

## Multistage interconnection networks without internal cell loss

Ako uz nazov napoveda, budeme sa zaoberat typom switching fabrics, u ktorých nemožu nastať kolízie buniek vo vnútornej štruktúre switching fabric, čo však nevylučuje kolízie na vstupe a výstupe z nej. Rozlišujeme 2 kategórie:

- s interným buffrovaním - musia nejako riešiť situáciu, keď sa zaplní buffer, napr. tak, že bude posielať správu predchádzajúcemu elementu o tom, že bunku neprijal ( backpressure mechanism ), napr. St. Louis switching fabric.
- bez interneho buffrovania- tu musíme zabezpečiť, aby žiadne 2 bunky nevyžadovali ten istý zdroj, napr. výstup, ale ani vnútorný element, čo riešia napr. topologie typu Batcher - Banyan. Podľa spôsobu riešenia tejto situácie rozlišujeme rôzne MINs, my sa budeme zaoberať konkrétnymi Starlite a Moonshine.

### St. Louis switching fabric

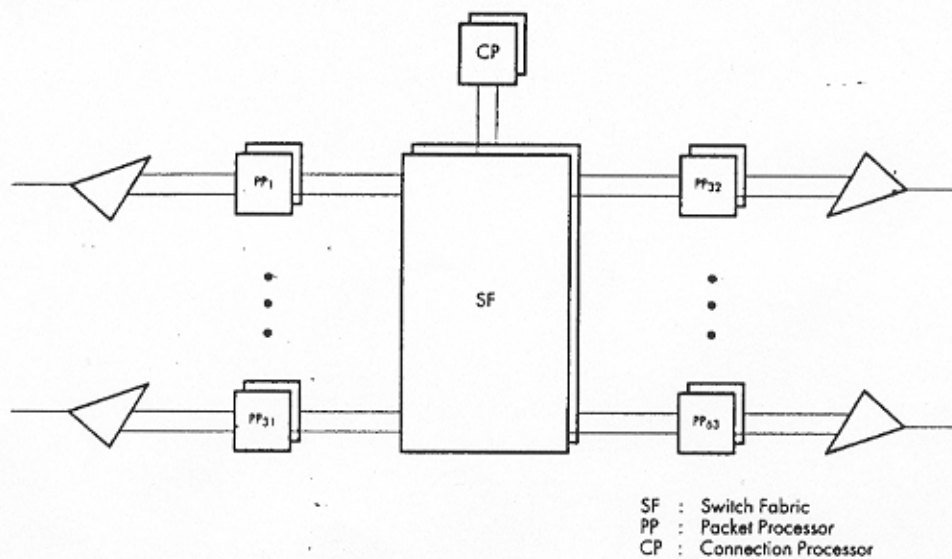


Fig. 4.34. – St. Louis Switching Fabric

PP - ukončuje externý protokol a pridáva informáciu pre vnútorné routovanie ( routing tag )

SF - uskutočňuje vnútorné routovanie

CP - vytvára point-to-point, ale aj multicast spojenia

Kvôli spoľahlivosti sú všetky tieto systémy duplikované, jedna vrstva je štandardne aktívna a druhá je v pohotovosti pre prípad zlyhania primárnej vrstvy. Táto switching fabric je typu II, čo znamená, že routovanie závisí len od informácie v bunkách a siete sú packet oriented. Všetky interne siete sú samoroutovacie.

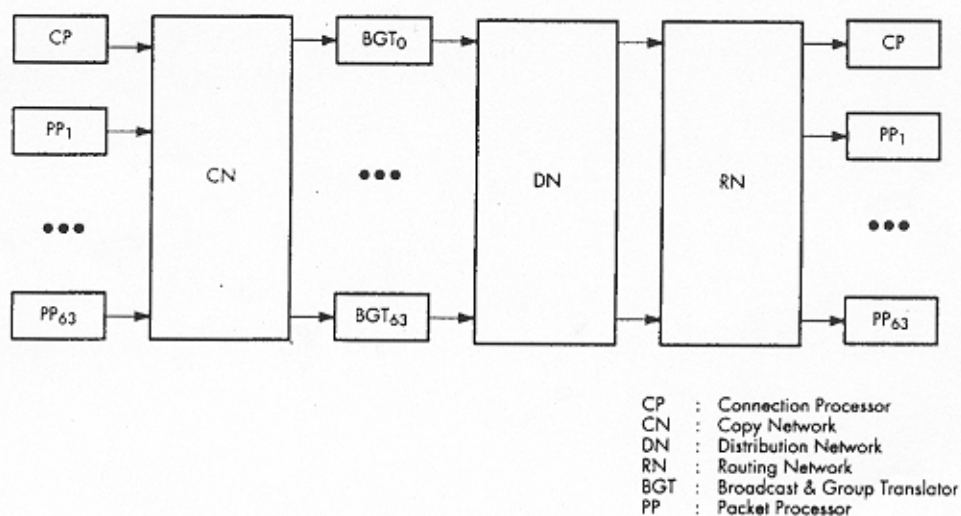


Fig. 4.35. – St. Louis Switching Fabric Building Blocks

CN - vytvara kopie prichadzajucich buniek, vyuziva sa len pri multicast alebo broadcast spojeniach

BGT - preklada a vyplna routing tags, tiez sa vyuziva najma pri multicast spojeniach

DN - nahodne rozmiestni prichadzajuce bunky kvoli rovnomernemu zatazeniu RN

RN - posle bunky na skutocne pozadovany vystup

CN, DN, RN sa skladaju z identickych 2x2 basic switching blocks, teda 2 Input controllerov (IC) a 2 Output Controllerov (OC). IC podla routing tagu vyberie vystup, na ktory bunku odosle. Kazdy IC ma aj buffer, ktory pouzije, ak je cielovy OC obsadeny. Ak sa tento buffer zaplni a pride dalsia bunka, posle predoslemu elementu negativny grant signal, aby sa vyhol strate tejto bunky. Ak OC dostane pozitivny grant signal, vyberie pomocou svojej internej logiky niektoru z nabuffrovanych buniek a odosle ju dalsiemu elementu. Ak dostane negativny signal, "pocaka jedno kolo". Ked je buffer prazdny, umoznuje tento prenos aj priamo ( virtual cut through ). Vnutorna prenosova rychlost je rovna dvojnasoisku vonkajsej, co znamena, ze pri externom zatazeni 80% je vnutorne 40%, co vyzaduje iba maly buffer a este stale je k dispozicii backpressure mechanism, takze sa pouziva v kazdom controlleri buffer na 2 bunky. Aby sa backpressure mechanism vyuzival, co najmenej, pouziva sa DN, ktora bunky nahodne rozmiestni na vstupy RN.

CN schopna vytvorit z jednej  $2^n$  buniek sa sklada z n vrstiev elementov a vyuziva pri kopirovani buniek nasledovny algoritmus:

```

if FANOUT > 2s then
  begin duplicate
    if multicast-connection is even then
      upper : FANOUT = | (FANOUT + 1) / 2 |
      lower : FANOUT = | FANOUT / 2 |
    else
      upper : FANOUT = | FANOUT / 2 |
      lower : FANOUT = | (FANOUT + 1) / 2 |
  end
end

```

else distribute traffic as in DN

, kde  $s$  je číslo vrstvy číslované zprava od 0 a  $|x|$  je dolná celá časť  $x$ . Tento algoritmus zabezpečí rovnomerné rozloženie zátazy tým, že pri rozdeľovaní počtu buniek spojenia s parným číslom, pošle viac buniek hornému elementu a pri nepárnom dolnému. Keď bunky opustia CN, ešte majú rovnaké routing tagy, takže ich treba nejako zmeniť, čo riešia BGT.

### Batcher - Banyan based MINs

Nepripústa žiadne spory o výstupy medzi bunkami, takže nepotrebuje žiadny buffering, ani backpressure mechanism. Takéto MINs sa volajú non-blocking SF, musia splniť podmienku, že žiadne dve bunky nesmú smerovať na ten istý výstup. Keďže táto možnosť existuje, musí ju riešiť nejaká rozhodovacia logika a ukladanie informácií na vstupe do switching fabric. Najznámejšie switching fabrics tejto triedy sú založené práve na Batcher - Banyan sieťovej topológii:

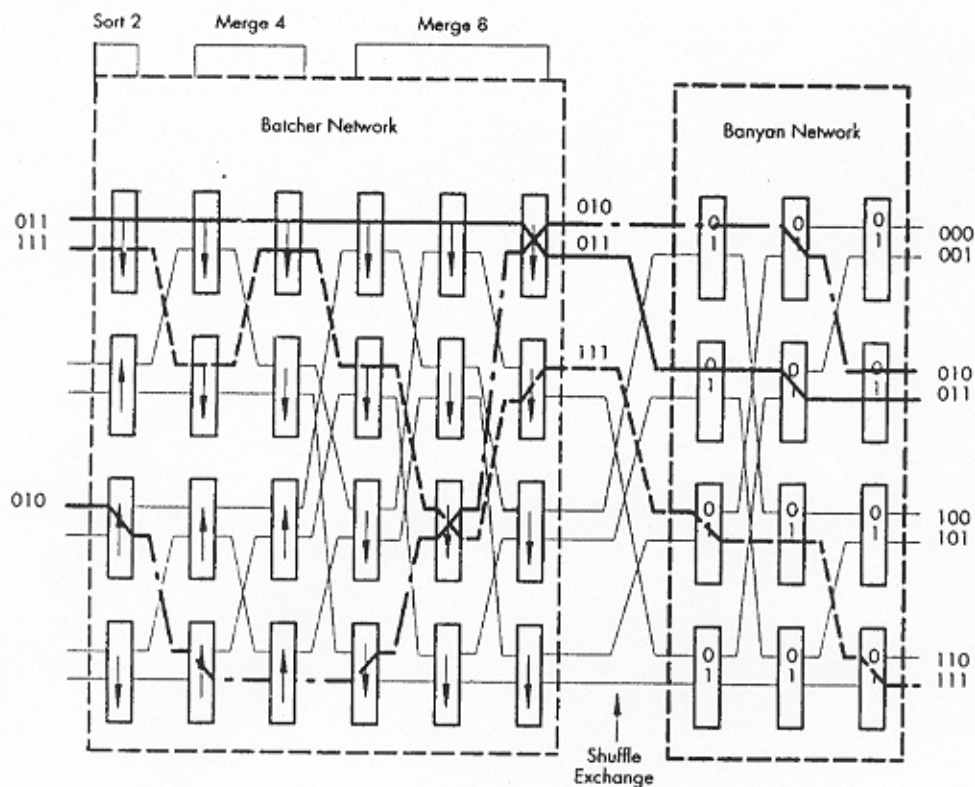


Fig. 4.38. - Batcher-Banyan Network Topology

Batcher network triedi bunky podľa cieľovej adresy tak, že bunka s nižšou adresou sa dostane na jej vyšší výstup ako bunka s vyššou adresou. Tiež sa skladá z elementov  $2 \times 2$ , ktoré adresy porovnávajú a bunku pošlú buď do pass alebo exchange vetvy "rozhodovacieho stromu". Triedenie využíva tzv. bitonic sorters, ktoré dostanu 2 utriedené postupnosti a spájajú ju do jednej utriedenej postupnosti.

### **Banyan network**

Ide o jednoduchu samoroutovaciú sieť. Prave fakt, že bunky na vstupe sú utriedené umožňuje jednoduchosť tejto siete, lebo sa netreba zaoberať ani vnútornými kolíziami.

### **Starlite**

Ide o prvú switching fabric založenú na Batcher - Banyan topológii. Je dizajnovaná na prenos s konštantným zdržaním. Skladá sa z routovacej siete (Batcher - Banyan) a z ďalších 2 elementov trap (pasca) a concentrator. Tu použité Banyan siete sa nazývajú aj expander, lebo umožňuje rast počtu výstupov a šírkou pásma na užívateľa. Starlite spracováva bunky s konštantnou veľkosťou, ku ktorým na vstupe pridáva routing tag. Aby sa vyhol konfliktom na výstupoch, pridá medzi Batcher a Banyan sieť pascu, ktorá zachytí bunky s rovnakou cieľovou adresou a odosle všetky okrem jednej späť na vstup Batcher siete. Pasca sa skladá z jednej vrstvy porovnávacích, za ktorými nasleduje sieť Banyan. Jedna vrstva porovnávacích stáci, pretože Batcher sieť zabezpečí, aby sa bunky s rovnakou cieľovou adresou dostali na susediace výstupy. Keď sa zistí kolízia, bunky, ktoré prejdú pascou ďalej sa dostanú na ľavé výstupy a tie, ktoré neprejdú, na prave a teda späť do Batcher siete. Keďže je veľká pravdepodobnosť, že nie na každom výstupe bude nejaká bunka, pridala sa concentration vrstva, ktorá umožňuje zmenšenie routovacej siete a je implementovaná ako inverzná Banyan sieť. Routing je založený na running sum of the activity bit (ACT), ktorý hovorí, či je bunka prázdna alebo aktívna, prázdne bunky budú odoslané na nepripojený výstup, čím jednoducho zanikajú. Pred recyklovaním buniek z pasce je ešte buffer pre prípad, že by bolo málo voľných vstupov do Batcher siete. Ďalším konfliktom zamedzuje fakt, že bunky, ktoré sú už dlhšie vnútri switching fabric majú vyššiu prioritu ako nové.

Po pridaní ďalších dvoch sietí umožňuje aj broadcast a multicast, obr. Sort-to-copy sieť má na vstupe pôvodne bunky z vstupu a prázdne bunky na kopírovanie generované zvláštnou jednotkou. Prázdne bunky majú informáciu o tom, obsah ktorých buniek sa im má skopírovať, čo sa pri prechode touto sieťou stane.

### **Moonshine**

Umožňuje variable packet length, takže ide o všeobecnejší mechanizmus ako len pre ATM sieť. Na riešenie konfliktov na výstupoch používa 3-fázový algoritmus s input queueing, čo spôsobuje HOL (Head of line blocking). V rozhodovacej fáze pošle každý vstup testovací balík, ktorý obsahuje len informácie o zdroji a cieľi, z čoho Batcher sieť dovoľí poslať potvrdenie len tým, ktorý majú nad sebou inú cieľovú adresu, ostatné vstupy dostanú negatívnu odpoveď, čo je jednoducho implementovateľné, lebo pri prechode cez Batcher sieť si bunky vytvorili cestu, ktorou sa môže sprava dostať späť k vstupu. Po prijatí správ už všetky vstupy vedia, či môže byť ich bunka spracovaná a odosle len tie, ktoré dostali pozitívny signál a už prejdú cez Batcher-Banyan sieť jednoducho. Keďže v prvej a druhej fáze sa nedejú žiadne reálne prenosy dát, potrebujeme interne urychlenie, približne o 14%.