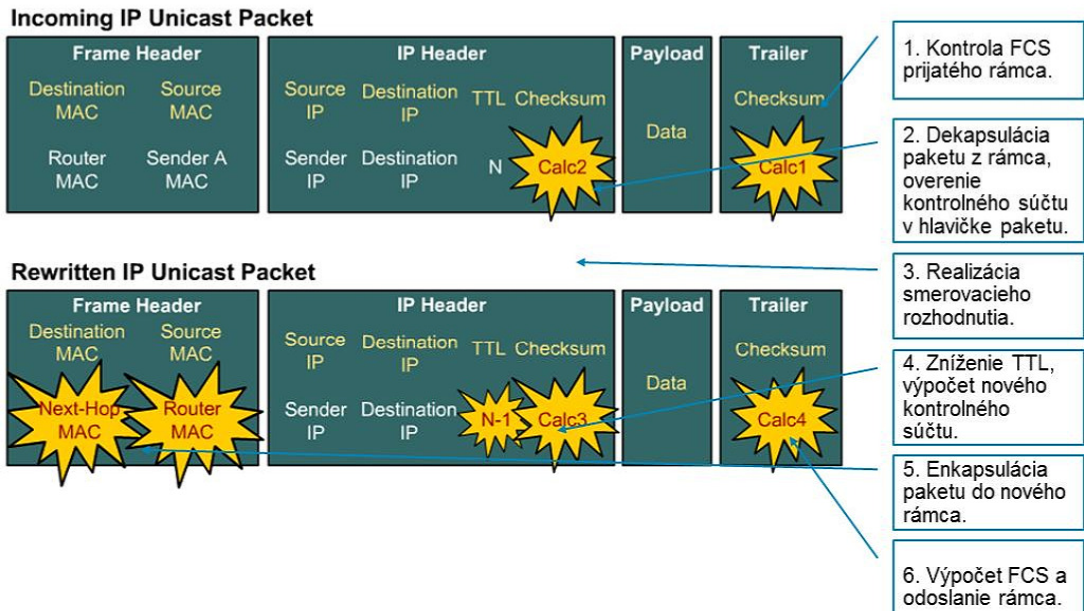
Stretnutie 4:

Inter VLAN routing pomocou L3 smerovača a Multilayer prepínača

SWITCH Modul 4

Ako vyzerá paket pred a po smerovaní?



Z akých krokov sa skladá unicastové smerovanie v IPv4 sieťach?

1. Má rámec korektnú veľkosť a je jeho FCS správne?
   - Ak áno, pokračujeme ďalšími krokmi
   - Ak nie, rámec zahodíme bez pokračovania
2. Má hlavička IP paketu správnu kontrolnú sumu?
   - Ak áno, pokračujeme ďalšími krokmi
   - Ak nie, paket zahodíme bez pokračovania
3. Je paket podľa IP adresy príjemcu určený pre lokálnu IP adresu?
   - Ak áno, patrí samotnému routeru – nebudeme ho smerovať
   - Ak nie, pokračujeme ďalšími krokmi
4. Je hodnota TTL v hlavičke IP paketu väčšia ako 1?
   - Ak áno, pokračujeme ďalšími krokmi
   - Ak nie, paket zahodíme a odosielateľa upozorníme ICMP správou

Z akých krokov sa skladá unicastové smerovanie v IPv4 sieťach?

5. K IP adrese príjemcu vyhľadáme v smerovacej tabuľke vyhovujúci záznam
   - Smerovacia tabuľka je usporiadaná zostupne podľa masiek
   - Hľadáme prvý riadok tabuľky, v ktorom platí:
     IP adresa príjemcu & Sieťová maska = Číslo siete
   - Ak nenájdeme zhodu, paket zahodíme a odosielateľa upozorníme ICMP správou
6. Ukazuje vyhľadaný riadok smerovacej tabuľky na východzie rozhranie, ktorým má paket odísť?
   - Ak áno, pokračujeme ďalším krokom
   - Ak nie, potom obsahuje IP adresu next hop routera. Tú si zapamätáme a s ňou sa vrátime na krok 5
7. K výstupnému rozhraniu a naposledy zapamätanej IP adrese next hop routera vyhľadáme Layer2 informáciu pre vytvorenie rámca
   - ARP tabuľka pre Ethernet, mapovacia tabuľka pre Frame Relay, ...

Z akých krokov sa skladá unicastové smerovanie v IPv4 sieťach?

8. V IP pakete dekrementujeme pole TTL a prepočítame kontrolnú sumu
9. IP paket enkapsulujeme do príslušného rámca pomocou informácií vyhľadaných v kroku 7 a odošleme rozhraním vyhľadaným v kroku 6

Z akých krokov sa skladá unicastové smerovanie v IPv4 sieťach?

8. V IP pakete dekrementujeme pole TTL a prepočítame kontrolnú sumu
9. IP paket enkapsulujeme do príslušného rámca pomocou informácií vyhľadaných v kroku 7 a odošleme rozhraním vyhľadaným v kroku 6

- Uvedený proces sa opakuje pre každý paket
- Takýto spôsob smerovania sa nazýva „**Process switching**“ a je najpomalším spôsobom smerovania
- Rýchlosť spracovania paketu pri smerovaní závisí od architektúry smerovača a od zaťaženia CPU
- Veľký objem smerovaných paketov vedie na veľkú záťaž na procesore

Efektívnosť smerovania

- Opísaný algoritmus pre jednoduchosť nerieši špecifické situácie
  - Fragmentácia IP paketov
  - ACL
  - Tunelovanie, šifrovanie/dešifrovanie
  - Preklad adries a mnohé ďalšie osobitné operácie
- Jeho najzdĺhavejšie kroky sú 5. – 7.
  - Vyhľadávanie rôznych informácií v rôznych databázach pre každý jeden paket, ktorý musíme smerovať
  - Na kontrolné sumy sa dá navrhnúť špecializovaný integrovaný obvod
- Veľké úsilie sa v posledných rokoch venovalo práve tomu, ako tieto operácie zjednodušiť resp. urýchliť
  - Fast switching, Autonomous switching, Silicon switching engine (SSE) switching, Optimum switching, Distributed fast switching, Cisco Express Forwarding (CEF), Distributed Cisco Express Forwarding (dCEF)

Fast Switching

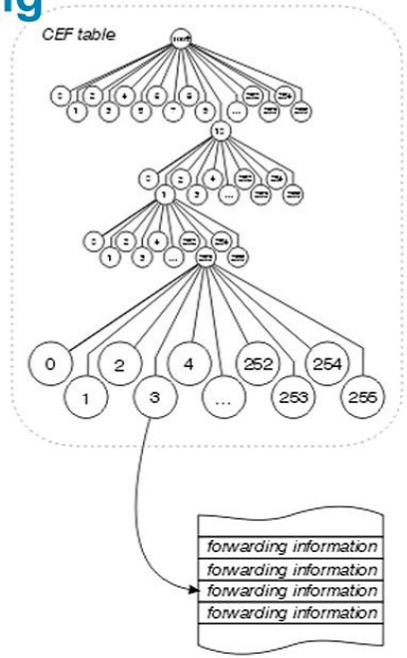
- Cisco Document ID: 13706, „How to Choose the Best Router Switching Path for Your Network“
- Fast Switching je takisto známy ako route cache
- Idea:
  - Prvý paket idúci do istého cieľa bude „process switched“
  - Výsledok tohto algoritmu (výstupné rozhranie, údaje potrebné na prepis rámca) sa zapamätá v tzv. route cache
    - Stromová dátová štruktúra s rýchlym prístupom a vyhľadávaním (binárny vyhľadávací alebo 256-cestný strom)
  - Každý ďalší paket do toho istého cieľa môže využiť predpripravené informácie z route cache
    - „Route once, forward many times“
- Nevýhody:
  - Route cache sa tvorí iba tokom paketov
  - Jej položky je potrebné priebežne nechávať expirovať
  - Zmena v ARP tabuľkách znamená prebudovanie časti route cache

Cisco Express Forwarding (CEF)

- CEF je ďalším vylepšením route cache
- Idea:
  - Zorganizovať si položky zo smerovacej tabuľky do samostatnej dátovej štruktúry, tzv. Forwarding Information Base (FIB), v ktorej sa dá veľmi rýchlo vyhľadávať (trie – znakový strom)
  - Informácie o prepise rámca si predpripraviť ihneď, ako je to možné, a organizovať si ich v tzv. adjacency dabatáze
  - Jednotlivé položky vo FIB budú pomocou smerníkov ukazovať na položky v adjacency DB
- Vlastnosti:
  - FIB a adjacency DB sa vytvárajú z existujúcich položiek (hotová smerovacia tabuľka, hotové L2 informácie)
  - Zmena v adjacency DB si nevyžaduje zmeny vo FIB

Cisco Express Forwarding

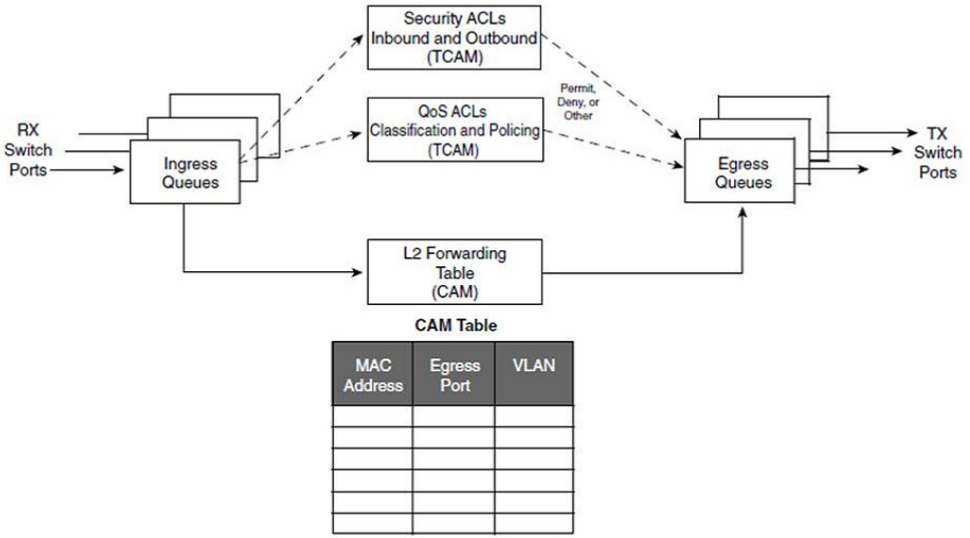
- Existujú dve realizácie CEF
  - Softvérová
  - Hardvérová
- Softvérová realizácia CEF je naznačená na obrázku
  - Stromová časť je FIB, principiálne sa jedná o tzv. znakový strom (trie resp. mtrie)
  - Tabuľka obsahuje adjacency DB
  - Obe štruktúry sú v RAM
  - Informácia o prepise sa priamo vo FIB nenachádza
  - Softvérové CEF využívajú najmä menšie Cisco smerovače



Cisco Express Forwarding

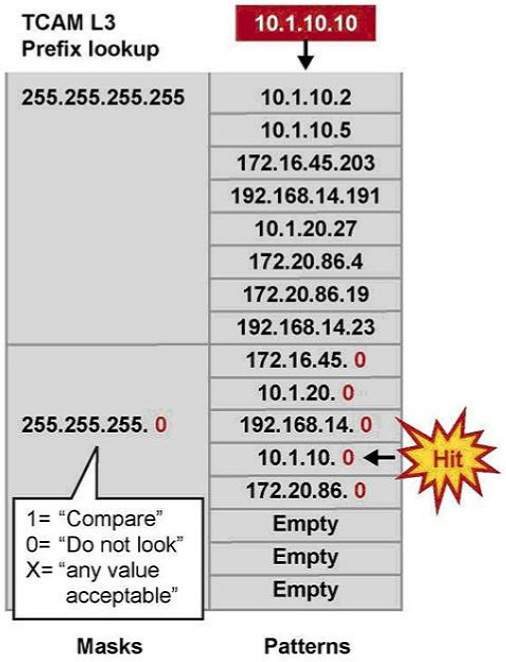
- Hardvérová implementácia CEF využíva špecializované integrované obvody na uchovávanie FIB
  - Ternary Content Addressable Memory (TCAM)
- CAM a TCAM sú dva typy cache pamätí
  - CAM obvykle obsahuje informácie pre L2 switching
  - TCAM obsahuje informácie pre L3 switching, ACL, QoS
- CAM hľadá presne zadaný reťazec (úplná zhoda)
  - Používa bity 0 a 1
- TCAM hľadá buď najdlhší alebo prvý zhodný reťazec
  - Je možné povedať, ktoré bity reťazca nás nezaujímajú
    - Vstup 0, 1, a X (don't care)
  - Môže byť rozdelená na regióny s rôznou politikou vyhľadávania
- Využitie TCAM na udržiavanie FIB je typickou doménou multilayer prepínačov (MLS) a high-end smerovačov

CAM a TCAM používajú aj L2 prepínače

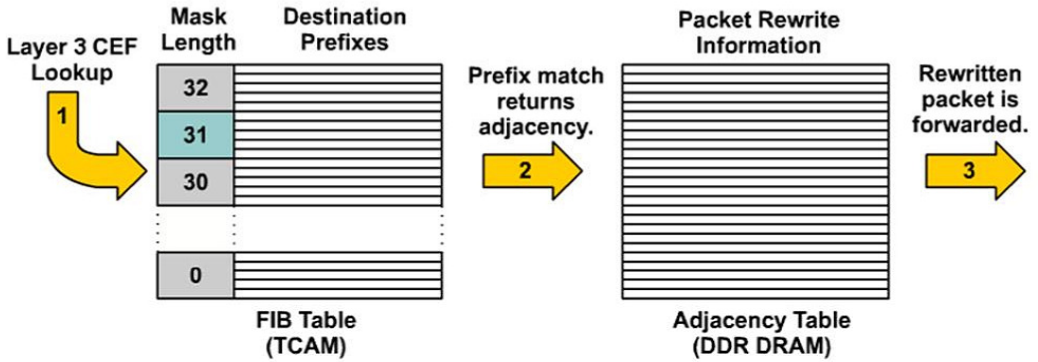


Vyhľadávanie v TCAM

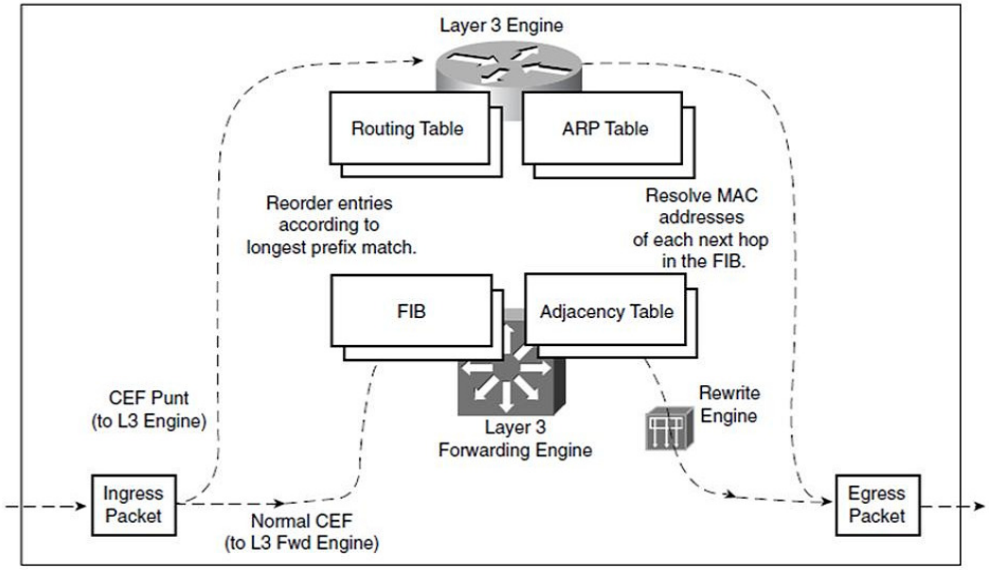
- TCAM je rozdelená na regióny s rôznou politikou
  - Exact Match
  - Longest Match
  - First Match
- Veľkosť týchto regiónov je na niektorých platformách možné meniť



Cisco Express Forwarding



Tok paketov cez CEF MLS



Cisco Express Forwarding

- Zapnutie CEF:

```
Switch(config)# ip cef
```

  - Na multilayer prepínačoch nie je možné CEF vypnúť
- Aktivácia/deaktivácia CEF na rozhraní

```
Switch(config)# int fa0/1
Switch(config-if)# [no] ip route-cache cef
```

  - CEF sa zásadne aktivuje alebo deaktivuje na vstupnom rozhraní
  - Route-cache sa aktivuje alebo deaktivuje na výstupnom rozhraní
- Zobrazenie informácií vo FIB a ADB

```
Switch# show ip cef
Switch# show adjacency
```

Príklad CEF

```
sw-vd-FRI# show ip cef 194.160.136.5 detail
194.160.136.0/24, epoch 1
  nexthop 158.193.26.1 Vlan26

sw-vd-FRI# show adjacency 158.193.26.1 detail
Protocol Interface                 Address
IP       Vlan26                    158.193.26.1(11)
                                   2 packets, 116 bytes
                                   epoch 0
                                   sourced in sev-epoch 88
                                   Encap length 14
                                   00E04C38C6D5001B8F8FDE410800
                                   ARP
sw-vd-FRI# show ip arp 158.193.26.1
Protocol  Address          Age (min)  Hardware Addr   Type   Iface
Internet  158.193.26.1            0   00e0.4c38.c6d5  ARPA   Vlan26
```

Príklad CEF

```
! Zobrazí obsah FIB daného prepínača
sw-vd-FRI# sh ip cef
Prefix              Next Hop            Interface
0.0.0.0/0           158.193.7.158       Vlan709
0.0.0.0/32          receive
158.193.7.84/30     158.193.7.158       Vlan709
158.193.7.92/30     158.193.7.158       Vlan709
158.193.7.152/29    attached            Vlan709
158.193.7.152/32    receive             Vlan709
158.193.7.157/32    receive             Vlan709
158.193.7.158/32    attached            Vlan709
158.193.7.159/32    receive             Vlan709
158.193.26.0/24     attached            Vlan26
158.193.26.0/32     receive             Vlan26
158.193.26.1/32     attached            Vlan26
158.193.26.6/32     receive             Vlan26
158.193.26.20/32    attached            Vlan26
158.193.26.21/32    attached            Vlan26
 --More--
```

Default route FIB

Príklad ADJ

```
! Zoznam všetkých next hop routerov a pripojených staníc
sw-vd-FRI# sh adjacency
Protocol Interface                 Address
...
...
IP       Vlan139                   158.193.139.189(8)
IP       Vlan139                   158.193.139.190(8)
IP       Vlan139                   158.193.139.191(8)
IP       Vlan139                   158.193.139.192(8)
IP       Vlan139                   158.193.139.198(8)
IP       Vlan139                   158.193.139.199(8)
IP       Vlan139                   158.193.139.219(8)
IP       Vlan139                   158.193.139.247(8)
IPV6     Vlan139                   2001:4118:300:122:20A:5EFF:FE64:90CA(8)
IPV6     Vlan139                   2001:4118:300:122:20C:29FF:FE59:1C48(8)
IPV6     Vlan139                   2001:4118:300:122:214:85FF:FEC8:ADE5(8)
IPV6     Vlan139                   2001:4118:300:122:250:4FF:FE2A:475F(8)
IPV6     Vlan139                   FE80::20A:5EFF:FE64:90CA(3)
...
...
```

Príklad ADJ

```
! Info o všetkých známych Next Hop IP vo VLAN 26
sw-vd-FRI# sh adjacency vlan 26 detail
Protocol Interface                 Address
IP       Vlan26                    158.193.26.1(13)
                                   2 packets, 116 bytes
                                   epoch 0
                                   sourced in sev-epoch 0
                                   Encap length 14
                                   0011D81BBE8F001B8F8FDE410800
                                   L2 destination address byte offset 0
                                   L2 destination address byte length 6
                                   Link-type after encap: ip
                                   ARP
IP       Vlan26                    158.193.26.20(8)
                                   0 packets, 0 bytes
                                   epoch 0
                                   sourced in sev-epoch 0
                                   Encap length 14
                                   001B549640AE001B8F8FDE410800
                                   L2 destination address byte offset 0
                                   L2 destination address byte length 6
                                   Link-type after encap: ip
                                   ARP
IP       Vlan26                    158.193.26.21(8)
                                   1 packets, 58 bytes
```

CEF položky – príznaky (flags)

- Attached
  - Položka pre cieľ s maskou /32, ktorý je priamo pripojený k MLS či smerovaču (typicky výsledok vyhľadania po Glean adjacency)
  - Next hop ukazuje na jedno z rozhraní smerovača
- Connected
  - Položka pre cieľovú sieť, z ktorej rozsahu ma MLS či router pridelenú IP adresu
  - Položka je aj Attached
- Receive
  - Položka pre vlastnú IP adresu MLS alebo routera.
  - MLS musí prijať a spracovať tento paket
- Recursive
  - Rekurzívny lookup
- Default Route Handler
  - Položka pre default routing

Generické CEF položky

- 0.0.0.0/32 receive
  - Položka umožňuje prijať pakety posielané na 0.0.0.0 bcast adresu
- 224.0.0.0/4 drop
  - Zahadzovanie multicastových paketov
- 224.0.0.0/24 receive
  - Spracovanie paketov v link-local rozsahu 224.0.0.0-224.0.0.255
  - Tento adresový rozsah používajú napr. smerovacie protokoly
- 255.255.255.255/32 receive
  - Spracovanie paketov posielaných na link-local broadcast adresu

Druhy adjacencies v CEF

- Auto
  - Normálne položky
- Punt
  - Pre pakety, ktoré nie je možné spracovať v CEF
  - Spracované ďalšou najlepšou metódou (Fast Switching, Process Switching)
- Glean
  - Zastupuje zatiaľ neznáme stanice na priamo pripojenej sieti (sieť je priamo pripojená, no nepoznáme MAC adresu konkrétnej stanice)
  - Prefix má adjacency Glean
  - ARP throttling
- Drop
  - Pakety sa majú zahodiť, no môže sa poslať ICMP správa o ich zahodení
- Discard
  - Pre IP adresy, ktoré sú súčasťou siete na našom loopbacku
  - Pakety sa majú v tichosti zahodiť bez ICMP správy
- Null
  - Pakety sa majú preposlať rozhraním Null0, t.j. zahodiť
- No route
  - V prípade chýbajúcej default route, ICMP generované
- Cached and uncached
- Unresolved
  - Chýba adjacency pre next hop

Obsluha Glean adjacency typu

- Glean adjacency zastupuje všetky stanice v priamo pripojenej sieti, pre ktoré ešte nepoznáme L2 prepis
  - Typicky IP/MAC mapovanie v ARP tabuľke
- Čo sa deje, ak prichádza paket na takúto adresu?
  - CEF lookup vráti záznam na sieť typu Glean
  - MLS pošle ARP žiadosť a zároveň nainštaluje tzv. ARP throttling položku (adjacency typu Drop) pre hľadanú adresu
  - Po získaní ARP odpovede MLS vytvorí L2 prepis a nahradí Drop položku bežnou Attached položkou

Príklad druhov ADJ

```
sw-vd-FRI# sh ip cef adjacency ?
  adj-null           Null Adjacency
  discard            Discard Adjacency
  drop               Drop Adjacency
  fcpa               Fiber Channel
  glean              Glean Adjacency
  punt               Punt Adjacency


sw-vd-FRI# sh ip cef adjacency glean
Prefix              Next Hop            Interface
158.193.7.152/29    attached            Vlan709
158.193.26.0/24     attached            Vlan26
158.193.128.0/24    attached            Vlan128
```

Cisco Express Forwarding

- Nie všetky pakety môžu byť spracované v CEF
- Cisco Express Forwarding nepodporuje:
  - Pakety, pre ktoré v CEF neexistuje platný záznam
  - Pakety určené pre samotný router resp. switch
  - Broadcasty a multicasty
  - IP pakety, ktoré využívajú voliteľné časti hlavičky (options)
  - IP pakety, ktoré musia byť fragmentované
  - IP pakety, ktorým expiruje TTL
  - Network Address Translation
  - Šifrovanie počas prenosu
- Ak sa do TCAM nezmestia všetky položky, namiesto chýbajúcich sa nainštaluje wildcard záznam ukazujúci na spracovanie v softvéri
  - Hrozí riziko preťaženia CPU

Cisco Express Forwarding

- Podľa verzie Cisco zariadenia môže CEF pracovať v **centralizovanom** alebo **distribuovanom** režime (dCEF)
  - V **centralizovanom** režime sa všetky rozhodnutia o forwardovaní paketu realizujú nad spoločnou centralizovanou databázou
  - V **distribuovanom** režime sa vybrané časti CEF štruktúr nahrávajú do procesorov na zásuvných moduloch, ktoré potom vedia realizovať forwarding vo svojej vlastnej réžii
  - CEF v distribuovanom režime sa nazýva **dCEF**

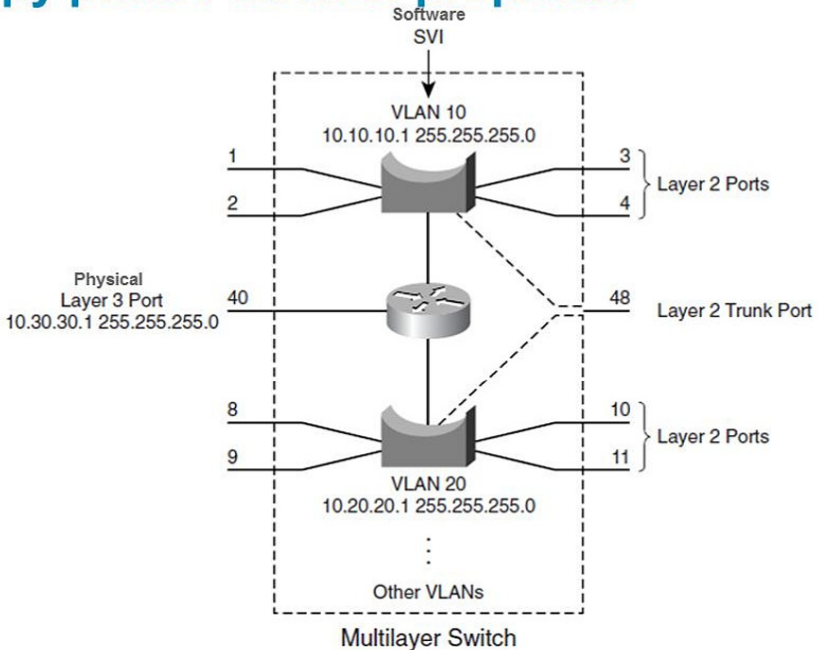
Smerovanie medzi VLAN pomocou externého routera

Konfigurácia multilayer switchov pre routing

Smerované rozhrania na MLS

- MLS switche používajú dva druhy smerovaných rozhraní
  - Switched Virtual Interface (SVI)
  - Fyzické smerované rozhrania (routed)
- SVI je jednoducho „interface VLAN X“
  - Jedná sa o virtuálne rozhranie, ktoré prepája route processor (RP) multilayer switcha s danou VLAN
  - Implicitne je vytvorené rozhranie vo VLAN 1, ostatné softvérové rozhrania môžeme podľa ľubovôle vytvárať či odstraňovať
  - Na rozdiel od L2 switchov, kde má zmysel vytvárať spravidla iba jedno takéto rozhranie, MLS switche môžu mať pre každú VLAN samostatné SVI
  - SVI je rozhraním route processora v MLS do danej VLAN siete

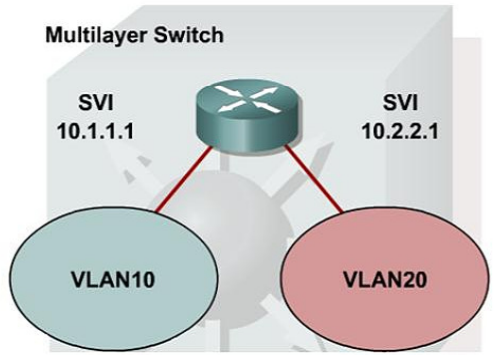
Typy portov na MLS prepínači



SVI rozhrania na MLS

```
Switch(config)# ip routing
Switch(config)# vlan 10,20
Switch(config-vlan)# exit
Switch(config)# int vlan10
Switch(config-if)# ip address 10.1.1.1 255.255.255.0
Switch(config-if)# int vlan20
Switch(config-if)# ip address 10.2.2.1 255.255.255.0
```

- Príkaz **`ip routing`** aktivuje podporu pre L3 switching
- Po zadaní tohto príkazu je možné na MLS pracovať ako na routeri
  - Smerovacia tabuľka
  - Smerovacie protokoly



SVI rozhrania na MLS

- SVI nie je „večne živé“ rozhranie, aj keď je virtuálne
- SVI je v stave „up, line protocol up“ práve vtedy, keď
  1. VLAN, do ktorej SVI patrí, je vytvorená a aktívna
  2. Rozhranie je zapnuté príkazom „no shutdown“
  3. Na switchi sa nachádza živý port v danej VLAN
     a) Buď access port v danej VLAN
     b) Alebo trunk, na ktorom je táto VLAN povolená
- Toto správanie je potrebné, aby sa v smerovacej tabuľke MLS a v jeho smerovacom protokole neudržiavali siete, ku ktorým route processor v MLS nemá pripojenie
  - Ak vo výnimočných prípadoch nechceme, aby nejaký konkrétny port mal vplyv na stav SVI, môžeme na ňom použiť príkaz

```
Switch(config-if)# switchport autostate exclude
```

Fyzické smerované rozhrania na MLS

- Fyzické L3 rozhrania sa používajú rovnako ako fyzické rozhrania na bežných smerovačoch s pár výnimkami
  - Nepatria do nijakej konkrétnej používateľskej VLAN
  - Nepodporujú trunking
  - Nepodporujú subinterfejsy
- Vytvorenie fyzického L3 rozhrania:

Príkazom switchport definujeme, či je port L2 alebo L3. Vypnutím na MLS robíme z L2 fyzického portu L3 port

```
Switch(config)# ip routing
Switch(config)# int fa0/5
Switch(config-if)# no switchport
Switch(config-if)# ip address 10.0.3.1 255.255.255.0
```

Fyzické smerované rozhrania na MLS

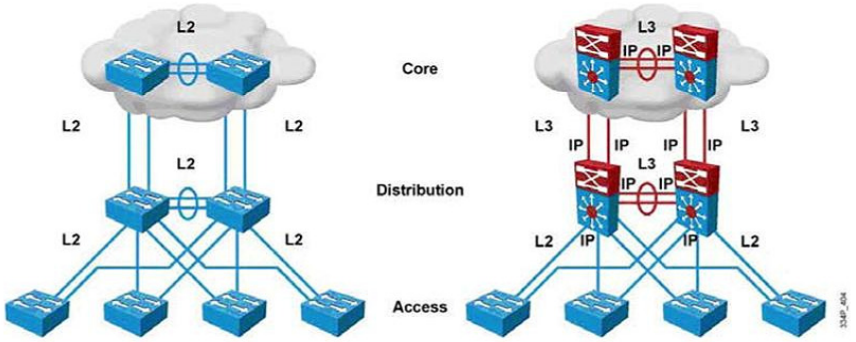
- Každé smerované fyzické rozhranie je interne v MLS switchi reprezentované špeciálnou internou VLAN
  - Tzv. Internal Usage VLAN

```
Switch# show vlan internal usage
```

- Interné VLAN sú alokované obvykle od 1006 nahor
  - Na niektorých platformách je možné toto poradie prepnúť v zmysle od 4094 nadol
  - Pre zníženie pravdepodobnosti konfliktu s internou VLAN sa odporúča vlastné extended VLANs alokovať od opačného konca extended priestoru

Posun od L2 k L3 dizajnu

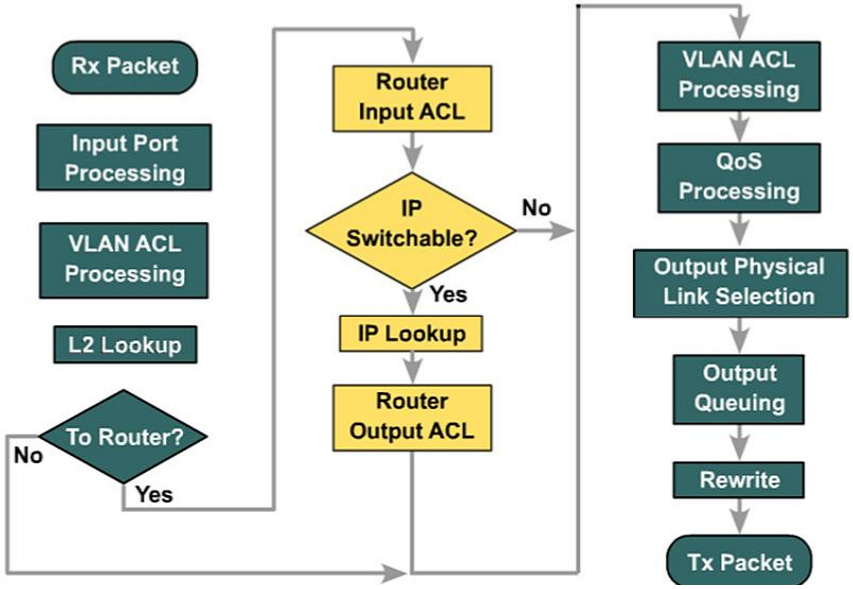
- Smerované fyzické porty alebo L3 EC umožňujú izolovať VLAN na distribučnej vrstve
  - L3 prepoje medzi Distro a Core vrstvou dovoľujú zmenšiť L2 domény, využiť multipathing či výhody rýchlej konvergencie smerovacích protokolov



Návrat portu do režimu L2 switchingu

```
Switch(config)# int fa0/5
Switch(config-if)# no ip address
Switch(config-if)# switchport
```

Poradie činností pri prechode paketu MLS



Switch Database Management

SDM Templates

- Kľúčovým stavebným komponentom L3 switcha je TCAM
  - TCAM sa využíva pre mnohé aplikácie
  - Miesto v TCAM je obmedzené a drahé
  - Podľa toho, aké funkcie bude MLS dominantne vykonávať, je možné prideliť rôznym aplikáciám rôzne veľkú časť TCAM
    - Rozdelenie TCAM na „partície“
- O toto rozdelenie sa starajú tzv. SDM Templates
- Rôzne MLS platformy môžu mať predpripravené rôzne SDM šablóny, napr. pre 3560 existujú:
  - Access – maximalizuje miesto pre ACL
  - Default – približne vyvážené rozdelenie miesta
  - Routing – maximalizuje miesto pre smerovacie záznamy
  - VLANs – maximalizuje miesto pre L2 switching
  - Dual-IPv4-and-IPv6 Default, Routing, VLAN

SDM Templates

- Niektoré funkcie na MLS switchoch nie sú podporované, pokým nie je TCAM vhodne rozdelená
  - Napr. IPv6 alebo Policy Based Routing
- Informácie o aktuálnej SDM šablóne, ako aj o iných dostupných šablónach je možné získať v príkazovom riadku MLS

```
Switch# show sdm prefer
 The current template is "desktop default" template.
 The selected template optimizes the resources in
 the switch to support this level of features for
 8 routed interfaces and 1024 VLANs.

  number of unicast mac addresses:                  6K
  number of IPv4 IGMP groups + multicast routes:    1K
  number of IPv4 unicast routes:                    8K
    number of directly-connected IPv4 hosts:        6K
    number of indirect IPv4 routes:                 2K
  number of IPv4 policy based routing aces:         0
  number of IPv4/MAC qos aces:                      0.5K
  number of IPv4/MAC security aces:                 1K
```

Informácie o SDM templates

```
Switch# show sdm prefer ?
  access               Access bias
  default              Default bias
  dual-ipv4-and-ipv6   Support both IPv4 and IPv6
  routing              Unicast bias
  vlan                 VLAN bias
  <cr>

Switch# show sdm prefer routing
 "desktop routing" template:
 The selected template optimizes the resources in
 the switch to support this level of features for
 8 routed interfaces and 1024 VLANs.

  number of unicast mac addresses:                  3K
  number of IPv4 IGMP groups + multicast routes:    1K
  number of IPv4 unicast routes:                    11K
    number of directly-connected IPv4 hosts:        3K
    number of indirect IPv4 routes:                 8K
  number of IPv4 policy based routing aces:         0.5K
  number of IPv4/MAC qos aces:                      0.5K
  number of IPv4/MAC security aces:                 1K
```

Zmena SDM

- Zmena SDM sa realizuje v globálnom konfiguračnom režime

```
Switch(config)# sdm prefer ?
  access               Access bias
  default              Default bias
  dual-ipv4-and-ipv6   Support both IPv4 and IPv6
  ipe                  IPe bias
  routing              Unicast bias
  vlan                 VLAN bias

Switch(config)# sdm prefer routing
Changes to the running SDM preferences have been stored, but
cannot take effect until the next reload.
Use 'show sdm prefer' to see what SDM preference is currently
active.
Switch(config)# exit
Switch# write
Switch# reload
```