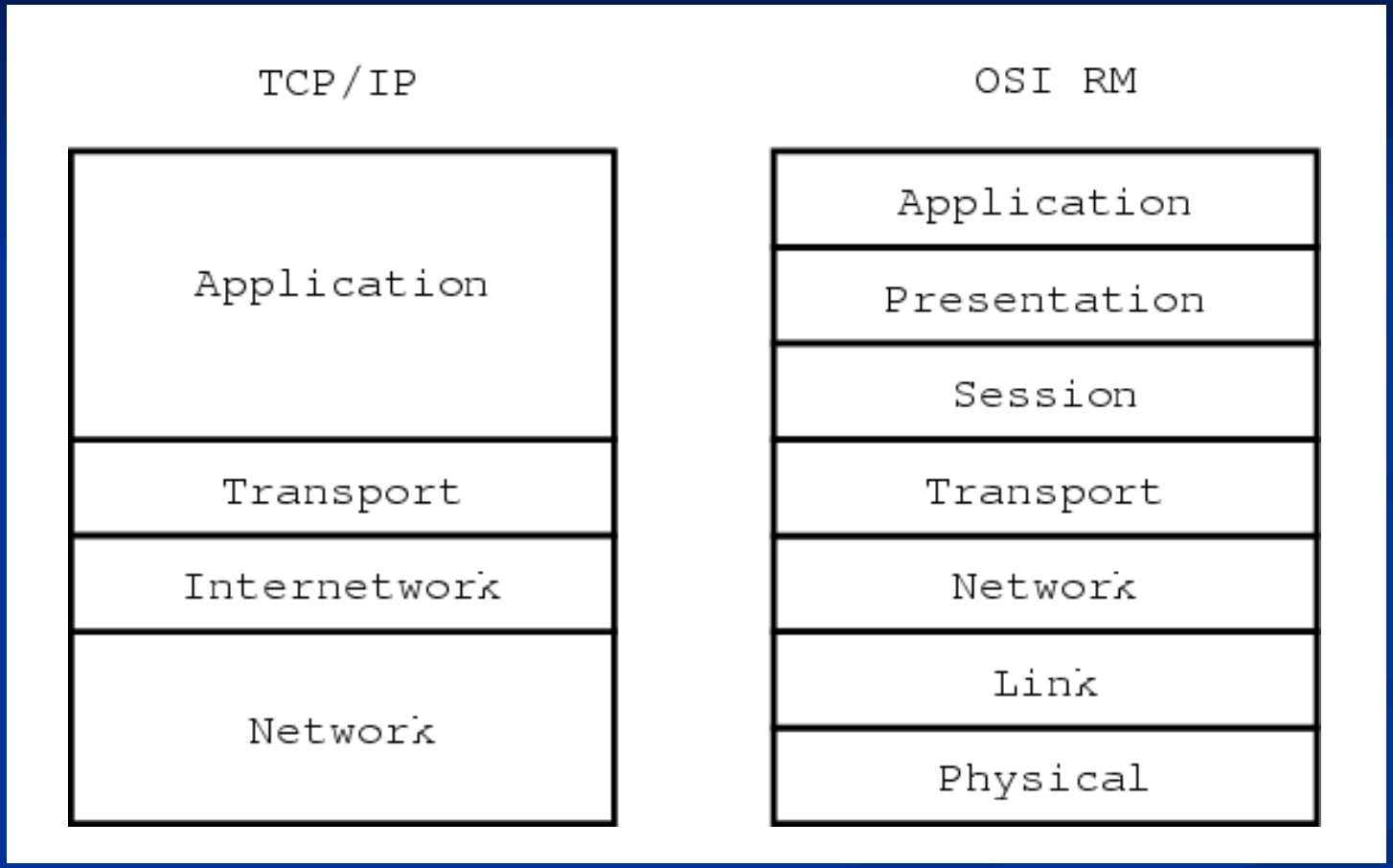
# Protokoly TCP/IP

Petr Grygárek

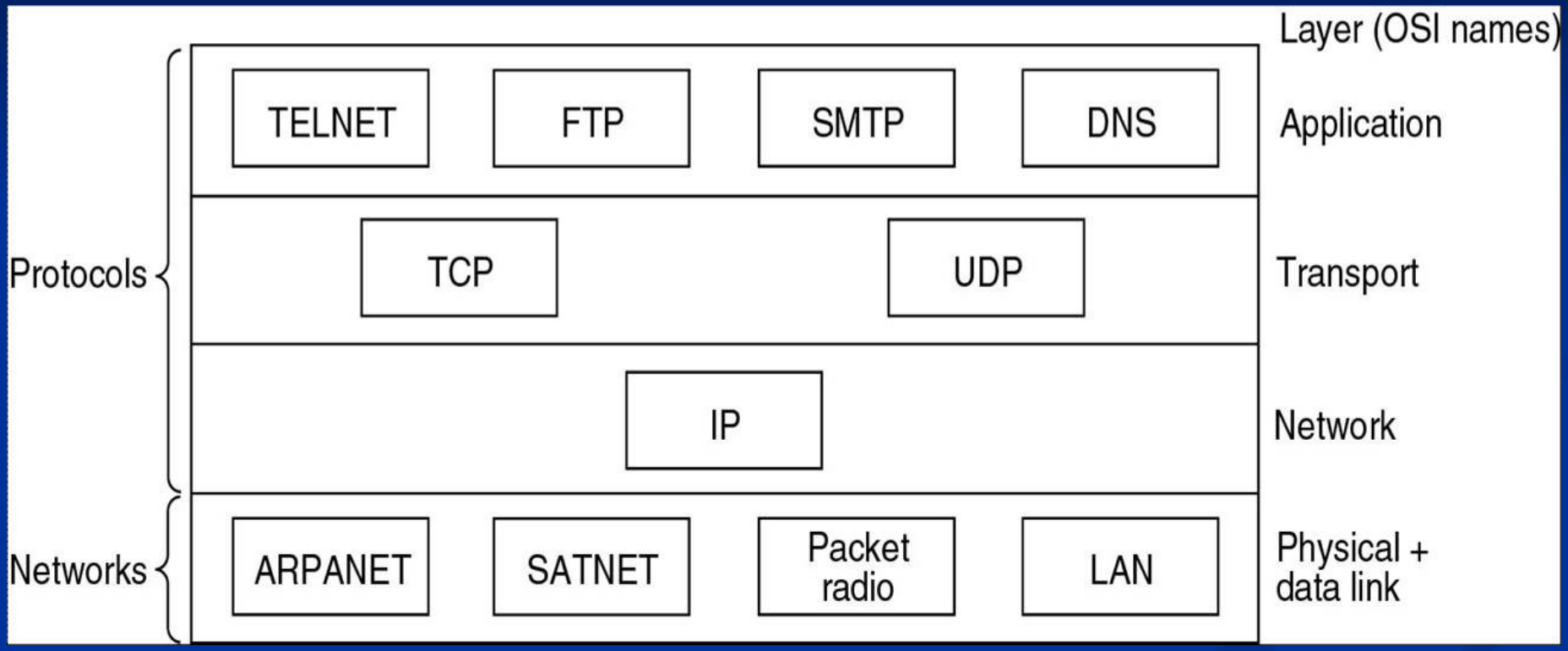
# TCP/IP

- standard pro komunikaci v Internetu
  - a stále více i v intranetech
- TCP – protokol 4. vrstvy (spolu s UDP)
- IP - protokol 3. vrstvy

# Vrstvený model a srovnání s OSI-RM



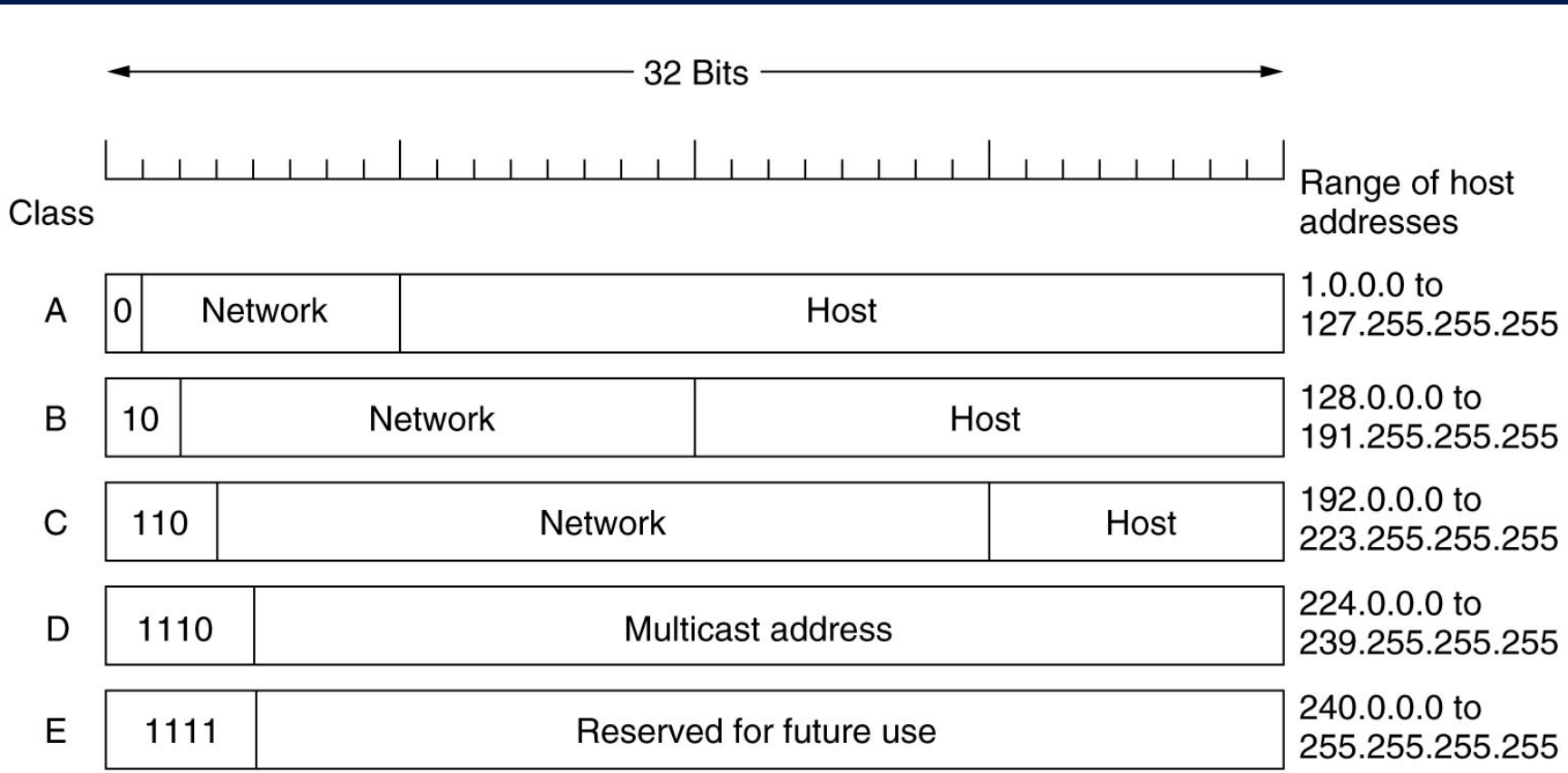
# Vrstvený model TCP/IP



# Adresace v protokolu IP

- adresy 32b (X.X.X.X)
  - každé rozhraní prvku rozumějícího 3. vrstvě OSI RM připojené do sítě musí mít jednoznačnou IP adresu
    - (stanice a rozhraní směrovačů)
- dělení na  adresu sítě + adresa uzlu v rámci sítě
  - adresy všech stanic na segmentu LAN (broadcast doméně 2. vrstvy) mají společnou část IP adresy (adresu sítě, prefix)
  - směrovače nemusí ukládat adresy všech stanic v sítí, pouze adresy jednotlivých sítí
    = > omezení rozsahu směrovacích tabulek

# Třídy IP adres (historie)



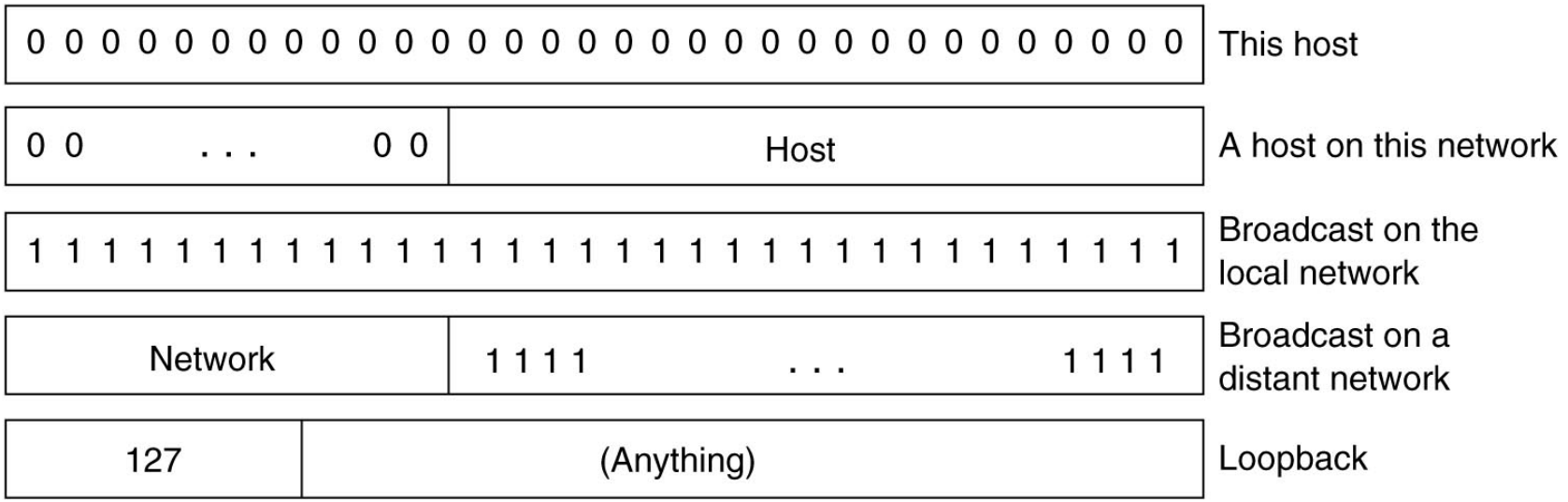
# Beztřídní (classless) adresy

- délka prefixu sítě přidělována podle potřeby
- k beztřídní adrese musí být specifikována maska podsítě (subnet mask) určující délku prefixu
- v poslední době se třídy adres prakticky přestaly používat
  - přechod na CIDR-Classless Inter-Domain Routing)
  - možnost agregování záznamů ve směrovací tabulce na základě společného prefixu (bez ohledu na třídy) - supernetting

# Přidělování IP adres

- adresy sítí přiděluje oblastní správce (pro Evropu RIPE)
  - vyřizování elektronickou cestou (zprostředkovává poskytovatel)
- adresy původně přidělovány bez ohledu na topologii a geografickou polohu
- v posledních letech snaha o hierarchickou adresaci (přidělování prefixu sítě s délkou podle potřeby)
  - případné další podsíťování (subnetting)
- soukromé izolované sítě mají vyhrazené rozsahy adres použitelné opakovaně, nesmí být přímo připojeny k Internetu
  - pokud jsou připojeny, tak přes proxy s překladem adres-NAT
  - 10.0.0.0, 172.16.0.0-172.31.0.0, 192.168.0.0-192.168.255.0

# Speciální IP adresy

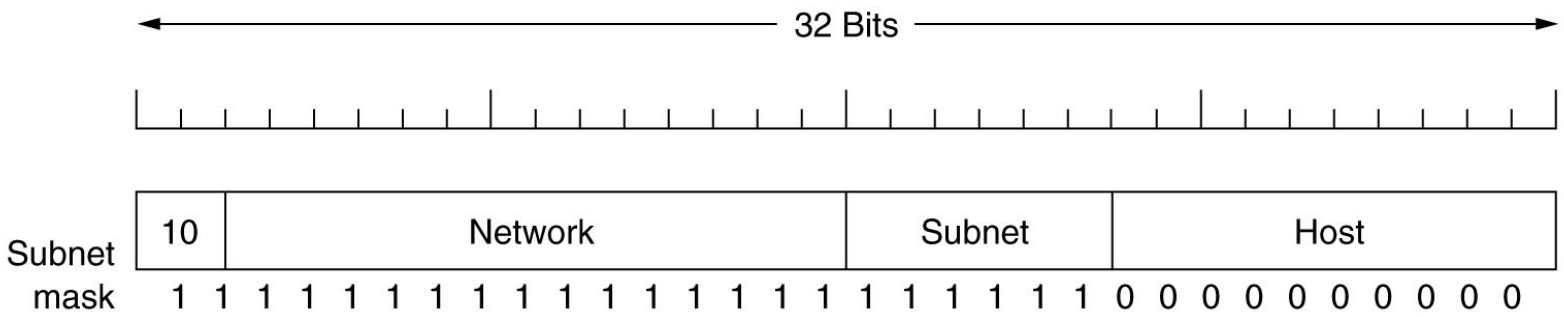



- Univerzální broadcast: 255.255.255.255
- Multicast: 224.x.x.x - 239.x.x.x

# Podsít'ování (subnetting)

- možnost rozdělení přiděleného adresního rozsahu mezi více segmentů
  - každý segment musí mít svou vlastní adresu podsítě
- Umožňuje efektivnější rozdělení adres vzhledem k reálným počtům stanic na segmentech
  - nejmarkantnější u třídních adres
- část adresy původně určené pro identifikaci uzlu sítě se rozdělí na adresu „podsítě“ a na adresu uzlu v této podsíti
- dělit možno po bitech s ohledem na skutečné počty uzlů v jednotlivých segmentech a počet segmentů

# Maska podsítě (subnet mask)

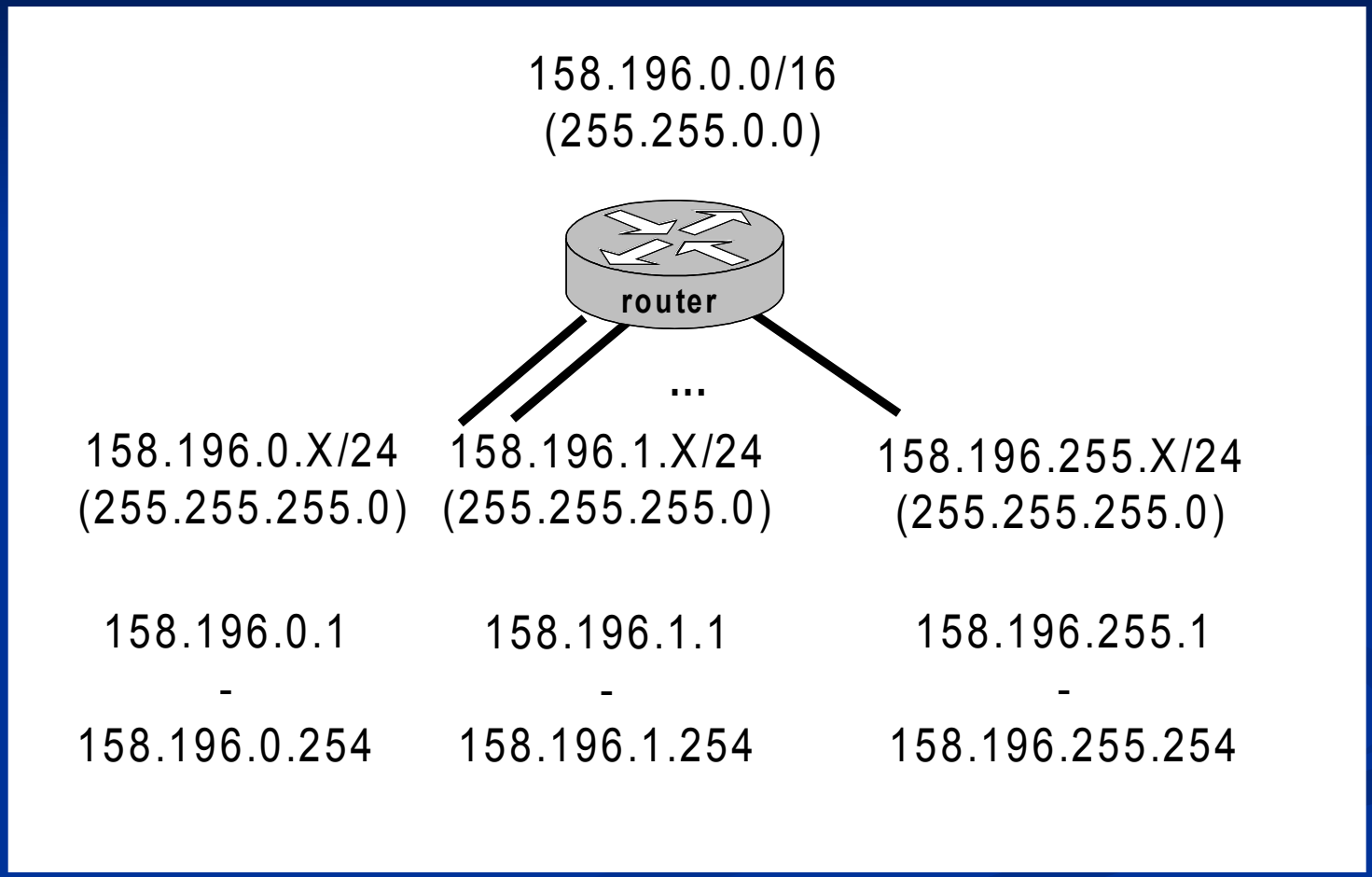



- pro každou (podsíťovanou) adresu nutno udat, kolik bitů zleva představuje sít+subsíť a kolik uzel.
- jednička na příslušném bitovém místě masky podsítě znamená, že odpovídající bit adresy patří do adresy sítě resp. podsítě, nula zařazuje bit do adresy uzlu

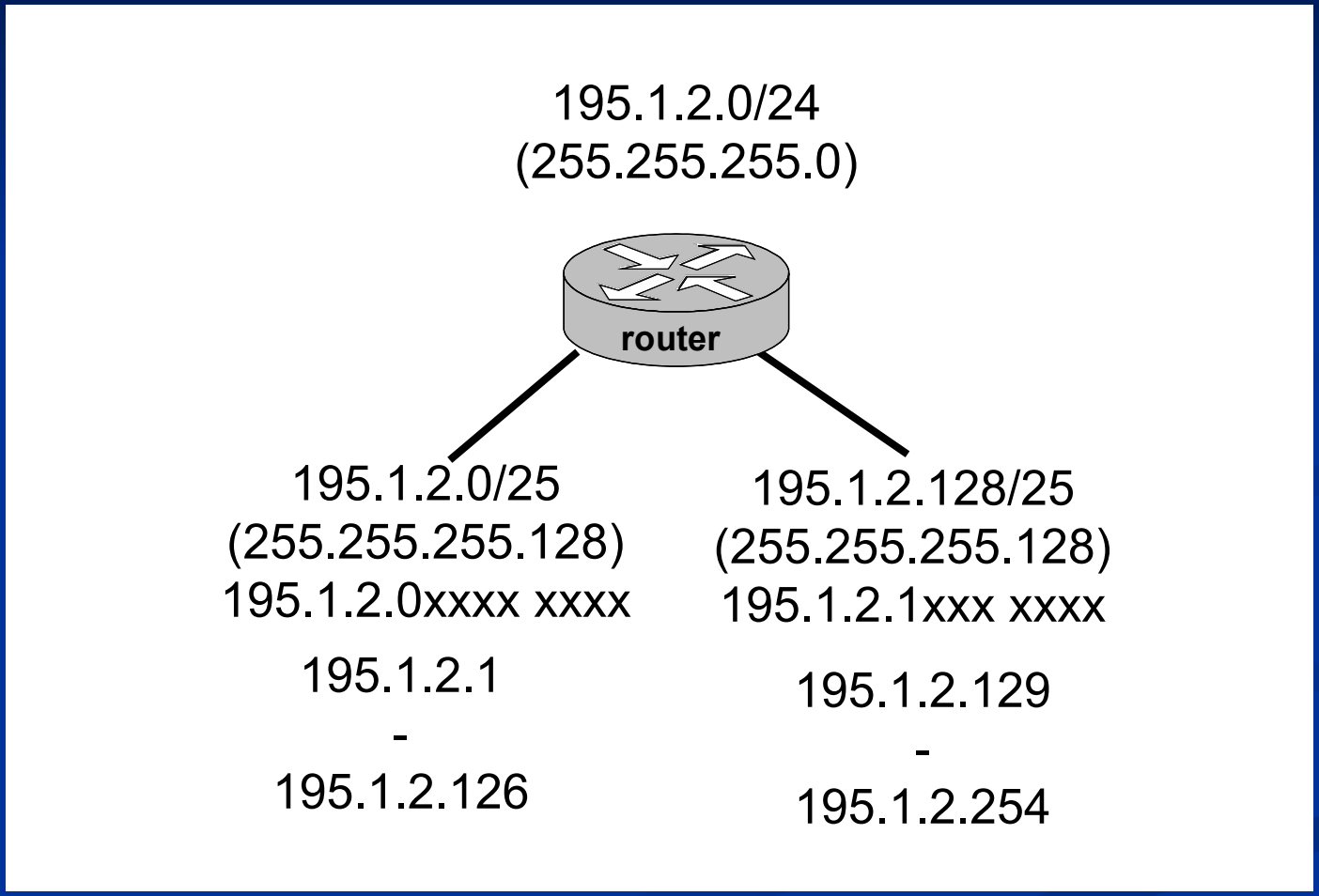
# Praktické použití podsíťování

- Rozdělení prefixu přidělené délky na daný počet podsítí (zadány maximální počty stanic na segmentech)
  - pozor na nepoužitelné adresy a adresu rozhraní routeru
- Stanovení maximální délky pevně přiděleného prefixu (požadovaného od ISP) pro požadovaný počet podsítí a požadované počty stanic na jednotlivých segmentech
- Vytvoření adresního plánu sítě WAN
  - zadaná topologie dvoubodových spojů, u jednotlivých směrovačů připojeny LAN, zadány požadované počty stanic na segmentech LAN

# Rozdělení rozsahu (1)

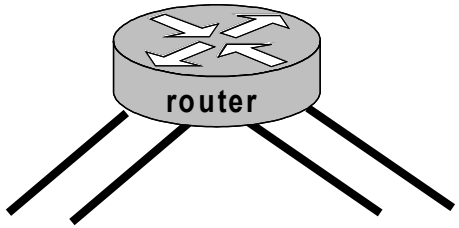


# Rozdělení rozsahu (2)



# Rozdělení rozsahu (3)

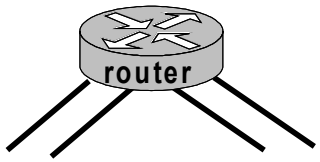
195.1.2.0/24
(255.255.255.0)




| 195.1.2.0/26 | 195.1.2.64/26 | 195.1.2.128/26 | 195.1.2.192/26 |
|---|---|---|---|
| (255.255.255.192) | (255.255.255.192) | (255.255.255.192) | (255.255.255.192) |
| 195.1.2.00xx xxxx | 195.1.2.01xx xxxx | 195.1.2.10xx xxxx | 195.1.2.11xx xxxx |
| 195.1.2.1 | 195.1.2.65 | 195.1.2.129 | 195.1.2.193 |
| - | - | - | - |
| 195.1.2.62 | 195.1.2.126 | 195.1.2.190 | 195.1.2.254 |

# Rozdělení rozsahu (4)

195.1.2.160/27
(255.255.255.224)
195.1.2.101x xxxx




195.1.2.160/29
(255.255.255.248)
195.1.2.1010 0xxx
195.1.2.161
-
195.1.2.166

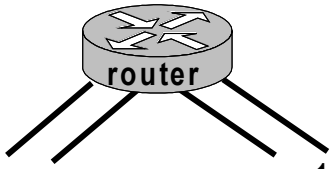
195.1.2.168/29
(255.255.255.248)
195.1.2.1010 1xxx
195.1.2.169
-
195.1.2.174

195.1.2.176/29
(255.255.255.248)
195.1.2.1011 0xxx
195.1.2.177
-
195.1.2.182

195.1.2.184/29
(255.255.255.248)
195.1.2.1011 1xxx
195.1.2.185
-
195.1.2.190

# Rozdělení rozsahu (5)

195.1.8.0/21
(255.255.248.0)
195.1.0000 1xxx. xxxx xxxx



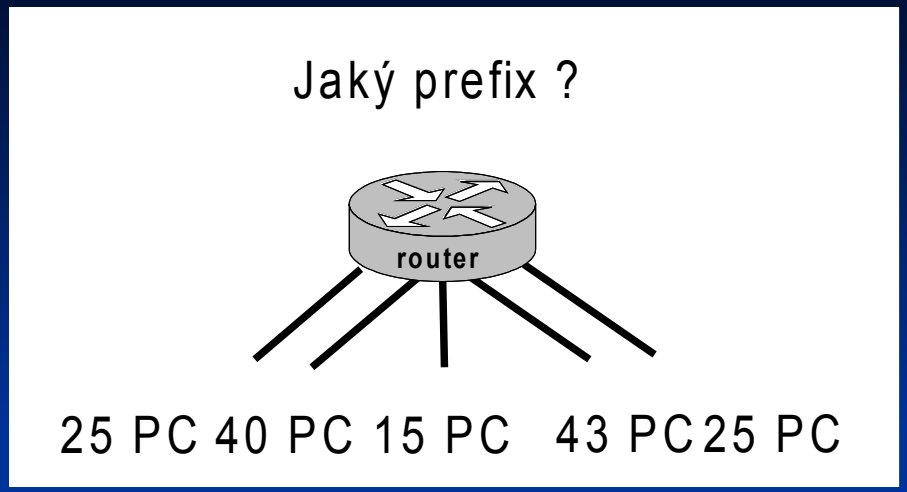

195.1.8.0/23
(255.255.254.0)
195.1.0000 100x. xxxx xxxx
195.1.8.1
-
195.1.9.254

195.1.12.0/23
(255.255.254.0)
195.1.0000 110x. xxxx xxxx
195.1.12.1
-
195.1.13.254

195.1.10.0/23
(255.255.254.0)
195.1.0000 101x. xxxx xxxx
195.1.10.1
-
195.1.11.254

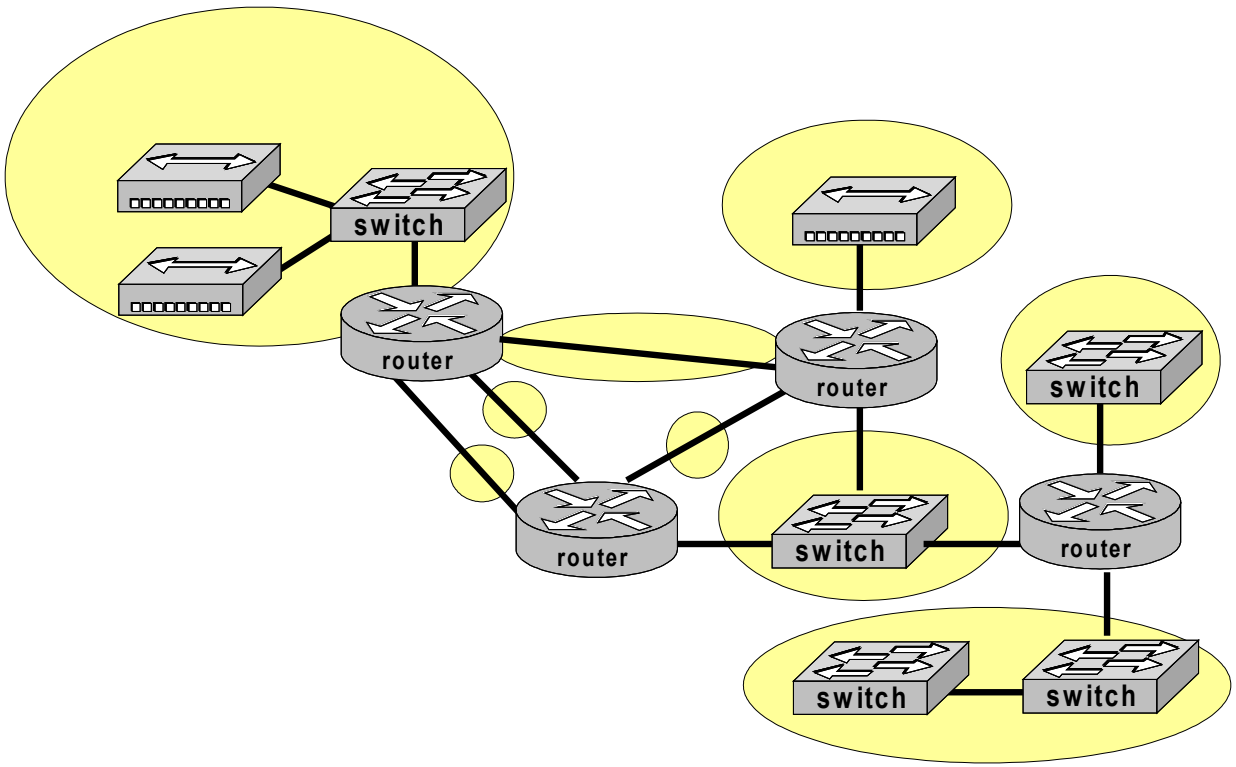
195.1.14.0/23
(255.255.254.0)
195.1.0000 111x. xxxx xxxx
195.1.14.1
-
195.1.15.254

# Jak dlouhý prefix vyžádat od ISP ?




- Maximální počet stanic 43
  - +1 adresa na rozhraní směrovače = 44
- K adresování 44 kombinací nutných 6b (64)
- 5 podsítí – k jejich adresování nutné 3b (8)
- Potřebujeme 6+3=9b, vyžádáme prefix 32-9-23b (/23)
  - Použijeme masku podsítě /26

# Adresní plán WAN




- Podsítě omezeny zařízeními pracujícími na 3. vrstvě OSI RM
  - směrovače, stanice – ne přepínače a rozbočovače
- Adresní prostor pro jednotlivé podsítě rozdělíme stejně jako v předchozím případě

# Podmínky pro podsíťování

- Minimální počet bitů pro adresu uzlu v podsíti je 2
  - Musíme umět zaadresovat podsíť jako takovou (adresa uzlu nuly) a všechny stanice na podsíti (adresa uzlu jedničky), takže maximální počet uzlů v podsíti je vždy o 2 menší, než odpovídá počtu bitů ponechaných pro adresu uzlu.
- Podsíť určená bitovou kombinací samých nul("subnet zero") se z historických (formálních) důvodů dříve nepoužívala, dnes se používá běžně.
  - Na některých směrovačích je nutné použití subnet zero explicitně povolit.

# Příklady podsíťovaných adres

- 151.88.19.103/255.255.255.0:
  třída B, podsíť 151.88.19 sítě 151.88.0.0, broadcast pro podsíť 151.88.19.255

- 151.88.19.103/255.255.255.224:
  třída B, 8 bitů předposledního a 3 bity posledního oktetu použito pro podsíť,
  podsíť 151.88.19.96 sítě 151.88.0.0, broadcast pro podsíť 151.88.19.127

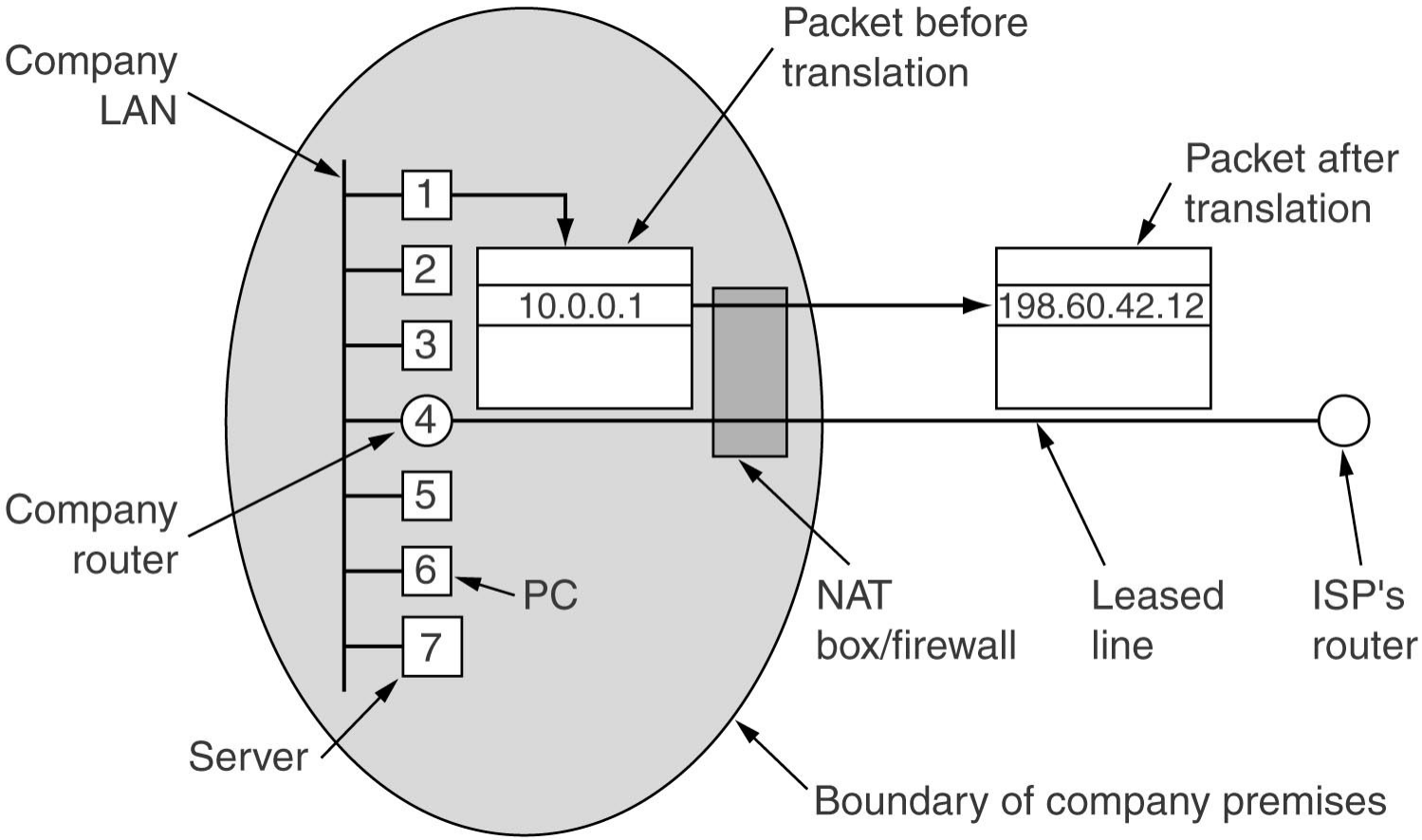
- 10.0.0.239/255.255.255.240:
  broadcast adresa na síti 10.0.0.224 (!)

# Překlad adres (NAT)

# Network Address Translation (NAT)

- překlad zdrojové nebo cílové IP adresy
  - probíhá na směrovačích/fierewallech (L3 prvcích)
- použití překladové tabulky
  - záznamy buďto konfigurovány staticky nebo se vytvářejí dynamicky automaticky
- typicky se překládá mezi "vnitřní" síti s privátními adresami a "vnější" síti s veřejnými (globálně jednoznačnými) adresami

# Scénář použití NAT



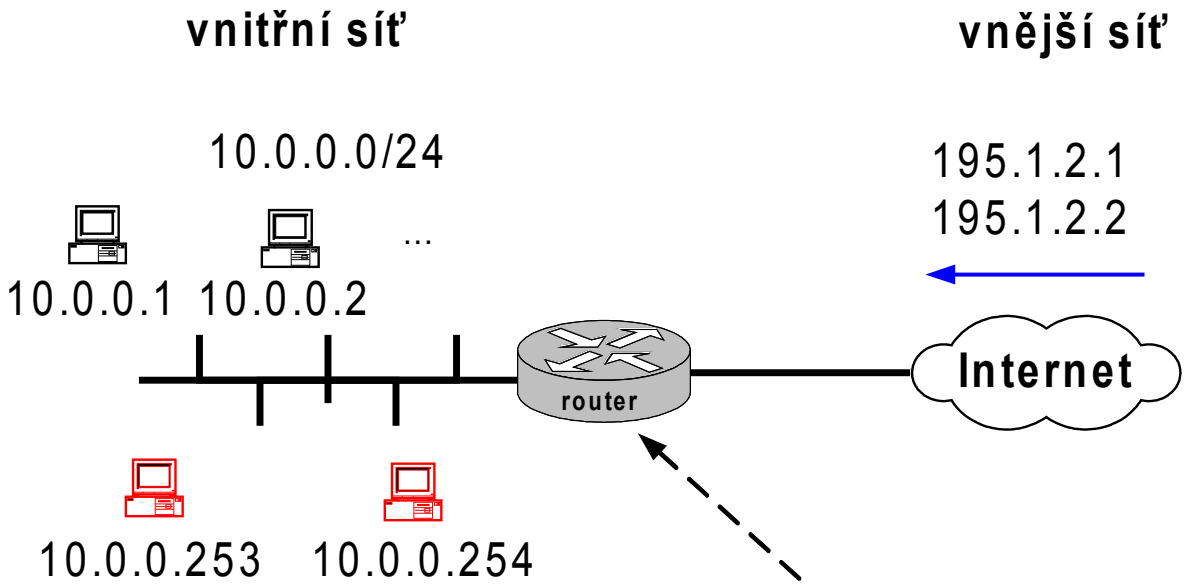
# Statický a dynamický NAT

- Statický NAT
  - překladová tabulka konfigurována staticky
- Dynamický NAT
  - překladová tabulka vzniká za provozu dynamicky
  - adresy se propůjčují z rezervoáru adres (pool)

# Způsob použití statického NAT

- statický překlad konkrétní zdrojové adresy vnitřní sítě (obvykle privátní) na konkrétní adresu směrovatelnou ve vnější síti
- statický překlad konkrétní cílové adresy (směrovatelné ve vnější síti) na konkrétní adresu vnitřní sítě (často privátní)

# Statický NAT - příklad



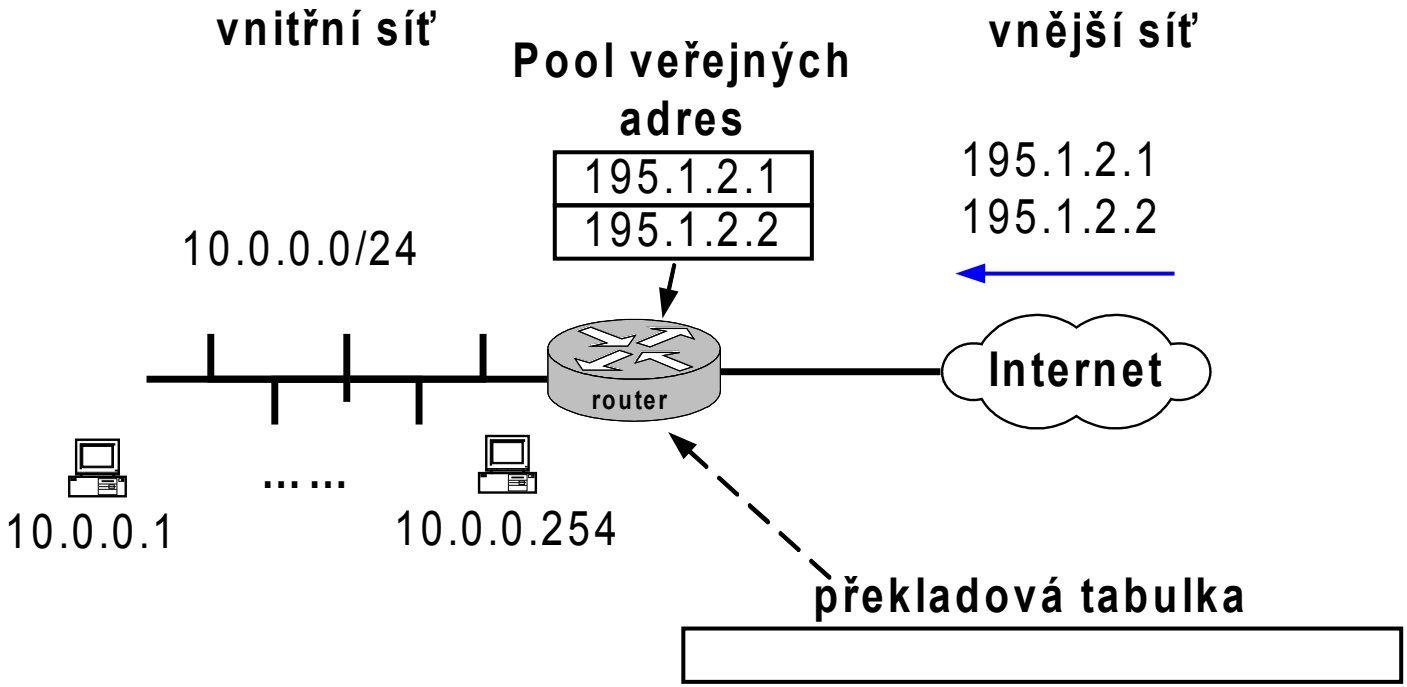

| |
|---|
| src 10.0.0.253 -> src 195.1.2.1 |
| src 10.0.0.254 -> src 195.1.2.2 |
| dst 195.1.2.1 -> dst 10.0.0.253 |
| dst 195.1.2.2 -> dst 10.0.0.254 |

**Statická překladová tabulka**

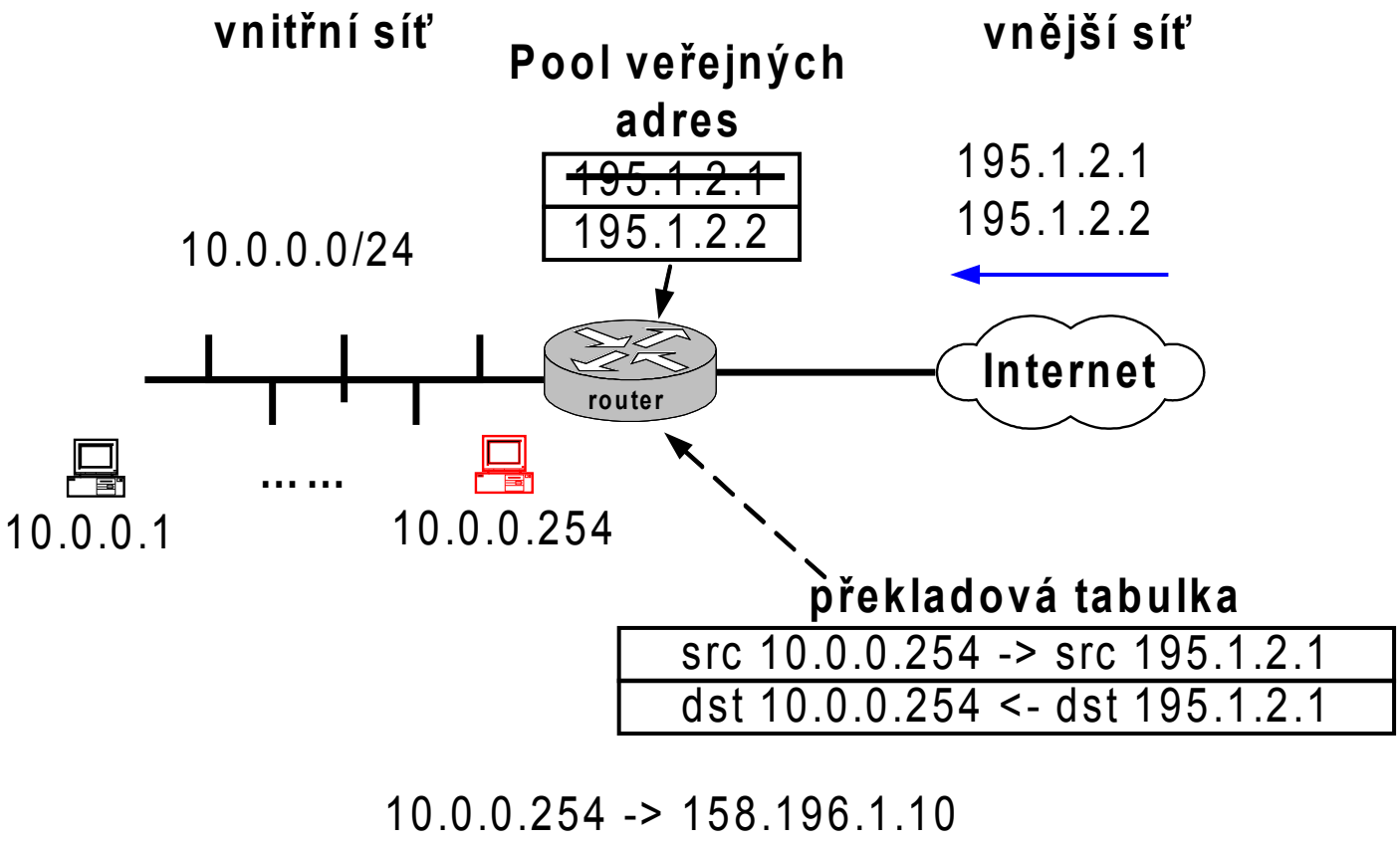
# Způsob použití dynamického NAT

- uživateli je přiděleno M veřejných adres
- uživatel chce ve vnitřní síti provozovat N>M strojů a umožnit jim přístup do vnější sítě (vždy nejvýše M strojům současně)
- dosud nevyužité veřejné adresy směrovač udržuje v poolu
- jestliže stanice S z vnitřní sítě pošle paket do vnější sítě, je jí dočasně přidělena  některá adresa V z poolu veřejných adres (pokud v něm nějaká zbývá)
  - v překladové tabulce se vytvoří záznam mapující IP adresu stanice S na adresu V
  - v odchozím paketu se přepíše (zdrojová) adresa stanice S na adresu V (ta je ve vnější síti jednoznačná a směrovatelná)
  - při příchodu odpovědi na adresu V se v překladové tabulce najde, že se cílová adresa V má přeložit na adresu S, což se provede a paket se odešle do vnitřní sítě

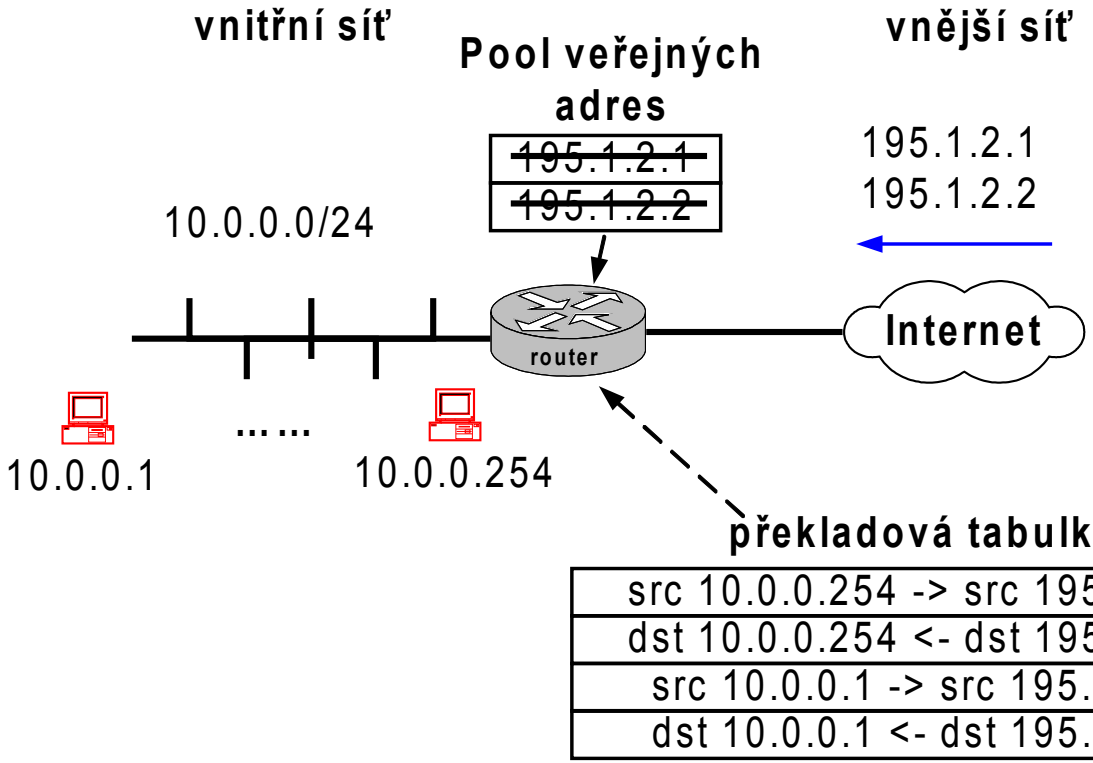
# Dynamický NAT –příklad (1)



# Dynamický NAT –příklad (2)



# Dynamický NAT – příklad (3)




10.0.0.1 -> 158.196.1.10

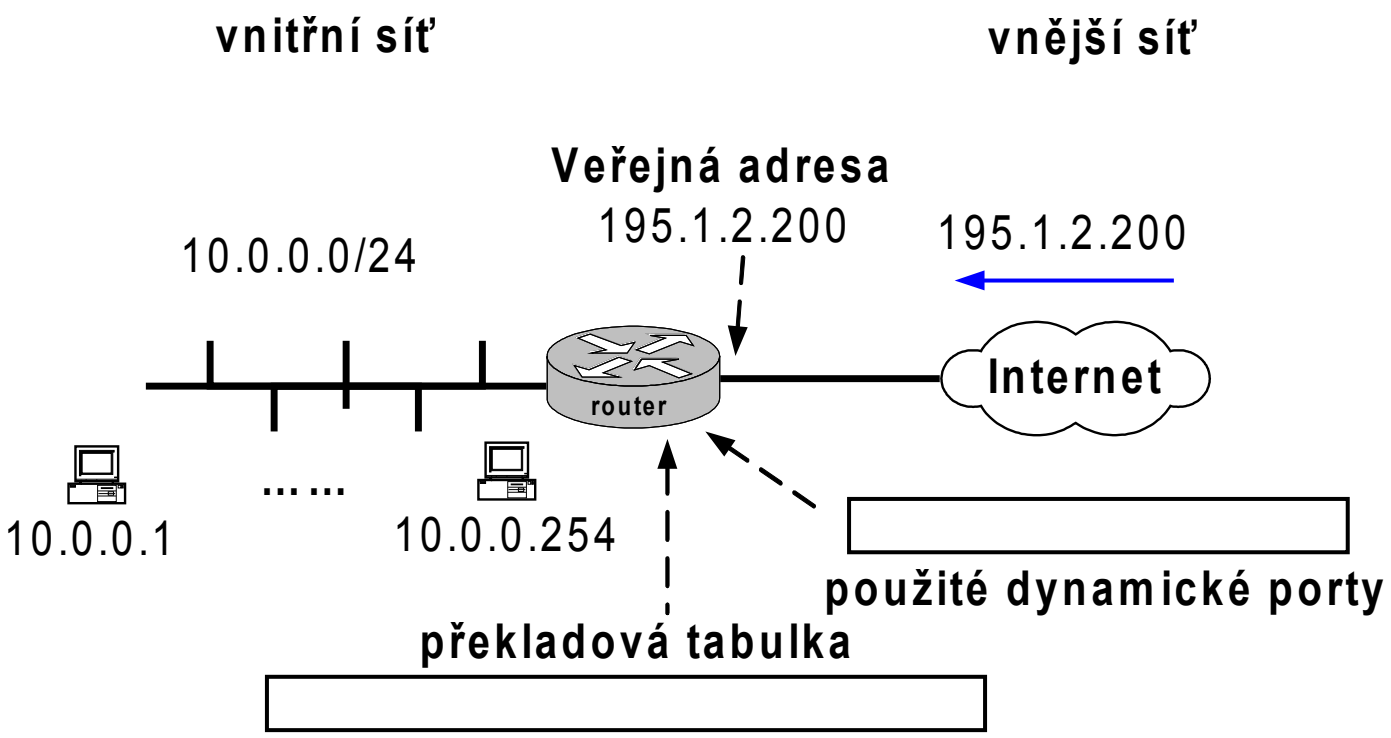
# Časové omezení dynamického NAT

- aby mohlo N strojů sdílet M adres, mají dynamicky vytvořené záznamy překladové tabulky časově omezenou platnost (timeout od posledního použití)
- při odstranění expirované položky se veřejná adresa vrátí zpět do poolu

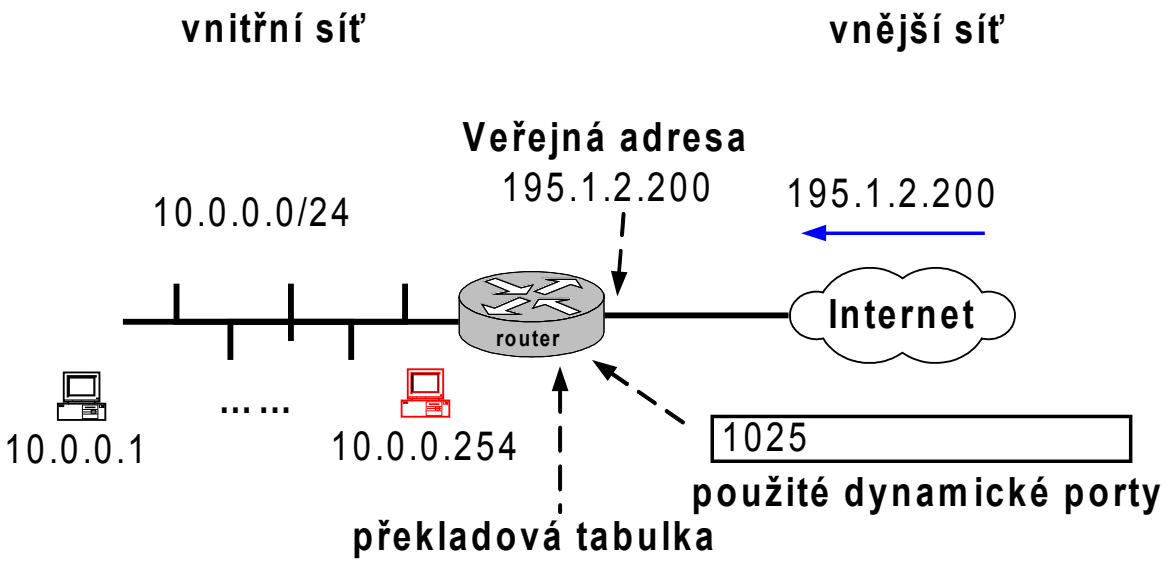
# Port Address Translation

- V terminologii Linuxu „Masquarading“
- Ukrytí více stanic za jedinou IP adresu, rozlišení pomocí různých zdrojových portů
  - Zdrojové porty přidělovány dynamicky, přičemž vzniká tabulka mapující jednotlivé porty na vnitřní IP adresy

# PAT – příklad (1)



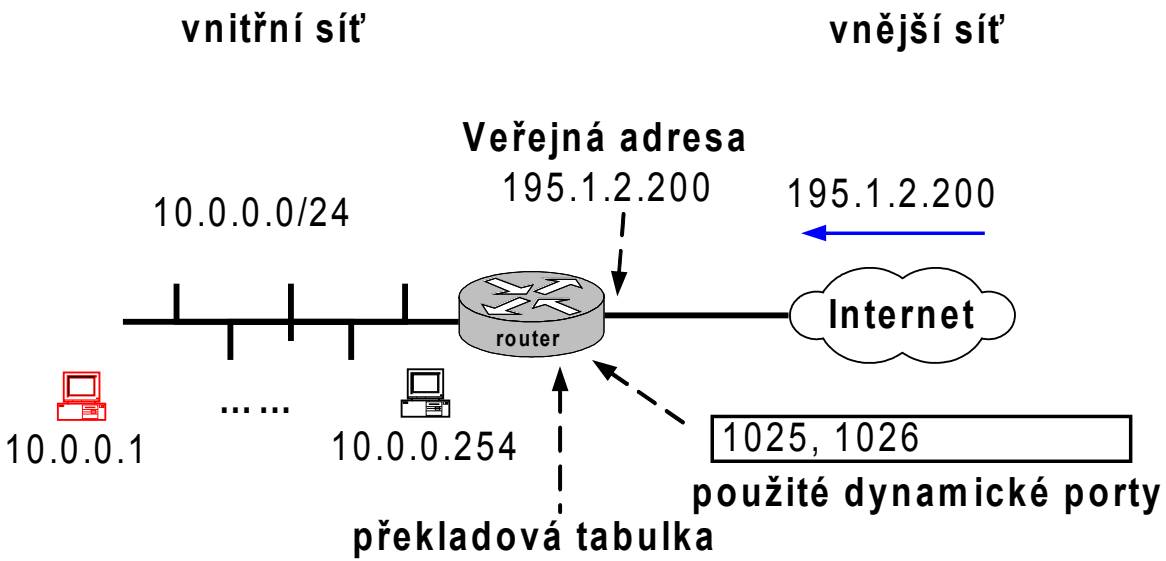
# PAT – příklad (2)




| src 10.0.0.254:2000 -> src 195.1.2.200:1025 |
| --- |
| dst 10.0.0.254:2000 <- dst 195.1.2.200:1025 |

10.0.0.254:2000 -> 158.196.1.10:80

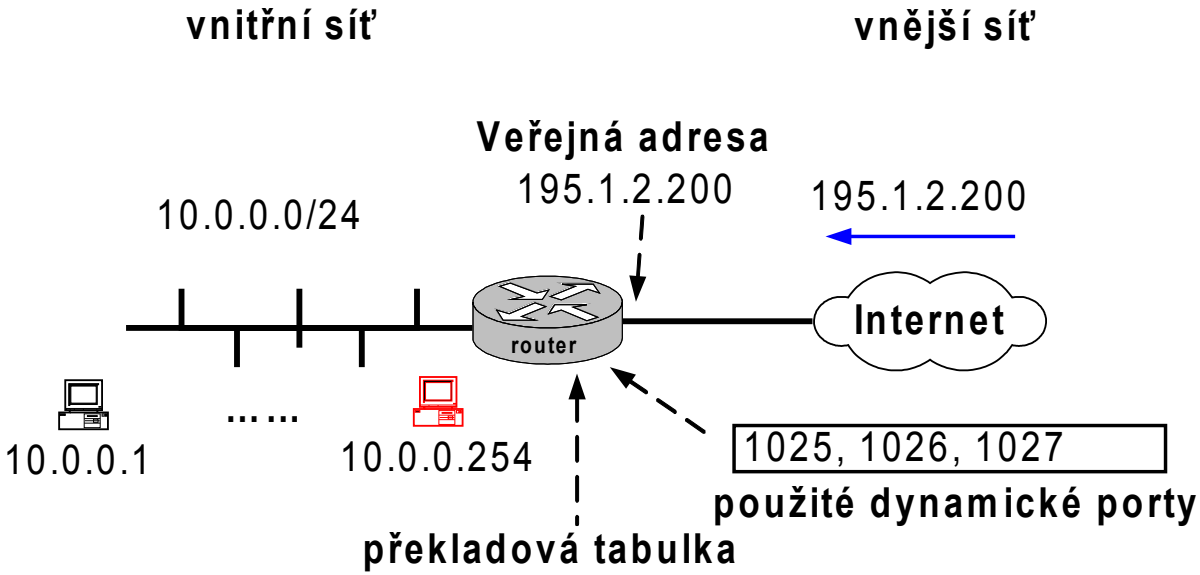
# PAT – příklad (3)




| |
|---|
| src 10.0.0.254:2000 -> src 195.1.2.200:1025 |
| dst 10.0.0.254:2000 <- dst 195.1.2.200:1025 |
| src 10.0.0.1:3000 -> src 195.1.2.200:1026 |
| dst 10.0.0.1:3000 <- dst 195.1.2.200:1026 |

10.0.0.1:3000 -> 158.196.1.10:80

# PAT – příklad (4)




| |
|---|
| src 10.0.0.254:2000 -> src 195.1.2.200:1025 |
| dst 10.0.0.254:2000 <- dst 195.1.2.200:1025 |
| src 10.0.0.1:3000 -> src 195.1.2.200:1026 |
| dst 10.0.0.1:3000 <- dst 195.1.2.200:1026 |
| src 10.0.0.254:2001 -> src 195.1.2.200:1027 |
| dst 10.0.0.254:2001 <- dst 195.1.2.200:1027 |

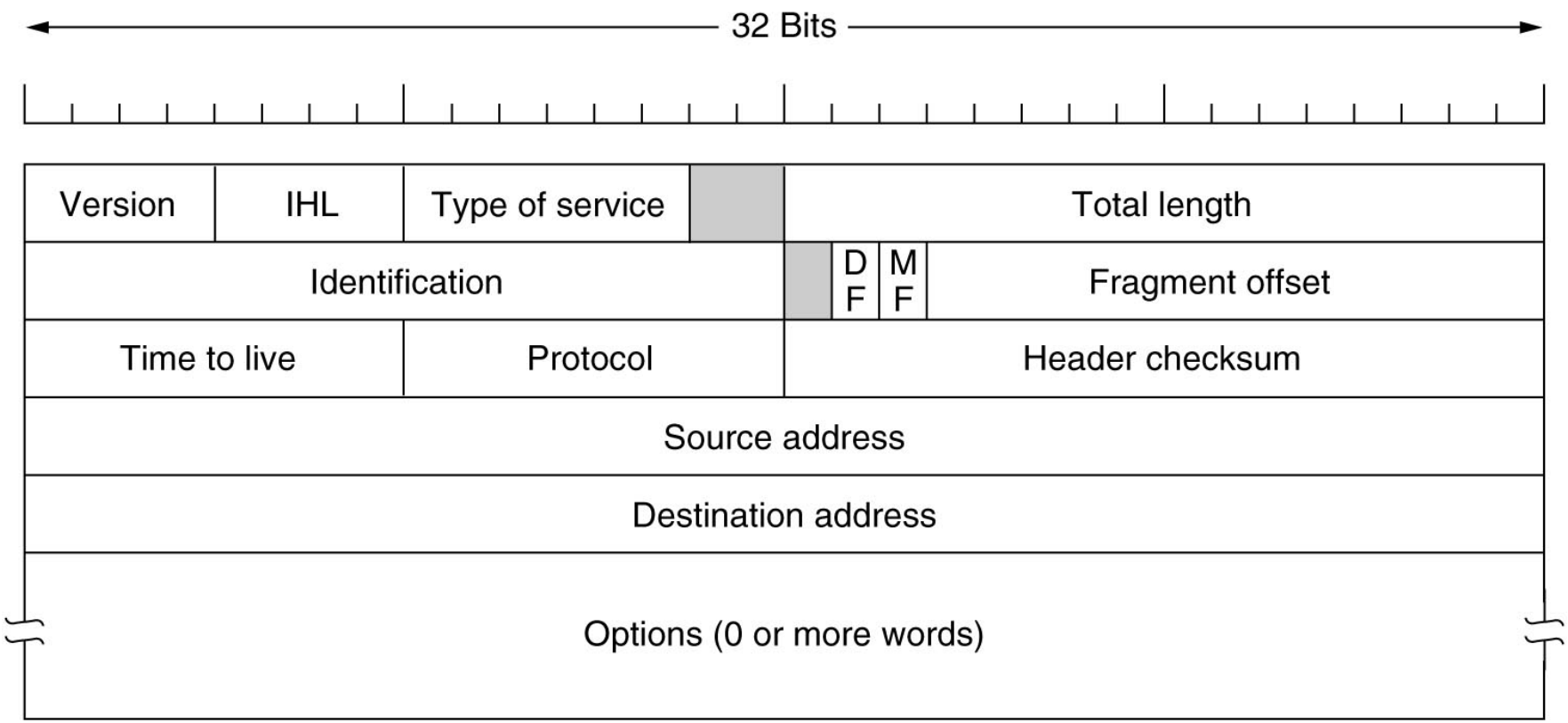
10.0.0.254:2001 -> 158.196.1.10:80

# Protokol IP

# IP - Internet Protocol

- 3. vrstva, síťová služba posílání nezávisle směrovaných paketů bez spojení
- RFC 791, 1042, 894, v současné době verze 4, chystá se verze 6

# Hlavička IP


32 Bits
Version
IHL
Type of service
Total length
Identification
D F
M F
Fragment offset
Time to live
Protocol
Header checksum
Source address
Destination address
Options (0 or more words)

# Fragmentace paketů

- Rozdělení paketu při průchodu linkami s nedostatečným MTU (Maximum Transfer Unit = max. délka datové části rámce)
  - fragmentace ve směrovačích nebo na zdroji
  - skládání až v cílovém uzlu
    - fragmenty mohou jít různými cestami
    - Skládání podle Identification, pořadí dle Fragment Offset, poslední fragment nemá nastaven More Fragments flag
- Podle konvence musí každý segment Internetu být schopen přenést paket o délce 576 B

# Podpůrné protokoly IP

# ARP - Address Resolution Protocol

- RFC 826, 1027
- mapování IP adres na MAC adresy
- Při potřebě zjistit MAC adresu k IP adrese se generuje ARP request (broadcast), ten obsahuje požadovanou IP adresu. Stanice s touto adresou odpoví svou MAC adresou (ARP reply).
- Zdroj ARP dotazu si výsledek uloží do ARP cache
  - (lokální cache jednotlivých stanic udržující známá mapování IP-MAC adres)
- Navíc se do requestu vkládá dvojice < zdrojová IP, zdrojová MAC >, každý počítač sleduje všechny ARP broadcasty a doplňuje informace ve své ARP cache.

# ICMP - Internet Control Message Protocol

- RFC 792
- protokol služebních řídících a informačních zpráv
- ohlašování chyb a zvláštních stavů při přenosu paketů
- šíří se v datové části IP paketů

# Zprávy ICMP (1)

- Echo request , echo reply
- Destination unreachable
  (network, host, port, protocol unreachable, zakázaná, ale nutná fragmentace)
  - + administratively prohibited
- Time exceeded (TTL=0 nebo vypršel čas pro refragmentaci)
- Redirect
- Parameter problem

# Zprávy ICMP (2)

**Novější (a ne vždy podporované) zprávy**

- Source quench - žádost cílové stanice o snížení rychlosti generování zpráv zdrojem (přeplňují se buffery)
- Address mask request, Address mask reply - zjištění síťové masky rozhraní
- Router solicitation, Router advertisement

# Zjišťování cesty sítí - traceroute

- většina OS
- zjištění všech směrovačů na cestě k cílové stanici
- využívá pole TTL, začíná se od 1, stále se zvyšuje, sledují se IP adresy, ze kterých přijde ICMP Time Exceeded
- testovací paket buďto ICMP (Microsoft) nebo UDP na neexistující port (Unix)

# Transportní vrstva TCP/IP: UDP a TCP

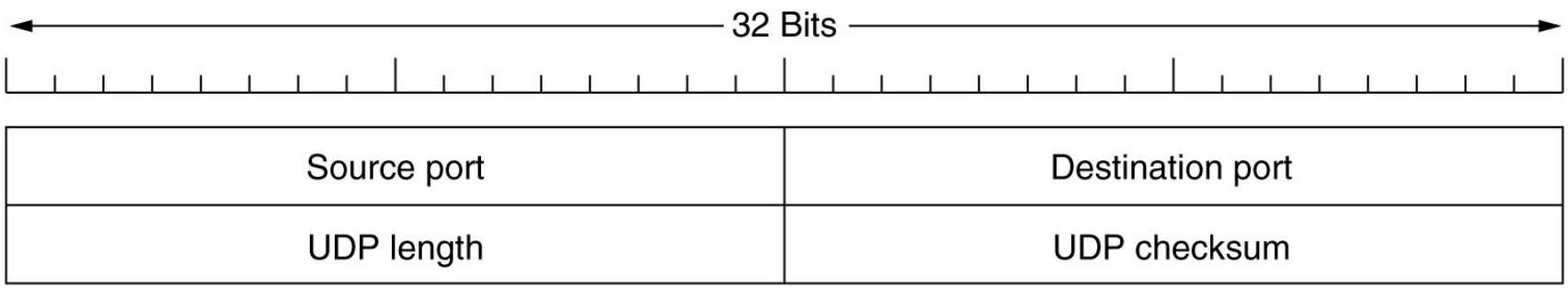
# Porty

- Spolu s IP adresou identifikují konkrétní proces (službu) na konkrétním zařízení v Internetu
  - (transportní entitu)
- 16bit (0-65535), zvlášť pro TCP a UDP
  - 0-1023: Veřejně definované služby (well-known)
  - >1024 (4096) – klientské porty, obvykle přidělování volných portů operačním systémem
- Vždy uveden cílový i zdrojový port

# UDP - User Datagram Protocol

- nepotvrzovaná datagramová služba
- podpora všesměrového a skupinového vysílání (na daném portu)
- porty identifikují proces odesílatele, resp. příjemce na vysílající, resp. přijímající stanici
- kontrolní součet zahrnuje datovou část (na rozdíl od IP, tam jen hlavičku)
  - není však povinný

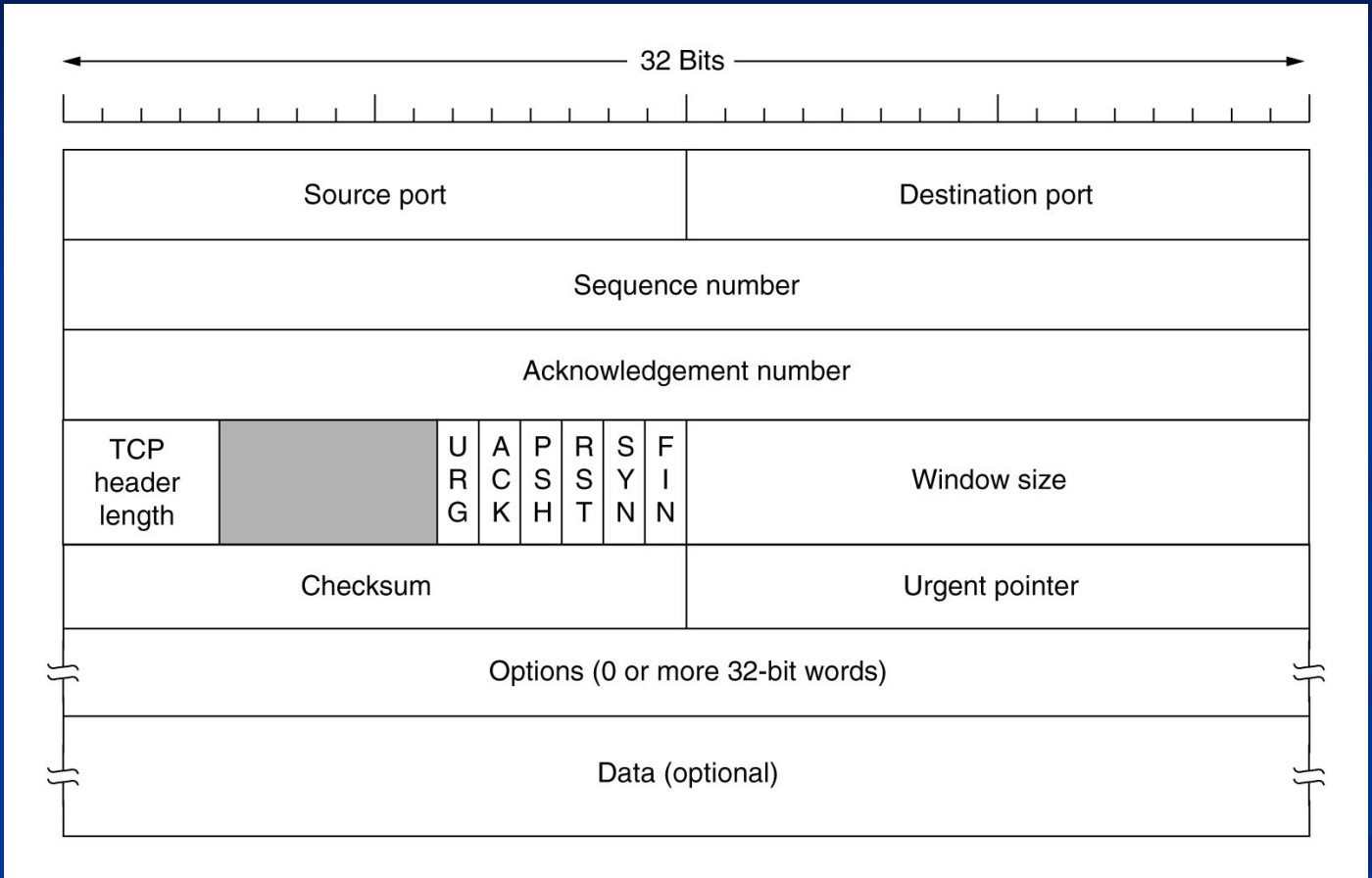
# (Pseudo)hlavička UDP



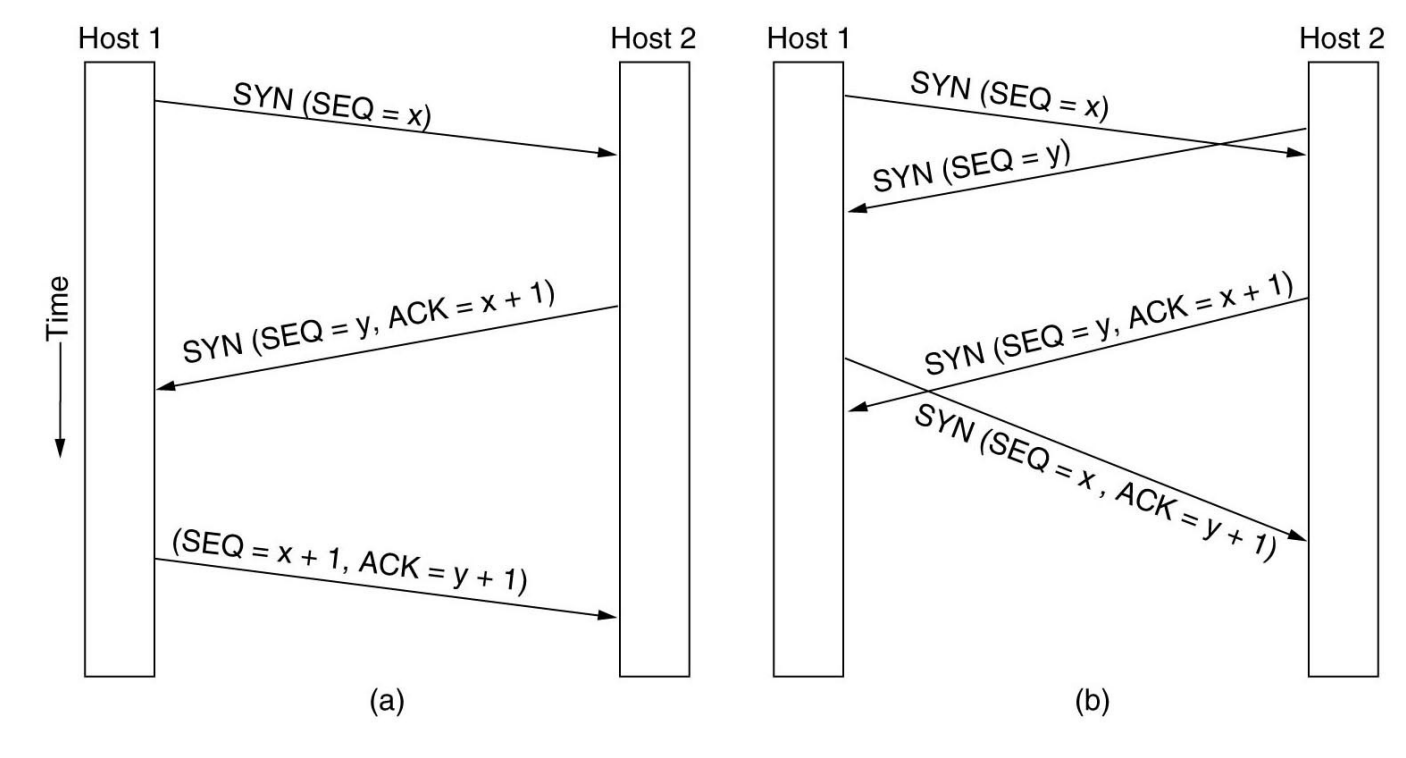
# TCP: Transmission Control Protocol

- duplexní spolehlivý logický kanál
  - v prostředí se ztrácením, duplikací, a přehazováním pořadí
- segmentování dat (rozdělení proudu dat do částí vhodných pro přenos v paketech), číslování oktetů proudu dat
- algoritmus Sliding window (go-back-N), pozitivní (inkluzivní) potvrzování, piggybacking, adaptivní změna časového limitu pro retransmisi
- řízení toku dat inzerováním aktuální kapacity přijímacích bufferů, vysílací okno se dynamicky přizpůsobuje přijímacímu
- robustní protokol navazování spojení a ukončování spojení

# (Pseudo)hlavička TCP



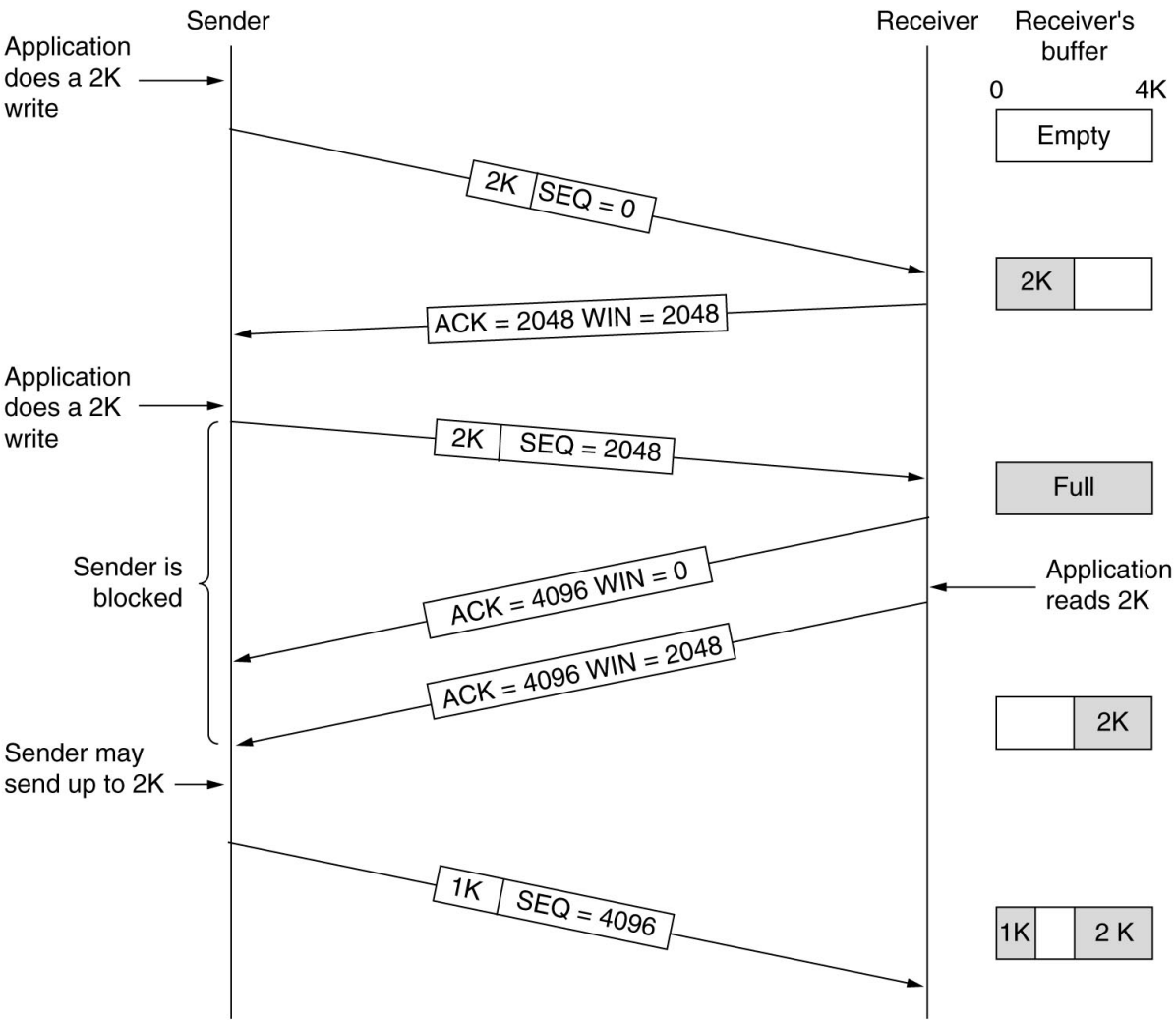
# Navazování TCP spojení (1)



# Navazování TCP spojení (2)

- three way handshake: SYN, SYN+ACK, ACK
  - dohoda o startovacím sekvenčním čísle (zvlášť pro oba směry)
  - počáteční sekvenční čísla náhodná, aby se zabránilo případnému ovlivnění zbloudilými pakety ze zavřeného a poté brzy opět znovu otevřeného spojení mezi týmiž entitami
- řeší i problémy pokusu o aktivní navázání spojení oběma stranami současně

# Průběh TCP spojení - řízení toku dat



# Uzavření spojení

- Uzavírá se zvlášť z obou stran
  - Možnost "polovičního" uzavření spojení (half-close)
  - FIN+ACK z obou stran
- První může uzavřít kterákoli strana