

# Routing pre ISP

Od L2 siete k routovanej infraštruktúre



Telekomunikačná únia  
Slovenskej republiky

**Radovan Jakubko**

# Switching verzus routing

|                 | Switching (L2)           | Routing (L3)                   |
|-----------------|--------------------------|--------------------------------|
| Rozhoduje podľa | Cieľovej MAC adresy      | Cieľovej IP adresy a prefixu   |
| Prepája         | Zariadenia v jednej VLAN | Rozličné IP siete              |
| Broadcast       | Šíri v rámci VLAN        | Oddeluje broadcast domény      |
| Tabuľka         | MAC adresa a port        | Prefix a ďalší smer (next-hop) |

**Aj L3 switch môže routovať. Rozhoduje funkcia, nie názov zariadenia.**

# Prečo má ISP výhodu v routovanej sieti

## Menší dosah poruchy

Slučka alebo broadcast búrka zostáva v menšej L2 doméne.

## Využitie záložných liniek

ECMP (Equal-Cost Multi-Path) umožňuje súbežne využívať rovnocenné cesty.

## Jednoduchšie škálovanie

Adresný plán a sumarizácia obmedzujú rozsah tabuliek.

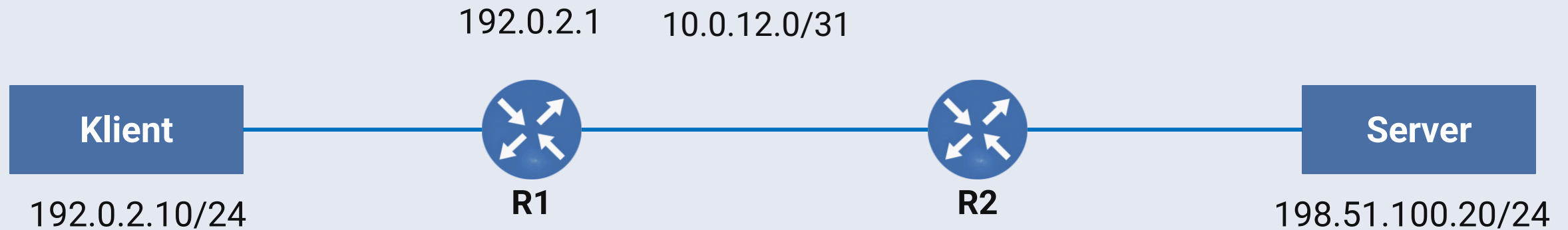
## Kontrola prevádzky

Operátor vie určiť cestu, politiku a hranice služieb.

**L2 má miesto v prístupe. Medzi lokalitami je routing často lepší základ.**

# Ako router doručí paket do inej siete

Cieľ paketu: 198.51.100.20



Klient použije bránu

R1 vyberie next-hop R2

R2 doručí do cieľovej siete

Každý router rozhoduje samostatne podľa svojej tabuľky.

# Routovacia tabuľka a longest-prefix match

Cieľová IP adresa: 10.20.30.42

| Prefix               | Next-hop           | Zhoduje sa?                                              |
|----------------------|--------------------|----------------------------------------------------------|
| 0.0.0.0/0            | Brána do internetu | Áno                                                      |
| 10.20.0.0/16         | R2                 | Áno                                                      |
| <b>10.20.30.0/24</b> | <b>R3</b>          | <b>Áno, najdlhší prefix /<br/>najmenší rozsah adries</b> |

**Paket pôjde cez R3. /24 je konkrétnejšie než /16 aj /0.**

V reálnom zariadení paket používa FIB (Forwarding Information Base).

# Statický a dynamický routing

## Statický routing

Cestu zadá správca.

Vhodný pre jednoduchú prípojku, default route alebo malú sieť.

Zálohovanie potrebuje detekciu poruchy.

**Oba prístupy sa bežne kombinujú.**

## Dynamický routing

Routre si vymieňajú informácie.

Pri zmene dostupnosti aktualizujú cesty.

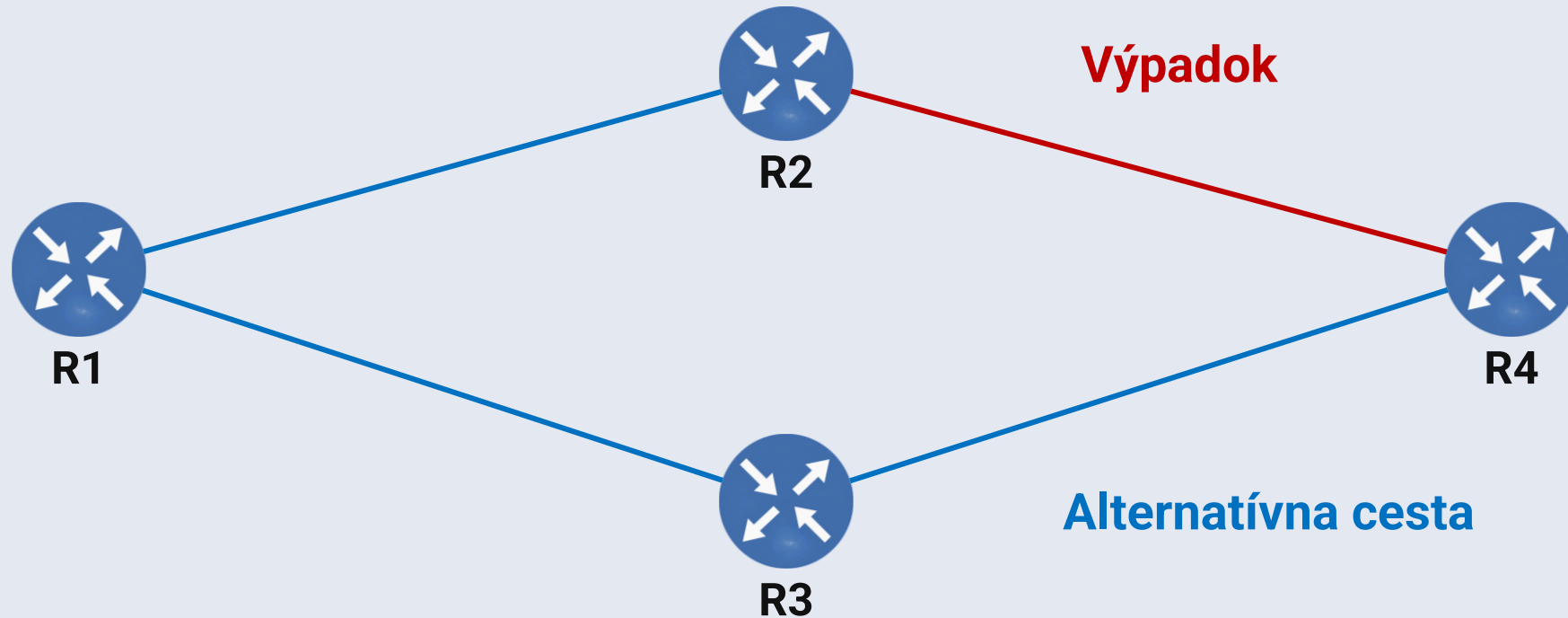
Vyžaduje správny návrh a filtre.

# Výpadok linky, konvergencia a ECMP

1 Detekcia poruchy

2 Aktualizácia ciest

3 Zmena preposielania



Rýchlosť obnovy závisí aj od detekcie poruchy a zápisu do FIB.

# RIP, OSPF, IS-IS a BGP

| Protokol | Princíp         | Výber cesty           | Úloha                     |
|----------|-----------------|-----------------------|---------------------------|
| RIP      | Distance-vector | Počet hopov, max. 15  | Malé alebo staršie siete  |
| OSPF     | Link-state      | Najnižšia suma cost   | IGP vo vlastnej sieti     |
| IS-IS    | Link-state      | Najnižšia suma metrík | IGP v infraštruktúre      |
| BGP      | Path-vector     | Politika a atribúty   | Internet aj interný návrh |

**IGP (Interior Gateway Protocol) rieši dosiahnuteľnosť infraštruktúry. BGP pridáva smerovaciu politiku.**

# OSPF a IS-IS vo vnútri ISP

## OSPF

Link-state a výpočet SPF. Vznikol pre IP siete.

Hierarchia oblastí s backbone Area 0.

OSPFv2 pre IPv4, OSPFv3 aj pre IPv6.

Približne 50 routerov na oblasť bolo historické odporúčanie, nie pevný limit. Platí aj pre Area 0.

Veľkosť oblasti závisí od výkonu routerov, topológie a požadovanej konvergenencie.

## IS-IS

Link-state a výpočet SPF. Pôvodne pre OSI siete, neskôr rozšírený o IP.

Nevznikol ako reakcia na limity OSPF.

Level 1: smerovanie v oblasti. Level 2: chrbtica medzi oblasťami.

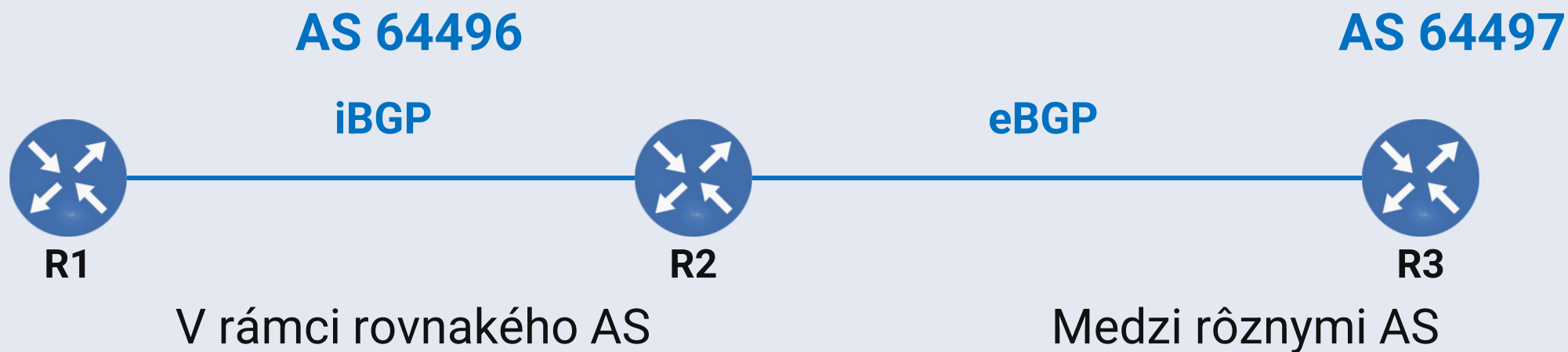
Oblasť môže byť viac. Celá sieť môže fungovať aj ako level-2-only.

Rozšírenia TLV pre IPv4 aj IPv6.

**Voľbu ovplyvňuje platforma, rozsah siete a skúsenosť tímu.**

# BGP, autonómny systém, eBGP a iBGP

AS je doména so spoločnou smerovacou politikou a číslom ASN.



BGP prenáša prefixy a atribúty. AS\_PATH pomáha odhaliť slučku.

# BGP vyberá cestu podľa politiky

Rovnaký cieľový prefix dostupný cez dva výstupy

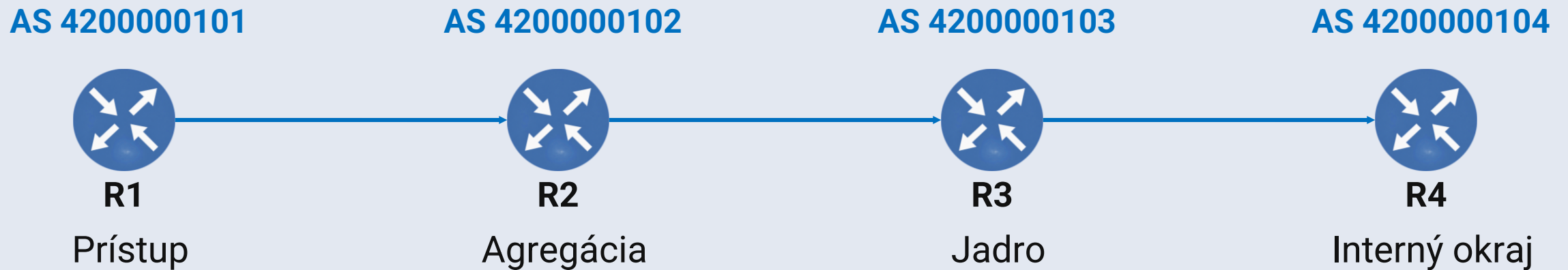
| Príklad výstupu | LOCAL_PREF | Dĺžka AS_PATH |
|-----------------|------------|---------------|
| Peering         | 200        | 3 AS          |
| Transit         | 100        | 2 AS          |

Pri ostatných rovnakých podmienkach vyhrá vyšší LOCAL\_PREF.

Kratší AS\_PATH sa uplatní až pri rovnosti skorších kritérií.

# Hop-by-hop eBGP s privátnymi 32-bitovými AS

BGP môžeme použiť aj na vnútorné smerovanie a vo vhodnom návrhu ním nahradiť IGP.



**R1 oznamuje 10.20.0.0/24**

R4 prijme AS\_PATH: 4200000103 4200000102 4200000101

**Každý hop má eBGP susedstvo. Hraničný router musí odstrániť privátne ASN-ka.**

Privátne 32-bitové ASN: 42000000000–4294967294 (RFC 6996).

# Praktický návrh routovanej siete ISP

|                          |                                                            |
|--------------------------|------------------------------------------------------------|
| <b>Prístup</b>           | Lokálne VLAN a jasná L3 hranica pre zákazníkov.            |
| <b>Infraštruktúra</b>    | Point-to-point linky, loopback adresy a záložné trasy.     |
| <b>Smerovanie</b>        | IGP + BGP alebo premyslený interný eBGP návrh.             |
| <b>Internetový okraj</b> | Explicitné filtre, limity prefixov a kontrola oznamovania. |

**Výpadok aj obnovu treba overiť na reálnej platforme.**

# Komunikácia routovacích protokolov so susedmi

| Protokol      | Ako prenáša správy                           | Bežná cieľová adresa                                     | Vytvorenie susedstva                                              |
|---------------|----------------------------------------------|----------------------------------------------------------|-------------------------------------------------------------------|
| <b>RIPv2</b>  | UDP port 520                                 | Multicast 224.0.0.9                                      | Nemá formálne nadväzovanie susedstva cez Hello                    |
| <b>OSPFv2</b> | Priamo cez IPv4, protokol 89                 | Multicast 224.0.0.5 a 224.0.0.6                          | Automaticky pomocou Hello                                         |
| <b>OSPFv3</b> | Bežne priamo cez IPv6, Next Header 89        | Multicast FF02::5 a FF02::6                              | Automaticky pomocou Hello                                         |
| <b>IS-IS</b>  | Priamo cez linkovú vrstvu, bez IP, TCP a UDP | Na ethernetovej LAN multicastové MAC adresy podľa úrovne | Automaticky pomocou IS-IS Hello                                   |
| <b>BGP</b>    | TCP port 179                                 | Unicast IP adresa konkrétneho suseda                     | Spravidla nakonfigurovaný sused, potom TCP spojenie a správy OPEN |

# MPLS v skratke

MPLS preposiela paket podľa krátkeho labelu.



**L3VPN: oddelené IP siete zákazníkov**

**L2VPN: ethernetová služba cez jadro operátora**

MPLS samo osebe dáta nešifruje.

# Čo si odniešť do vlastnej siete

**Routing zmenšuje dosah L2 porúch  
a umožňuje využiť viac ciest.**

**OSPF / IS-IS riešia vnútornú dosiahnuteľnosť.  
BGP riadi výmenu prefixov a smerovaciu politiku.**

MPLS pridáva služby nad infraštruktúrou.

## Otázky?

# Záloha: kde sú limity RIP

| Vlastnosť                                 | Praktický dôsledok                               |
|-------------------------------------------|--------------------------------------------------|
| Metrika je počet hopov                    | Kapacita a latencia linky sa priamo nezohľadnia. |
| 15 hopov, hodnota 16 znamená nedostupnosť | Obmedzená veľkosť routovacej domény.             |
| Periodické a vyvolané aktualizácie        | Zmena topológie môže konvergovať postupne.       |
| Distance-vector                           | Split horizon a poisoned reverse tlmia slučky.   |

**Pre novú ISP infraštruktúru je RIP spravidla nevhodná voľba.**

# Záloha: cesta je v tabuľke, no paket neprejde

## Prefix a FIB

Je zvolená správna cesta a naozaj nainštalovaná?

## Next-hop

Je sused dosiahnuteľný a funguje ARP / ND?

## Návratová cesta

Vie cieľová strana odpovedať späť?

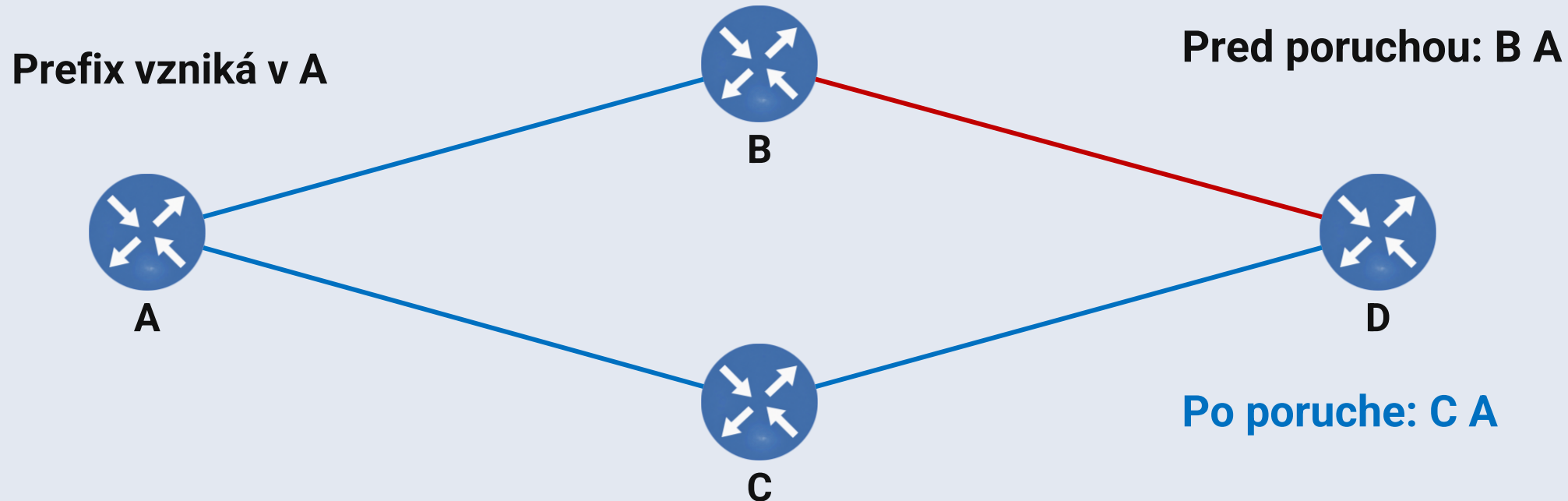
## Prenos a politika

Nezastaví paket MTU, ACL, firewall alebo uRPF?

**Overenie z oboch strán odlíši problém smerovania od filtrácie.**

# Záloha: eBGP a alternatívna cesta

Skrátenie v schéme: A, B, C, D označuje samostatné privátne AS.



Vlastné ASN v AS\_PATH štandardne vedie k odmietnutiu cesty.

# Záloha: BGP filtre a ochrana pred route leakom

## Import aj export

Každé eBGP susedstvo má explicitnú politiku.

## Prefix-list a max-prefix

Povolíme očakávané siete a obmedzíme ich počet.

## RPKI origin validation

Kontroluje pôvod prefixu, nie celý AS\_PATH.

## Verejný internet

Privátne prefixy a privátne ASN zostávajú mimo exportu.

**Fungujúce BGP susedstvo ešte nepotvrdzuje správnu politiku.**

# Záloha: L3VPN a L2VPN nad MPLS

## L3VPN

ISP routuje zákaznícke IP prefixy vo VRF.

Zákazníci môžu mať prekrývajúce sa adresy.

Jadro nemusí poznať všetky VPN prefixy.

## L2VPN

ISP prenáša ethernetovú službu medzi lokalitami.

Zákazník si riadi vlastný routing.

Rozsah L2 poruchy závisí od návrhu služby.

**Výber služby závisí od toho, kto má riadiť zákaznícky routing.**