího přenosu
  - uzel A předá paket své vrstvě síťového rozhraní, k doručení uzlu S

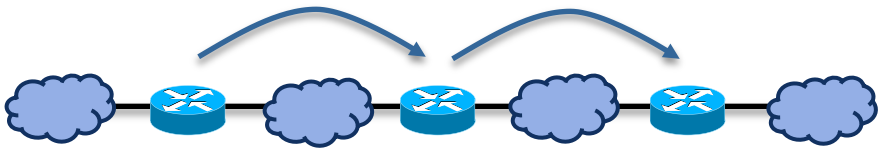
co je směrování (routing)?

- v užším slova smyslu:
  - volba směru pro další předání paketu/datagramu do jiné sítě
    - algoritmy směrování
- ve skutečnosti zahrnuje:
  - výpočet optimální cesty
    - je to kombinatorický problém hledání nejkratší cesty v grafu
    - výsledkem jsou "podklady pro volbu směru"
  - uchovávání směrovacích informací (podkladů pro rozhodování)
    - vedení směrovacích tabulek
  - předávání paketů (forwarding)
    - využívání výsledků výpočtů ("podkladů")
  - udržování směrovacích informací
    - aktualizace údajů pro výpočty cest, reakce na změny

- v širším slova smyslu:
  - koncepce IP adres
  - celková koncepce směrování
    - které uzly se účastní směrování
      - a do jaké míry
  - celková koncepce internetu/ Internetu
    - katenetový model
    - historický vývoj
  - metody optimalizace směrovacích tabulek
  - řešení směrování v opravdu velkých systémech
    - hierarchické směrování
    - autonomní systémy
  - směrovací politiky
  - směrovací protokoly
    - RIP, OSPF, BGP, EIGRP, ……..

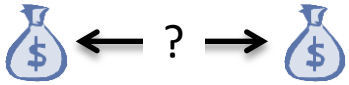
principy směrování (v IPv4)

- princip katenetu
  - „svět“ je tvořen soustavou sítí, vzájemně propojených pomocí směrovačů
    - mezi každými dvěma sítěmi vždy existuje souvislá cesta
      - při směrování se hledá posloupnost přeskoků přes směrovače
- hop-by-hop routing
  - směruje se „per hop“: v každém směrovači se rozhoduje znovu
    - znovu se hledá „další hop“
      - nezávisle na ostatních IP datagramech



- destination-based routing
  - směruje se (jen) na základě cílové adresy
    - zdrojová adresa při směrování nehraje roli
  - směruje se na základě příslušnosti k síti
    - na základě síťové části cílové IP adresy
      - a pouze v cílové síti se bere v úvahu také relativní část IP adresy
- least-cost routing
  - optimální cesta se volí podle nejnižší „ceny“
    - ve smyslu používané metriky

  - není podporováno více cest se stejnou cenou
    - možnost jejich současného využití



celková koncepce směrování (v IPv4)

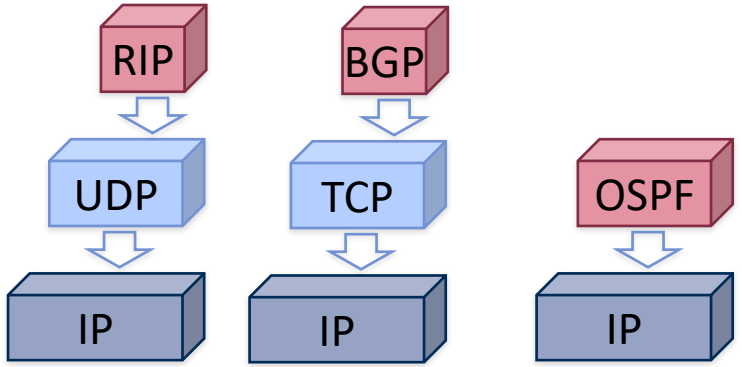
- směrování je bezestavové
  - rozhodování o dalším směru (hop-u) je nezávislé na historii a předchozích datagramech
- možné alternativy:
  - koncept toků (flows)
    - objevuje se v IPv6
    - jednotlivé pakety/datagramy nějak „patří k sobě“ a podle toho jsou směrovány
  - „tag switching“
    - obdoba toků (flows)
      - s přechodem z L3 na L2

- směrování je nezávislé na obsahu a zdroji
  - algoritmy směrování se neptají na to, co jednotlivé pakety obsahují a odkud pocházejí
    - souvisí s principem best effort a absencí QoS
- možné alternativy:
  - content switching
    - snaha rozhodovat se při směrování také podle obsahu/charakteru dat
      - alespoň podle čísel portů
  - source-based routing
    - algoritmy směrování se rozhodují i podle toho, odkud data pochází
  - policy-based routing
    - ještě obecnější koncept: směrování bere v úvahu celou řadu faktorů
      - včetně komerčních zájmů provozovatele sítě

jak je řešeno směrování?

- představa:
  - ke směrování (nepřímému doručování) dochází na síťové vrstvě (L3)
    - směrovače fungují na síťové vrstvě (L3), nemají vyšší vrstvy
- přitom:
  - rozhodování o volbě dalšího směru vychází z informací dostupných na síťové vrstvě (L3)
    - konkrétně z IP adres
  - samotná manipulace s pakety či datagramy probíhá na síťové vrstvě
- ale:
  - většina souvisejících činností se odehrává na vyšších vrstvách
    - zejména: výměna a aktualizace směrovacích informací, hledání cest

- například:
  - protokoly RIP a BGP jsou aplikačními protokoly
    - fungují na aplikační vrstvě (L7)
      - RIP využívá služby transportního protokolu UDP (port 520)
      - BGP využívá služby transportního protokolu TCP (port 179)



  - protokol OSPF vkládá svá data přímo do IP datagramů
    - IP Protocol Type 0x59 = OSPF
    - patřil by proto na transportní vrstvu
      - ale je také spíše aplikační - má vlastní (zabudovanou) transportní vrstvu

směrovací tabulky

- jsou datovou strukturou, ve které jsou uchovávány (některé) podklady pro směrování
  - pro „logické činnosti“: hledání nejkratších cest a výměnu směrovacích informací
  - pracují s nimi protokoly jako RIP a OSPF
- položky obsahují:
  - cílovou síť s maskou
    - nebo s prefixem
  - „next hop IP“
  - odchozí rozhraní
  - ohodnocení (v metrice)

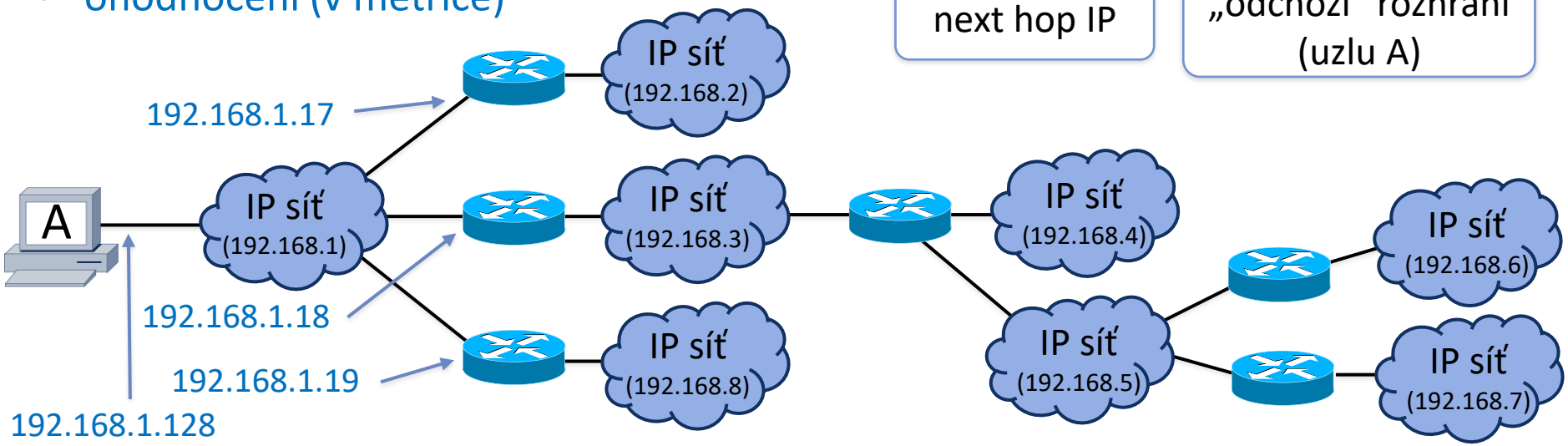
směrovací tabulka uzlu A

Správce: Příkazový řádek

```
IPv4 Směrovací tabulka
===========================================================================
Aktivní směrování:
          Cíl v síti   Síťová maska          Brána       Rozhraní Metrika
      192.168.2.0    255.255.255.0    192.168.1.17    192.168.1.128     11
      192.168.3.0    255.255.255.0    192.168.1.18    192.168.1.128     11
      192.168.4.0    255.255.255.0    192.168.1.18    192.168.1.128     20
      192.168.5.0    255.255.255.0    192.168.1.18    192.168.1.128     20
      192.168.6.0    255.255.255.0    192.168.1.18    192.168.1.128     30
      192.168.7.0    255.255.255.0    192.168.1.18    192.168.1.128     30
      192.168.8.0    255.255.255.0    192.168.1.19    192.168.1.128     11
===========================================================================
```

next hop IP

„odchozí“ rozhraní (uzlu A)



adaptivní vs. neadaptivní směrování

- neadaptivní (též: statické) směrování
  - nesnaží se reagovat na změny v soustavě vzájemně propojených sítí
    - nepotřebuje aktualizace ani spolupráci s ostatními uzly
      - tedy ani protokoly jako RIP, OSPF, BGP atd.
  - obsah směrovacích tabulek je dopředu a pevně dán
    - je statický, nemění se v čase
  - výhody:
    - vyhovuje to zvýšeným požadavkům na bezpečnost
      - nelze napadnout skrze šíření aktualizací
    - není režie na aktualizaci
    - lze vyhovět i speciálním požadavkům na směrování
  - nevýhoda:
    - nereaguje na změny, pokud k nim dojde
- adaptivní (dynamické) směrování
  - snaží se reagovat na změny
    - vyžaduje aktualizaci informací o stavu celé soustavy sítí
    - vyžaduje průběžné hledání nejkratších cest
  - potřebuje protokoly jako RIP, OSPF, BGP, …..
    - které dynamicky aktualizují obsah směrovacích tabulek
      - základ jejich obsahu může být dán staticky/dopředu
  - nevýhoda:
    - vysoká režie zejména na aktualizaci informací
      - s velikostí soustavy sítí rychle roste
        - otázka škálování

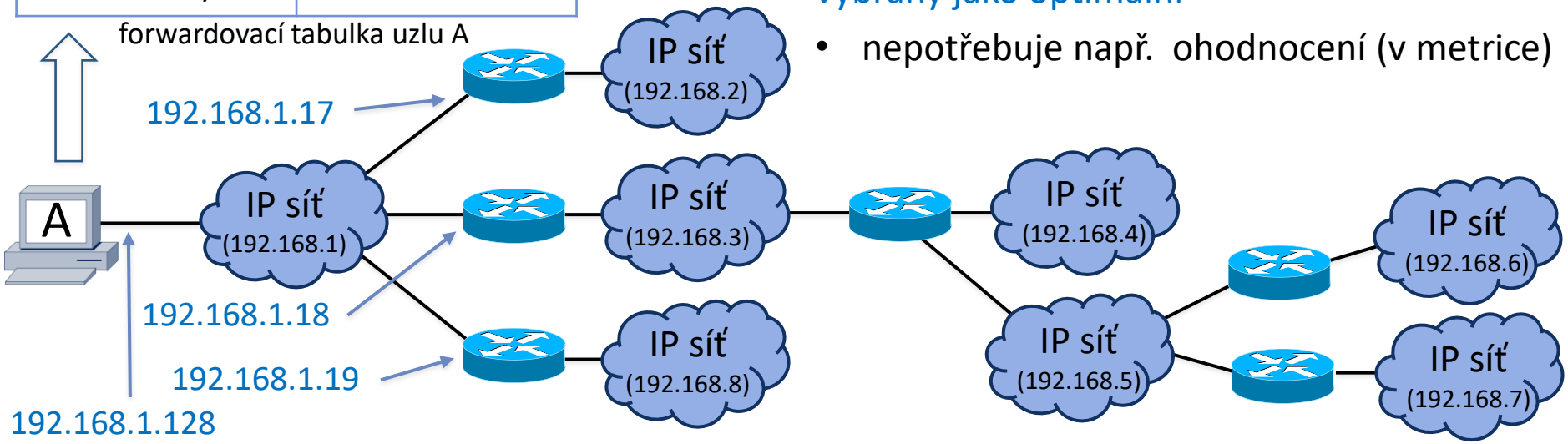
forwardovací tabulky

- směrovací tabulky jsou „informačně bohaté“
  - hodí se pro rozhodování, ale nikoli pro samotnou manipulaci s IP datagramy
    - která musí být co možná nejrychlejší – a u které se již „nepřemýšlí“

cílová síť	next hop IP
192.168.2/24	192.168.1.17
192.168.3/24	192.168.1.18
192.168.4/24	192.168.1.18
192.168.5/24	192.168.1.18
192.168.6/24	192.168.1.18
192.168.7/24	192.168.1.18
192.168.8/24	192.168.1.19

forwardovací tabulka uzlu A

- pro samotnou manipulaci s datagramy se používají menší a rychlejší tabulky
  - tzv. forwardovací tabulky (forwarding tables)
- představa:
  - forwardovací tabulka je „výcucem“ ze směrovací tabulky
    - obsahuje pouze ty cesty, které již byly vybrány jako optimální
    - nepotřebuje např. ohodnocení (v metrice)



agregace položek

- počet položek v tabulkách lze snižovat agregací CIDR bloků
  - jde o supernetting

cílová síť	next hop IP
192.168.2/24	192.168.1.17
192.168.3/24	192.168.1.18
192.168.4/24	192.168.1.18
192.168.5/24	192.168.1.18
192.168.6/24	192.168.1.18
192.168.7/24	192.168.1.18
192.168.8/24	192.168.1.19

tabulka uzlu A

192.168.00000100 (192.168.4)
192.168.00000101 (192.168.5)
192.168.00000110 (192.168.6)
192.168.00000111 (192.168.7)

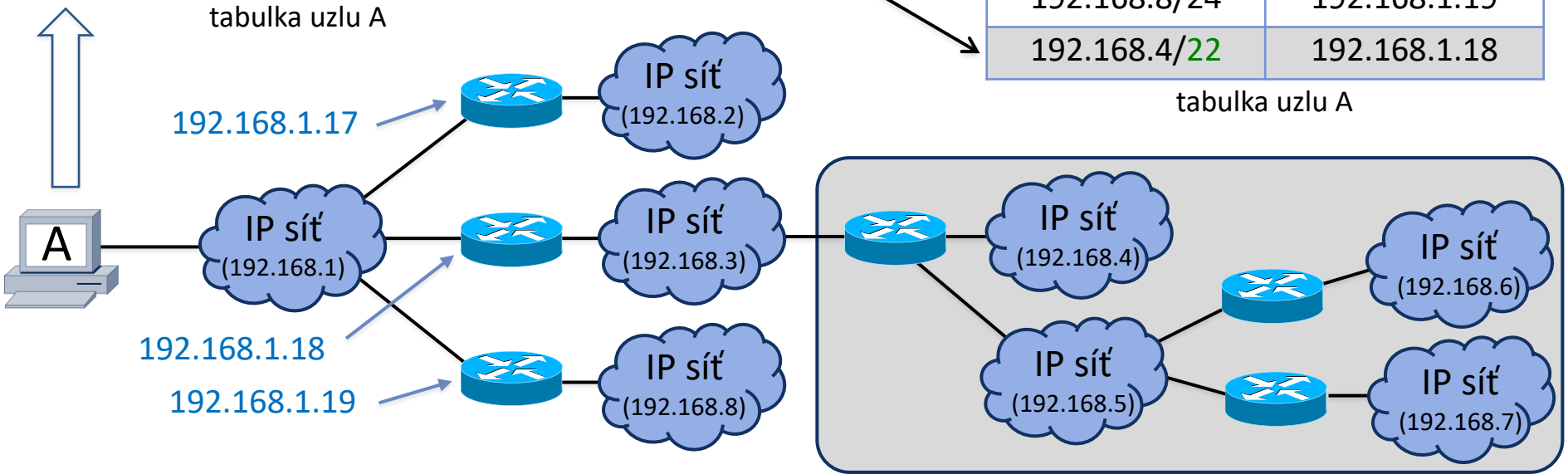
192.168.000001xx (192.168.4)

22 bitů

24 bitů

cílová síť	next hop IP
192.168.2/24	192.168.1.17
192.168.3/24	192.168.1.18
192.168.8/24	192.168.1.19
192.168.4/22	192.168.1.18

tabulka uzlu A



implicitní cesta (default route)

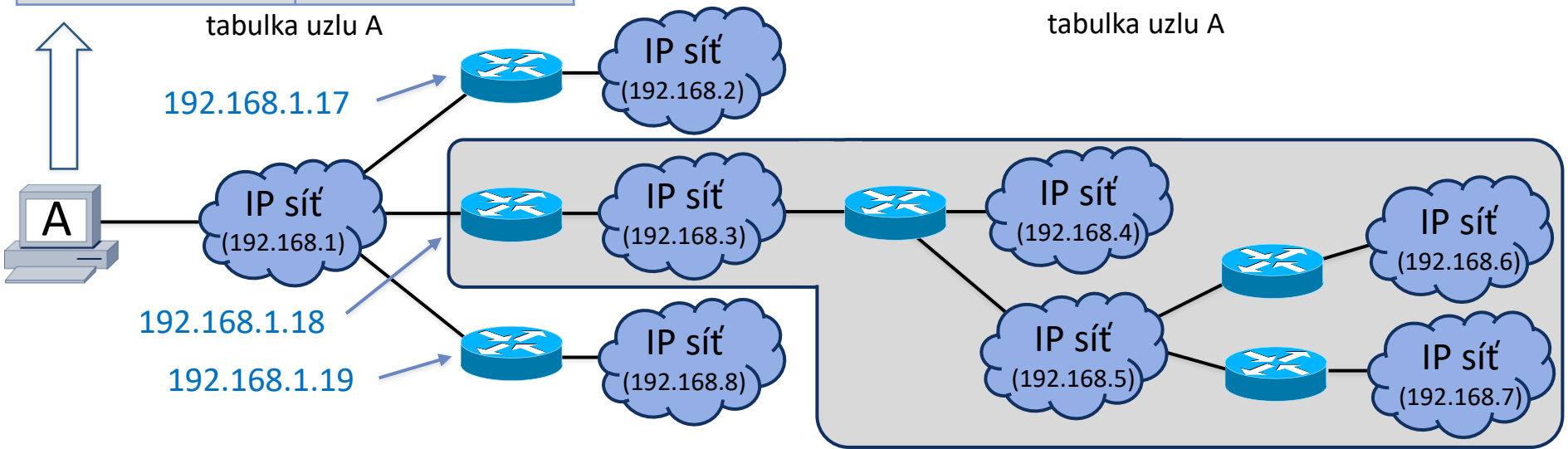
- další možností, jak redukovat počet položek v tabulce, je zavedení tzv. implicitní cesty (default route)
  - tj. explicitně se vyjmenuje jen to, co „jde jinudy“
  - a všechno ostatní se směruje „implicitně“, přes default route
    - ta má prefix 0

cílová síť	next hop IP
192.168.2/24	192.168.1.17
192.168.3/24	192.168.1.18
192.168.8/24	192.168.1.19
192.168.4/22	192.168.1.18

tabulka uzlu A

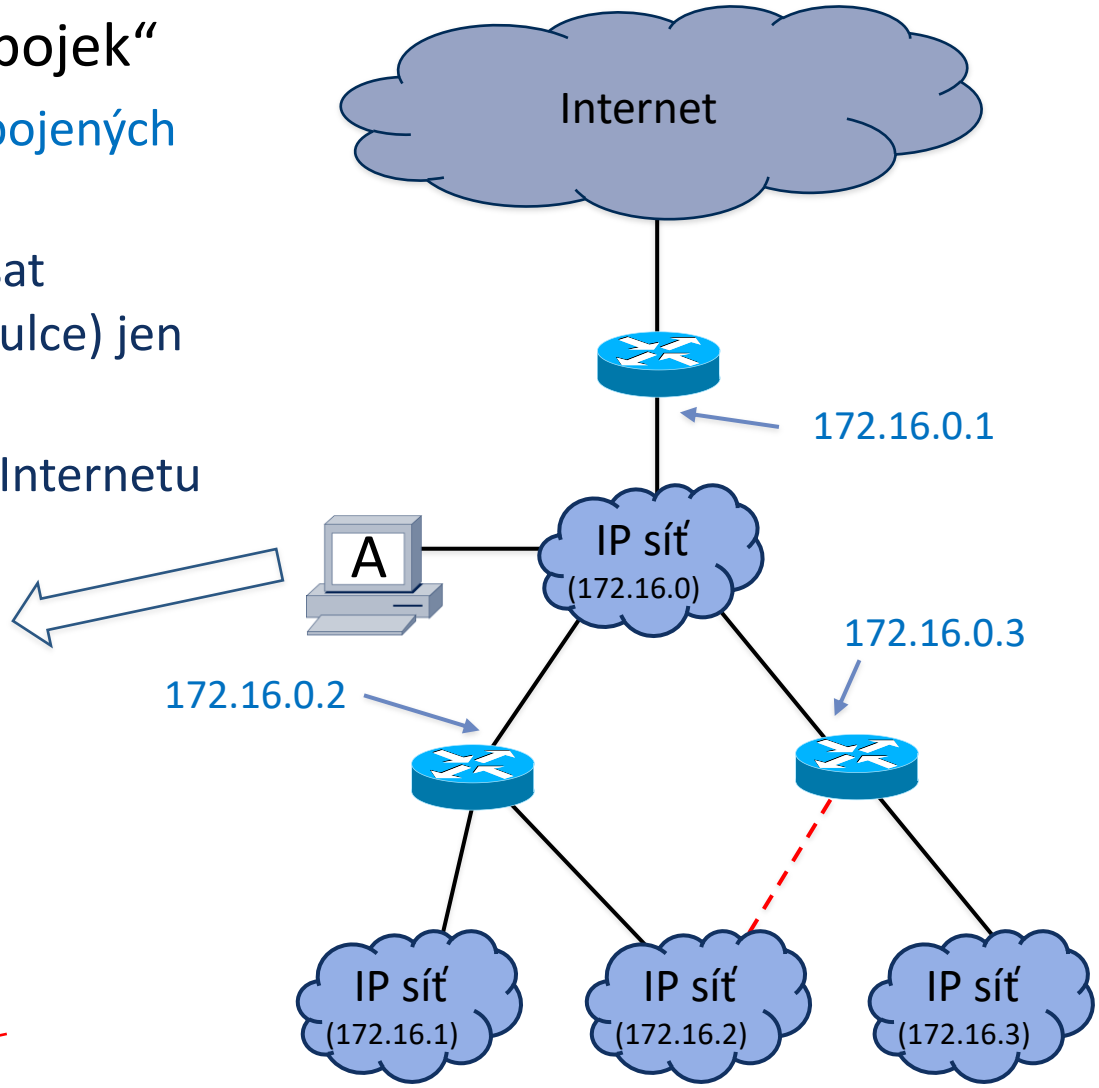
cílová síť	next hop IP
192.168.2/24	192.168.1.17
192.168.8/24	192.168.1.19
0/0	192.168.1.18

tabulka uzlu A



implicitní cesta (default route)

- implicitní cesta přináší největší efekt „na konci internetových přípojek“
  - u soustav (koncových) sítí, připojených k Internetu
    - stačí „znát“ (explicitně popsat položkami ve směrovací tabulce) jen „nižší“ sítě
    - vše ostatní lze směrovat do Internetu přes implicitní cestu



cílová síť	next hop IP
172.16.1/24	172.16.0.2
172.16.2/20	172.16.0.2
172.16.3/16	172.16.0.3
0/0	172.16.0.1

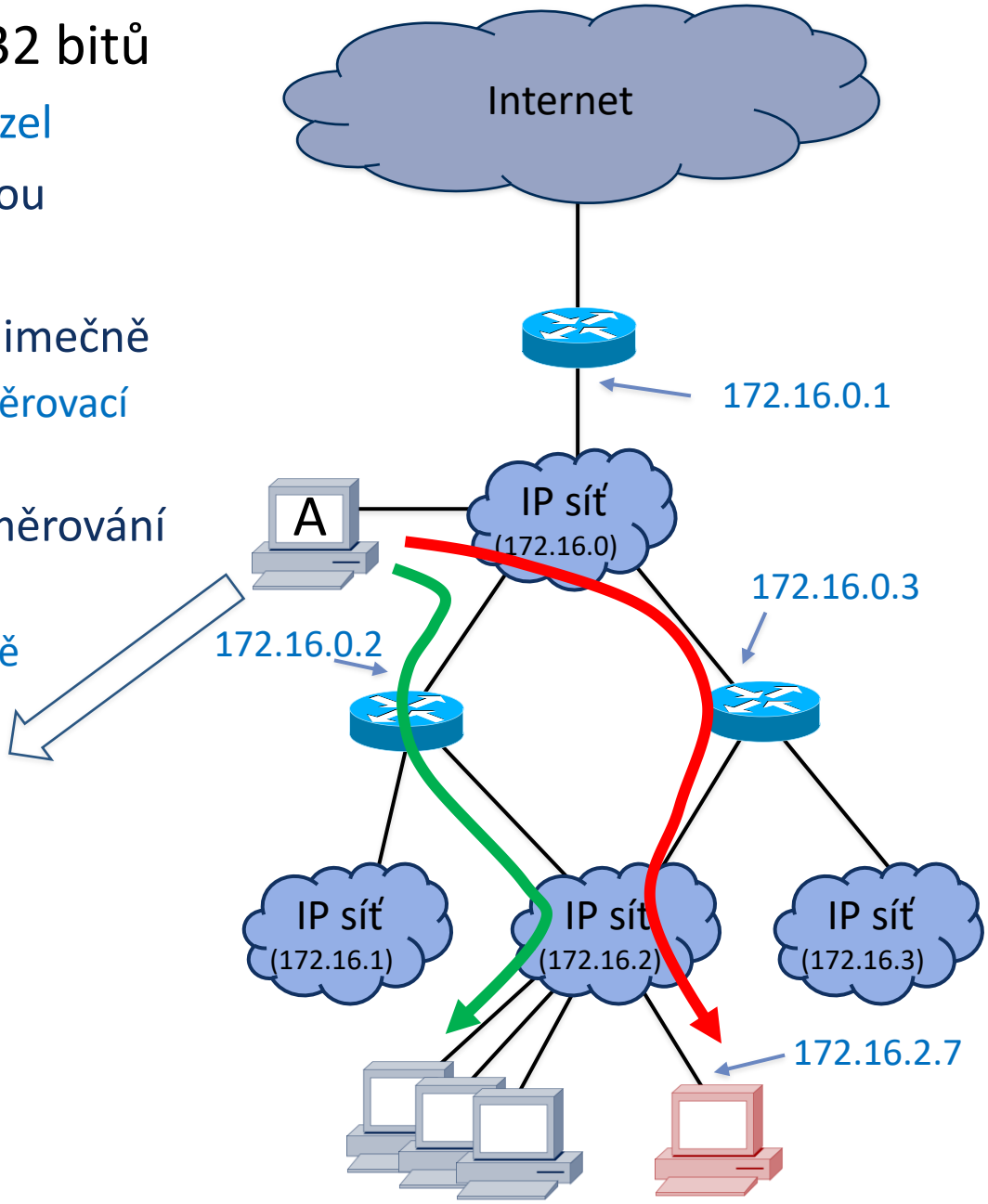
směrovací tabulka uzlu A

172.16.2/24	172.16.0.3

další možnou cestu nelze využít

host specific route

- cílová síť může mít prefix až 32 bitů
  - pak je cílovou sítí jeden jediný uzel
    - s onou 32-bitovou IPv4 adresou
  - jde o tzv. host-specific route
    - měla by být používána jen výjimečně
      - protože enormně zvětšuje směrovací tabulky
    - jde o výjimku z pravidla, že směrování vychází z příslušnosti do sítě
      - tj. že ke všem uzlům stejné sítě vede stejná cesta



cílová síť	next hop IP
172.16.2.7/32	**172.16.0.3**
172.16.1/24	172.16.0.2
172.16.2/20	**172.16.0.2**
172.16.3/16	172.16.0.3
0/0	172.16.0.1

tabulka uzlu A

algoritmus směrování IP protokolu

1. nejprve je třeba zjistit, zda jde o přímé či nepřímé doručování
   - v případě přímého doručování algoritmus končí
     - datagram je předán vrstvě síťového rozhraní (k doručení přímo cílovému uzlu)
2. jde o nepřímé doručování (směrování)
   - je nutné projít (forwardovací) tabulku a nalézt takovou položku, která určí další směr přenosu
   - na pořadí záleží: tabulka se prochází od větších prefixů směrem k nižším

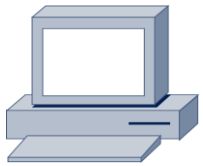
1. nejprve se hledají host-specific route
   - pokud se najde, je datagram odeslán a algoritmus končí
2. pak se prohledají „běžné“ směry do koncových sítí
   - pokud se najde vhodný směr, datagram je odeslán a algoritmus končí
3. použije se implicitní cesta (default route)
   - pokud implicitní cesta není definována, datagram je zahozen, je odeslána ICMP zpráva (Destination Unreachable) a algoritmus končí

cílová síť	next hop IP
172.16.2.7/32	172.16.0.3
172.16.1/24	172.16.0.2
172.16.2/24	172.16.0.2
172.16.3/24	172.16.0.3
0/0	172.16.0.1

hostitelské počítače a směrování

- role směrovačů a hostitelských počítačů v rámci směrování se liší
- směrovače
  - zajišťují všechny činnosti, spojené se směrováním
    - včetně hledání optimálních cest, aktualizací směrovacích informací atd.

- hostitelské počítače
  - neúčastní se hledání optimálních cest ani aktualizací směrovacích informací
  - pouze se *„chovají tak, jak jim někdo jiný řekne“*
    - mají směrovací (forwardovací) tabulky
      - a používají je při odesílání datagramů
    - ale samy si je neaktualizují
      - nepotřebují a nemají „vyšší vrstvy“, neimplementují protokoly RIP, OSPF, BGP ……
    - obsah směrovacích tabulek může být statický i dynamický
      - pokud je dynamický, o aktualizaci se starají směrovače
        - princip: pokud host odesílá svá data po špatné (neoptimální) cestě, nejbližší směrovač ho na to upozorní
          - a řekne, kudy vede „lepší“ cesta
          - hostitelský počítač by si podle toho měl aktualizovat svou tabulku



cílová síť	next hop IP
172.16.2.7/32	172.16.0.3
172.16.1/24	172.16.0.2
172.16.2/24	172.16.0.2
172.16.3/24	172.16.0.3
0/0	172.16.0.1

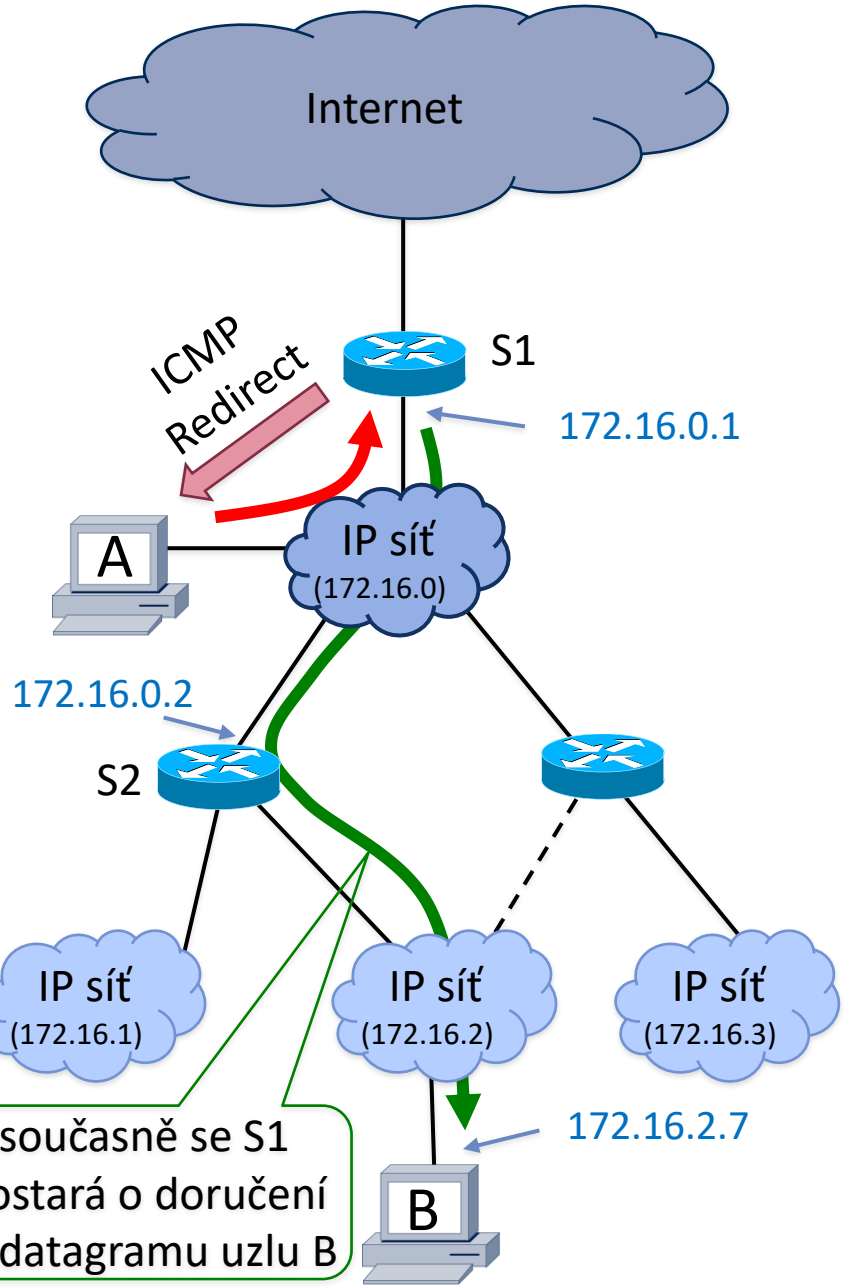
zpráva ICMP Redirect

- zpočátku:

cílová síť	next hop IP
0/0	172.16.0.1

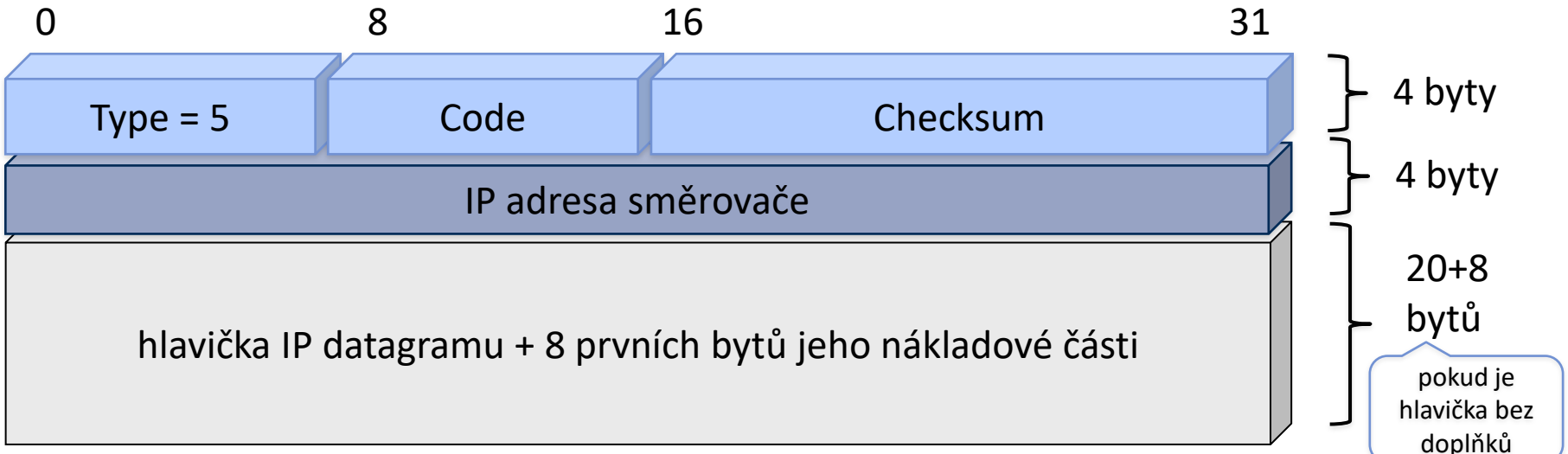
  - uzel A „zná“ pouze směrovač S1 s IP adresou 172.16.0.1
    - jako „implicitní směrovač“, přes který vede implicitní cesta (default route)
- když
  - uzel A potřebuje něco odeslat uzlu B
    - pošle to přes implicitní směrovač
  - směrovač S1 zjistí, že jde o nesprávný (neoptimální) směr a pošle uzlu A zprávu ICMP Redirect, ve smyslu:
    - *cesta k síti 172.16.2/24 vede přes směrovač S2, s adresou 172.16.0.2*
  - uzel A by si měl aktualizovat svou tabulku

cílová síť	next hop IP
172.16.2/24	172.16.0.2
0/0	172.16.0.1



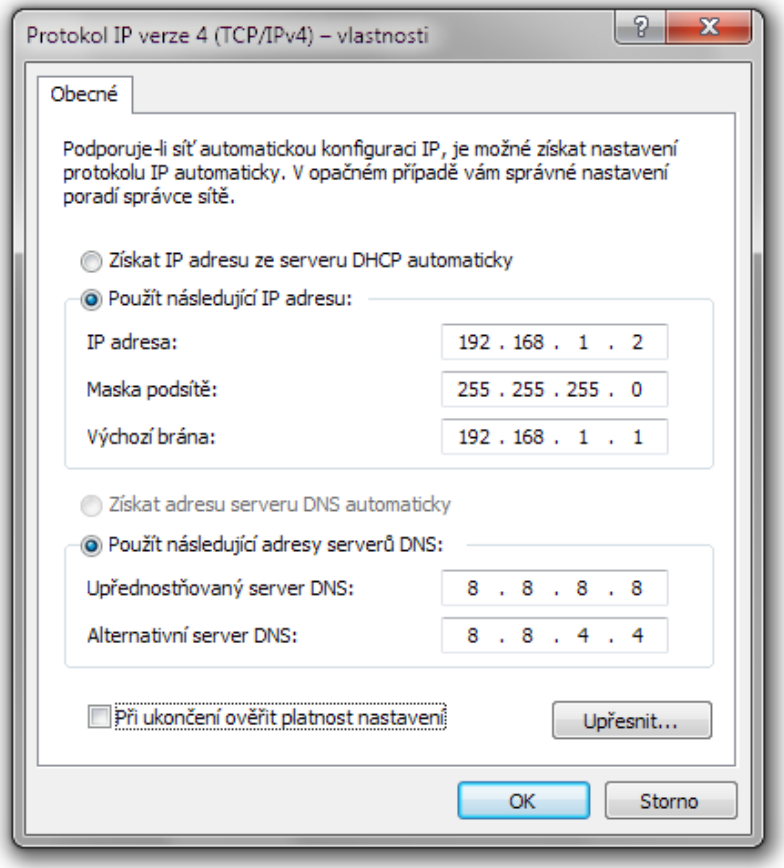
zpráva ICMP Redirect

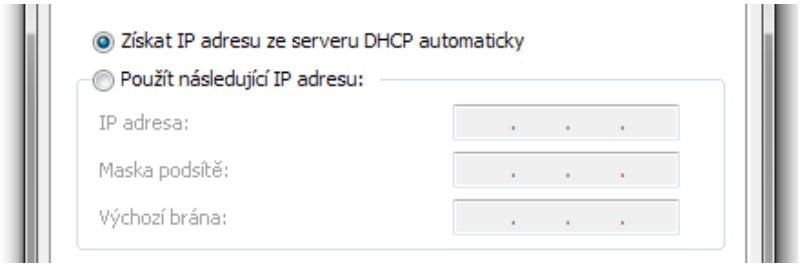
- odesílá ji směrovač
  - není považována za chybovou zprávu, ale za informační zprávu
  - kromě odeslání zprávy Redirect má směrovač povinnost postarat se o správné doručení dat, která vše „způsobila“
- příjemcem je hostitelský počítač
  - měl by na zprávu reagovat tím, že se „poučí“ (že si vhodnější cestu zanese do své směrovací tabulky)
    - ale také nemusí – například z bezpečnostních důvodů, nebo kvůli omezené velikosti své směrovací tabulky
- položka CODE blíže určuje, čeho se zpráva ICMP Redirect týká:
  - 0: jde o směrování IP datagramů do celé cílové sítě
  - 1: jde o směrování IP datagramů ke konkrétnímu cílovému uzlu (host-specific)

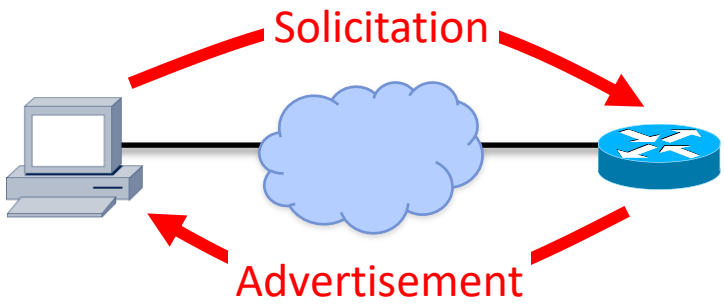


implicitní směrovače

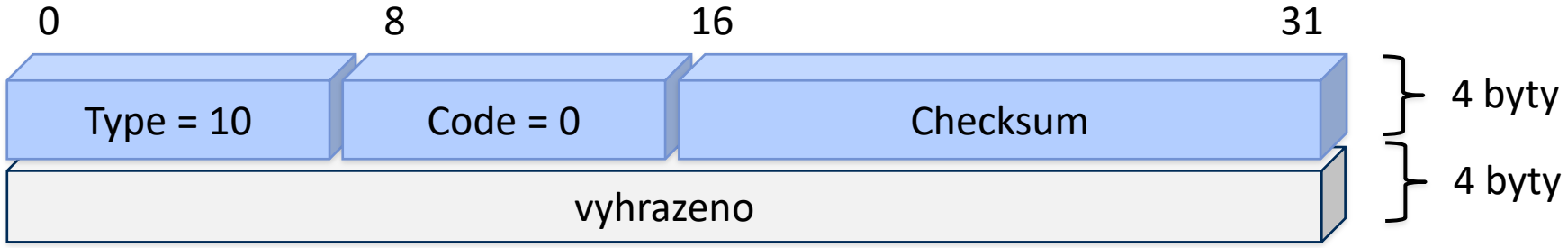
- skrze zprávy ICMP Redirect se uzel dozvídá o „dalších“ směrovačích
- ale:
  - *jak se dozví o „prvním“ směrovači?*
    - pozor: obvykle nesprávně překládáno jako „výchozí brána“
- možnosti:
  - má jej pevně nastaven ve své konfiguraci
  - od DHCP serveru
  - může si jej sám (proaktivně) zjistit
    - pomocí ICMP zpráv Router Solicitation
  - z „inzerátů“ jednotlivých směrovačů
    - pomocí ICMP zpráv Router Advertisement







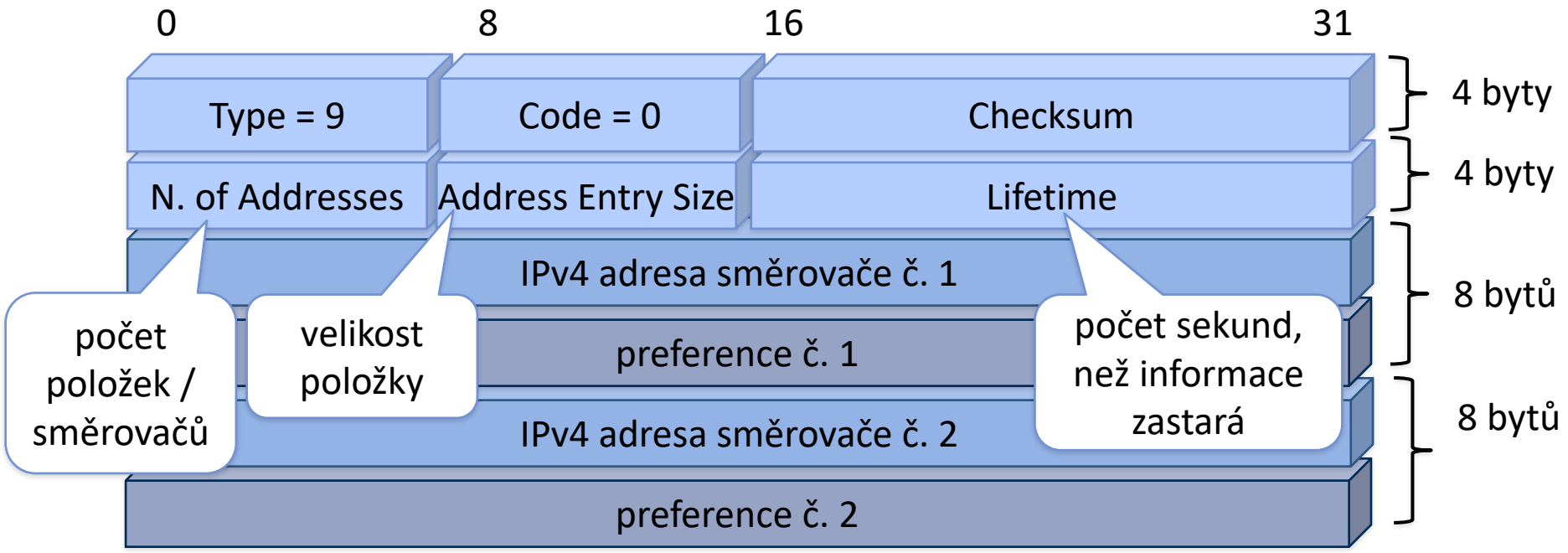
ICMP zpráva Router Solicitation



- ICMP Router Solicitation:
  - je „výzvou“, kterou vysílá hostitelský počítač, ve smyslu:
    - „je zde (v síti) nějaký směrovač? Ozvěte se …..“
      - pokud takový směrovač existuje, odpoví zprávou ICMP Advertisement
        - pokud je takových směrovačů více, odpoví všechny
- vysílá se:
  - když je k dispozici multicast: na adresu 224.0.0.2 (všechny směrovače v dané síti)
  - pokud není k dispozici multicast: na „místní“ broadcast adresu (255.255.255.255)
- nejde o:
  - mechanismus určený ke vzájemné komunikaci směrovačů
    - ale jen o komunikaci mezi hostitelskými počítači a směrovači
    - v praxi se příliš nevyužívá

ICMP zpráva Router Advertisement

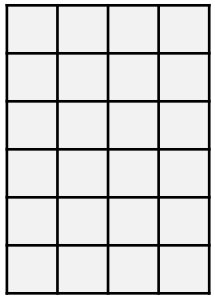
- tuto zprávu vysílá směrovač
  - v odpovědi na ICMP zprávu Router Solicitation, nebo
  - z vlastní iniciativy (jednou „za delší čas“)
    - směrovač jakoby sám upozorňuje na svou existenci
- vysílá se:
  - když je k dispozic multicast
    - na adresu 224.0.0.1
      - „všechny uzly v dané síti“
  - když není k dispozici multicast
    - pomocí broadcastu
- součástí zprávy může být informace i o dalších směrovačích



připomenutí: jak je řešeno směrování?

- k „logickým“ činnostem dochází na vyšších vrstvách
  - k hledání nejkratších cest
    - aplikují se algoritmy pro hledání nejkratších cest
      - např. Bellman-Ford, Ford-Fulkerson, …..
    - výsledkem jsou naplněné směrovací tabulky
      - podle kterých jsou předávány „podklady“ na L3
  - k aktualizaci směrovacích informací (u dynamického směrování)
    - používají se konkrétní „směrovací algoritmy
      - ve skutečnosti spíše algoritmy pro aktualizaci směrovacích informací
        - např. na principu distance-vector či link-state
- na síťové vrstvě (L3) dochází pouze k samotné manipulaci k pakety či datagramy
  - na základě předem připravených „podkladů“
    - obsažených v tzv. forwardovacích tabulkách
      - zjednodušená verze směrovacích tabulek, optimalizovaná pro rychlost

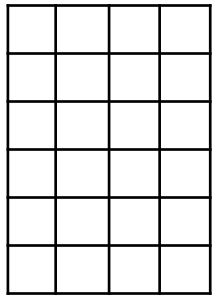
řeší protokoly jako RIP, OSPF, BGP, ….



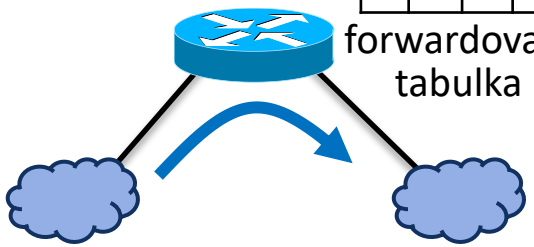
směrovací tabulka

„podklady“

řeší protokol IP



forwardovací tabulka

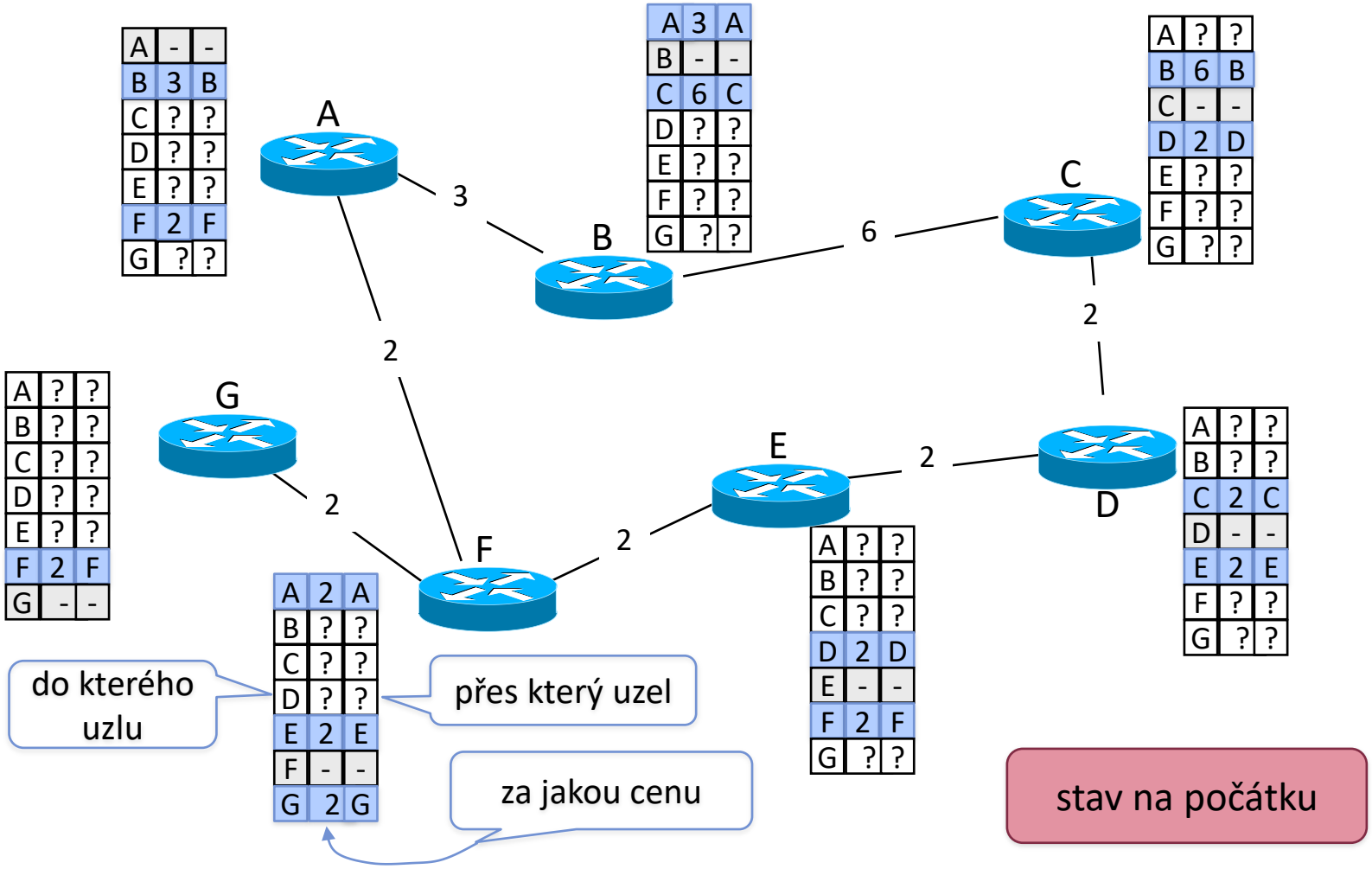


klasifikace (algoritmů) směrování

- izolované směrování
  - směrovače vzájemně nespolupracují, fungují nezávisle na sobě (proto: izolované)
  - například:
    - záplavové směrování: rozesílá se do všech směrů (kromě příchozího)
    - metoda horké brambory: odesílá se nejméně vytíženým směrem
    - náhodné směrování: odesílá se náhodně zvoleným směrem
    - ….
- centralizované směrování
  - existuje centrální autorita (route server) která rozhoduje o všech cestách
    - jednotlivé „směrovače“ mají jen forwardovací tabulky
- distribuované směrování
  - směrovače navzájem spolupracují
    - na hledání nejkratších cest
    - na výměně směrovacích informací
      - pro potřeby aktualizací
  - vše se odehrává „v jednom prostoru“
    - každý směrovač má k dispozici směrovací informace o celé soustavě propojených sítí
      - každý dostává všechny aktualizace
  - možnosti:
    - princip distance-vector
      - vyměňují si vektory vzdáleností
    - princip link-state
      - vyměňují si info o stavu přenosových cest
- hierarchické směrování

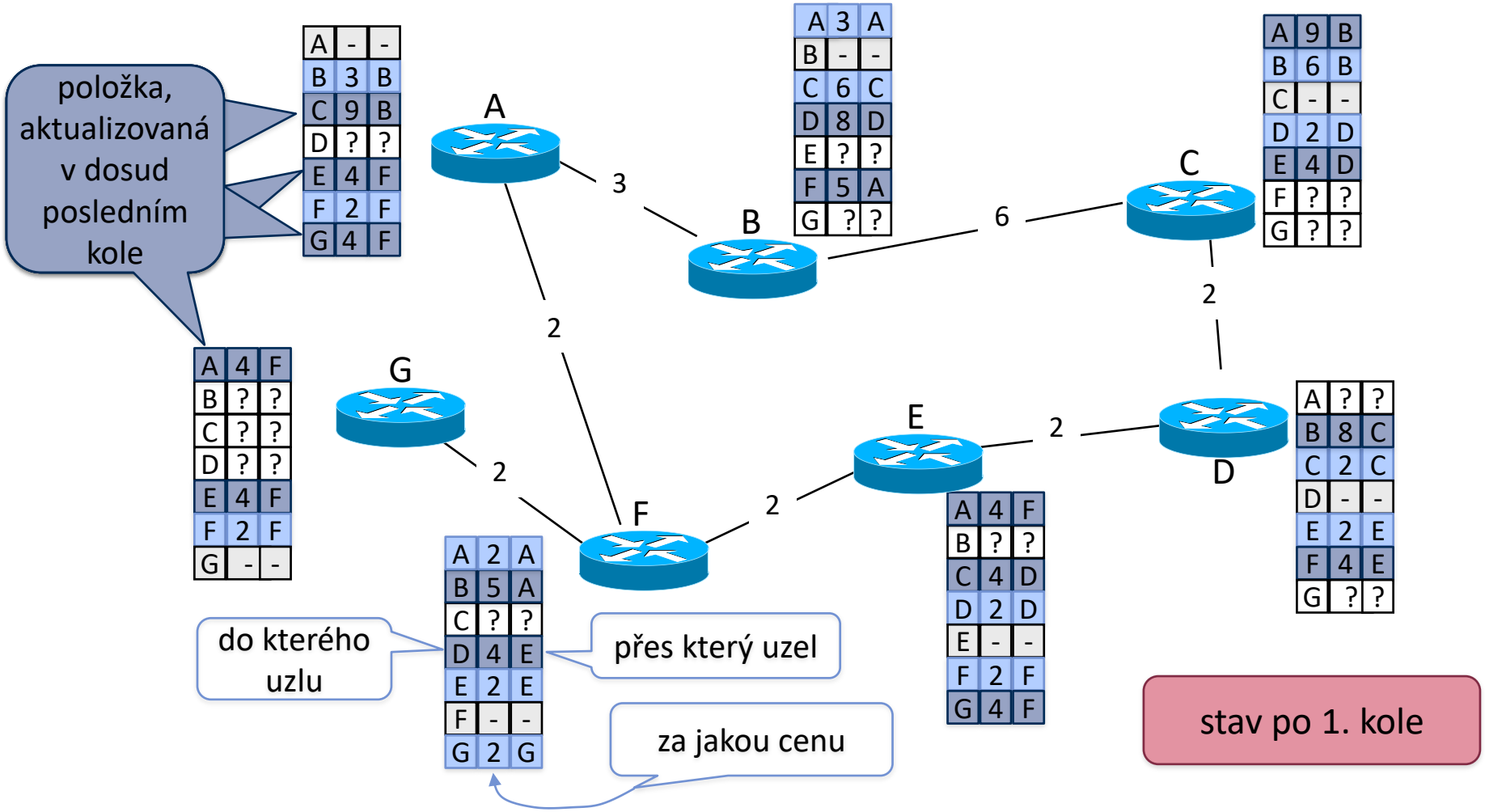
směrování „distance vector“

- je založené na výměně celých směrovacích tabulek
  - na počátku:
    - každý uzel má ve své směrovací tabulce uvedeny jen své přímé sousedy



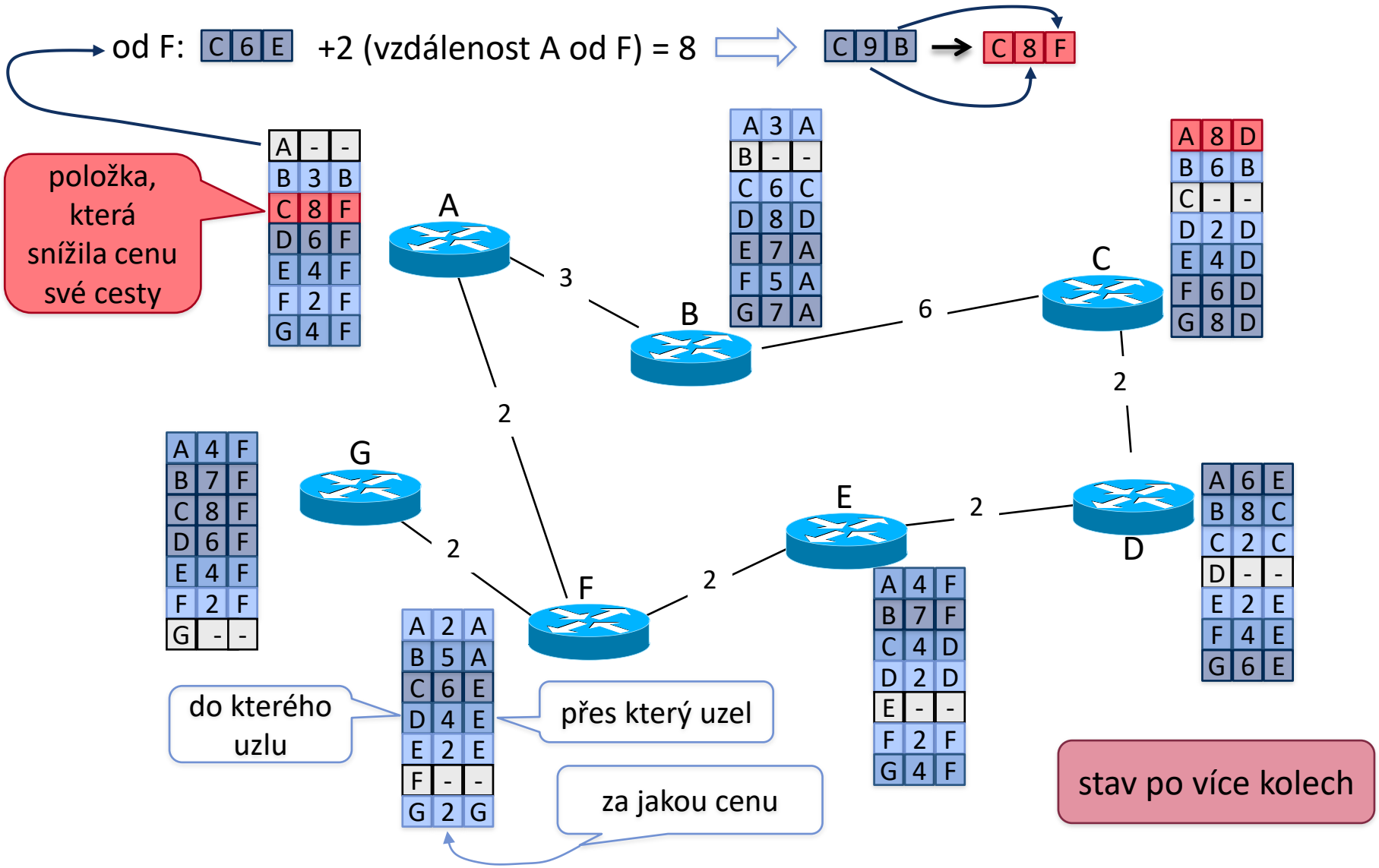
směrování „distance vector“

- v prvním kole každý směrovač předá svou tabulku všem svým přímým sousedům
  - každý směrovač si aktualizuje svou směrovací tabulku podle tabulky souseda



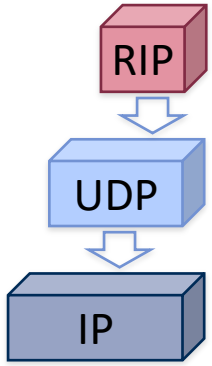
směrování „distance vector“

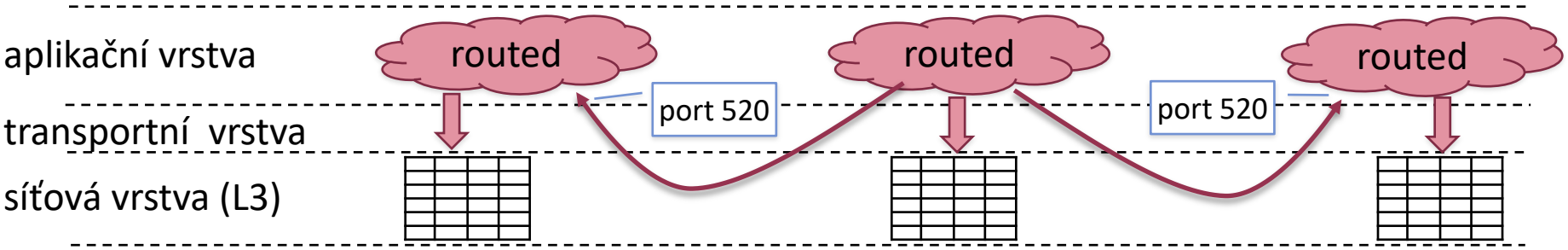
- v každém dalším kole si každý směrovač aktualizuje svou tabulku
  - podle tabulky svého souseda: pokud ten „zná“ kratší cestu k cíli, převezme ji ….



protokol RIP

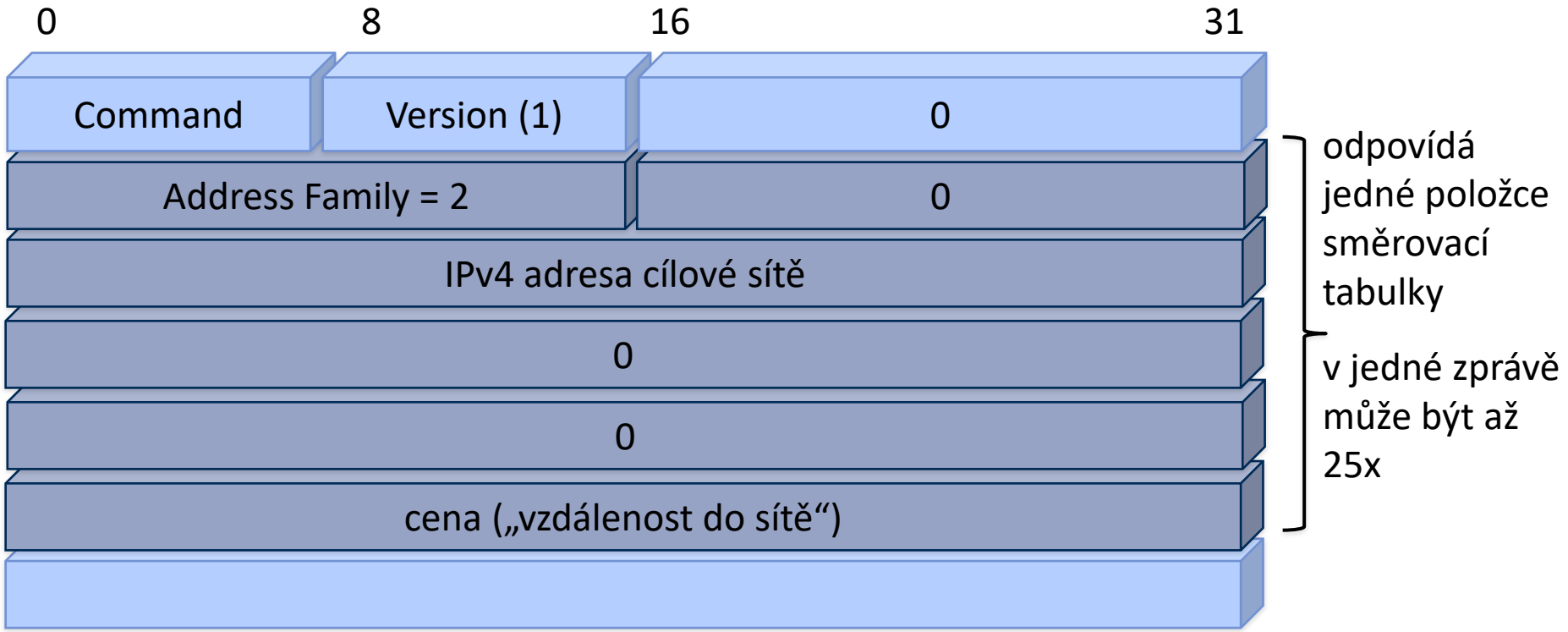
- RIP: Routing Information Protocol (nikoli: Rest In Peace ;-)
  - směrovací protokol (protokol pro výměnu směrovacích informací)
    - fungující na principu distance vector
  - je velmi starý, vznikl pro potřeby malých soustav vzájemně propojených sítí
    - do „vzdálenosti“ (počtu přeskoků, ceny) 15
      - na „vzdálenost“ má vyhrazeny pouze 4 bity: 15 možných hodnot, 16 = nekonečno
  - každý směrovač rozesílá svou směrovací tabulku svým přímým sousedům každých 30 sekund
    - na jejich port č. 520, pomocí transportního protokolu UDP
- RIP je aplikační protokol
  - nejčastěji jej implementuje démon routed (route demon)
    - který „poslouchá“ na portu č. 520
      - tento démon má také na starosti aktualizaci forwardovacích tabulek na síťové vrstvě





zprávy protokolu RIP

- představují buď žádosti (Command=1) nebo odpovědi (Command=2)
  - standardně uzel rozesílá pouze odpovědi (každých 30 sekund)
    - žádost posílá (svým sousedům) jednorázově uzel, který byl právě spuštěn
- Version je rovno 1 nebo 2 (pro RIPv1, RIPv2)
  - Address Family = 2 (pro IP adresy, dnes konstanta)



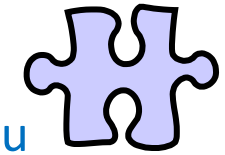
směrování „link state“

- snaha:
  - minimalizovat režii s aktualizací směrovacích informací
  - dosáhnout lepší škálovatelnosti (možnost nasazení ve větších soustavách sítí)
- řešení:
  - neposílají se celé směrovací tabulky, ale pouze info o dostupnosti
    - rozesílají se informace o dostupnosti sousedních uzlů
      - fakticky informace o stavu spoje mezi 2 uzly (proto: link state)
        - je to informace o 2 možných hodnotách (ano/ne)
      - v podobě zpráv LSA (Link State Advertisement)
  - informace (o dostupnosti) stačí posílat při změně
    - plus „po dlouhé době“ kvůli připomenutí
  - informace (o dostupnosti) se musí posílat všem směrovačům !!!
    - nestačí je posílat jen sousedům, jako u distance vector
      - musí se řešit na principu laviny (chytré záplavové směrování)
  - výpočet již není distribuovaný
    - každý uzel má úplnou informaci o celé soustavě vzájemně propojených sítí
    - každý uzel si sám počítá nejkratší cesty

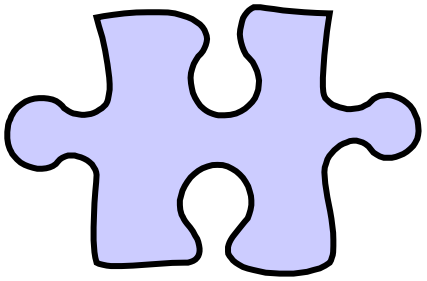
zde je limit škálovatelnosti !!!

„vývoj“ směrování v čase

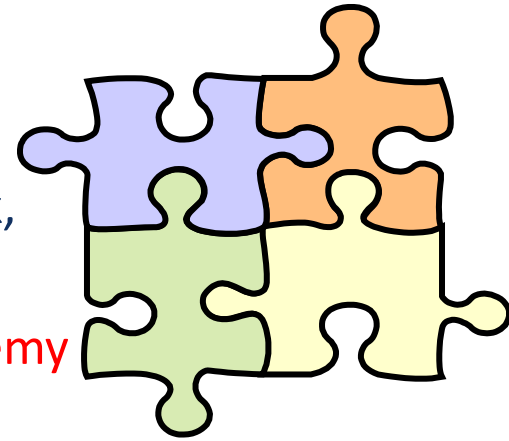
- zpočátku:
  - soustavy vzájemně propojených sítí (internety) byly (relativně) malé
    - objemy směrovacích informací byly (relativně) malé a „zvládnutelné“
  - aktualizace směrovacích informací mohla být založena na metodách typu distance vector (např. protokol RIP)
    - problémem byla špatná škálovatelnost



- poté:
  - soustavy vzájemně propojených sítí jsou větší
    - stejně jako objemy směrovacích informací
  - snaha přejít na metody typy link-state (např. protokol OSPF)
    - které jsou lépe škálovatelné

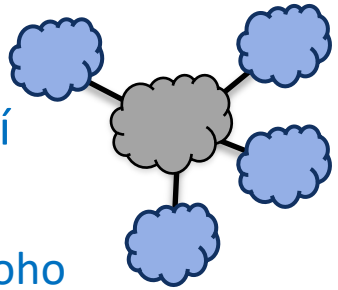


- později:
  - soustavy vzájemně propojených sítí jsou již příliš velké
    - velký objem směrovacích informací nelze zvládnout jinak, než dekompozicí a „lokalizací“ směrovacích informací
  - nutnost přejít na hierarchické směrování a autonomní systémy
    - protokoly IGP a EGP, BGP, peering atd.

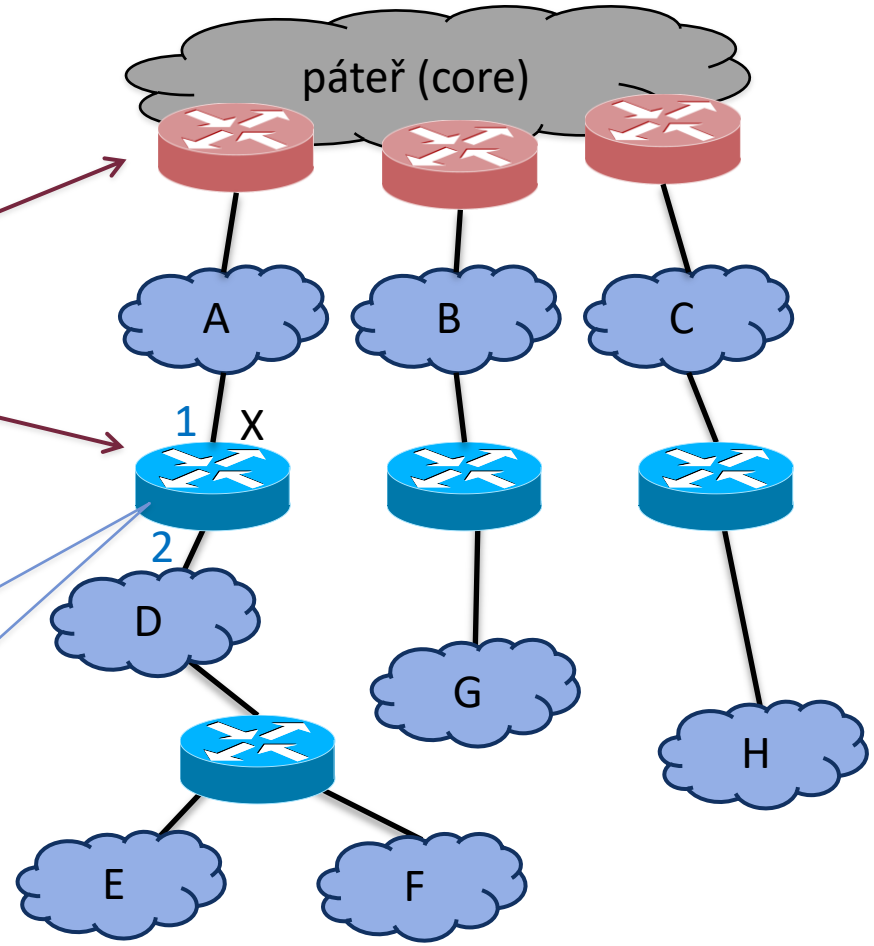


vývoj směrování v rámci Internetu

- úplně na počátku:
  - Internet byl jednou jedinou soustavou (vzájemně propojených) sítí
    - každý směrovač měl úplnou informaci o topologii celého Internetu
      - časem se to stalo neúnosné – směrovacích informací bylo příliš mnoho



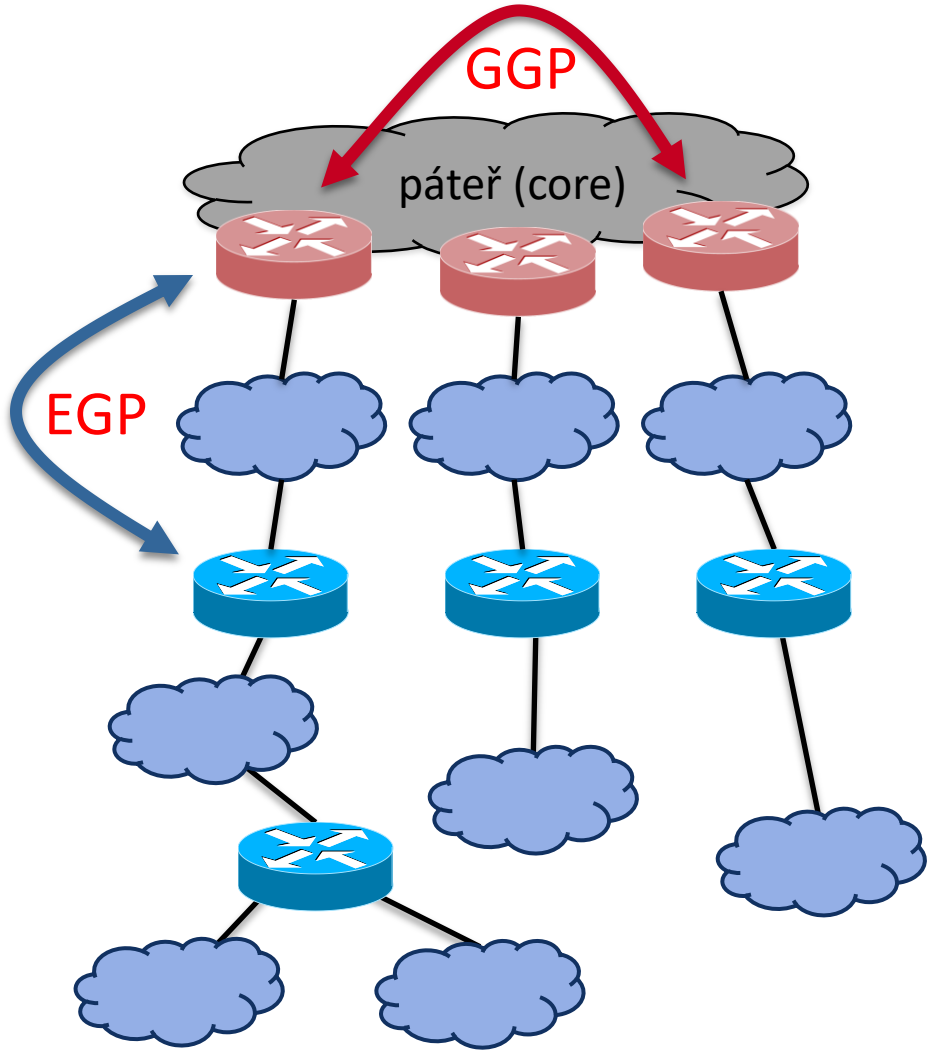
- později:
  - Internet byl rozdělen na páteř (core) a ostatní (non-core)
    - směrovače v páteři (core gateways) měly úplné směrovací informace
    - směrovače mimo páteř (non-core gateways) měly podrobné směrovací informace jen o své „oblasti“
      - znaly cestu jen do „svých podsítí“
      - vše ostatní směrovaly pomocí implicitní cesty do páteře



směrovač X (mimo páteř) směruje:
- k sítím D, E a F přes rozhraní 2
- vše ostatní přes rozhraní 1

protokoly GGP, EGP a IGP

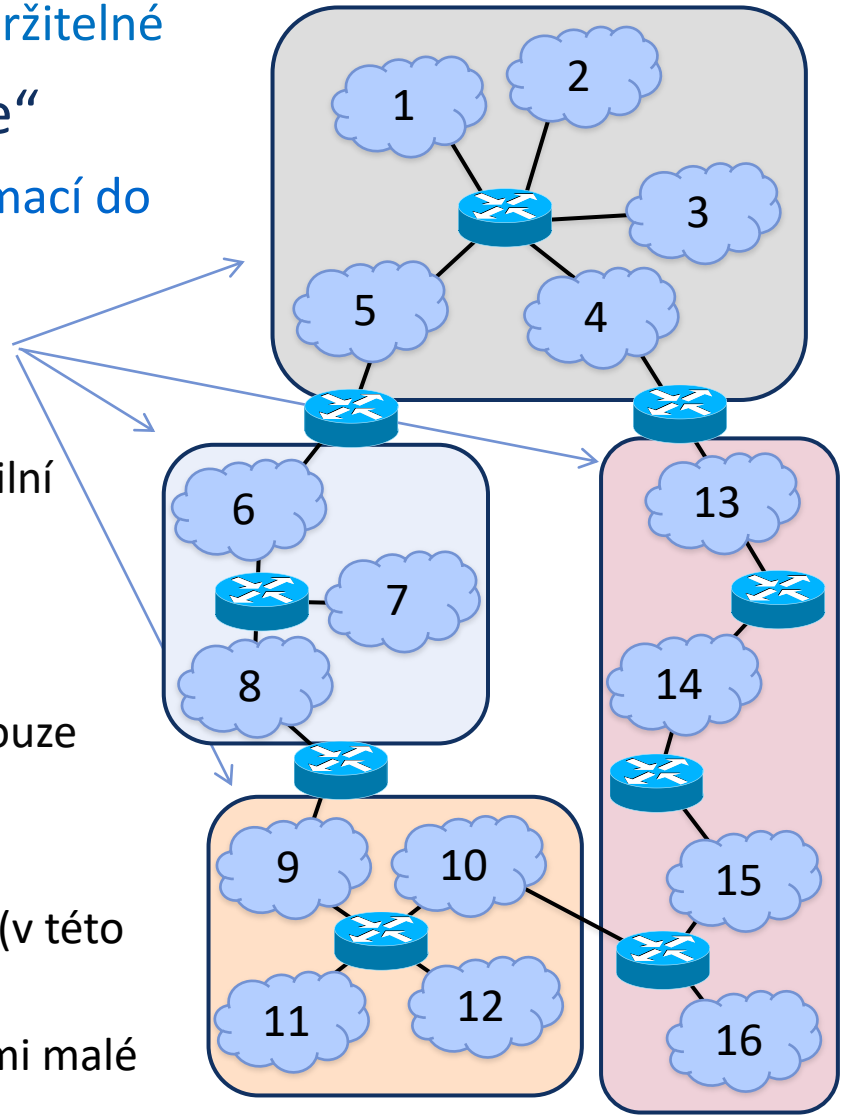
- předpoklad pro rozdělení na core / non-core:
  1. možnost vzájemné komunikace mezi směrovači v páteři (core gateways)
     - k tomu byl vytvořen protokol GGP (Gateway to Gateway Protocol)
  2. možnost komunikace mezi směrovači v páteři a směrovači mimo páteř
     - k tomu sloužila celá skupina protokolů, označovaná jako EGP (Exterior Gateway Protocol)



- jde o „dvouúrovňové“ řešení
  - které vydrželo jen po určitou dobu rozvoje Internetu
  - ale časem se také stalo neúnosné
    - kvůli velkému objemu směrovacích informací v páteři
      - se kterými musely pracovat páteční směrovače

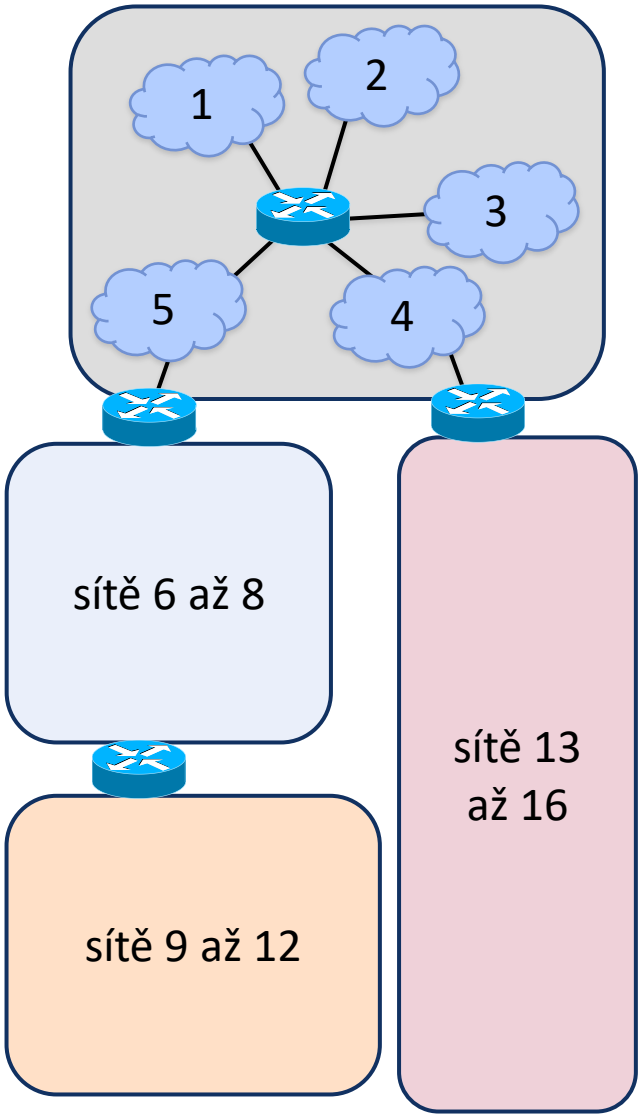
směrovací domény

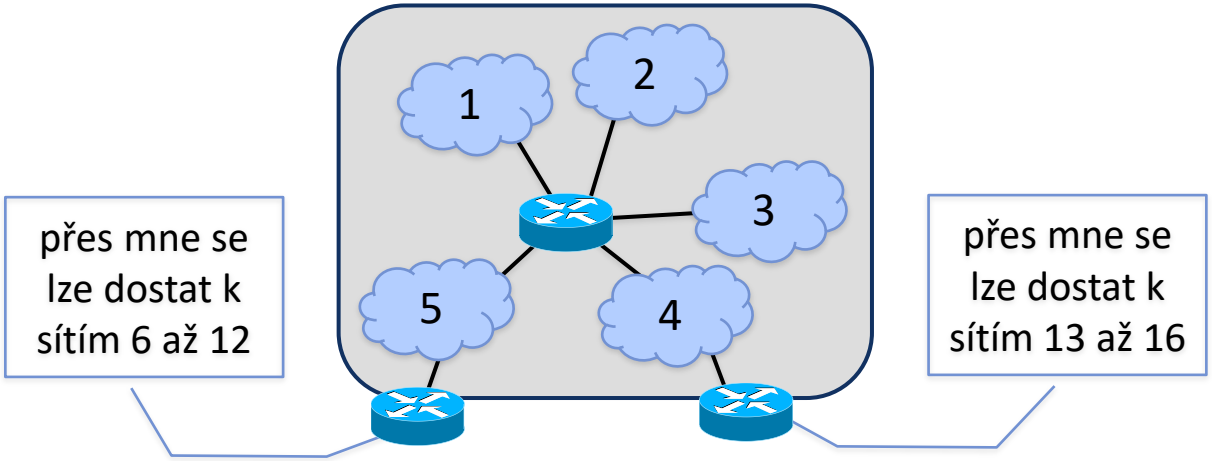
- ani „dvouúrovňové řešení“ (s core a non-core) není dostatečně škálovatelné
  - při určité velikosti Internetu se stalo neudržitelné
- řešením je pouze důsledná „lokalizace“
  - soustředění detailních směrovacích informací do jednotlivých oblastí
    - tzv. směrovacích domén (routing domains)
  - princip:
    - v rámci směrovací domény jsou šířeny detailní směrovací informace
      - těchto informací může být hodně – ale domény lze volit dostatečně malé
    - mezi směrovacími doménami jsou šířeny pouze informace o dostupnosti
      - tzv. reachability information
        - „intervalové“ informace typu OD-DO (v této doméně jsou sítě od A/x do B/y)
          - informace o dostupnosti jsou velmi malé (obvykle jen síťové prefixy)



směrování na základě dostupnosti

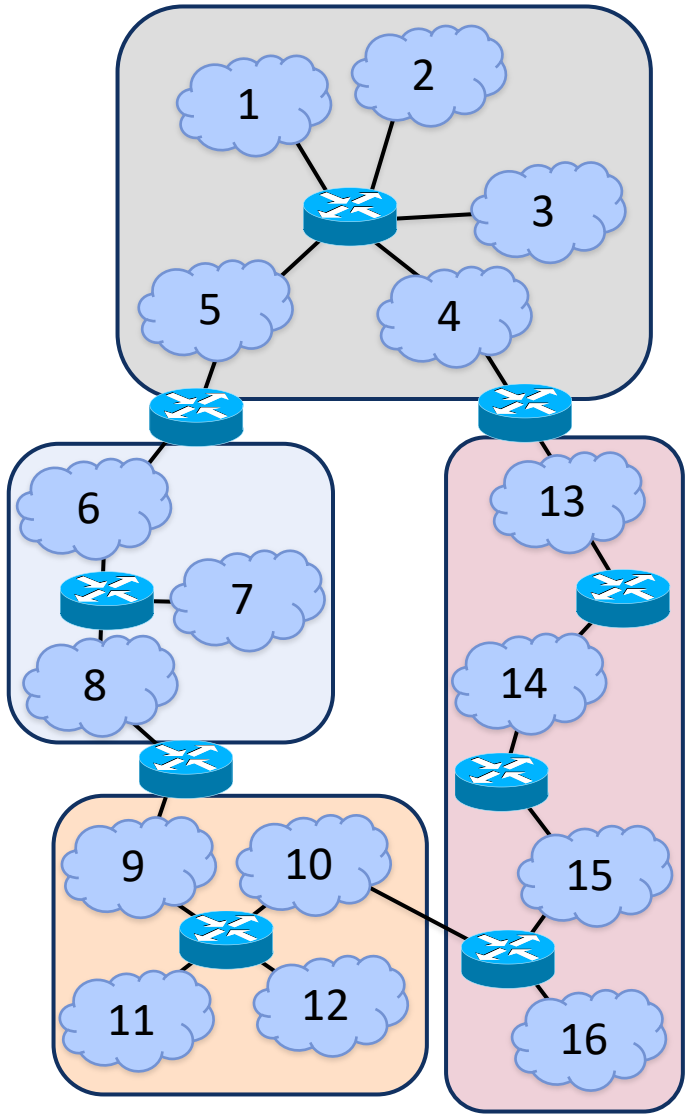
- mění se pravidla směrování mezi doménami
  - místo „optimální“ (nejlevnější) cesty se hledá „alespoň nějaká“ cesta
- důvod:
  - algoritmy distance vector a link state zde nejsou použitelné
    - protože se nepracuje s ohodnocením hran
      - není známa „cena“ cesty přes směrovací domény
        - nejčastěji je jen jedna možná cesta
  - používají se algoritmy typu path vector
    - které pracují s celými cestami
      - „průchody“ přes směrovací domény





autonomní systémy a protokoly IGP

- směrovacím doménám se obvykle říká autonomní systémy (AS)
  - kvůli tomu, že si mohou samy (autonomně) rozhodovat o detailním směrování uvnitř sebe sama
  - mohou si vybrat takový způsob směrování, jaký chtějí / považují za nejvhodnější
    - obvykle: na principu distance vector, link state
    - nezávisle na tom, jaký způsob směrování si vyberou v jiném autonomním systému
  - jsou identifikovány čísly, které přiděluje IANA
    - čísly AS: například AS 2852 (AS CESNET-u)
- protokoly IGP (Interior Gateway Protocols)
  - jde o souhrnné označení pro všechny protokoly, které se používají pro směrování uvnitř autonomních systémů (směrovacích domén)
  - v praxi například:
    - protokol RIP
    - protokol OSPF



směrovací politiky, protokoly EGP, BGP

- autonomní systémy (AS)
  - mají obvykle jednoho vlastníka (komerčního provozovatele), který rozhoduje:
    - o směrování v rámci svého AS (směrovací domény)
      - včetně volby konkrétního protokolu z „množiny“ IGP
    - o vazbách na ostatní AS
      - kudy a jak směrovat provoz do ostatních AS
        - roli zde mohou hrát i další faktory, včetně komerčních vztahů
      - co a jak inzerovat ostatním AS
        - jakou dostupnost konkrétních sítí propagovat (inzerovat)

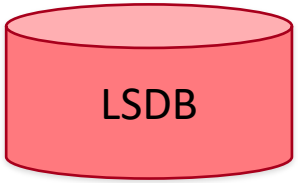
toto určuje tzv. směrovací politika

  - pro tyto účely musí existovat vhodné protokoly:
    - protokoly EGP (Exterior Gateway Protocol)
      - obecné označení pro protokoly, umožňující definovat vazby mezi AS
        - tj. jde o skupinu protokolů
      - dnes je nejpoužívanějším protokolem ze skupiny EGP protokol BGP
        - Border Gateway Protocol
          - aktuálně verze 4

stejné označení, jako u core/non-core uspořádání

protokol OSPF

- OSPF: Open Shortest Path First
  - patří mezi Interior Gateway Protocols (IGP)
    - vznikl (1989) jako náhrada za RIP, je velmi komplikovaný
      - dnes se používá verze OSPF 2 (označovaná jen jako OSPF) z roku 1991,dle RFC 1247
  - používá se:
    - ke směrování (aktualizaci směrovacích informací) uvnitř AS
- je typu link-state
  - každý směrovač má úplnou informaci o topologii celé soustavy propojených sítí, ve které se nachází
    - tzv. LSDB: Link-State DataBase (“topologická databáze“)

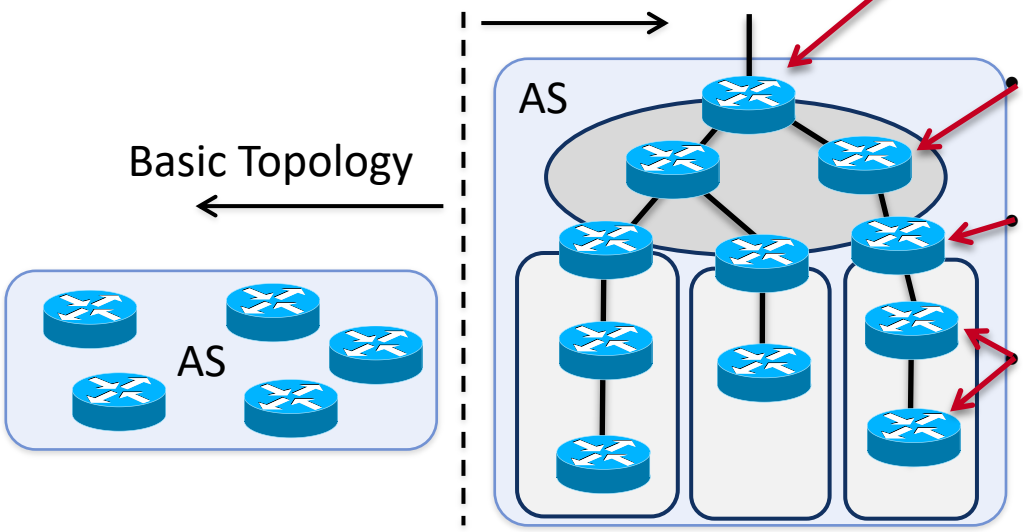


  - každý směrovač si sám počítá nejkratší cesty (Shortest Path)
    - podle údajů ve své topologické databázi (LSDB), pomocí Dijkstrova algoritmu
  - každý směrovač průběžně monitoruje dostupnost všech sousedních směrovačů
    - zjišťuje „stav linky“ (link state) ke svým sousedům
  - každou zjištěnou změnu zanese do své topologické databáze
    - a oznámí (rozešle) informaci o změně všem ostatním směrovačům
      - které si také upraví svou LSDB a přepočítají optimální cesty

protokol OSPF

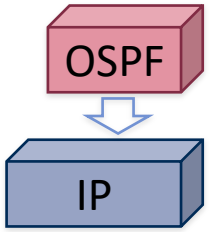
- může fungovat ve dvou různých režimech

- Basic Topology
  - celý AS (autonomní systém) je homogenní
    - jedna LSDB „pokrývá“ celý AS
  - každý směrovač „zná“ celý AS
    - má jen jednu LSDB
  - je to vhodné jen pro menší AS
    - hůře škálovatelné
  - všechny směrovače jsou si rovny



- Hierachical Topology
  - celý AS je rozdělen na oblasti (Areas)
    - každá oblast má svou LSDB
      - oblast = *autonomní systém v malém*
    - jedna oblast je páteřní, ostatní ne-páteřní
      - obdoba „core | non-core“
  - role směrovačů se mohou lišit
    - hraniční směrovač (boundary router)
      - propojuje páteřní oblast s „vnějším světem“
    - páteřní směrovač (backbone router)
      - uvnitř páteřní oblasti
    - hraniční směrovač (area border router)
      - propojují oblasti mezi sebou,
    - vnitřní směrovač (internal router)
      - uvnitř oblastí, pracují jen s 1 LSDB

zprávy protokolu OSPF

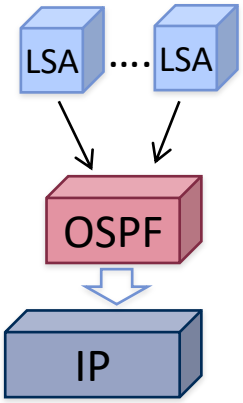


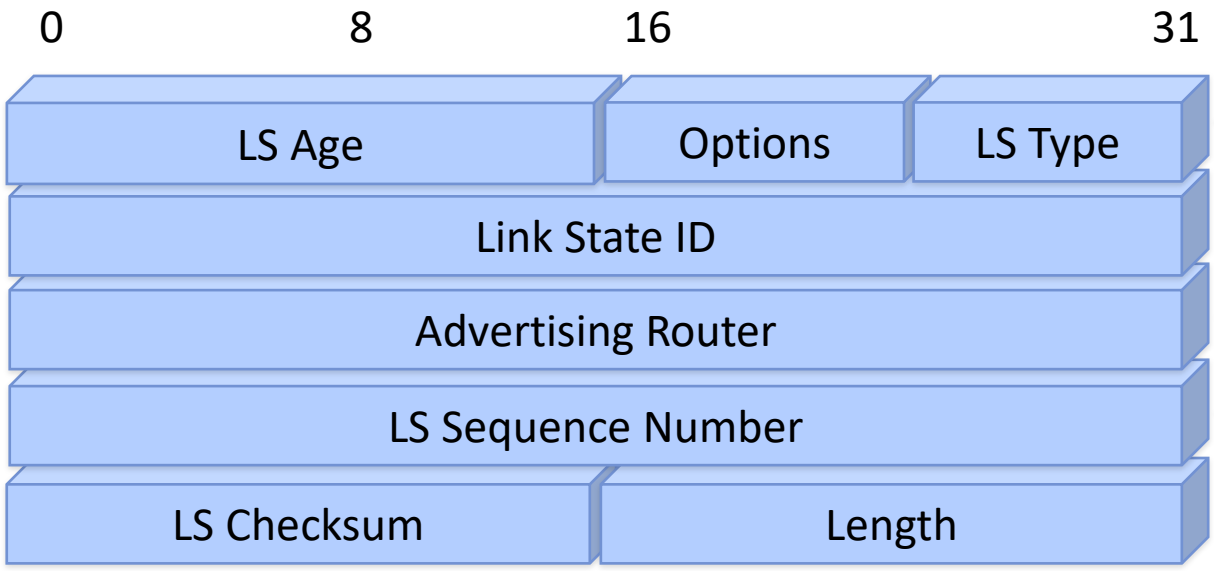
- protokol OSPF používá 5 druhů zpráv
  - které vkládá přímo do IP datagramů (Protocol ID= 89)
- Hello
  - zpráva pro „navazání kontaktu“ se sousedním směrovačem
- Database Description
  - přenos obsahu topologické databáze (LSDB)
- Link State Request
  - žádost o zaslání části/celé LSDB
- Link State Update
  - informace o změně v topologii (změně stavu linky)
- Link State Acknowledgement
  - potvrzení zprávy Link State Update

- směrovač po spuštění:
  - „naváže kontakt“ se všemi sousedními směrovači
    - pomocí zpráv HELLO
  - vyžádá si zaslání LSDB od svých sousedů
    - pomocí zpráv Link State Request a Database Description
- když směrovač zjistí změnu
  - informuje o tom všechny ostatní směrovače (ve své oblasti)
    - pomocí zpráv Link State Update
      - šíří se pomocí multicastu
  - všechny směrovače potvrdí změnu
    - pomocí Link State Acknowledgement

OSPF LSA: Link State Advertisement

- LSA je položka, která popisuje jeden spoj a jeho stav (link state)
  - představa: jde o jednu položku „topologické databáze“
  - některé OSPF zprávy (např. Database Description či Link State Update) obsahují několik takovýchto položek LSA
    - například: když směrovač zjistí změnu dostupnosti svého souseda,
      - sestaví podle toho položku LSA (1 nebo více)
      - vloží ji do zprávy Link State Update,
      - rozešle ji (záplavově) všem směrovačům ve své oblasti/oblastech





- cíl je směrovač
- cíl je jednou sítí
- cílem jsou sítě od..do
- ……

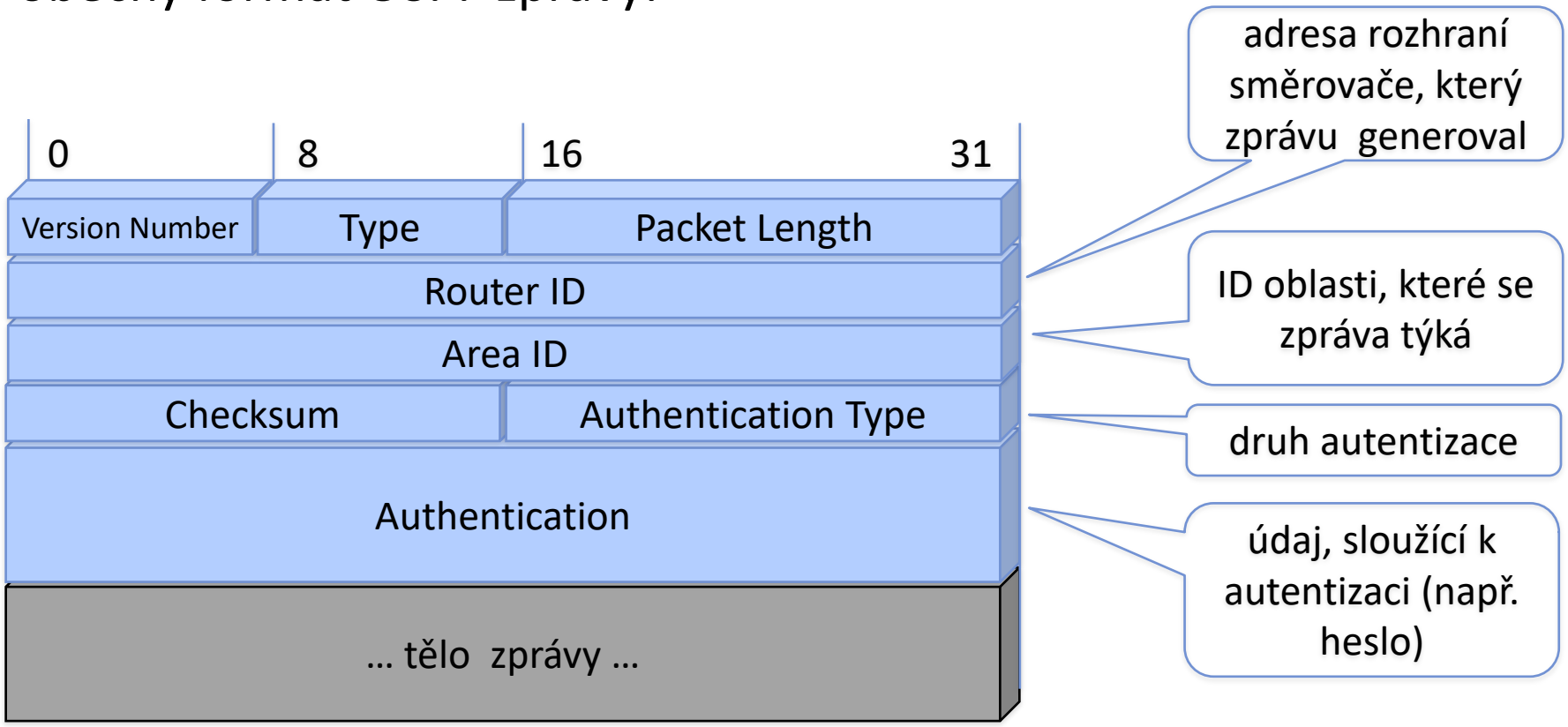
kam spoj vede
(IP adresa cíle)

kde spoj začíná
(IP adresa zdroje)

pořadové číslo LSA
(pro detekci duplicit)

další vlastnosti protokolu OSPF

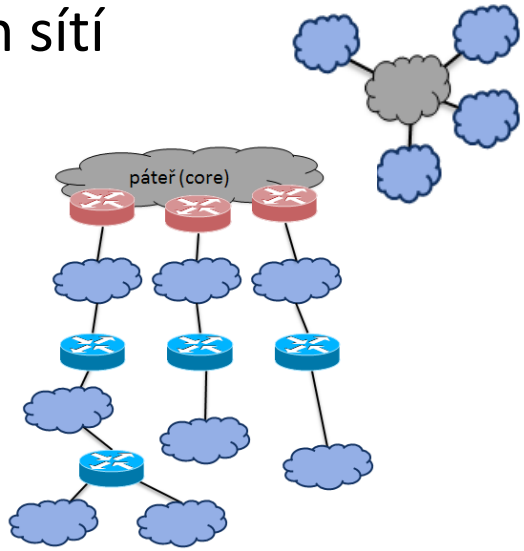
- jednotlivé zprávy OSPF mohou být autentizovány
  - aby příjemce měl (rozumnou) jistotu, že jsou autentické a nikoli nějak podvržené
- možností autentizace je více (s různou „silou“)
  - od jednoduchého hesla až po použití asymetrické kryptografie
- obecný formát OSPF zprávy:

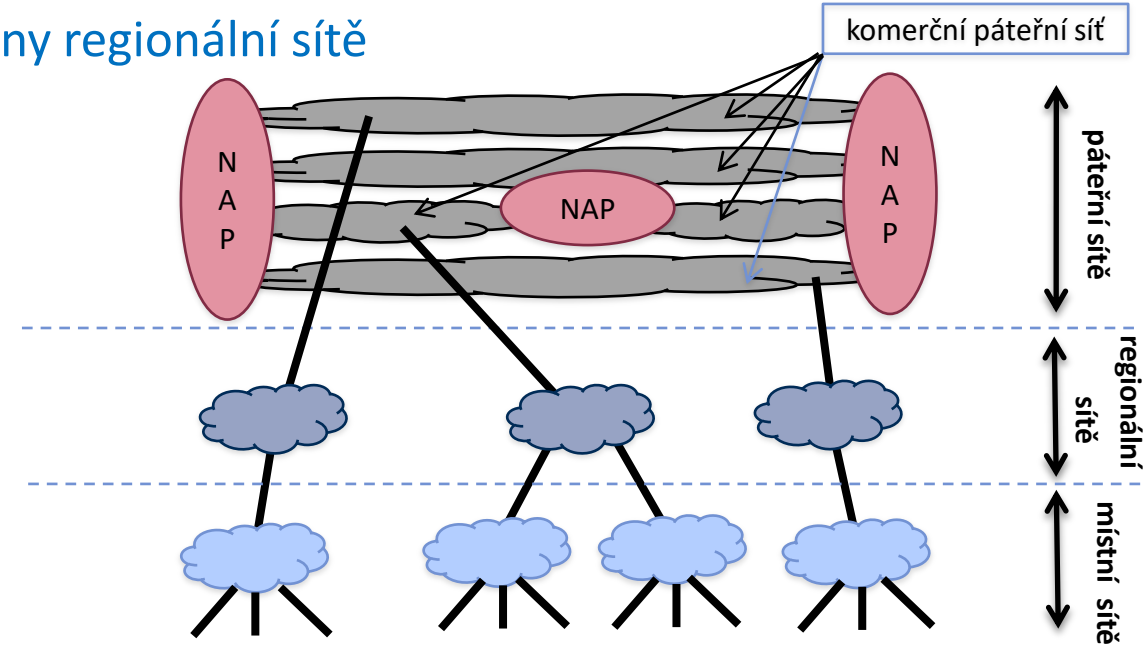


vývoj směrování v rámci Internetu

- nejprve: „plochá“ soustava vzájemně propojených sítí
  - bez nějaké diferenciace, všechny sítě a směrovače jsou si rovny, všichni znají vše
- poté: 2-úrovňové uspořádání
  - s páteřní (core) sítí a nepáteřními (non-core) sítěmi
- následně (nástup komerčních páteřních sítí)
  - páteřních sítí je více, jsou vzájemně alternativní
    - jsou propojeny pomocí bodů NAP (Network Access Points)
  - na páteřní sítě jsou napojeny regionální sítě
  - a na ně zase místní sítě

počátek 90. let

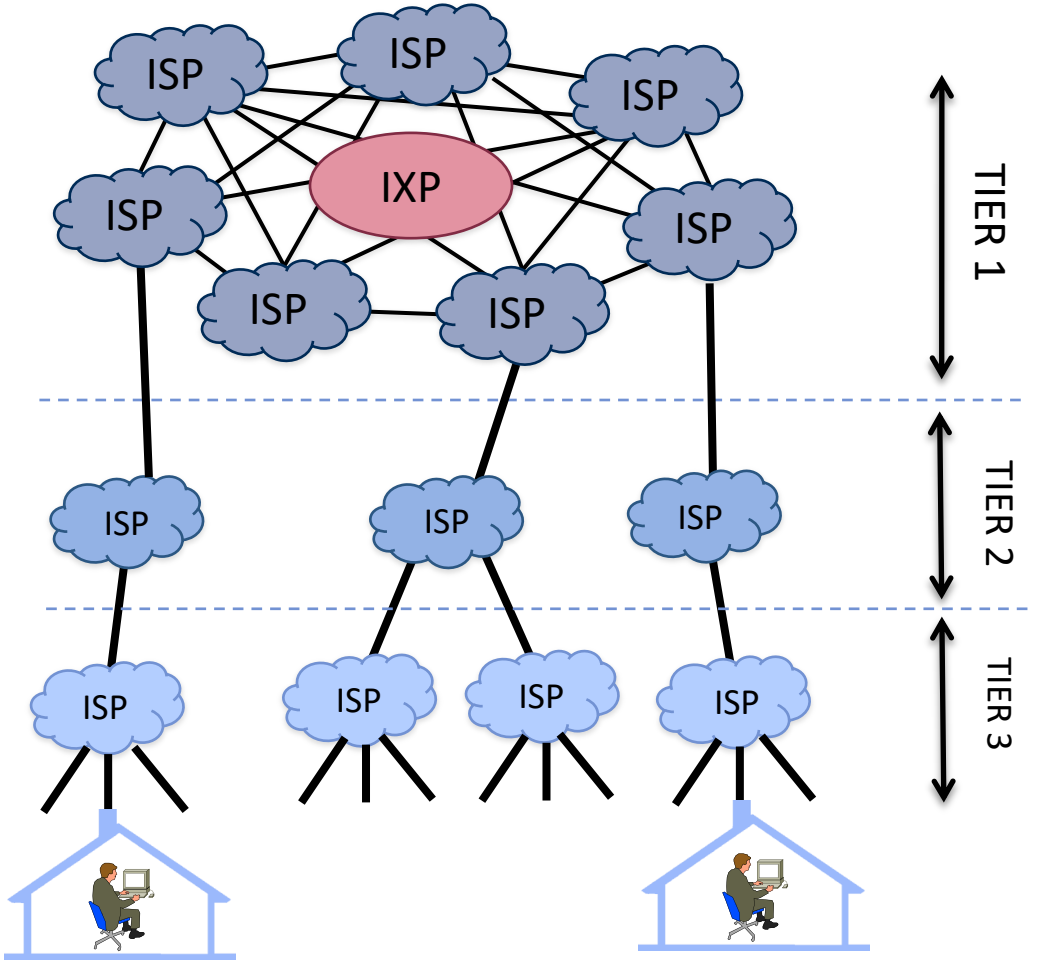




toto uspořádání se používalo od roku 1995 cca do roku 2000

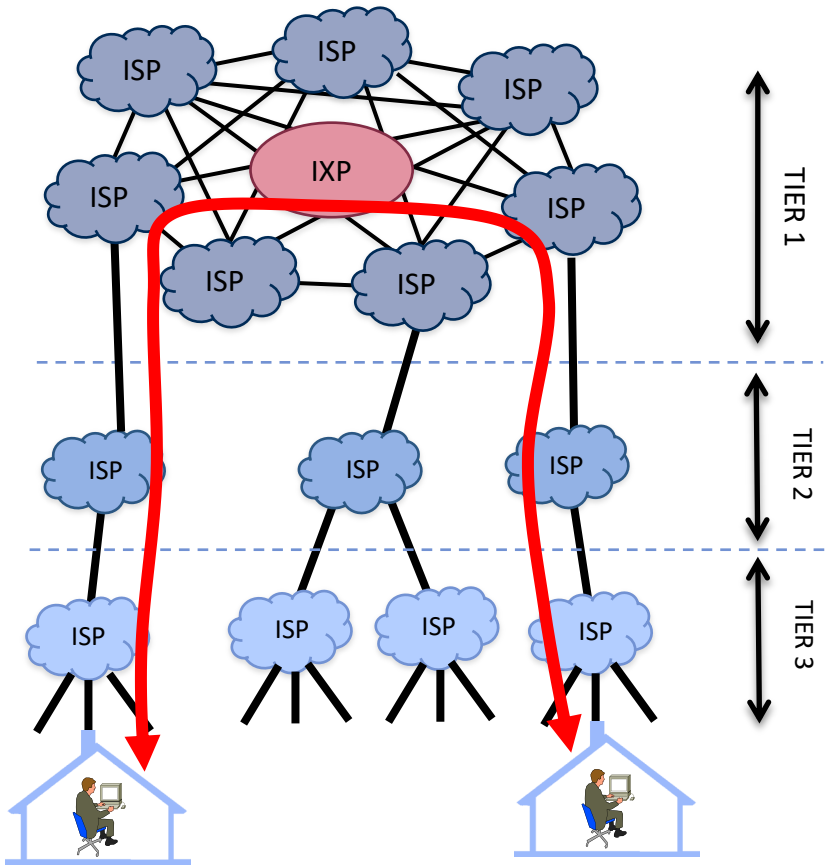
dnešní struktura Internetu

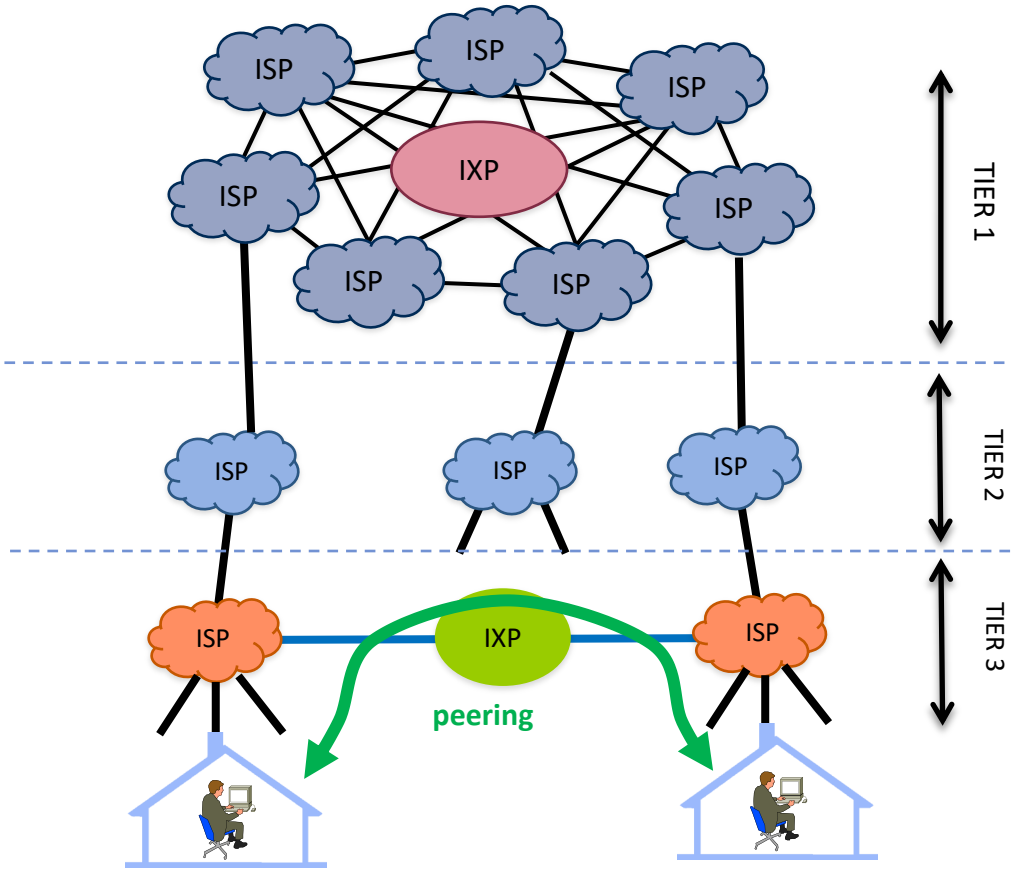
- body NAP byly časem nahrazeny centry IXP (Internet eXchange Points)
  - charakteru peeringových bodů
- z páteřních, regionálních a místních sítí se staly „běžné“ (komerční) jednotlivých ISP, rozdělené do úrovní (tiers)
- Tier 1:
  - takové sítě ISP, které nejsou napojeny na žádné „vyšší“ sítě
    - obdoba dřívějších páteřních sítí
- Tier 2:
  - sítě ISP, které jsou napojené na sítě Tier 1 a získávají od nich (kupují si) tzv. upstream (páteřní, tranzitní) konektivitu
- Tier 3:
  - sítě ISP, napojené na sítě Tier2
  - obvykle již připojují koncové uživatele



peering v rámci Internetu

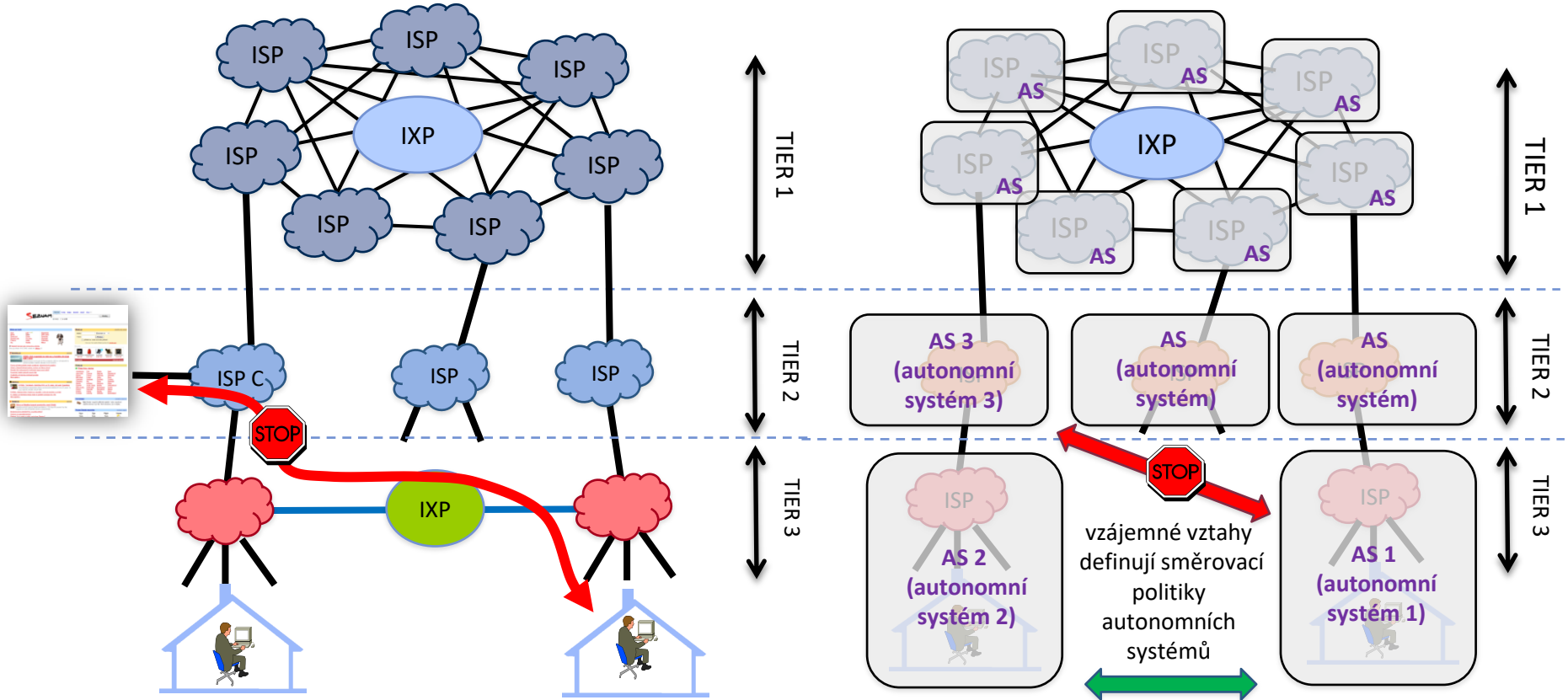
- standardně (bez peeringu) provoz mezi různými sítěmi (Tier 3) prochází přes vyšší úrovně (Tier 2 či Tier 1)
  - což je drahé a pomalé
- při (lokálním) peeringu provoz nemusí „stoupat“ na vyšší úroveň
  - ale může se předávat na dané úrovni, přes (lokální) peeringové body (IX Points)





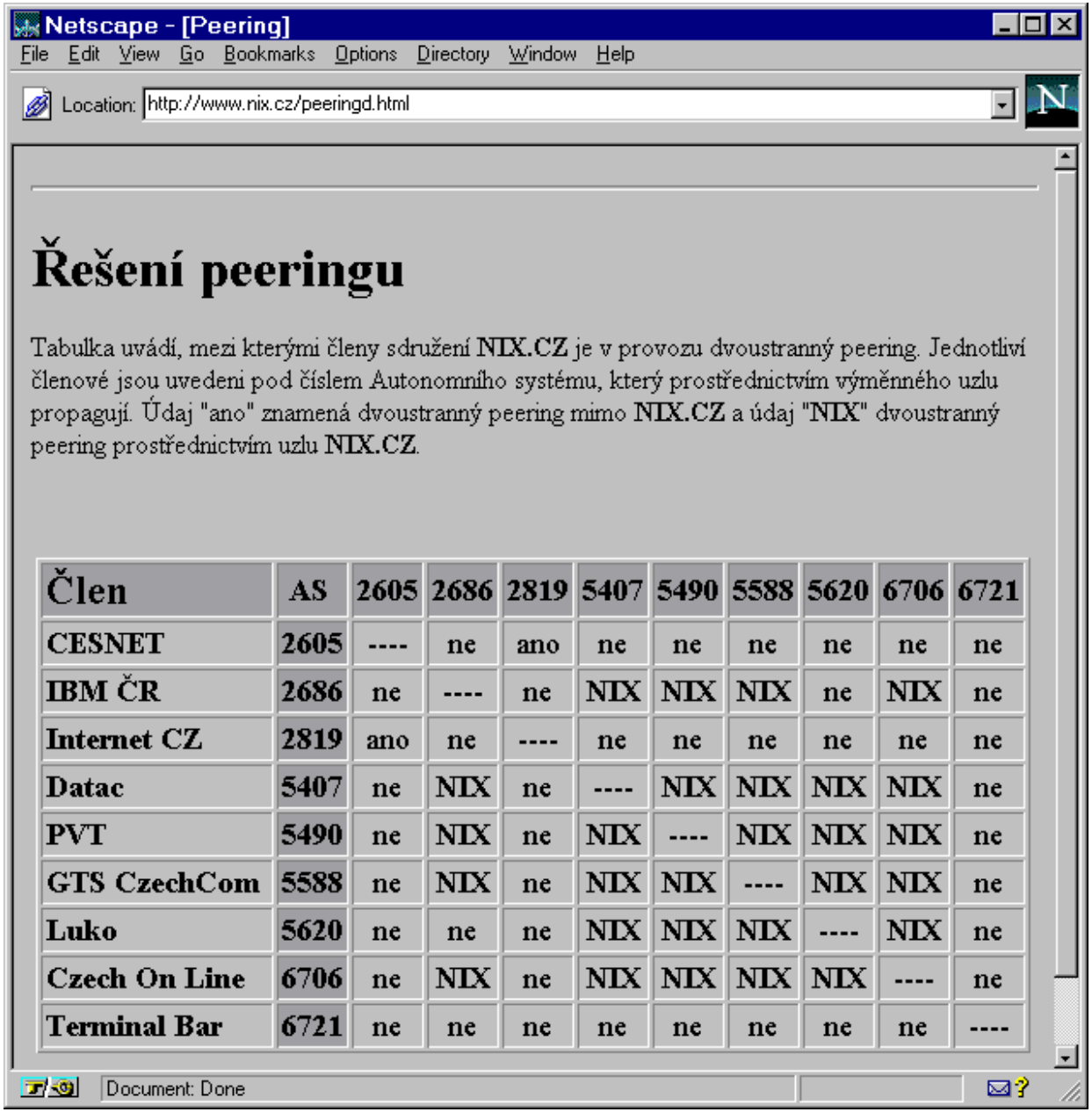
podmínka peeringu

- podmínkou pro zavedení peeringu je možnost vyhnout se situaci, kdy jeden ISP využívá (zneužívá) drahou upstream konektivitu jiného ISP
  - řeší se pomocí autonomních systémů a směrovacích politik
    - síť každého ISP je samostatným autonomním systémem (AS)
    - každý ISP si (v rámci své směrovací politiky) určuje, s kterým AS jiných ISP si jeho AS chce vyměňovat provoz (včetně: „odkud“ a „kam“)



peeringová matice (NIX.CZ)

- peeringová matice popisuje, mezi kterými AS existuje peering, a mezi kterými nikoli
  - příklad:
    - peeringová matice tabulka NIX.CZ z května 1997
  - v roce 2013
    - stejná matice má přes 100 řádek a sloupců



Řešení peeringu

Tabulka uvádí, mezi kterými členy sdružení NIX.CZ je v provozu dvoustranný peering. Jednotliví členové jsou uvedeni pod číslem Autonomního systému, který prostřednictvím výměnného uzlu propagují. Údaj "ano" znamená dvoustranný peering mimo NIX.CZ a údaj "NIX" dvoustranný peering prostřednictvím uzlu NIX.CZ.

Člen	AS	2605	2686	2819	5407	5490	5588	5620	6706	6721
CESNET	2605	----	ne	ano	ne	ne	ne	ne	ne	ne
IBM ČR	2686	ne	----	ne	NIX	NIX	NIX	ne	NIX	ne
Internet CZ	2819	ano	ne	----	ne	ne	ne	ne	ne	ne
Datac	5407	ne	NIX	ne	----	NIX	NIX	NIX	NIX	ne
PVT	5490	ne	NIX	ne	NIX	----	NIX	NIX	NIX	ne
GTS CzechCom	5588	ne	NIX	ne	NIX	NIX	----	NIX	NIX	ne
Luko	5620	ne	ne	ne	NIX	NIX	NIX	----	NIX	ne
Czech On Line	6706	ne	NIX	ne	NIX	NIX	NIX	NIX	----	ne
Terminal Bar	6721	ne	ne	ne	ne	ne	ne	ne	ne	----