

1. Multiple Access (TDMA, FDMA, CDMA) alapfogalmak. Multiple Access vs. Multiplexing (TDM, FDM, CDM) vs. Duplexing (TDD, FDD).
2. Telefonhálózat (PSTN) elemei. PCM moduláció. Miért lassú a dial-up, mit változtat ezen az xDSL?
3. Frekvencia felosztás ADSL-nél. Miért aszimmetrikus? Mi befolyásolja az átvitel minőségét különböző csatornákon? Hogyan választjuk ki, melyik csatornákat használjuk?
4. Mi a szerepe egy ADSL modemnek, egy splitternek, egy DSLAM-nak, egy BRAS-nak?
5. ADSL2 újdonságai – teljes digitális mód, SRA. Frekvenciatartomány változása ADSL2+ esetén – milyen következményei vannak? Ugyanez VDSL-re. Mire használják emiatt leginkább a VDSL-t? Mi az alapvető újdonság a VDSL2-ben?
6. Miért kell szimmetrikus hozzáférési technológiákban is gondolkodni? Milyen alkalmazási példák lehetnek? Hogyan lehet ezt biztosítani xDSL-nél?
7. G.fast újdonságai – frekvenciatartomány, TDD, profilok, energiatakarékosság.
8. DSL lefedettség. Miért kell szabályozni a szélessávú internet szolgáltatást? Mit jelent a monopolhelyzet, milyen hátrányai vannak, hogyan lehet tenni ellene? LLU megoldások összehasonlítása.
9. Miben más egy HFC hálózaton kábeltévé szolgáltatást vagy szélessávú internetet nyújtani? Mi a szerepe egy fiber node-nak? Hogyan történik a spektrumkiosztás a TV, rádió, upstream és downstream internet között, és miért pont így?
10. Mi volt a Telco return lényege a DOCSIS 1.0-ban? QoS támogatás az 1.1-ben: miért, hogyan? Hogyan növelik a sebességet a későbbi verziókban (S-CDMA, channel bonding, OFDM)?
11. Mi a ranging, miért van rá szükség? Versenyhelyzetes feltöltés kábelneten – FDD/TDMA. Sáv szélesség igénylés különböző QoS osztályoknál (CBR – UGS, rtVBR – RTPS, UGS-AD, nrt-VBR – nRTPS, BEG). Miben más a letöltés kezelése? Összegzésként, mitől aszimmetrikus a kábelnet? (csatornák száma, moduláció, versenyhelyzet)
12. Ethernet címezés (MAC, OUI), csomag formátumok – Ethernet II, 802.3/LLC/SNAP. Főbb hálózati elemek, interfészek. Learning bridge - cím tanulás, flood. Spanning Tree Protokollok: szerepük, változatok (STP, RSTP, MSTP - különbségek!). VLAN-ok: mi az, működés: szűrés ki/bemeneten. 802.1Q tag. Tagging módok/hozzárendelés. Trunk port. Prioritás kezelés. Ethernet multicast, problémák, megoldások (IGMP snooping).
13. Egymódusú vs. többmódusú optikai szál. PON vs. aktív node. PON le- és feltöltés. ATM cellák, SAR, előnyök és hátrányok. APON vs. EPON. FTTs szolgáltatási modellek – saját hálózat vs. nyílt hozzáférés.
14. Ethernet szolgáltatások: L2VPN-ek. Szolgáltatás bemutatása: e-line, e-lan, evc, uni. Felhasználási módok: switch/router interconnect. Megvalósítások – Ethernet kiterjesztésével: Q, Provider Bridges (Q-in-Q), Provider Backbone bridges (MAC-in-MAC), PBT. MPLS alapú megvalósítások: VPLS, H-VPLS. OAM kérdések.
15. IP címezés: cím, fejléc. IP címek felépítése (classful). IP cím - netmask, default gw. Speciális IP címek (broadcast, mcast, „fehér”). ICMP, DHCP-csak hogy mi az.
16. Routing alapfogalmak – router feladata, kapacitása, routing metrikák. IP route lookup, longest prefix match. RIB vs. FIB. Útválasztási szemantika – unicast, multicast, anycast, geocast, broadcast. IGP vs. EGP protokollok (két előadásban is). Statikus vs. dinamikus routing. Egyutas vs. többutas. Lapos vs. hierarchikus. Intra-domain vs. inter-domain. Hop-by-Hop vs. source routing.
17. Távolság-vektor alapú routing működése. Bellman-Ford algoritmus. Mit tárol és mit hirdet a RIP? Végtelenig számolás problémája és megoldása.

18. Kapcsolat-állapot alapú routing működése. Dijkstra algoritmus. OSPF tartományok. DR és BDR választás – miért, hogyan?
19. Internet topológia, tranzit vs. peering kapcsolat ISP-k között.
20. BGP protokoll jellemzői, működési alapelvek. iBGP vs. eBGP célja.
21. Multicast alkalmazás példák. Megbízhatóság kérdése miben más, mint unicast esetén? Ethernet multicast alapelv. Hálózati rétegű multicast célja, a multicast fa fogalma. Multicast címzés (IPv4-ben). Multicast scoping.
22. Az IGMP protokoll feladata, működési alapelvei. Query – Report üzenetek, Host Suppression, Unsolicited Report. IGMP v1, v2 és v3 közötti különbségek.
23. MOSPF protokoll működési alapelve, miben más, mint az OSPF? Hátrányok.
24. DVMRP protokoll működési alapelve – elárasztás-metszés.
25. PIM protokoll működési alapelve. Mit jelent az, hogy „protocol independent”?
26. A hagyományos ASM modell korlátai, az SSM modell előnyei. Az SSM modell lényegi különbsége – multicast csatorna.
27. IGMP/MLD snooping switch feladata, működése.
28. Explicit multicast működése, előnyei és hátrányai.
29. Az ALM multicast alapelve, előnyei és hátrányai. (Narada és TBCP nem kell)
30. IPv6 címzés, scoping. IPv6 fejléc fontosabb elemei. Flow label. Kiegészítő fejlécek összeláncolása. Hop-by-Hop, Routing, Fragment, Destination Option fejlécek.
31. Áttérés IPv6-ra: Dual Stack, alagutak, protokoll fordítás – alapelvek szintjén.
32. MPLS: Label Swapping mechanizmus, terminológia: LSP, LER, LSR, FEC. Label distribution: mi az, hogyan csináljuk?
33. Traffic Engineering: mire jó, MPLS hogy teszi lehetővé? Védelem, Fast reroute. GMPLS.
34. L4: Portok fogalma, TCP vs UDP - mikor melyiket érdemes használni?
35. TCP - kapcsolat kiépítés (3 way handshake), MSS, ablak csúszási mechanizmus, hogyan garantálja az átvitelt. Torlódás vezérlés: slow start, fast retransmit, congestion avoidance, Bandwidth-delay product, RTT és RTO.
36. Példa feladatok:
 - a) Mekkora az ideális ablakméret ha az RTT= 250ms és a BW=200Mbps?
 - b) A következő szekvenciaszámokat kapja egy állomás: 23,24,25, 30,31,32,33,34. RTT=100ms, RTO=500ms. Rajzold fel a csomagok és válaszok szekvenciáját!

Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. szeptember 7.



Előadók

Vida Rolland

egyetemi docens, tárgyfelelős

IE 325, vida@tmit.bme.hu



Moldován István

IB 229, moldovan@tmit.bme.hu



Adminisztratív részletek

Tárgy honlapja:

<http://www.tmit.bme.hu/vitmac05>

Jegyzet nincs, de (viszonylag) részletes fóliák

Elvileg 28 alkalom (21 előadás + 7 gyakorlat)

- 2 alkalom elmarad (szünnapok)

Órákra járni nem kötelező (de ajánlott)

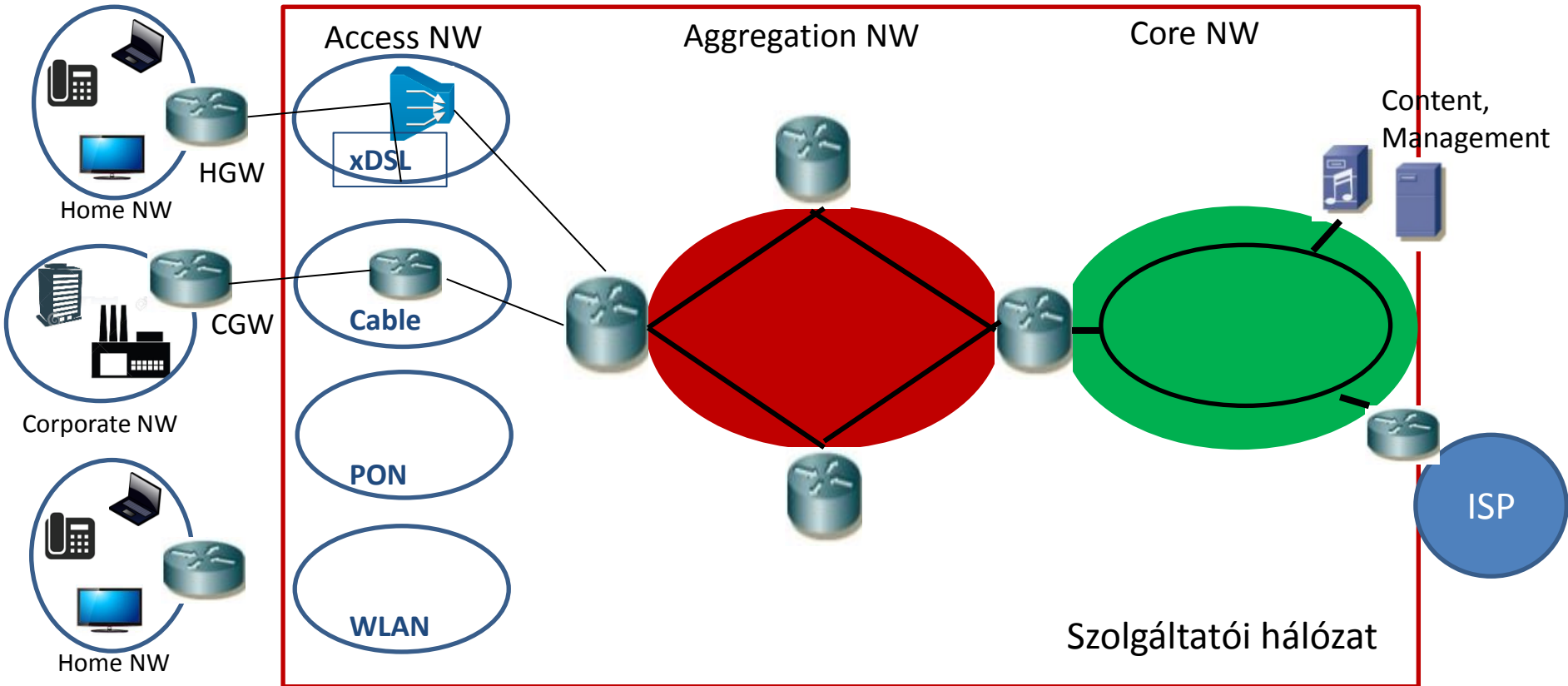
A gyakorlatok legalább 70%-ra kötelező bejárni

- 6 gyakorlatból 5-re
- Gyakorlatok nem feltétlenül órarend szerint, hanem az anyagihoz kötődően, de előre be lesznek jelentve

Számonkérés

- 1 db nagy ZH, kb. 10. héten
- 1 pót ZH az utolsó héten
- 1 PPZH a pótlási héten
 - ZH nem számít be a jegybe, aláíráshoz kell
 - Ugyanabból az anyagból a ZH, PZH, PPZH
- Írásbeli vizsga

Big picture



Hálózatok

- **Internet**
 - „Hálózatok hálózata”, hálózatok közötti kapcsolat
 - Szabad világ
 - Elosztott viselkedés
 - Alacsony hálózati intelligencia
 - Nincs általános menedzsment megoldás
 - Nincs lehetőség garanciákra (Best Effort)
- **Szolgáltatói hálózatok**
 - Kötött világ
 - Központosított
 - Általános menedzsment megoldások
 - Garantált szolgáltatások

Hozzáférési hálózatok - alapfogalmak

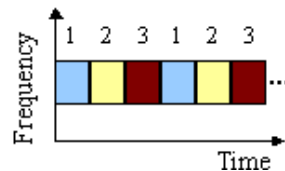
- A telekommunikációs hálózat azon része, mely közvetlenül összeköti a felhasználót a szolgáltatóval
 - Ethernet, WLAN, FTTx, xDSL, kábelnet, ...
- Gyakran **osztott átviteli közeg** (*shared transmission medium*)
 - Többen hallanak engem, és én is több mindenkit hallok
 - Nem lehetséges fizikailag, vagy nem éri meg anyagilag minden felhasználónak dedikált átviteli csatornát biztosítani
- A megoldandó feladat az átviteli közeghez (csatornához) való **hozzáférés szabályozása**
 - A felhasználók nem tudják egymásról, hogy ki mikor szeretne adni
 - A küldéseket koordinálni kell

Többszörös hozzáférés (*Multiple Access*)

- **Fix kiosztásra alapuló megoldások**

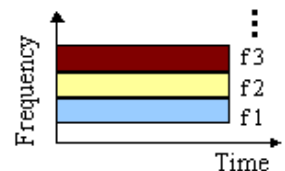
- **TDMA – Time Division Multiple Access**

- Minden felhasználónak saját időszelete amikor küldhet
 - A teljes frekvenciatartományt használhatja



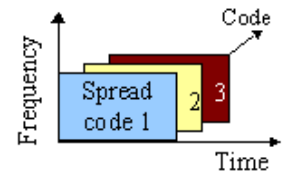
- **FDMA – Frequency Division Multiple Access**

- A spektrumot frekvenciacsatornákra vágjuk
 - Minden felhasználó a saját frekvenciáján kommunikál



- **CDMA – Code Division Multiple Access**

- Minden felhasználó a teljes csatornán, egyfolytában kommunikál
 - Kódelmélet segítségével különítjük el a forgalmakat
 - Az adó megszorozza a jelet egy kóddal (spreading code), és az eredményt küldi el
 - A vevő a vett jelet újra megszorozza ugyanazzal a kóddal, reprodukálva az eredeti jelet
 - Minden felhasznált kód ortogonális
 - » Két különböző kód összeszorozása 'nullák' sorozata lesz



Multiple Access vs. Multiplexing vs. Duplexing

- **Multiple Access (TDMA, FDMA, CDMA)**
 - A csatorna-hozzáférés szabályozása több forrás esetén
 - Alapvetően uplink irány
- **Multiplexing (TDM, FDM, CDM, ...)**
 - A csatorna-hozzáférés szabályozása több küldeni kívánt jel esetén, ugyanattól a forrástól
 - Uplink vagy downlink irány
- **Duplexing (TDD, FDD)**
 - A downlink és uplink forgalom közötti megosztás
 - FDD – Frequency Division Duplexing
 - „Párba állított” frekvenciák, elkülönített uplink és downlink csatornák
 - TDD – Time Division Duplexing
 - Pár nélküli frekvenciák, rugalmasan megosztott uplink és downlink csatornák

Többszörös hozzáférés (*Multiple Access*)

- A fix kiosztás nem hatékony ha kevés és bősztös a forgalom
- **Versengésre alapuló csatorna-hozzáférés**
 - **Lekérdezések** (*polling*)
 - **Erőforrások lefoglalása és ütemezése** aktuális igények alapján
 - **Véletlen hozzáférés** (*random access*)
 - Egy csomópont akkor küld amikor akar, előzetes egyeztetés nélkül
 - Ha két vagy több csomópont egyszerre beszél, ütközés, majd később újraküldés
 - ALOHA, Slotted ALOHA, CSMA/CD

Hozzáférési hálózatok

- (Vezetékes) hálózatok zöldmezős kiépítése nagyon drága lehet
 - Nem a vezetékek a drága, hanem a munkálatok
 - Ásás, épületeken belüli munkák
- Megoldás: **igénybe kell venni a már meglévő hálózatokat**
 - Nyilvános kapcsolt telefonhálózat
 - Public Switched Telephone Network (PSTN)
 - Kábel TV hálózatok
 - Elektromos hálózat
 - Gázvezeték hálózat (?)
 - Ultra Wideband rádiós kommunikáció
 - Szennyvízcsatorna hálózat (?)
 - Optikai kábelek
- De bizonyos esetekben lehet azért újat is építeni...

Internet a gázvezetéken?

- NetherComm ötlete 2005-ben
- **Ultra Wideband**
 - Nagy frekvenciasáv (>500 Mhz), nagy átviteli sebességek (100 Mbps)
 - Nagy teljesítményű adók esetén túl nagy interferencia más vezeték nélküli technológiákkal, ezért csak kis hatótávolságra engedélyezve
 - A föld alatti gázvezetékekben ez nem gond, lehet nagyobb teljesítménnyel adni
- Az UWB technológia ígéretesnek tűnt, de ...
 - Szigorú szabályozás, lassú szabványosítás, az ígértnél lassabb sebességek
 - 2008-2009-ben az ipar nagy része kihátrált mellőle
 - A NetherComm is eltűnt...

PSTN

- A telefonhálózat elemei:
 - Előfizetői hurok
 - Csavart réz érpár
 - A háztól vagy az irodától a helyi kapcsolóközpontig („local exchange”)
 - „Local loop”, „last mile”
 - Optical local loop, wireless local loop
 - Kapcsolóközpontok
 - Tronkók
 - a kapcsolóközpontokat összekötő szálak
 - gerinchálózat (törzshálózat)
- A kezdeti hálózat teljesen analóg
 - Fokozatos áttérés a digitális átvitelre, főleg a kapcsolóközpontok között (gerinchálózat)

PSTN



Beszédcsatorna

- 4kHz sávzélességű beszédcsatorna
 - A beszédjel átviteli tartománya 0.3 – 3.4 kHz között
 - Védősávokkal kiegészítve
- Az emberi fül által érzékelhető frekvenciatartomány: 20Hz – 15-20 kHz
 - A beszédhangok átvitele volt a cél
 - Nem kell minden hallható hangot átvinni
 - Gazdasági megfontolások

PCM

- Pulse Code Modulation
 - Az analóg jelek digitalizálására
- Nyquist tétel alapján 4kHz-es jelhez 8kHz-es mintavételezés
 - 256 jelszintre kvantálva
 - 8 biten kódolva
 - Átviteli sebesség: $8\text{bit} \times 8\text{kHz} = 64\text{ kbit/s}$

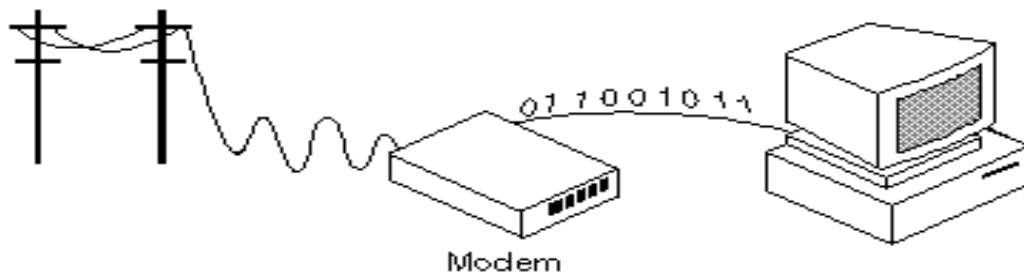


Digitális hangátvitel

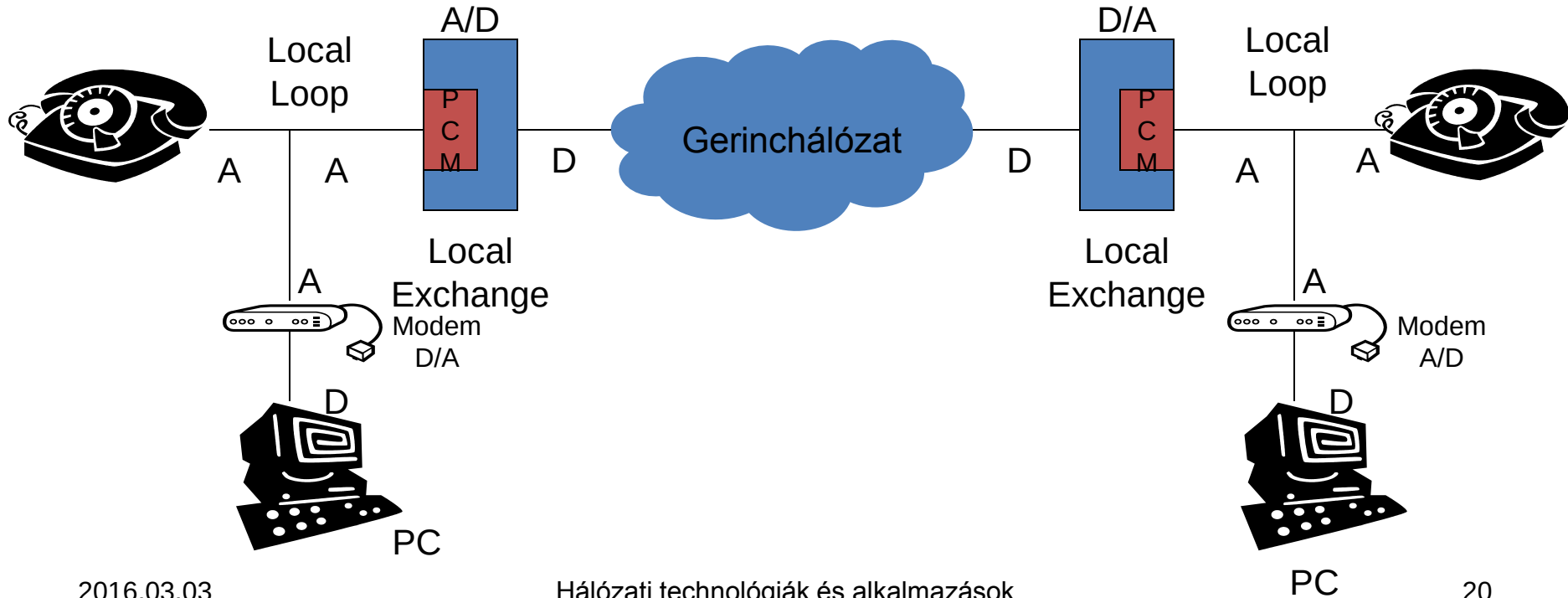


Dial-up Access

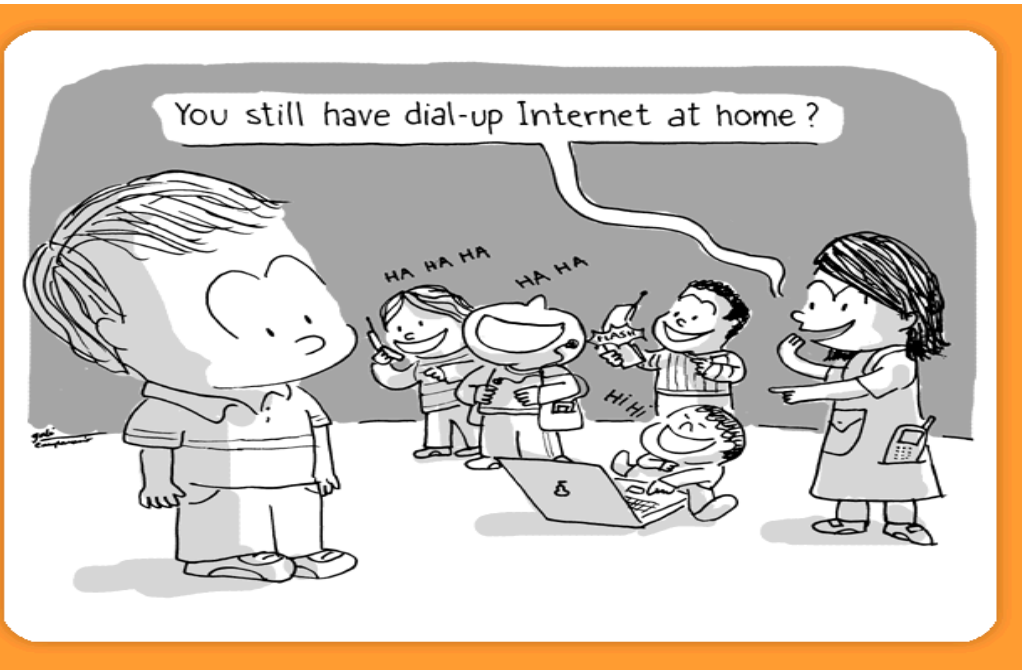
- „Betárcsázós internet”
- A computerek digitális információi analóg jellé alakíthatóak, és átvihetők a hagyományos telefonhálózaton
 - „Modem” – **m**odulator-**d**emodulator



Dial-up modem



Kihalófélben a dial-up



Miért DSL?

- Telefon ipar (dial up) – 56 Kbps
 - Kábeltévé ipar – 10Mbps osztott kábeleken
 - Műholdas cégek – 50 Mbps ajánlatok
 - Lépni kellett az internetezők megtartása érdekében
- Megjelenik a „szélessávú” (broadband) hozzáférés
 - Inkább reklám mint valóság
 - Nem egyértelmű mit értünk szélessávon
- xDSL – különféle DSL változatok
 - Digital Subscriber Line

Mitől gyors a DSL?

- **Miért lassú a dial-up?**
 - A telefonhálózatot beszédátvitelre optimalizálták
 - A helyi központban egy sávszűrő
 - Csak a 4 KHz-es beszédsáv marad
 - Az adatok is ezt a sávot használhatják csak
- Az **xDSL** előfizető vonalát egy olyan kapcsolóra kötik át, amelyen **nincs szűrő**
 - Kihasználhatóvá válik az előfizetői hurok teljes kapacitása
 - Függ a hurok hosszától, a kábelköteg vastagságától, és a minőségétől
 - Optimális viszonyok: új vezetékek, vékony kötegek, rövid hurok
- Ha nagy sebességet akarunk, sok helyi központot kell telepíteni
 - Ha valaki túl messze lakik, költözzön közelebb
 - Minél alacsonyabb a sebesség, annál nagyobb a hatótávolság – több lehetséges előfizető
 - Minél alacsonyabb a sebesség, annál kevesebb érdeklődő

Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. szeptember 9.

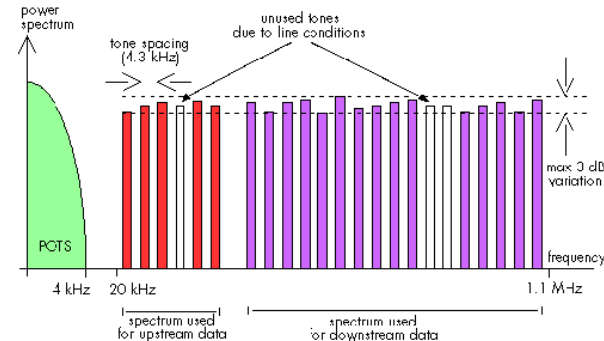
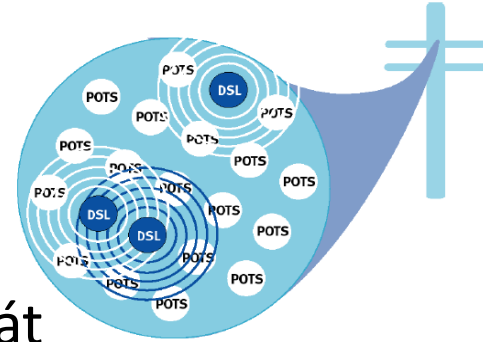


ADSL

- **DMT - Discrete Multitone Modulation**
 - 1.1 MHz-es frekvenciatartomány
 - 256 csatorna, egyenként 4.3125kHz
 - 0 csatorna – POTS (hang)
 - 1-5 csatorna – biztonsági sáv (üres)
 - A hang és adatátvitel közötti interferenciák elkerülésére
 - a maradék 250 csatornából 1 az upstream, 1 a downstream jelzése
 - a többi a felhasználói forgalomé
- Frekvenciák felosztása ADSL-nél
 - 0-4 kHz – hang
 - 4-25 kHz – biztonsági sáv
 - 25-160 kHz – upstream sáv
 - 200 kHz - 1.1 MHz – downstream sáv

ADSL DMT

- Átvitel minden csatornán, párhuzamosan, az átviteli paraméterek függvényében
 - Csillapítás a magasabb frekvenciákon
 - Interferenciák
 - Áthallás (crosstalk) a kábelkötegben
- A kapcsolat felépítésénél tesztel minden csatornát
 - A jel/zaj viszony alapján több/kevesebb bit/csatorna
 - Esetleg más moduláció (x-QAM)
 - Ha túl zajos a csatorna, nem küldünk rajta



ADSL architektúra

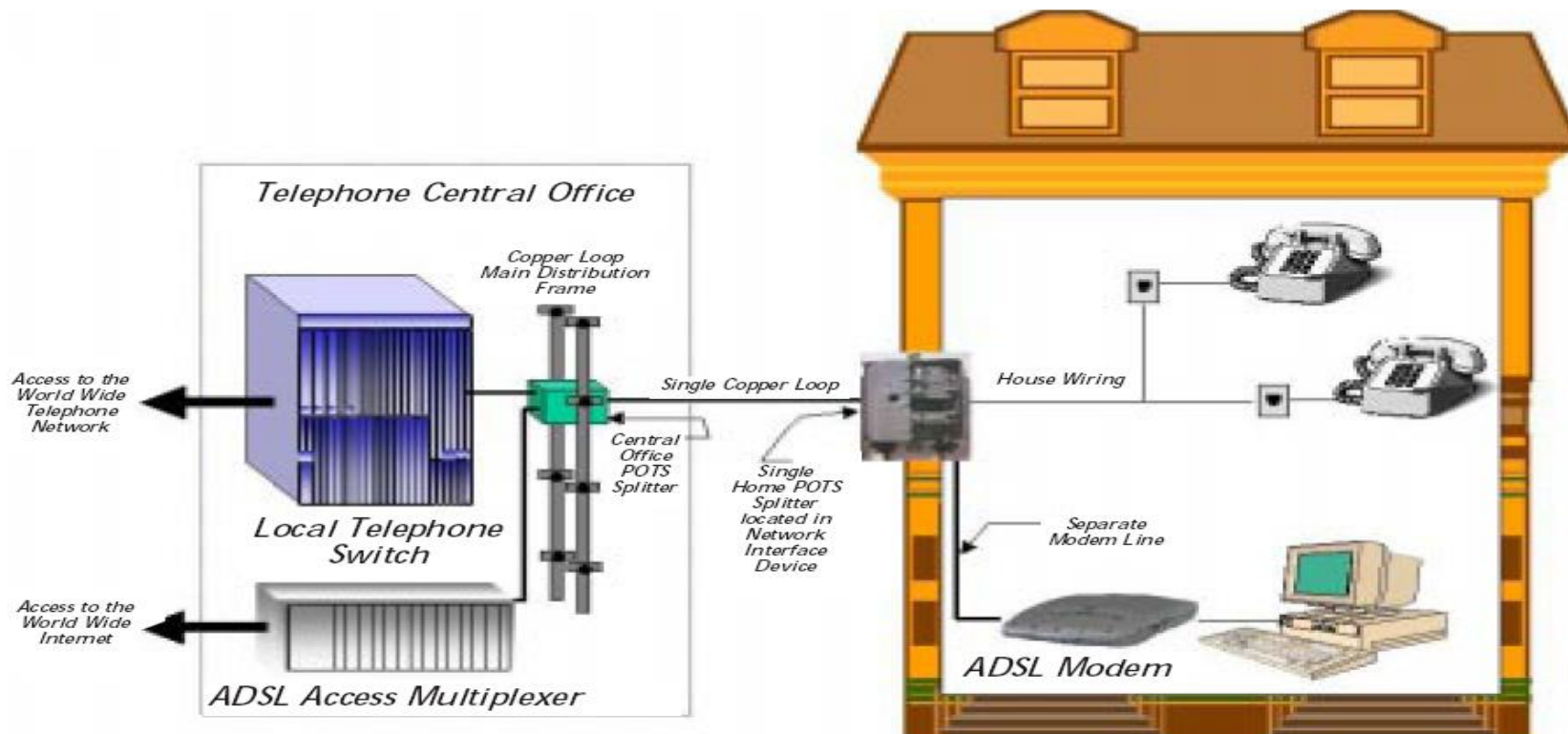
A szolgáltatónál

- **POTS Splitter**
 - Frekvenciaosztó a beszédjel és az adatok szétválasztására
 - A beszéd a hagyományos POTS switch-hez irányítva
 - A 25 KHz feletti rész a DSLAM-hoz
- **DSLAM – DSL Access Multiplexer**
 - AD / DA átalakító
 - Több előfizető adatforgalmát multiplexeli egy közös nagysebességű digitális kommunikációs csatornára (ATM vagy Ethernet)
- **BRAS – Broadband Remote Access Server**
 - Csatlakoztatja a DSLAM-okat egy internetszolgáltató hálózatához

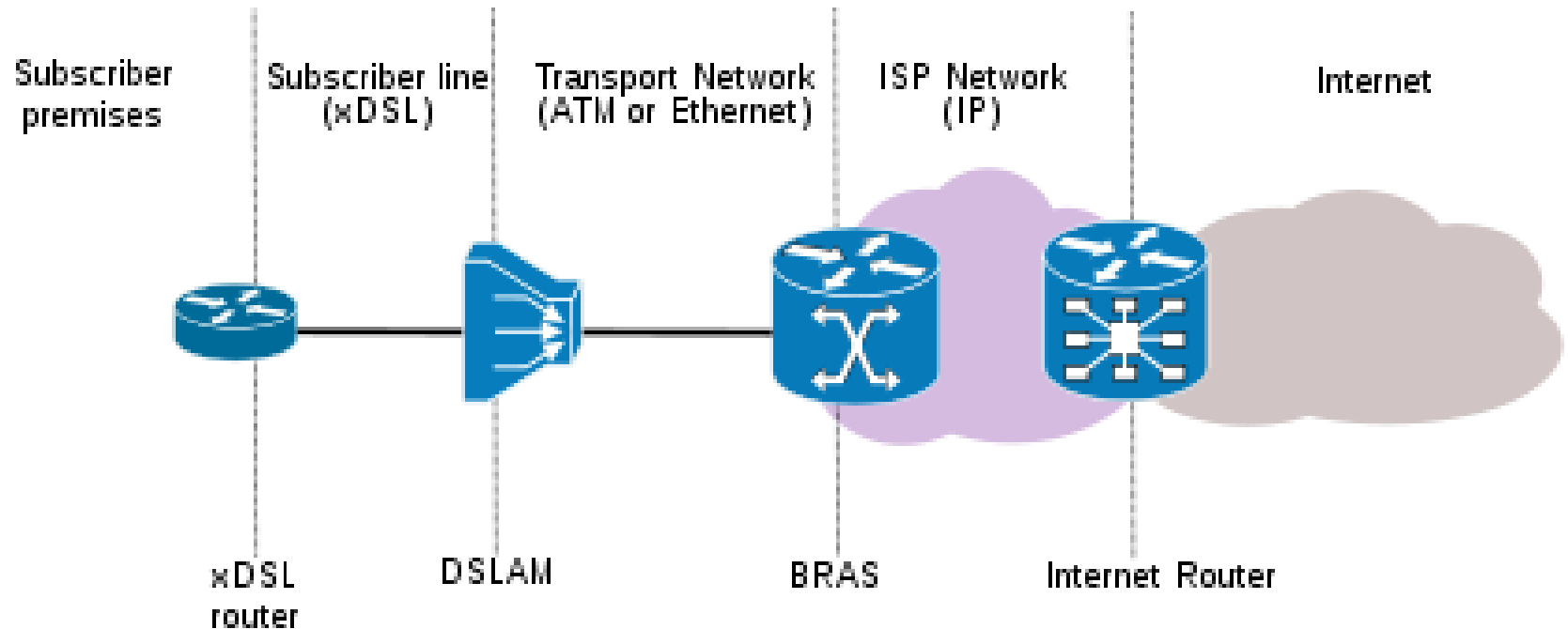
Az előfizetőnél

- POTS Splitter
- **ADSL modem**
 - Digitális jelfeldolgozó (DSP)
- Nagysebességű (Ethernet) összeköttetés a PC-vel

ADSL architektúra



ADSL architektúra



ADSL G.dmt

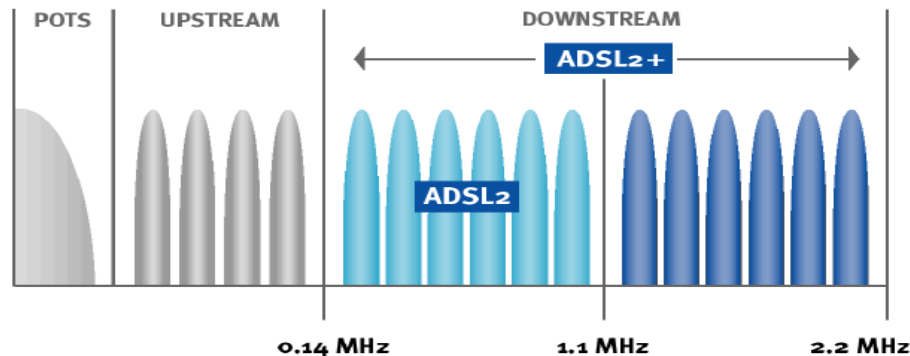
- ITU-T G.992.1 szabvány (1999)
- Lényegesen nagyobb a letöltésre elkülönített sávszélesség
 - a webes böngészés igényeire szabott technológia
 - maximális letöltési sebesség 8 Mbit/s
 - általában 512 Kbit/s – 1 Mbit/s
 - maximális feltöltési sebesség 1 Mbit/s
 - általában 64 Kbit/s – 256 Kbit/s
- A helyi központtól max. 3 km-es távolságig
- Ideális technológia lakossági felhasználásra
 - a hagyományos hangátvitellel közösen osztozik a már meglévő csavart érpáras vezetéken
 - a felhasználók egy időben telefonálhatnak és internetezhetnek ugyanazon a vezetéken keresztül

ADSL2

- ITU-T G.992.3 szabvány (2002)
- A hagyományos ADSL technológiát bővíti ki
 - A maximális adatátviteli sebesség 12 Mbit/s-ra nő
 - Az elérhetőségi távolság kb. 500 méterrel bővül
 - A javulás leginkább a hosszú vezetékeken tapasztalható interferenciák kiszűrésének tudható be
- Az ADSL2 átmenetileg átválthat „teljes digitális” módba
 - átadja a hangátvitelre elkülönített csatornákat is az adatátvitel számára
- Automatikus átviteli sebesség adaptáció
 - **Seamless Rate Adaptation (SRA)**
 - Menet közben tud változtatni a csatornákon, kiiktatja a zajosakat
 - Az ADSL-nél ez csak a kapcsolat megszakításával működött

ADSL 2+

- ITU-T G.992.5 szabvány (2003)
- Növeli a sávszélességet a használható frekvenciatartomány bővítése által
 - a hangátvitelre, illetve az adatfeltöltésre használt frekvenciák nem változnak
 - a letöltési csatorna maximális frekvenciája 1.1 MHz-ről 2.2 MHz-re bővül.
 - A maximális letöltési sávszélesség 8Mbit/s-ról 16 Mbit/s-ra nő
 - 1.5 km-es távolságon belül.



G.SHDSL

- Symmetric High-speed DSL
 - ITU-T G.991.2 (2001)
- 2.3 Mbit/s maximális átviteli sebesség mindkét irányban
 - egy második sodrott érpár hozzáadásával a kétirányú sebesség 4.6 Mbit/s-ra növelhető
 - A sebesség 3 km-es körzetben biztosítható
 - e távolságon felül az átviteli paraméterek fokozatosan gyengülnek
- Az alacsonyabb frekvenciák használata kizárja a hagyományos hangátvitelt
 - Jelentősen növeli a telepítési költségeket
 - Inkább üzleti, mintsem lakossági felhasználók

Üzleti SHDSL alkalmazások

- **Web hosting**
 - Olyan alkalmazások ahol a felhasználó egy web server-t üzemeltet egy DSL kapcsolaton keresztül
 - Nagy upstream sávszélességet igényel
- **Videokonferencia**
 - Egy videokonferencia szolgáltatás adat, text, hang és videó csomagok átvitelére épül
 - Mivel egy kétirányú szolgáltatás, egy szimmetrikus DSL kapcsolat (SHDSL) jobban megfelel
- **VPN (Virtual Private Network) szolgáltatások**
 - Magánhálózat a publikus telekommunikációs infrastruktúra felett
 - Az adatforgalom biztonsága (privacy) alagutazással és kódolással garantálva
 - VPN kapcsolatok SHDSL felett egy cégcsoport irodáinak összekötésére, ott ahol egy optikai kábeles megoldás nem elérhető, vagy túl drága
- **Remote LAN Access**
 - Távmunka (teleworking) vagy SOHO (Small Office Home Office) esetén a vállalati hálózat elérésére

Otthoni SHDSL alkalmazások

- **Internet Gaming**

- Egy otthoni felhasználó egy game szerver vagy más otthoni felhasználók ellen játszik
- Nagyon fontos a jó minőségű (upstream) kapcsolat

- **Residential Gateway Access**

- Egy olyan CPE (Customer Premises Equipment) melyen keresztül több otthoni szolgáltatás is elérhető (Internet hozzáférés, otthoni videofelügyelet, intelligens otthon)

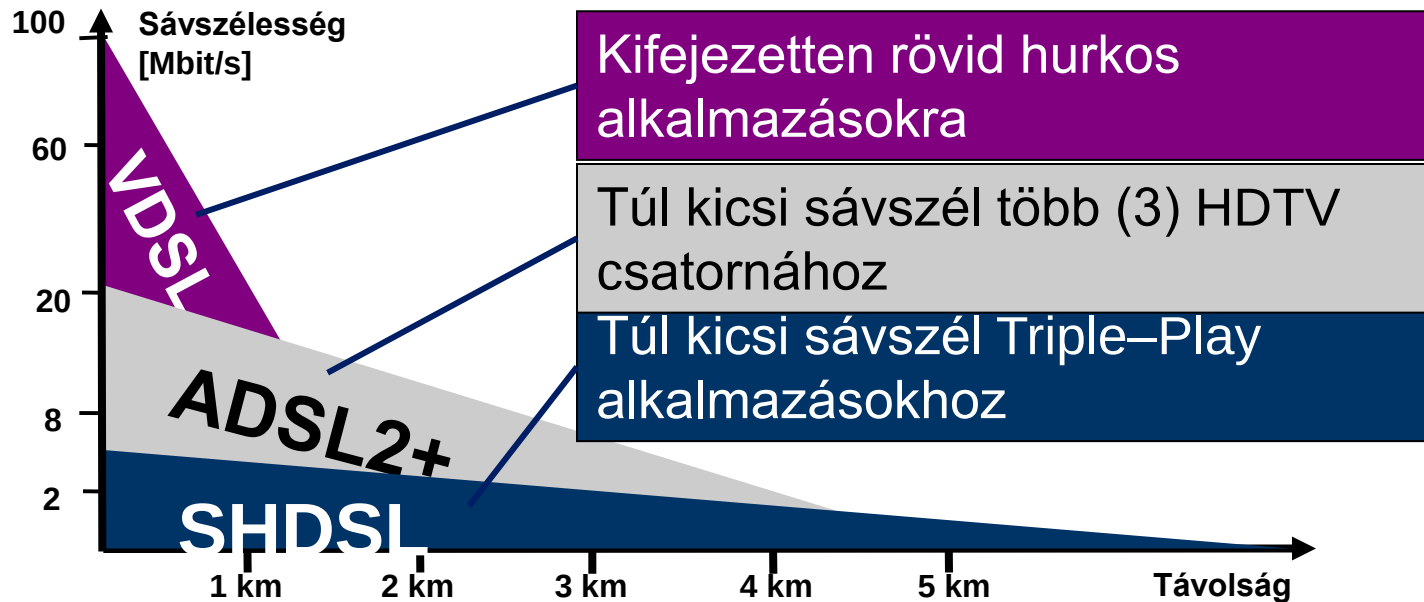
- **Peer-to-peer alkalmazások**

- Fájlcseré, alkalmazás rétegbeli multicast
- Szimmetrikus kapcsolat előnyt jelent a letöltési sebességnél
 - Ha te is tudsz feltölteni másoknak, hasznos peer leszel, jobb lesz a letöltésed

VDSL

- **HDSL (*High bit-rate DSL*)** – ITU-T G.991.1 (1998)
- **VDSL (*Very-high-data-rate DSL*)** - ITU-T G.993.1 (2004)
- Lényegesen nagyobb sebességű adatátvitel kis távolságokon
 - 52 Mbit/s downstream, 16 Mbit/s upstream
 - Lehet szimmetrikus is (26-26 Mbit/s)
 - 12 MHz sávszélesség
 - Max. 1 km hatótávolság
 - Inkább 300 méter
- Leginkább optikai hálózatok épületeken belüli kiterjesztésére javasolják, mintsem vidéki szétszórt felhasználócsoportok szélessávú bekötésére
 - Az optikai kábelek épületeken belüli telepítése a számos hajlítás szükségessége miatt nem ajánlott
 - A sodrott érpárt használó VDSL vonalak jó kiegészítést jelentenek

VDSL2

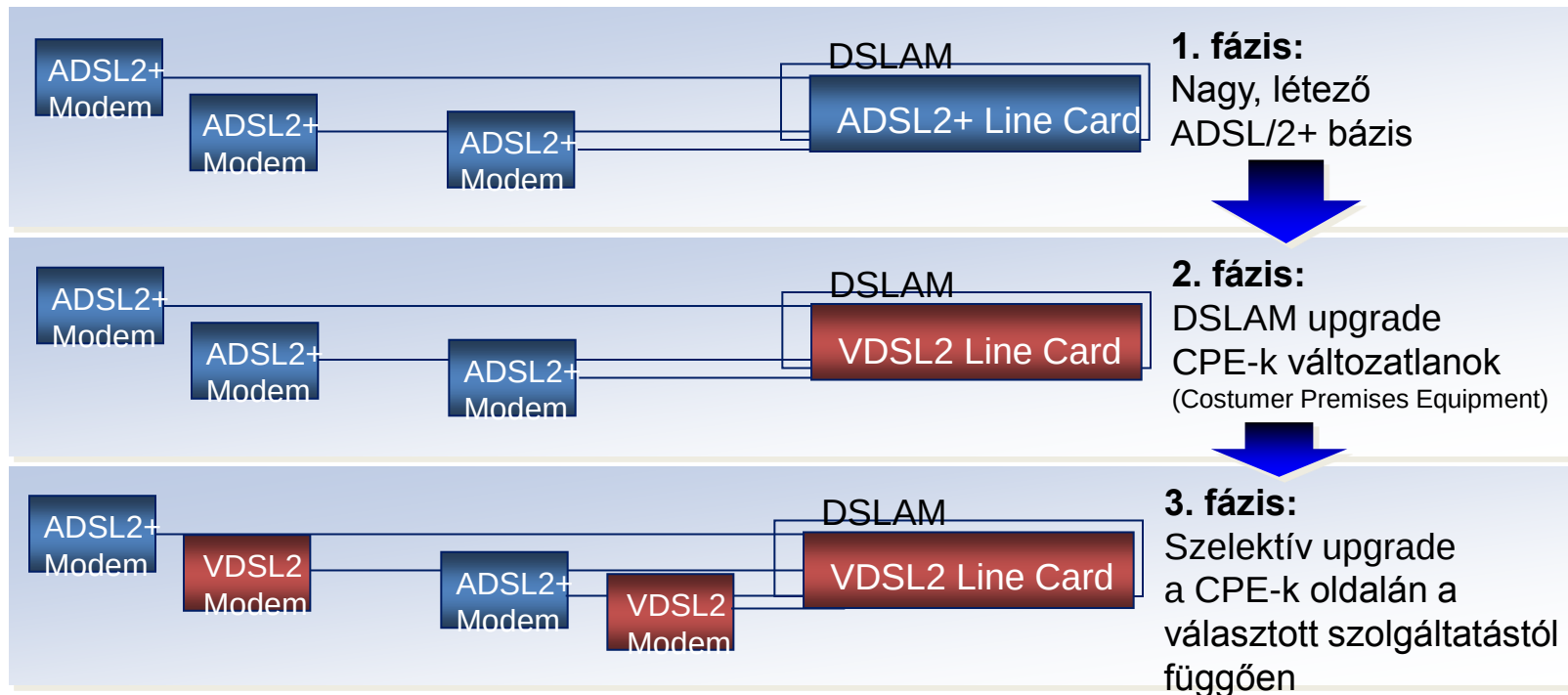


VDSL2 = VDSL sebesség ADSL/2+ hatótávolsággal

VDSL2

- ITU-T G.993.2 (2006)
 - 100 Mbit/s downstream és upstream
 - 30 MHz-es frekvenciatartomány
 - 3 km-es hatótávolság
 - A nagy sebesség és a nagy hatótávolság egyszerre nem teljesíthető
- 8 meghatározott profil, különböző szolgáltatási szinteknek
 - Más és más sáv szélesség igény régióként
- ADSL kompatibilis (a VDSL nem az)
 - Könnyen telepíthető, vonzó technológia a szolgáltatók részére

ADSL kompatibilitás

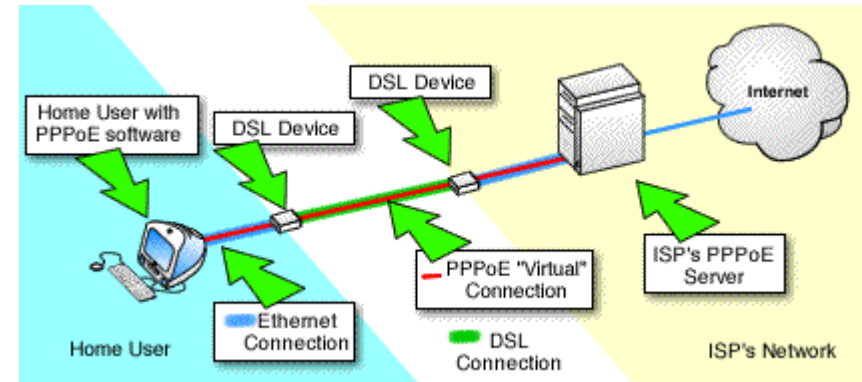


G.fast

- A legújabb DSL szabvány (2014)
 - 106 MHz-es frekvenciatartomány (később 212 MHz)
 - 150 Mbit/s-tól 1 Gbit/s-ig
 - Néhány száz méter, FTTB kiegészítésre
- A korábbi xDSL szabványoktól eltérően nem FDD-t hanem **TDD**-t használ az upstream és downstream szétválasztásra
 - 90/10 és 50/50-es profilok kötelezőek
 - Mivel a TDD nem egy folyamatos üzemmód, akár hosszabb időre is kikapcsolhatunk egy adót és egy vevőt
 - Ha nincs küldendő adat, lehet energiát spórolni

PPP – Point-to-Point Protocol

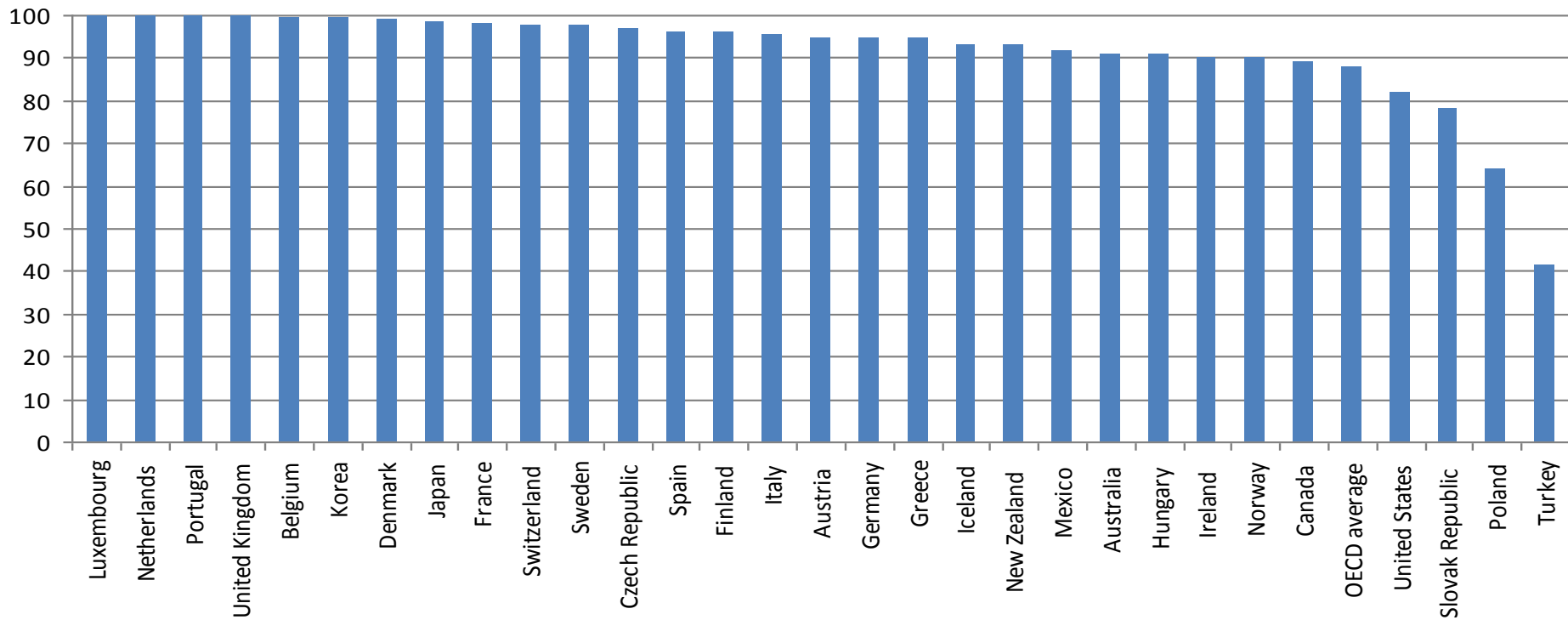
- Adatkapcsolati rétegbeli protokoll, két csomópont közvetlen összekötésére
 - xDSL-ben a felhasználó és a BRAS között
- Autentikáció, titkosítás, tömörítés
- PPPoE – PPP over Ethernet
- PPPoA – PPP over ATM



DSL lefedettség

- Egy adott ország területének hány százalékán érhető el DSL szolgáltatás, vagy a lakosság hány százaléka férhet hozzá?
 - A vezetékes telefonvonalak hány százaléka képes DSL szolgáltatásra
 - Ha túl messze a központ, a telefon OK de a DSL nem
 - Világviszonylatban elég nagy lefedettség, de (sok helyen) messze még a 100%
- Két fontos korlátozó tényező
 - Távolság
 - Nagyon gyéren lakott területeken, elszigetelt helyeken vagy nem gazdaságos, vagy nem megoldható technikailag
 - Nagy sáv szélességet igénylő alkalmazások
 - A hagyományos internetezés (web, e-mail) továbbra is fontos
 - Egyre jobban elterjednek a nagy sebességet igénylő alkalmazások (video, triple play)
 - Ezeknek a támogatása nagysebességű DSL technológiákkal fontosabb lehet mint a hagyományos „lassú” DSL megoldások terjesztése

DSL lefedettség (OECD, 2009)



Nemzetközi szabályozás

- A világ számos országában a kormány beavatkozott a piaci viszonyokba, az elérhető árú szélessáv elterjesztése érdekében
 - Közvetlen részvétel a hálózatok kiépítésében és üzemeltetésében
 - Szabályozások egy egészséges konkurenciára épülő szélessávú szolgáltatás piac megteremtésére
 - Megszüntetni a **monopolhelyzetben levő („incumbent”) szolgáltatók** privilégiumait (pl. Matáv)
- Egy új szolgáltató betörése egy tradicionálisan monopol-helyzetben lévő szolgáltató által dominált piacra nehéz
 - Korlátozott méretű piac
 - Viszonylag lassan megtérülő, nagyméretű befektetések

Nemzetközi szabályozás

- A versenyhelyzet kialakításához meg kell könnyíteni egy új szolgáltató betörését
 - A hosszú engedélyezési procedúrák egyszerűsítése
 - A külföldi befektetőkkel szembeni korlátozások feloldása
 - Adózási kedvezmények
 - Az új infrastruktúra kiépítését vállaló szolgáltatók támogatása
 - kedvező megoldás a hosszú távú piaci versenyhelyzet kialakulására
 - Szabad hozzáférés biztosítása a hálózathoz

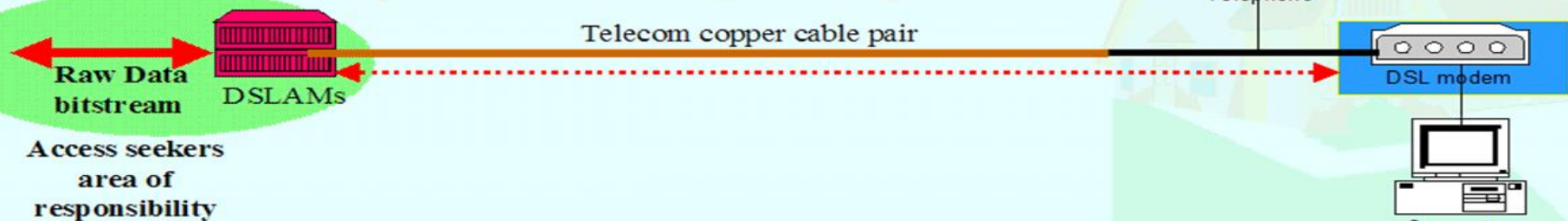
Local Loop Unbundling - LLU

- Az infrastruktúrával rendelkező szolgáltatókat kényszerítik, hogy biztosítsák más potenciális szolgáltatók részére a szabad hozzáférést a saját hálózatukhoz
 - legfőképpen a helyi előfizetői hurokhoz való hozzáférés
 - korrekt, non-diszkriminatív alapon és elérhető áron
- Többféle megoldás
 - *a helyi hurok teljes átengedése*
 - a konkurens szolgáltató teljes mértékben rendelkezik a vezetékekkel, úgy a hangátvitelt mint az adatátvitelt tekintve
 - *a helyi hurok megosztása*
 - a konkurens szolgáltató vagy a helyi hurok hangátvitelt biztosító részével, vagy az adatátvitelt biztosító résszel rendelkezik
 - *bitfolyam alapú hozzáférés*
 - az incumbent szolgáltató kiépít egy nagysebességű átvitelt biztosító vonalat a felhasználóhoz, és biztosítja a konkurens szolgáltatók hozzáférését ehhez a vonalhoz
 - A vonal technikai karbantartása és a szolgáltatás üzemeltetése továbbra is az „*incumbent*” szolgáltató hatáskörébe tartozik

LLU megvalósítások

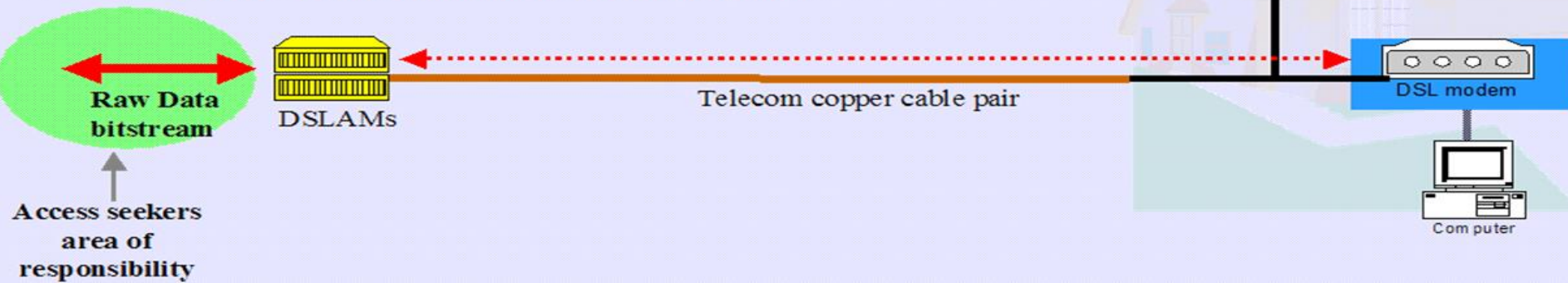
Option 1: Local loop unbundling

- Access seeker provides electronics (DSLAMs)



Option 2: Bitstream unbundling

- Telecom provides electronics (DSLAMs)
- Access seeker connects to the data stream



Local Loop Unbundling - LLU

- A világ számos országában alkalmazták, változó sikerrel
 - Japánban NTT West és NTT East az incumbent operátor
 - 97-ben bevezették az LLU-t, kevés sikerrel
 - Eredmény: 2000 végén még csak 70.000 ADSL vonal
 - 2000-ben megszigorítottak a szabályokat
 - Csökkentek az LLU-ra kiszabható árak
 - Csökkent az időrés melyen belül az „incumbent” szolgáltató köteles volt a hozzáférést lehetővé tenni
 - Eredmény: 2003 elején 6.5 millió ADSL vonal
 - » Az ADSL piac 70%-a a konkurens cégeknél
 - Az EU-ban nem volt ilyen sikeres

Hálózati Technológiák és Alkalmazások

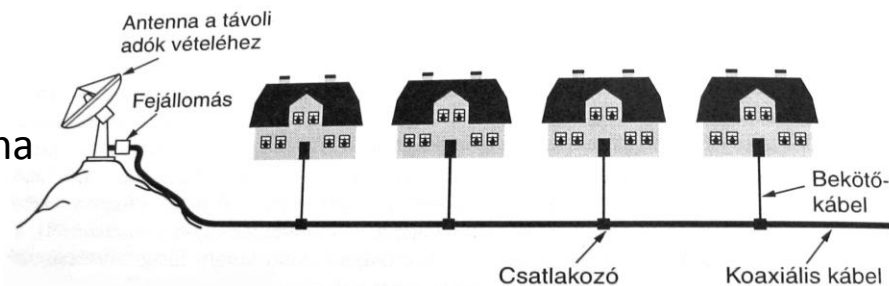
Vida Rolland
BME TMIT

2016. szeptember 16.



Kábel TV

- Ötlet a 40-es évek végén
 - Jobb vétel ott, ahol a hagyományos antennák nem nyújtottak megfelelő minőséget
- Községi antennás televízió
 - **Community Antenna Television – CATV**
 - Egy dombtetőn elhelyezett nagy antenna
 - Erősítő fejállomás (head end)
 - Koaxiális kábel
- Családias üzletág, bárki telepíthetett ilyen szolgáltatást
 - Ha több előfizető, újabb kábelek és erősítők
- Egyirányú átvitel, a fejállomástól a felhasználók felé



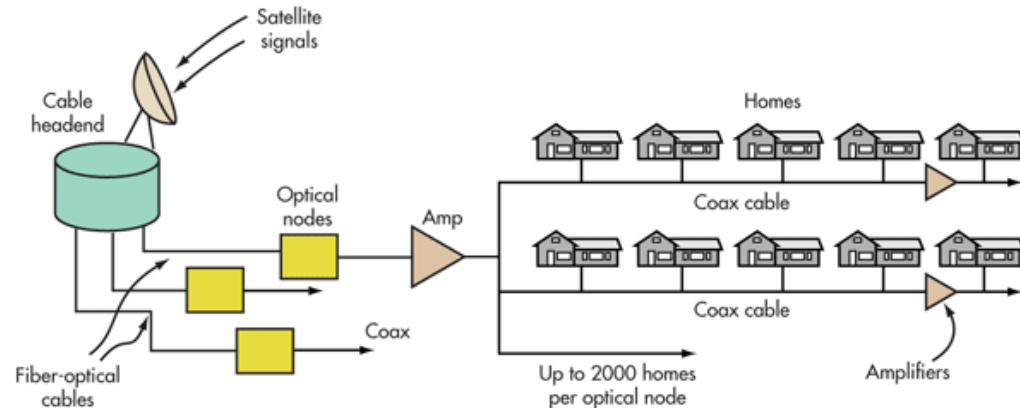
A kábeltévé fejlődése

- 1970-re több ezer független rendszer
- 1974-ben elindul az HBO, kizárólag kábelen
 - Több új kábeles csatorna – hírek, sport, főzés, stb.
- Nagyvállalatok elkezdik felvásárolni a létező kábelhálózatokat, új kábeleket fektetnek le
 - Kábelek a városok között a hálózatok egyesítésére
 - Hasonló ahhoz, ahogy a távközlő iparban a század elején összekötötték a helyi központokat a távolsági hívások miatt
- Később a városok közötti kábeleket nagy sáv szélességű fényvezető szálakra cserélik



HFC rendszer

- HFC - Hybrid Fiber Coax
 - Fényvezető-koax hibrid rendszer
 - Fényvezető szálak a nagy távolságok áthidalására
 - Koaxiális kábel a házakhoz
 - Fényvezető csomópont (fiber node)
 - Elektrooptikai átalakító
 - a fényvezető és villamos rész közötti csatolásnál
 - Egy fényvezető szál több koax kábelt is táplálhat
 - Sokkal nagyobb sáv szélesség



Internet a kábeltévéen



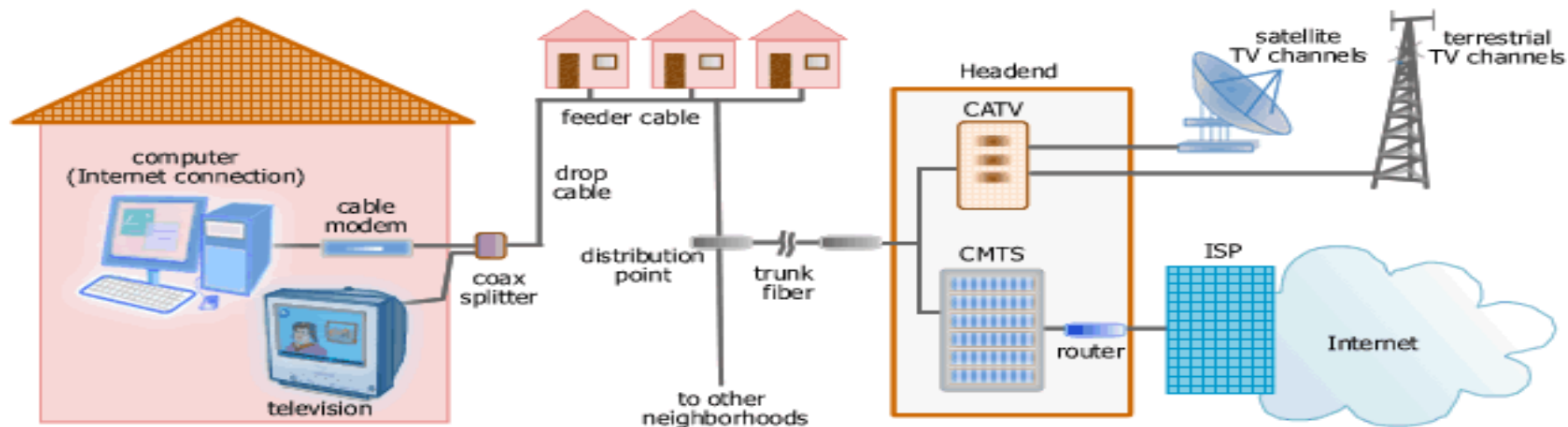
Internet a kábeltévéen

- A kábelhálózat üzemeltetők elkezdtek bővíteni a szolgáltatásaikat
 - Internetelés
 - Telefonszolgáltatás
- Át kell alakítani a hálózatot
 - Az egyirányú erősítőket kétirányú erősítőre kell cserélni mindenhol
 - A fejállomást fel kell fejleszteni
 - Egy buta erősítőből egy intelligens digitális számítógéprendszer
 - Nagysebességű optikai szálakat csatlakoztat egy internet szolgáltató (ISP) hálózatához
 - Cable-Modem Termination System (CMTS)
 - A koax kábel osztott közeg, több ház egyszerre használja
 - A telefonhálózatban mindenki rendelkezik saját érpárral (előfizetői hurok)
 - A TV műsorok szórásánál ez nem fontos
 - minden műsort ugyanazon a kábelen szórnak, mindegy hogy 10 vagy 10.000 ember nézi azt egyszerre
 - Internetezésnél óriási különbség ha 10 vagy 10.000 felhasználó
 - Ha valaki letölt egy nagy fájlt, a többieknek nem marad sávszél
 - Másfelől a koax kábel sokkal nagyobb sávszélt biztosít mint a sodrott érpár

Internet a kábeltévéen

- Megoldás: több darabra osztunk egy hosszú kábelt
 - Minden szakaszt közvetlenül egy fiber node-hoz kötünk
 - A fejállomás és a fiber node-ok között a sávszélesség lényegében végtelen
 - Ha nincs túl sok felhasználó egy szakaszon, a forgalom kezelhető marad
 - Tipikusan 500-2000 ház egy szakaszon
 - További felosztás várható ahogy nő az előfizetők száma és a forgalom

Internet a kábeltévén

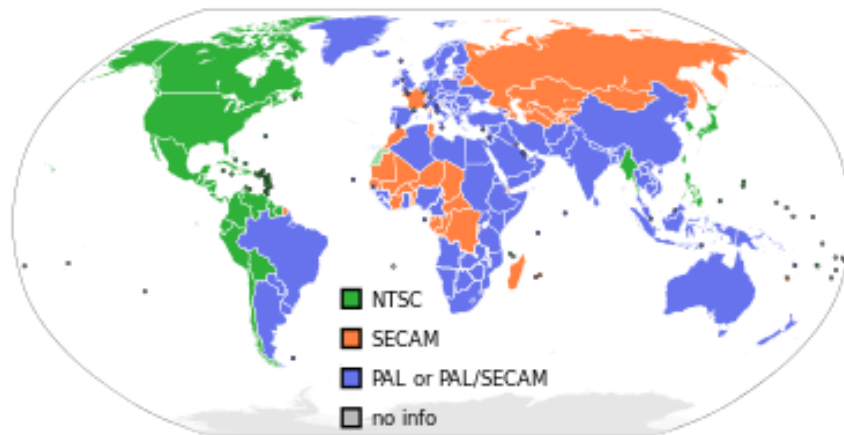


Spektrumkiosztás

- A kábelhálózatot nem lehet (egyelőre) kizárólag internetezésre használni
 - Több a tévénéző mint az internetező ügyfél
 - A városok szabályozzák mi mehet a kábelben, a tévészolgáltatás kötelező
 - Fel kell osztani a frekvenciákat a TV és az internet elérés között
- USA, Kanada
 - FM rádió: 88 – 108 MHz
 - kábeltévé-csatornák: 54 – 550 MHz
 - 6 MHz széles csatornák, védősávval együtt
 - **NTSC - National Television System Committee**
 - Felbontás: 720 x 480, 29.97 fps

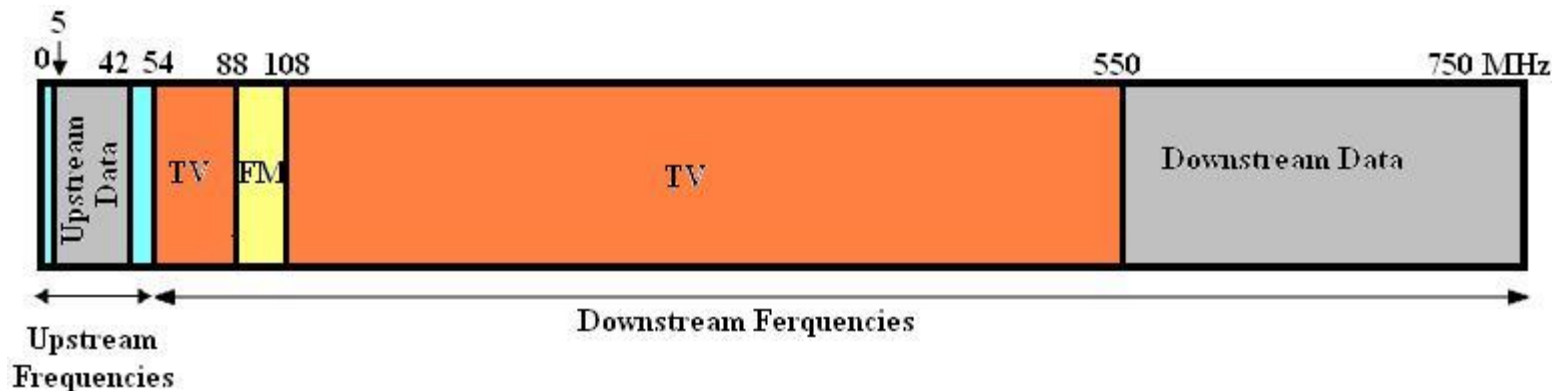
Spektrumkiosztás

- Európa
 - TV sávok alsó határa 65 MHz
 - 6-8 MHz széles csatornák
 - **PAL és SECAM** rendszerek nagyobb felbontási képessége miatt
 - PAL - Phase Alternating Line
 - SECAM - Système Electronique Couleur Avec Mémoire
 - Felbontás: 768 x 576, 25 fps
 - A sáv alsó részét nem használják



Spektrumkiosztás

- Modern kábeleken 550 MHz felett is lehetséges az adatátvitel, gyakran 750-800 Mhz felett is
 - Megoldás: feltöltés 5 – 42 MHz között (Európában 5 - 65 MHz)
 - A spektrum felső végén lévő frekvenciák a letöltéshez



Aszimmetrikus átvitel

- A TV és rádió mind lefele halad
 - A fejállomástól a felhasználó felé
 - Felfele olyan erősítők melyek az 5-42 MHz-es tartományban működnek
 - Lefele az 54 MHz feletti tartományban működő erősítők
 - Aszimmetrikus rendszer, nagyobb downstream mint upstream
 - Ezt itt műszaki okok befolyásolják, nem úgy mint az ADSL-nél

Moduláció

- Minden 6-8 MHz-es csatornát **64-QAM**-el modulálnak
 - Quadrature Amplitude Modulation
 - Ha kivételesen jó minőségű kábel, akkor 256-QAM
- 6 MHz-es csatornán 64-QAM-el → kb. 36 Mbps
 - A fejlécek nélküli sávszél 27 Mbps
 - 256-QAM-el kb. 39 Mbps
 - Európában magasabb sávszél, a 8 MHz-es csatorna miatt
- A feltöltési csatornán a 64-QAM nem ilyen jó
 - Túl sok zaj a felszíni mikrohullámú rendszerek, CB-rádiók, stb. miatt
 - Citizen Band – walky-talky
 - QPSK moduláció
 - Quadrature Phase Shift Keying
 - Csak két bit szimbólumonként (a 64-QAM-nél 6, a 256-QAM-nál 8)
 - Sokkal nagyobb az upstream és a downstream közötti különbség

Kábelmodem

- A kábelen jövő analóg jelet digitálissá alakítja és fordítva
 - **MO**dulál és **DE**Modulál
- Két interfész – egy a PC és egy a kábelhálózat felé
 - A modem és a PC között lehet Ethernet/USB/WLAN



"I'VE MET SOMEONE WITH A FASTER MODEM."

CN
COLLECTION

Kábelmodem

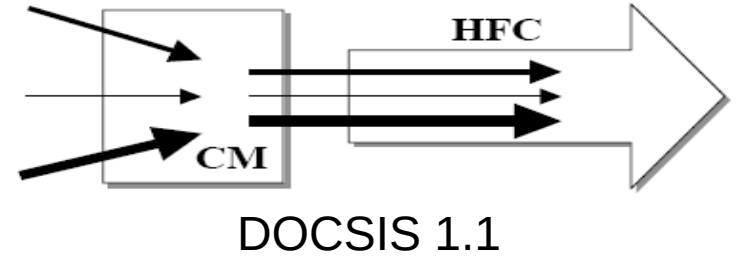
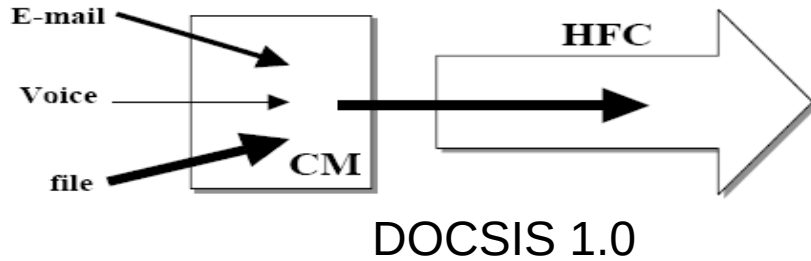
- A kezdetekben minden hálózatüzemeltetőnek saját modem-je, melyet egy technikus telepített
 - Nyílt szabvány kellett
 - Versenyhelyzethez vezet a modemek piacán
 - Csökkennek az árak
 - Ösztönzi a szolgáltatás terjedését
 - Ha a felhasználó telepíti a modemet, nem kell kiszállási költség
- CableLabs
 - A legnagyobb kábelszolgáltatók szövetsége
 - DOCSIS szabvány
 - Data Over Cable Service Interface Specification
 - EuroDOCSIS – európai változat
 - Sokan nem örültek neki
 - Nem tudták tovább drágán bérbe adni modemjeiket a kiszolgáltatót előfizetőknek



DOCSIS

- DOCSIS 1.0 (1997)
 - RF Return
 - Kétirányú kommunikáció biztosítása
 - Telco Return
 - Dial-up kapcsolat az upstream forgalomra
 - Nem kell módosítani az infrastruktúrát, egyirányú kommunikáció a kábelben
 - A modemárak 300\$-ról (1998) <30\$-ra estek
- DOCSIS 1.1 (1999)
 - VoIP, gaming, streaming
 - Kompatibilis a DOCSIS 1.0-val
 - Szolgáltatásminőségi osztályok (QoS) támogatása

DOCSIS

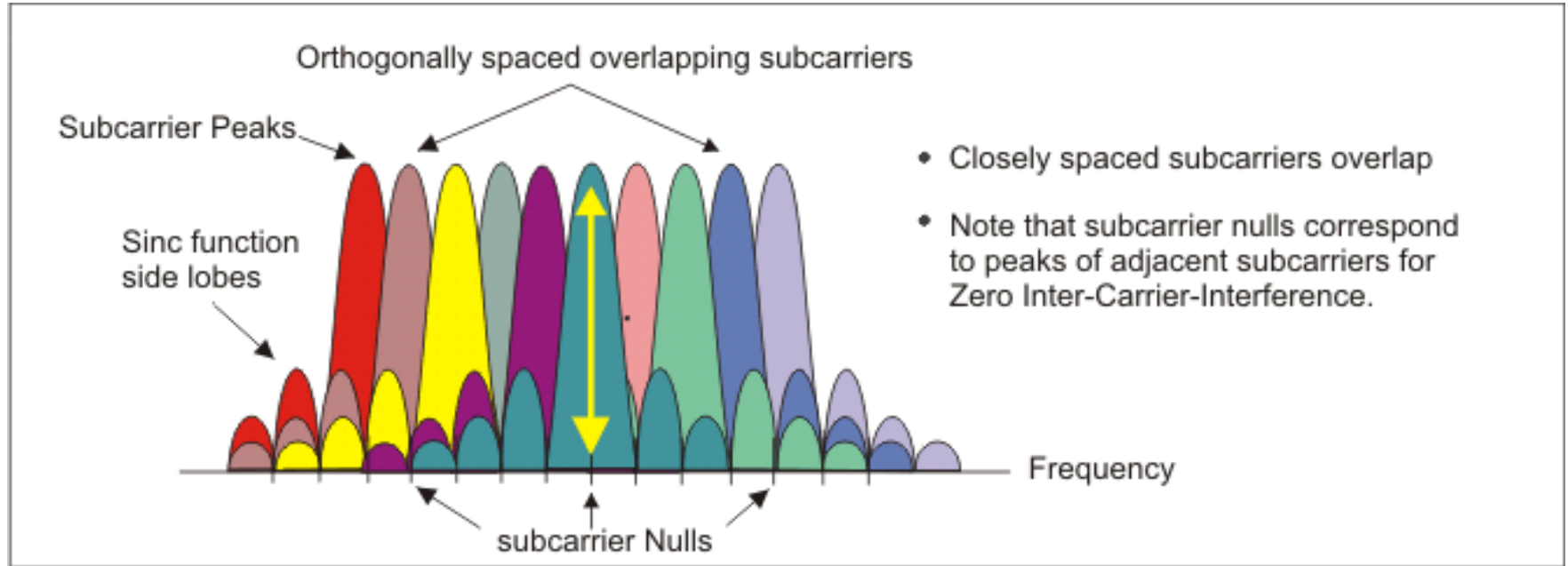


- A DOCSIS 1.0-ban minden szolgáltatás „best effort” alapon versenyez a feltöltési sávszélért
- A DOCSIS 1.1-ben minden szolgáltatáshoz QoS garanciákat lehet rendelni

DOCSIS

- DOCSIS 2.0 (2002)
 - Kapacitás szimmetrikus szolgáltatásokhoz
 - Nagyobb upstream kapacitás mint a DOCSIS 1.0-ban (x6) és a DOCSIS 1.1-ben (x3)
 - QPSK helyett 32-QAM, 64-QAM vagy 128-QAM az upstream részen is
 - TDMA helyett TDMA és S-CDMA a MAC rétegben
- DOCSIS 3.0 (2006)
 - 160 Mbps downstream, 120 Mbps upstream
 - Channel bonding
 - Több csatornát párhuzamosan használhat egy felhasználó
- DOCSIS 3.1 (2013)
 - 10 Gbps downstream, 1 Gbps upstream, 4096 QAM moduláció
 - 6-8 MHz széles csatornák helyett 20-50 KHz-s keskeny csatornák, OFDM (Orthogonal Frequency Division Multiplexing)
 - Channel bonding – akár 200 MHz széles spektrum

OFDM



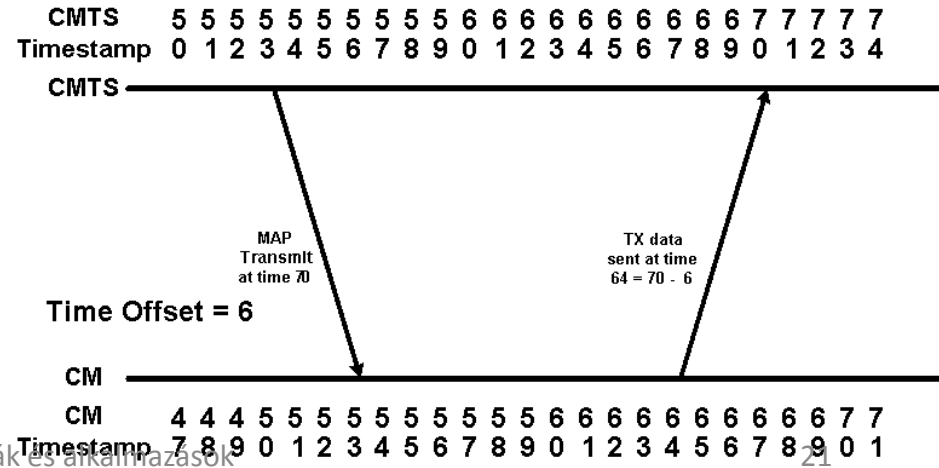
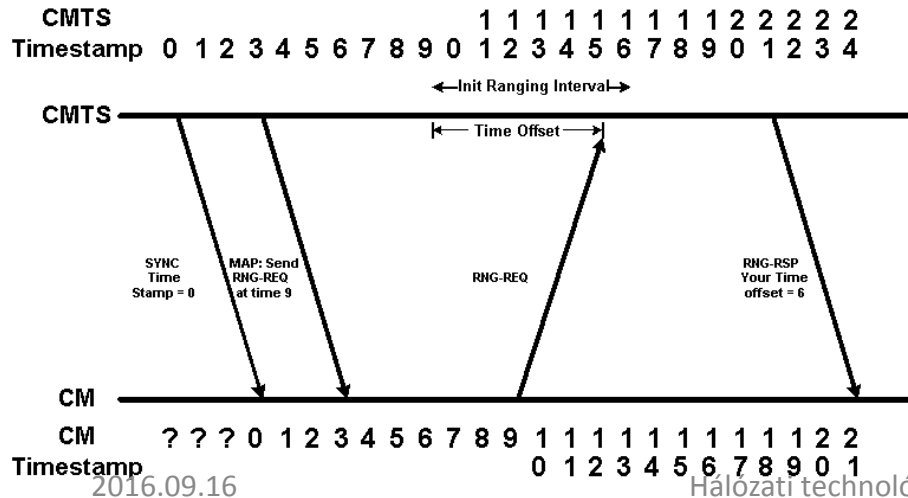
OFDM Signal Frequency Spectra

Csatlakozás

- Csatlakozásnál a modem pásztázni kezdi a letöltési csatornákat
 - A CMTS egy speciális csomagban időnként elküldi a rendszer paramétereit az újonnan kapcsolódó modemek részére
 - A modem bejelentkezik a CMTS-nél
 - A CMTS kijelöli az új modem feltöltési és letöltési csatornáit
 - Ezt később lehet változtatni, például a terhelés kiegyenlítése miatt
 - Több modem ugyanazon a feltöltési csatornán
 - Az első csomag a modemtől az ISP-hez megy
 - IP címet kér, DHCP protokollon keresztül
 - Dynamic Host Configuration Protocol
 - A pillanatnyi pontos időt is megkapja a CMTS-től

Versenyhelyzetes feltöltés

- A modem megméri milyen távol van a fejállomás
 - Távolságbecslés (ranging) – mint a ping
 - Szükség van rá az időzítések miatt



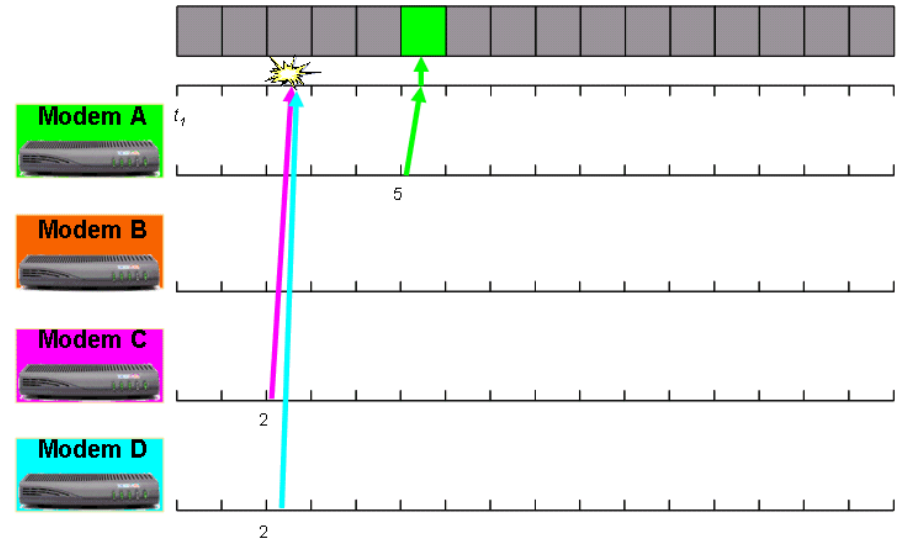
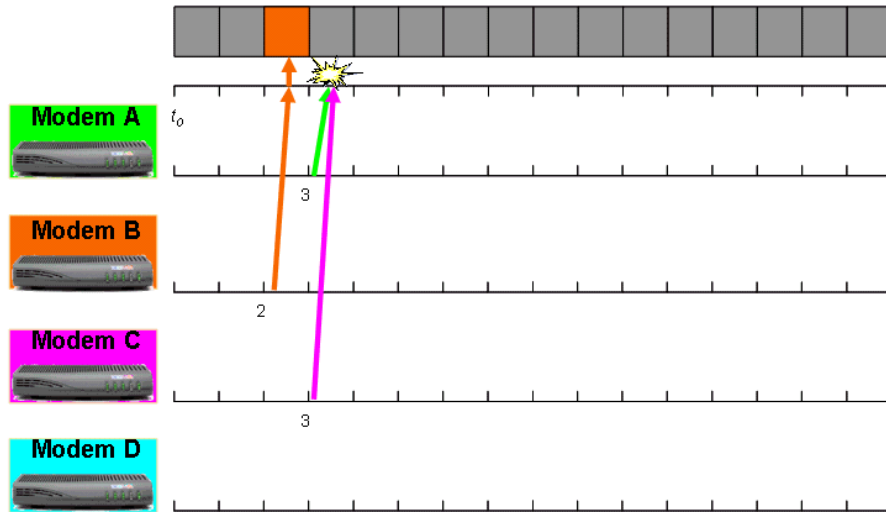
Versenyhelyzetes feltöltés

- A feltöltési csatornát mini időszkeletekre osztják (minislot) – **FDD/TDMA**
 - Minden felfele haladó csomag egy vagy több minislot-ban
 - A minislot-ok hossza hálózatonként más és más
 - Tipikusan 8 byte felhasználói adat egy minislot-ban
- A fejállomás rendszeresen bejelenti mikor új minislot-csoport kezdődik
 - A kábelen való terjedés miatt nem egyszerre hallják meg a modemek
 - Mindenki ki tudja számítani mikor volt az első minislot kezdete
 - Minden modemhez hozzárendelve egy speciális minislot (**Bandwidth Request Slot**) melyben feltöltési sáv szélességet igényelhet
 - Több modem lehet ugyanazon a minislot-on

Versenyhelyzetes feltöltés

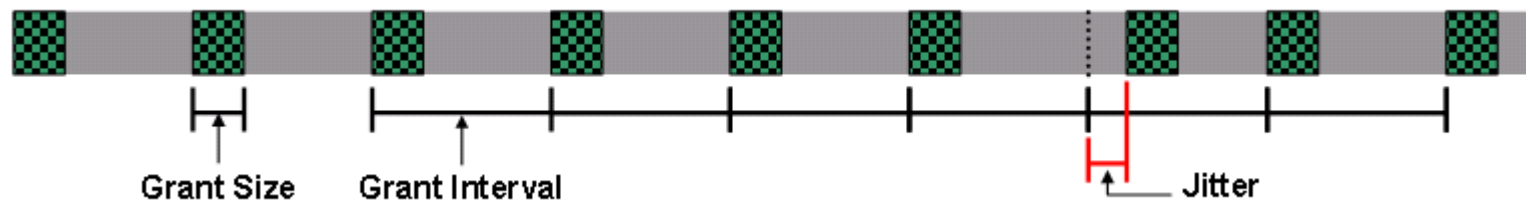
- Ha a modem csomagot akar küldeni, szükséges számú minislot-ot igényel
 - Ha a fejállomás elfogadja, a nyugtában megmondja mely minislot-okat jelölte ki
 - Ha további csomagokat akar küldeni, a fejlécben új minislot-okat kérhet
 - Ha az igényléskor ütközés, nincs nyugta
 - Vár egy véletlen ideig és újra próbálkozik
 - Minden egymás utáni kudarc után a max. idő duplázódik

Versenyhelyzetes feltöltés



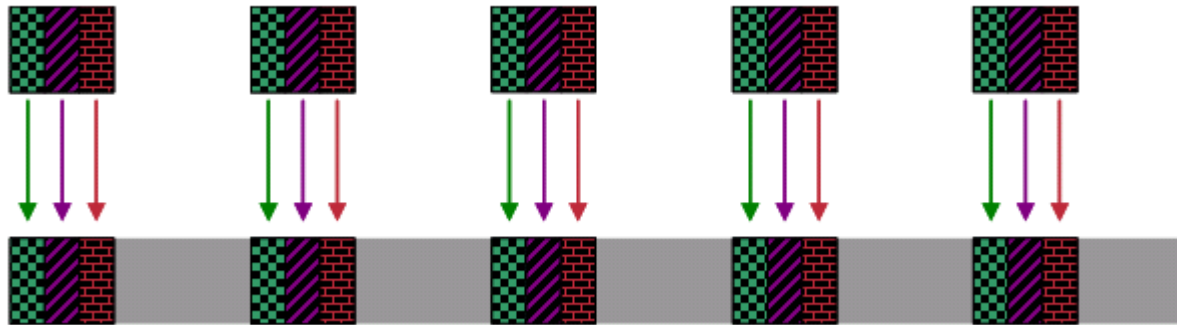
Szolgáltatásminőség biztosítása

- Különböző alkalmazásoknak különböző QoS követelmények
- CBR – Constant Bit Rate (pl. VoIP)
 - **Unsolicited Grant Services (UGS)**
 - Nem kell folyamatosan igényelni időkeretet



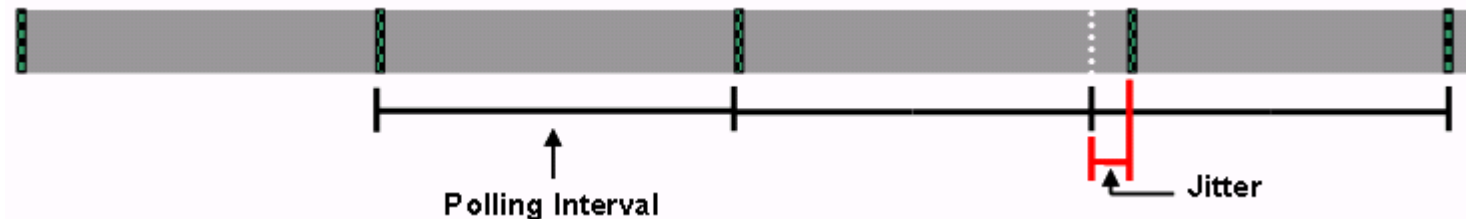
Admission Control

- UGS kéréseket csak a lehetőségek függvényében fogad el
 - Kellenek szabad időkeretek maradjanak másfajta forgalomnak



Szolgáltatásminőség biztosítása

- rt-VBR (Real Time Variable Bit Rate)
 - pl. live video stream
 - **Real Time Polling Service (RTPS)**
 - Csak az az alkalmazás/modem használhatja azt a Bandwidth Request Slot-ot
 - Biztosan tud igényelni, nincs ütközés



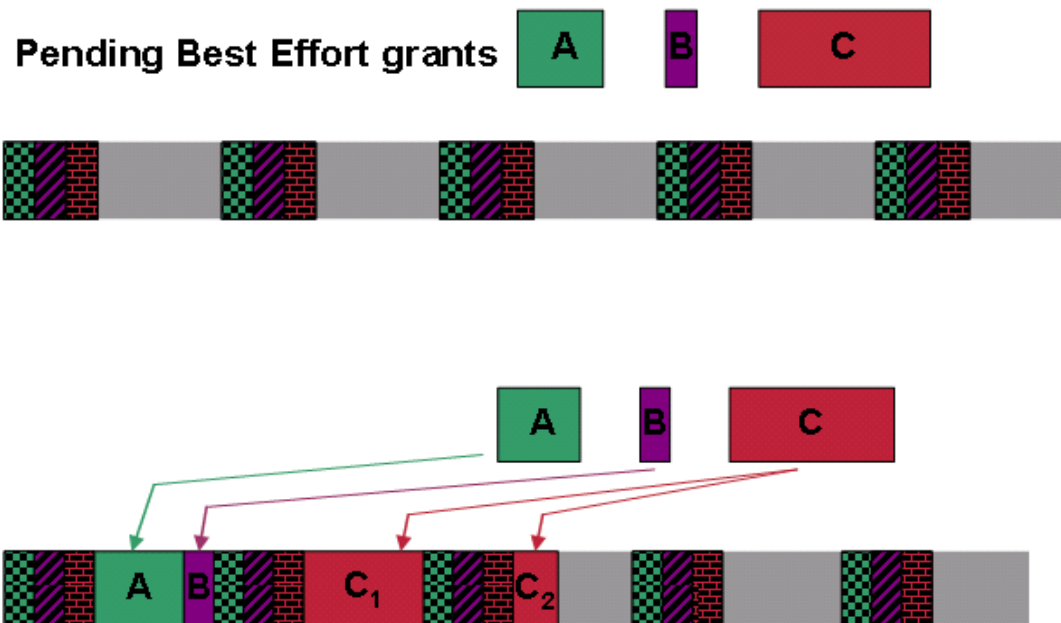
Szolgáltatásminőség biztosítása

- **Unsololicited Grant Service with Activity Detection (UGS-AD)**
 - Akkor működik UGS módban, ha van küldénivalója
 - Ha átmenetileg nincs, átvált RTPS módba
 - Ha újból szükség van rá, vissza tud váltani ismét UGS-be
 - Pl. VoIP with Voice Activity Detection (VAD)
- **Non-Real Time Polling Service (nRTPS)**
 - nrt-VBR forgalomhoz
 - A lekérdezési intervallumok nem folyamatosak

Szolgáltatásminőség biztosítása

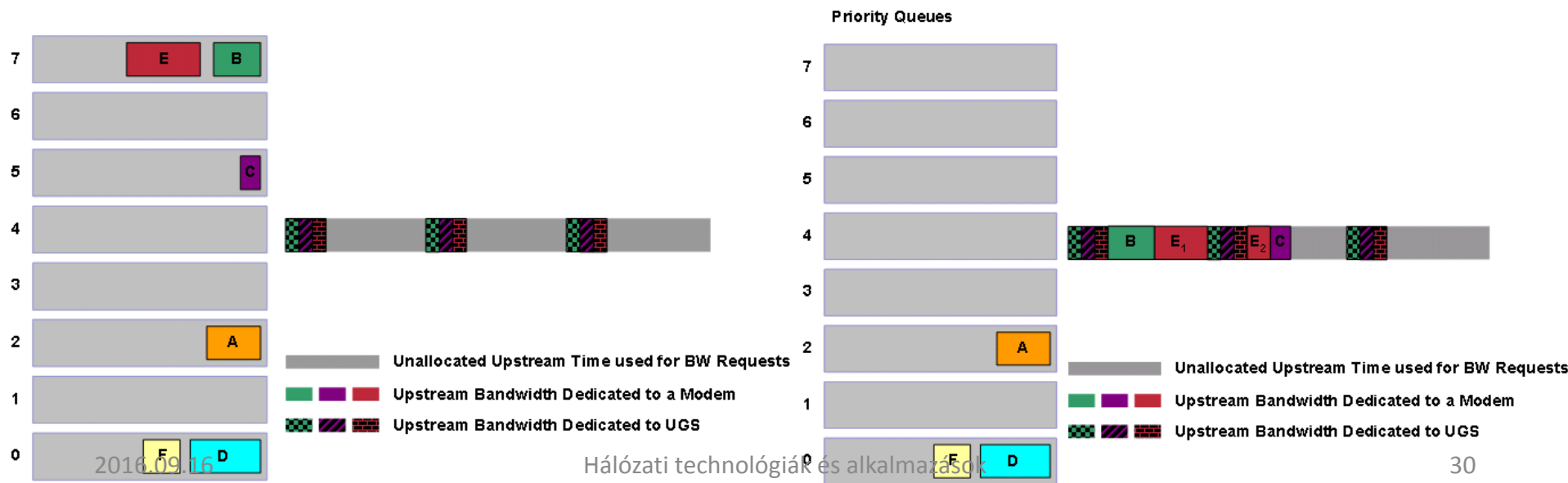
- **Best Effort Grants (BEG)**

- Nincsenek szoros követelmények a késleltetésre és a késleltetés ingadozásra
- Fragmentation – ha szükséges, az igényelt időkereteket lehet darabolni
 - Több fejléc, de (néha) megéri

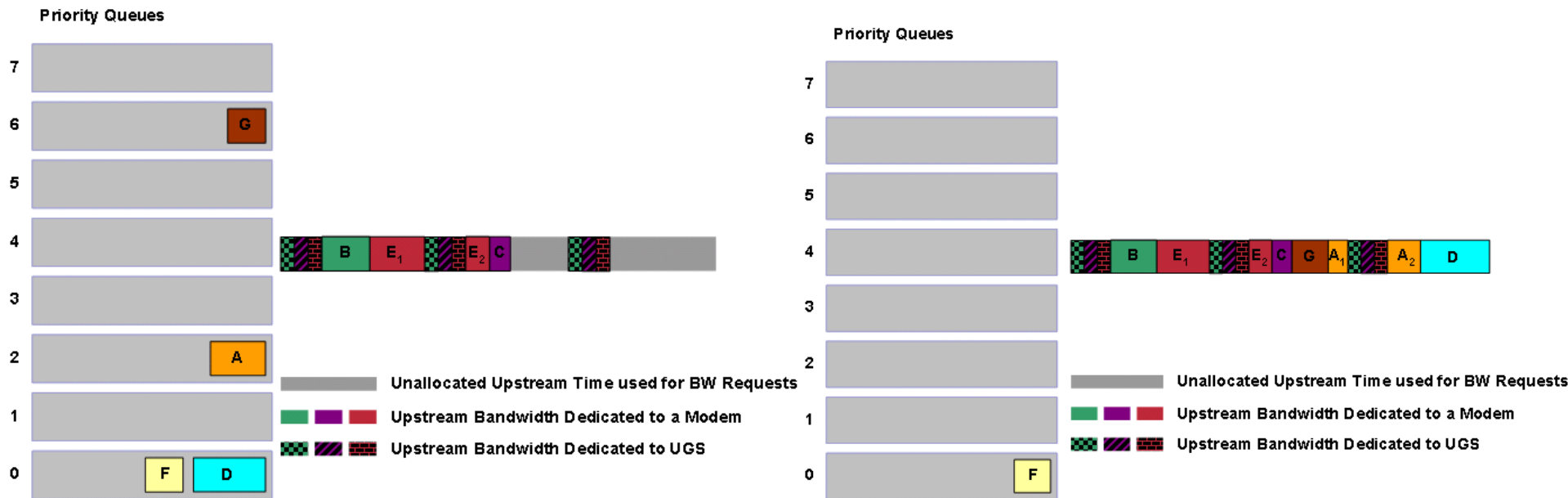


Scheduling (ütemezés)

- Prioritási sorok – alapból 8 (0-tól 7-ig)
 - A magasabb prioritási sorokat hamarabb szolgálja ki



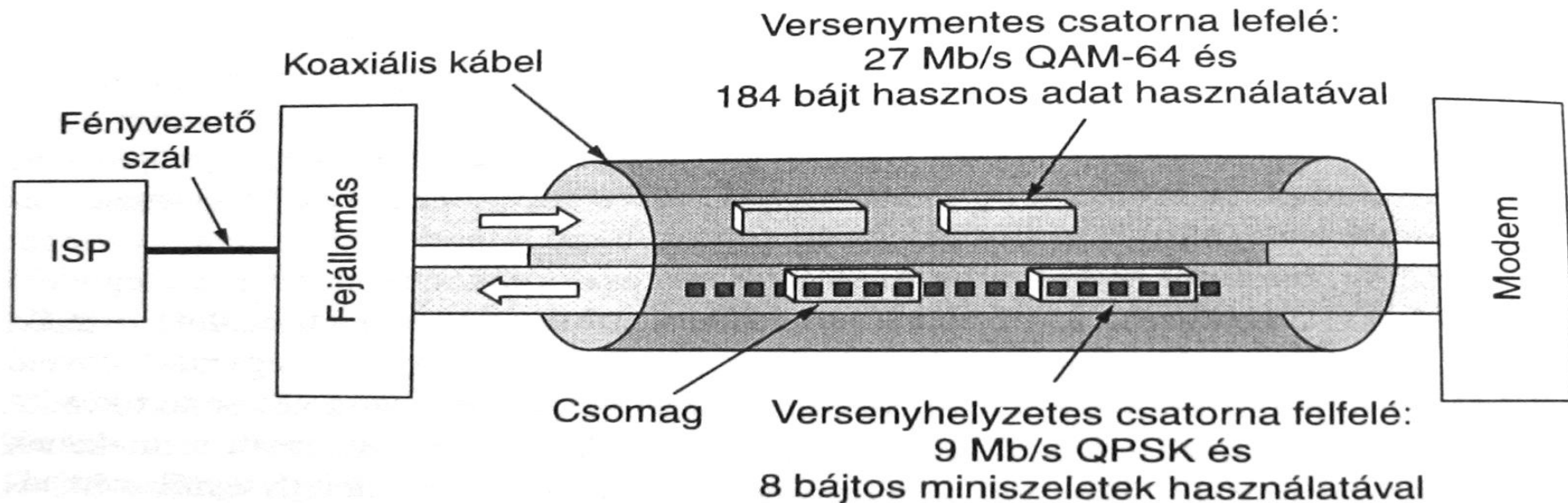
Scheduling (ütemezés)



Versenymentes letöltés

- Letöltésnél csak egy küldő, a fejállomás
 - Nincs versenyhelyzet, nincs szükség minislot-okra
 - Nagyméretű forgalom lefelé
 - Nagyobb, 204 byte-os rögzített csomagméret
 - Ebben Reed-Solomon hibajavító kód
 - 184 byte a felhasználói adatoknak

Le- és feltöltés a kábelen



Ethernet – első rész

Moldován István



Budapest University of Technology and Economics

Department of
Telecommunications and Media Informatics



- A *lokális hálózat* azonos szinten elhelyezkedő gépek összességét jelenti.
 - Ezek az adatkapcsolati szintű működés szempontjából azonos jogú egységek. – logikai szinten
- Ezeket szokás *többszörös hozzáférésű hálózatoknak* is nevezni, mert több, azonos joggal rendelkező egység fér hozzá egy adott, közös elérésű erőforráshoz.
 - Ez például a sín topológia esetén maga a sín.
- Annak érdekében, hogy ehhez az elosztott erőforráshoz mindenki igazságosan tudjon hozzáférni, *elosztott protokollok* alkalmazása szükséges.

Az Ethernet őse: Aloha



BME-TMIT

- Hawai szigetek közti rádiózásra fejlesztették
- Több állomás egymással való beszélgetésére
- Algoritmus:
 - Ha van adat, elküldi
 - Vár a nyugtára. A vevők minden csomagot nyugtáznak
 - Ha nem jön ACK, az ütközést jelent. Random idő múlva újraküldi a csomagot

Az Ethernet fejlődése



BME-TMIT

Aloha

Az ős.

**Slotted
Aloha**

Újítás: csak adott időpontokban küldhet (slots)

CSMA

CSMA = Carrier Sense Multiple Access
Újítás: Először ellenőrzi, hogy van-e adás, és csak akkor küld ha nincs

CSMA/CD

CD = Collision Detection

Újítás: Leállítja a küldést ha ütközést észlel (ilyen az Ethernet)

Hogyan kezdődött...



- 1972 Dr Robert Metcalfe

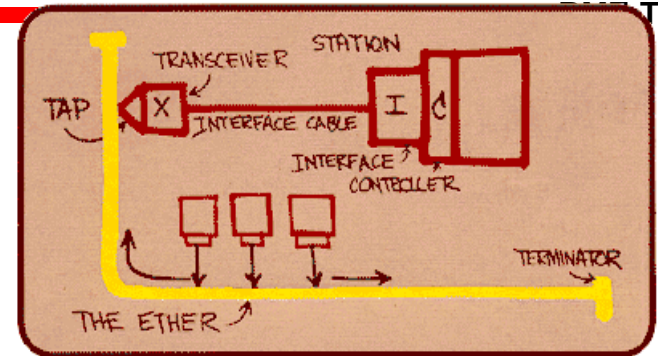
1976 az Ethernet nevet először használták

- Az eredeti DIX Ethernet V2 standard

- 1982 (DEC-Intel-Xerox)

- Az IEEE 802.3

- 10Base-5 - 1983
 - 10Base-2 - 1988
 - 10Base-T - 1990



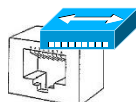
Az első Ethernet ábra

- Ethernet fejlődés –10 Mega után
 - 100BASE-TX (Fast Ethernet)
 - IEEE 802.3u: 1995
 - 1000BASE-X (Gigabit Ethernet)
 - IEEE 802.3z: Június 1998
 - 1000BASE-T (Gigabit on Copper)
 - IEEE 802.3ab Június 1999
 - 10 Gigabit Ethernet (IEEE 802.3ae)
 - IEEE 802.3ae 2002 nyarán

Ethernet az OSI Modellben



BME-TMIT

OSI MODEL		TCP / IP		Exchange Unit
7	 Application Layer Communication Type: E-mail, FTP, client/server...	FTP,	Application Protocol	APDU
6	 Presentation Layer Encryption, data conversion: ASCII to EBCDIC, BCD to binary...	HTTP,		PPDU
5	 Session Layer Starts, stops sessions. Maintains orders.	SMTP,	Presentation Protocol	SPDU
4	 Transport Layer Ensures delivery of entire file or message.	DNS,		Segments
3	 Network Layer Routes data to different LANs, WANs based on Network address.	Telnet	Session Protocol	Packet
2	 Data Link (MAC) Layer Transmit packets from node to node based on station address.	TCP, UDP		Frame
1	 Physical Layer Electrical signals and cabling.	IP (ICMP, ARP, RARP)	Transport Protocol	Bits
		Ethernet IEEE 802.3		

OSI = **O**pen **S**ystem **I**nterconnection

- Az Ethernet által lefedett rétegek:
 - Physical Layer (Layer 1) – teljesen lefedi
 - Data Link layer (Layer 2) – részlegesen lefedi

IEEE	Leírás
802.2	Logical Link Control (LLC) szabvány. Egy általános interfészt határoz meg a hálózati réteg (IP, IPX,...) és az adatkapcsolati réteg (Ethernet, Token Ring,...) közt
802.3	CSMA/CD hálózat specifikáció. Meghatározza a csomag formátumot, kábelevezést és a jelzési rendszert.

IEEE 802 Csoportok



BME-TMIT



IEEE
Standard Boards

IEEE 802
LAN/MAN
Standard Committee

802.1
Higher Layer
LAN Protocols
Working Group

...

802.3
Ethernet
Working Group

...

802.5
Token Ring
Working Group

...

802.17
Resilient
Packet Ring
Working Group

P802.3ah
Ethernet in
the first mile
Task Force

P802.3ae
10GbE
Task Force

P802.3af
DTE Power
via MDI
Task Force

P802.3ag
Maintenance
revision #6

P1802.3rev
Conformance Test
Maintenance #1

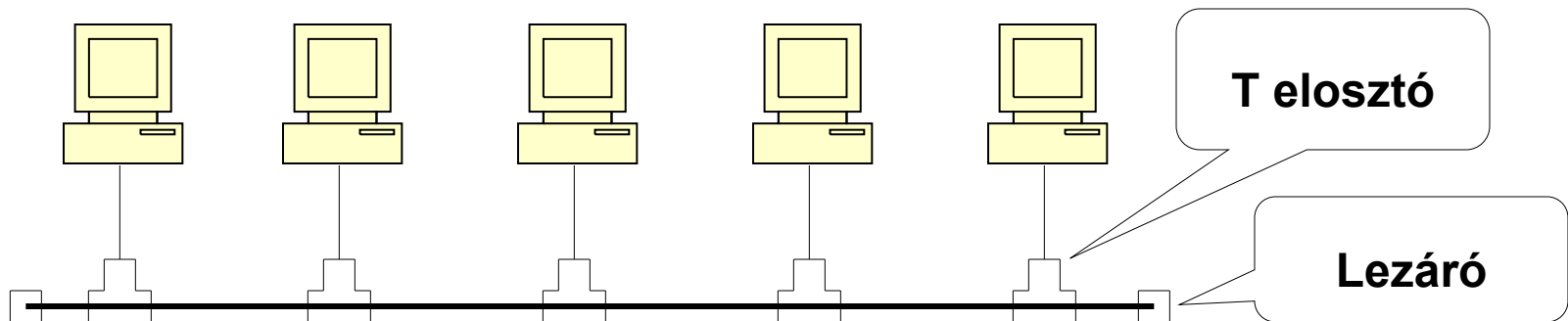
P802.3
Static
Discharge
Task Force

Fizikai kapcsolat típusok - 1



BME-TMIT

- Koax, vagy 10base2
 - 10: 10Mbps; 2: 200 méter max kábel hossz
 - Vékony koax kábelt használt, busz topológia
- Nagyobb távolság áthidalása:
 - repeater

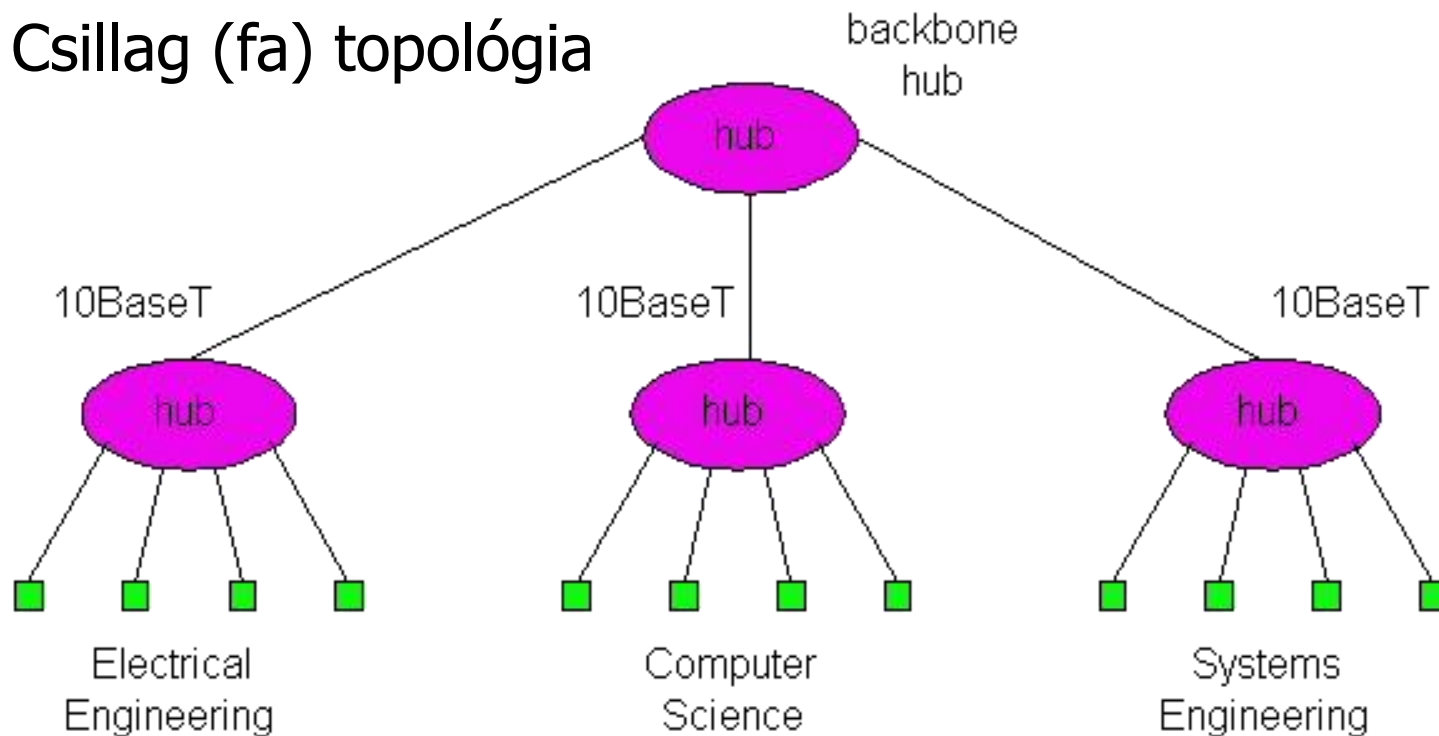


Fizikai kapcsolat típusok - 2

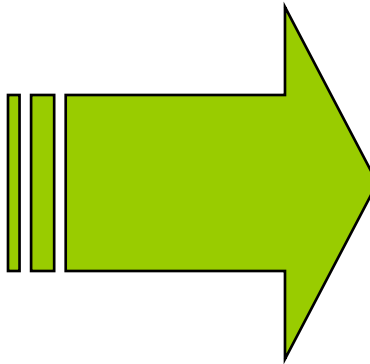
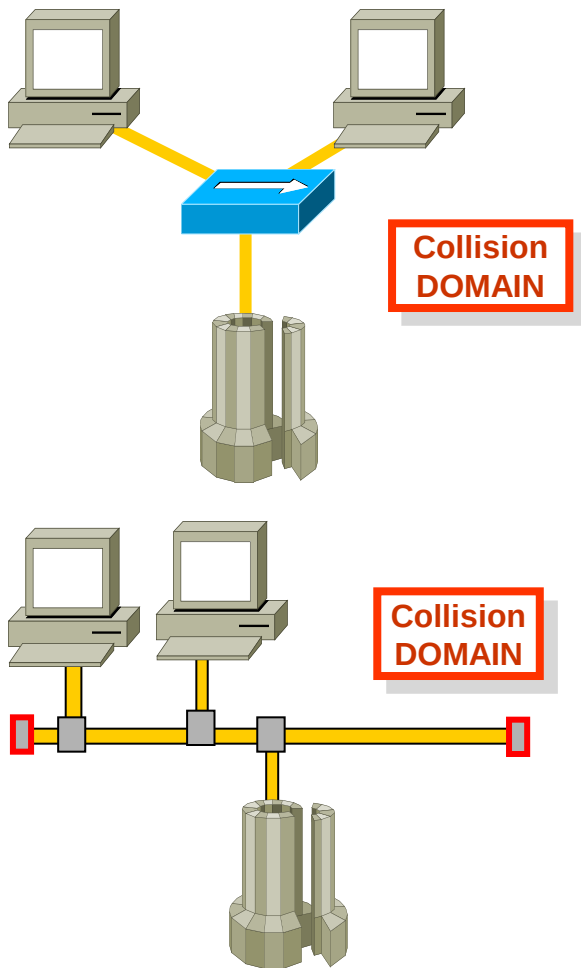


BME-TMIT

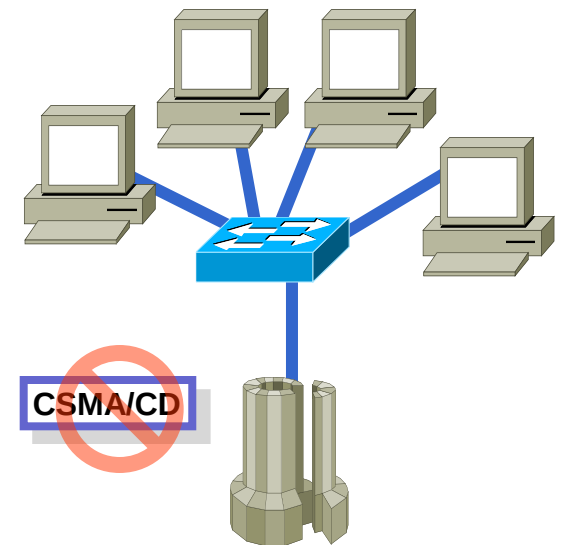
- 10BaseT és 100BaseT
 - 10 vagy 100 MBps
 - T: Twisted Pair, csavart érpár
 - Csillag (fa) topológia



Nincs több ütközés!



FDX & Microsegmentation
Nincs ütközés



*L2+ Switching - Full Duplex
CSMA/CD nem kell*

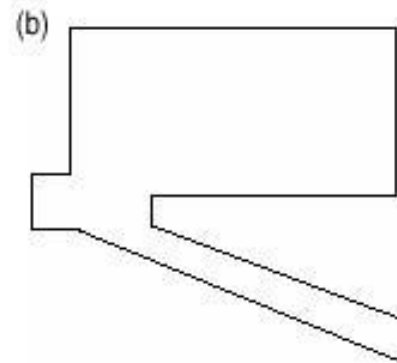
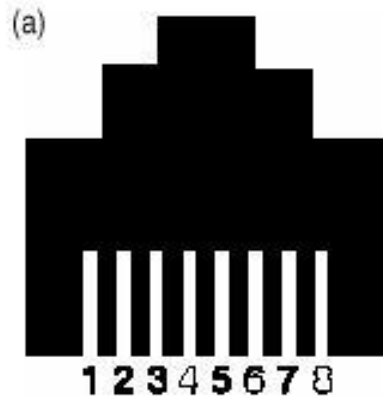
- GE: Gigabit
 - TX – csavart érpár
 - SX/LX/FX – üvegszál, különböző távolságok áthidalására
- 10GE
 - Csak üvegszál
- 802.11: WLAN
 - Ethernet az is!

UTP – Category 5



BME-TMIT

- RJ-45 dugasz



- **Láb kiosztás (10/100)**

- 1 TD+ (Transmit Data)
- 2 TD- (Transmit Data)
- 3 RD+ (Receive Data)
- 4 Nem használt

- 5 Nem használt
- 6 RD- (Receive Data)
- 7 Nem használt
- 8 Nem használt

- A GE mind a 8 szálát használja!

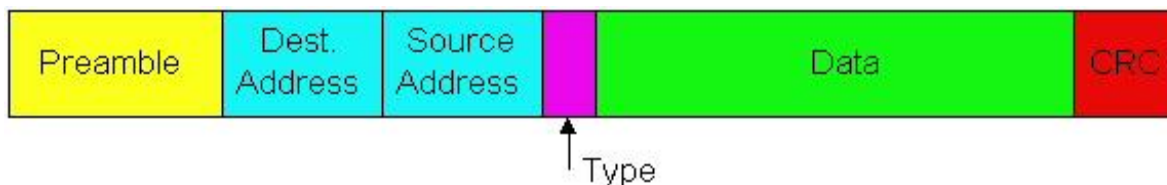
- Direkt kábel
 - A terminálok HUB-hoz való csatolására szolgál
 - A kábel mindkét vége ugyanúgy van bekötve
 - Figyelni kell hogy X- és X+ -t hasonló szín vigye
 - (pl. TD+ narancs-fehér, TD- fehér-narancs)
- Cross kábel
 - Két gép egymással csak keresztkábelben tud kommunikálni
 - A TD kivezetések az RD lábakra kell legyenek kötve

- Előnye a nagyobb távolságok áthidalásának lehetősége
 - Olcsóbb, gyors
 - Kiterjeszti a LAN hálózatot 100m –nél távolabbra
- Pont-pont típusú kapcsolat
- FX interfész, csatlakoztatás:
 - média átalakító
 - Bővíthető kapcsoló
- GE: GBIC/SFP

- Standard Ethernet formátumot használ
- Point-to-point és megosztott broadcast működést támogat
- Megosztott módban CSMA/CD-t használ
 - Csak rövid távolságot hidal át (<100m réz esetén.)
- Full-Duplex 1 Gbps sebességgel point-to-point linkeken
 - Általában üvegszálás média, nagyobb táv

- Általában a GE-képes eszközökben kiválasztható a fizikai médium
 - TX (réz), SX/LX/FX (üveg)
- GigaBit Interface Converter
 - Cserélhető, hot-swappable modul
- Small Factory Plug
 - Kisebb méretű átalakító

- Ethernet csomag fejlődés
 - I, II, 802.3 (802.2 SNAP az Ethernet II kompatibilitás miatt)
- IEEE 802.3 Data Link Control (DLC)



- A Preamble és CRC mezőket a hardver kezeli:
 - 7 bájt 10101010 melyet egy 10101011 bájt követ (szinkronizáció céljából szükséges)
- Az IEEE 802.3 kasznál LLC és SNAP keretezést

- Címek: 6 bájtosak
 - A csomagokat minden állomás fogadja, de eldobja ha nem neki címezték
- Type mező: 2 bájt
- CRC: 4 bájt, a vevő ellenőrzi és eldobja a csomagot, ha hibát detektál
- Data: maximum 1500 bájt, minimum 46 bájt
 - Maximum 9000 bájt GE esetén

Csomag formátum - 3



BME-TMIT

E.g. 0800 IPv4
86DD IPv6
0806 ARP
...

MTU	FR	4470
	FDDI	4500
	ATM	9180
	Ethernet	1500

Ethernet V2



Octets 7 4 1 6 6 2 from 46 to 1500 9.6 μ sec

IEEE 802.3



802.2 LLC
802.2 SNAP

Megkülönböztethető a V2 és 802.3

Maximum Frame Size az 1518 (decimal), vagy 0x05EE Hex
EthernetV2 Ethertype mindig nagyobb mint 0x05EF

<http://www.iana.org/assignments/ethernet-numbers>

- **Logical Link Control (LLC) Feladatai:**
 - **MAC szintű protokoll multiplexálás/demux**
 - **opcionálisan flow control, csomagvesztés esetén újraküldés kérések**
- **Service Access Point (SAP) mező szerinti szétválasztás**
 - Source és Destination SAP
 - Control
 - **Több működési mód**
- **Subnetwork Access Protocol (SNAP)**
 - Még több protokoll multiplexálására (8 bit nem elég)
 - Az 0xAA és 0xAB SAP SNAP fejlécet jelent
- Mivel az LLC/SNAP 8 byte overhead-et jelent, az IETF szabvány kimondja, hogy az IP/ARP csomagokat Ethernet esetében Ethernet II keretformátummal kell küldeni

Az LLC/SNAP fejléc

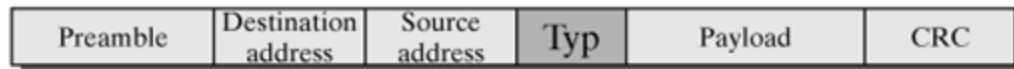


BME-TMIT

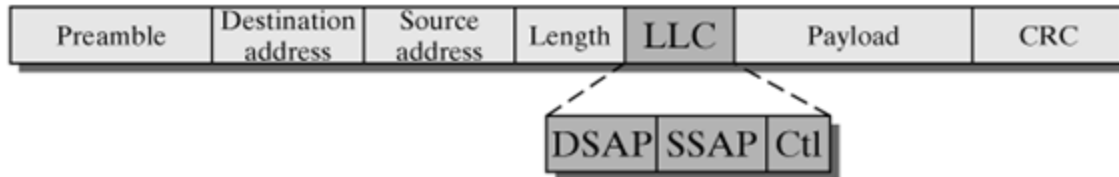
LLC variant

MAC and LLC frame formats

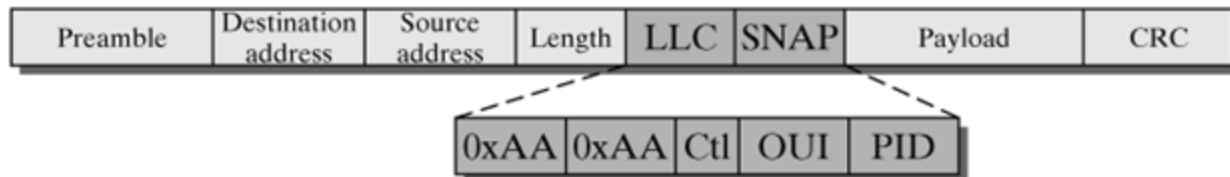
802.3:



802.2:



802.2/SNAP:



IEEE Organizationally Unique Identifier (OUI)

000000, a protocol ID = Ethernet type (EtherType)

más OUI érték más szervezet protokoll ID-je

Protocol Identifier – PID

Ethertype, vagy az OUI által kiosztott

- HUB
- Switching HUB
- Bridge
- Switch
 - Menedzselhető
 - Nem menedzselhető

- Fizikai szintű ismétlő eszköz
 - Bit szinten ismétli a csomagot
- Több egyidejű küldő: ütközés
 - Tehát a „collision domain” megmarad
- A HUB-okat általában hierarchikusan, fa topológiába kapcsolják
 - Uplink: általában cross-connect kábelén csatlakozik
 - Az összes csomagot minden állomás megkapja

HUB – előnyök, hátrányok



BME-TMIT

- Minden HUB portra kapcsolt LAN egy ***szegmens***
- Minden állomás ütközhet a kollíziós terület bármely tagjával
 - Ez rontja a hálózat teljesítményét
 - Csökkenti a skálázhatóságot
 - Mindenki látja a többiek forgalmát
- Különböző médiák nem kapcsolhatók össze
 - Pl. ha van 10Mbps állomás a rendszerben, a teljes sebesség visszaesik 10-re

- Egy „okosabb” HUB
- Megtanulja a hozzá kötött eszközök címeit (MAC vagy IP), és csak oda küldi a csomagokat
- Előnyök:
 - jobban skálázható
 - biztonságosabb
 - Az ütközési valószínűséget lecsökkenti, de nem szünteti meg

- Link Layer eszközök: megvizsgálják a MAC fejléceket és szelektíven továbbítanak
- Elválasztják az ütközési zónákat:
 - pufferelik a csomagokat
 - Csak a megfelelő szegmensre továbbítanak
- A célszegmensben CSMA/CD-t használnak a hozzáférésre
- A pufferelés lehetővé teszi különböző médiák/sebességek összekapcsolását

- Cél: traszparens működés
 - Automatikus, plug-n-play működés
 - Automatikus konfigurálás
 - A létező LAN-okkal való együttműködés
- Három fő funkcionális:
 1. Csomag továbbítás
 2. MAC cím tanulás
 3. Hurok elhárítás: Spanning Tree algoritmus

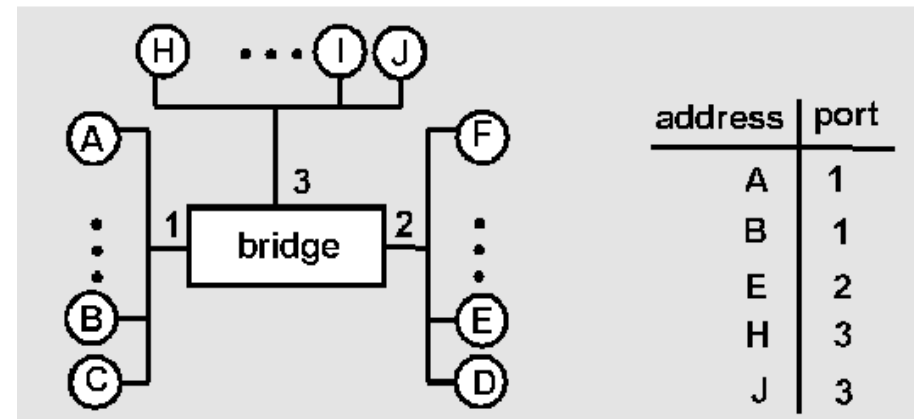
- Megtanulják, hogy mely MAC címet melyik porton érik el: szűrési táblák
 - A beérkező csomagnak kiolvassa a forrását
 - Bejegyzi a szűrési táblába a megfelelő port-al
- Minden bejegyzéshez tartozik egy időbélyeg
 - {MAC cím, port, idő}
 - A bejegyzések az idő lejártával törlődnek
- Működés:
 - Ha van bejegyzés, oda továbbít
 - Ha ugyanaz az interfész, eldobja a csomagot
 - Ha nincs bejegyzés, broadcast

MAC cím tanulás - példa



BME-TMIT

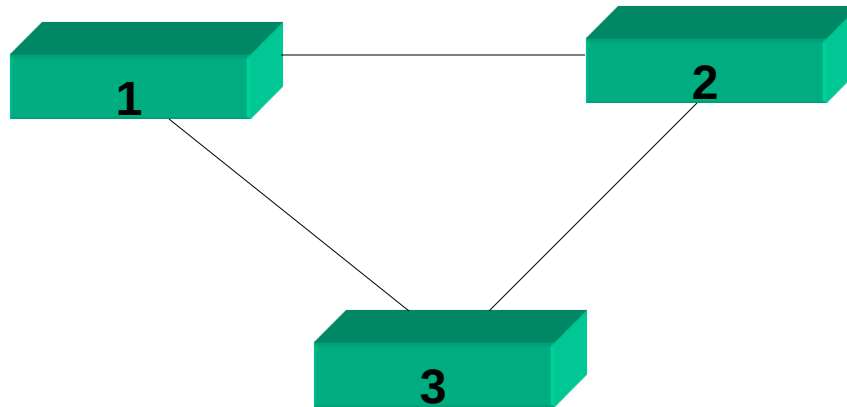
- C küld D-nek
 - A bridge broadcast-ol a 2 és 3 interfészeken
 - A 3. interfészen mindenki eldobja
- D válaszol
 - A bridge már tudja C helyét, csak az 1.-es szegmensre küldi



Redundancia - hurok



BME-TMIT



1. Az első bridge kap egy csomagot. Továbbítja 2 es 3 felé
2. 2 a csomagot továbbítja 3 felé,
3. ugyanakkor 3 továbbítja 2 felé
4. 2 es 3 a csomagokat továbbítják 1 felé
 - ez egy hurok, végtelen körforgás

- Célja a hurok elkerülése
 - Induláskor fa topológiára korlátozza a fizikai topológiát
- Tanuló bridge alapú
- A csomagok kizárólag a fa mentén közlekednek
 - a gyökér irányában, ameddig a cél MAC cím egy más interfészhez nem tartozik
- 802.1d

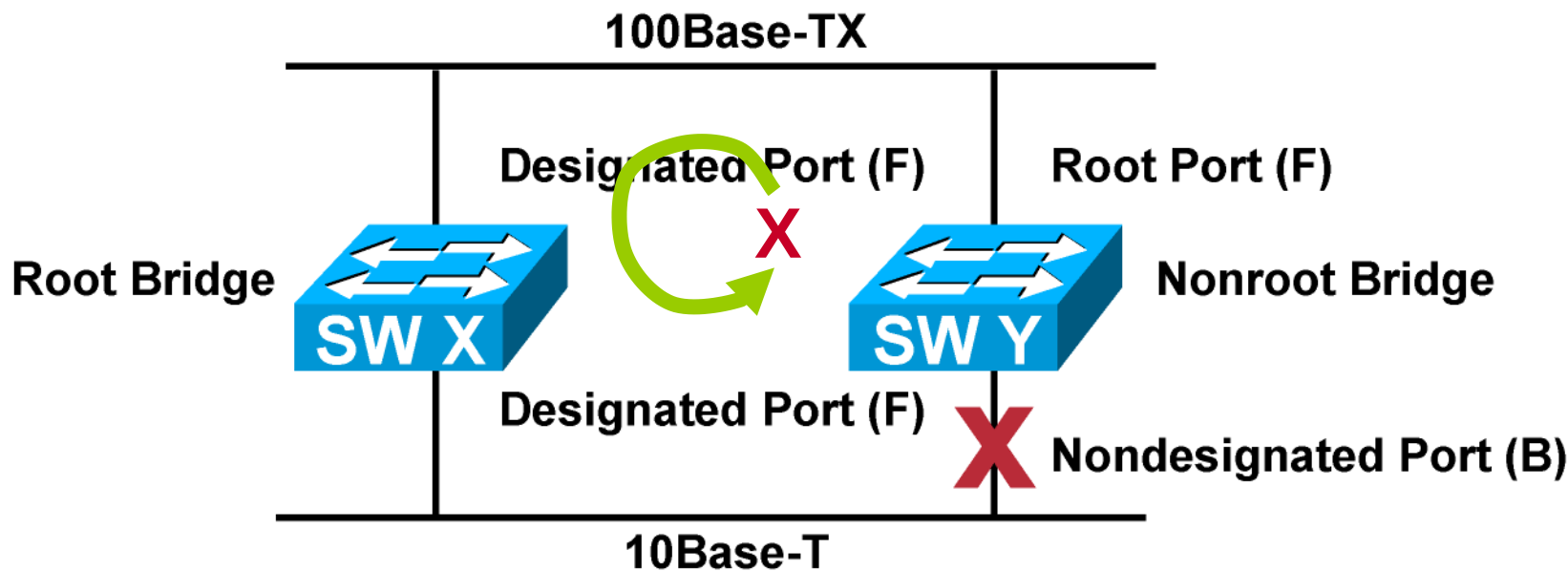
Spanning Tree Alapok



BME-TMIT

Hurok mentes

- Egy root bridge egy hálózatban
- Egy root port minden nem-root bridge-n
- Egy „designated port” minden szegmensre
- Az összes többi port blokkol

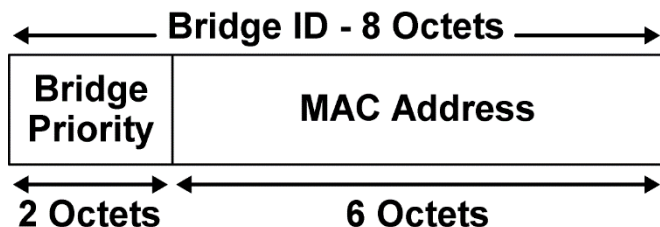


Bridge Protocol Data Unit



BME-TMIT

Bytes	Field
2	Protocol ID
1	Version
1	Message Type
1	Flags
8	Root ID
4	Cost of Path
8	Bridge ID
2	Port ID
2	Message Age
2	Maximum Time
2	Hello Time
2	Forward Delay



IEEE 802.1t - bővítések
Priority = Pri;VLAN ID

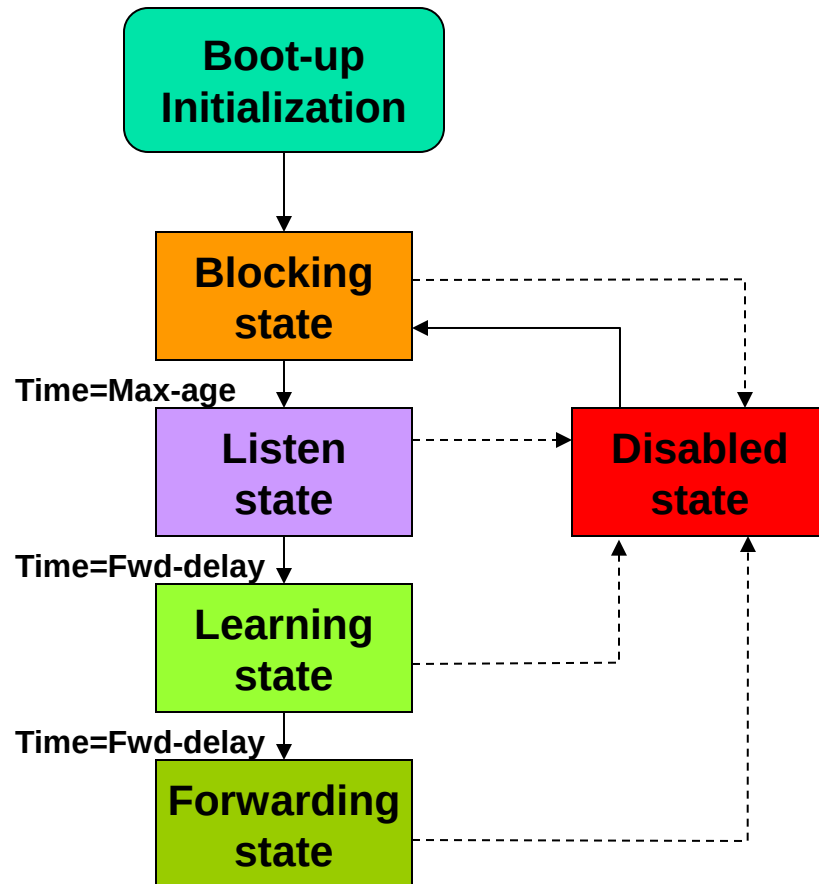
- a BPDU-k feladata:
 - root bridge választás
 - Hurok felderítés
 - Blokkolás a hurok elkerülésére
 - Értesíteni a hálózatot a változtatásokról
 - A spanning tree monitorozása

Link Speed	Cost (Revised IEEE Spec)	Cost (Previous IEEE Spec)
10 Gbps	2	1
1 Gbps	4	1
100 Mbps	19	10
10 Mbps	100	100

Egy port élete ...



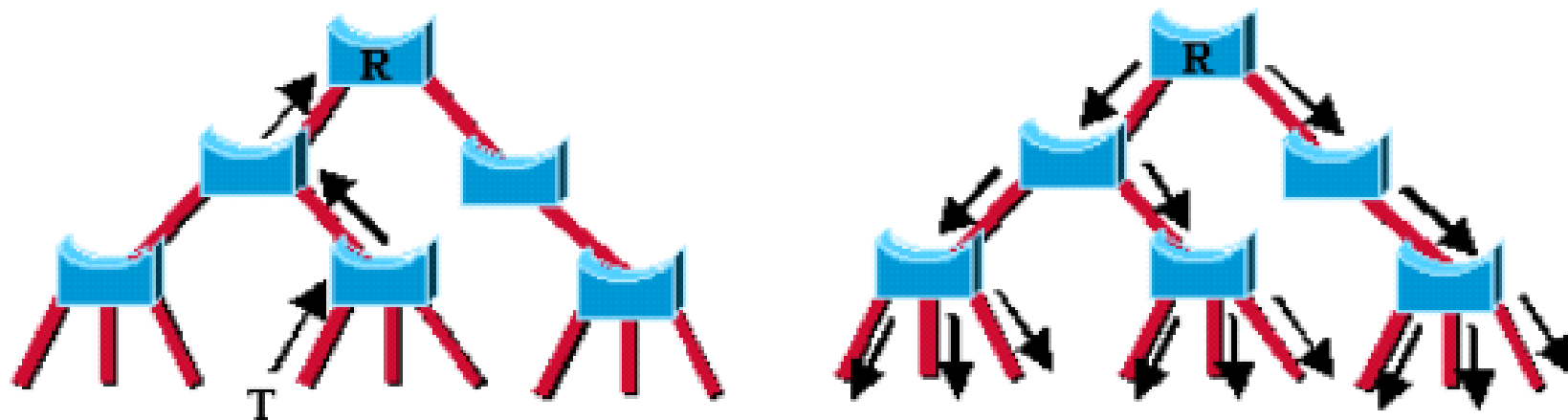
BME-TMIT



- Időzítő alapú
 - Max-age (def. 20 sec)
 - Forward Delay (def. 15 sec)
- Blocking
 - Frames discarded
- Listen
 - Root identified, Port role
- Learning
 - MAC cím tanulás
- Forwarding
 - Csomag továbbítás
 - MAC cím tanulás
- Disabled

802.1D-1998 Topology Change

BME-TMIT



- Cél: a MAC bejegyzések frissítése
 - Amikor egy port állapota változik:
 - A bridge egy **TCN BPDU**-t küld a **Root bridge-nek**
 - A TCN BPDU-t nyugtázzák míg a root-hoz ér
 - A Root beállítja a **TC flag-et** a BPDU-ban ($\text{fwd_delay} + \text{max_age}$) másodpercre, melynek hatására az L2 címek törlődnek

STP algoritmus - 1



BME-TMIT

- A kapcsolók először kiválasztanak egy gyökeret (*Root Bridge*)
 - a legkisebb MAC címmel vagy ID-vel rendelkezőt
- a gyökérből kiindulva kiépítik a fát
 - minden port rendelkezik egy árral (*Port Cost*)
 - a fa kiépítésekor a legkisebb árral rendelkező útvonalat választja
- a fa kiépülése után megtanulja a címeket
 - 15 másodperc tanulási idő

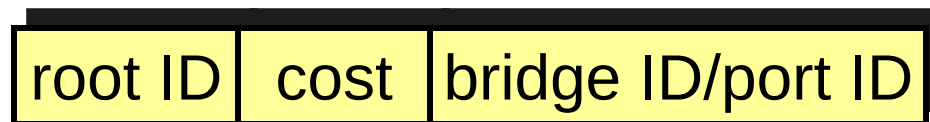
- Bridge ID
 - Alapértelmezett értéke a MAC cím
 - állítható
- Portonkénti ár (**Port Cost**)
 - Alapértéke a sebességtől függ
 - Manuálisan beállítható

STP algoritmus - 2



BME-TMIT

- Először minden bridge azt feltételezi hogy ő a root
 - BPDU üzenetet küld a következő tartalommal:



Root bridge ID (amit gondol)

Root cost

Saját bridge ID

- Első BPDU: (B, 0, B)

STP algoritmus - 3



BME-TMIT

- A Root bridge a legkisebb ID-vel rendelkező bridge lesz.
 - Ha egy kapcsoló kisebb ID-vel rendelkező BPDU-t kap R-től, elfogadja root-nak, és a következő BPDU-t továbbítja:



- Ahol B a saját ID és cost a *Port Cost*-ok összege R felé
- 15 másodperc van a topológia kialakítására

STP algoritmus - 4



BME-TMIT

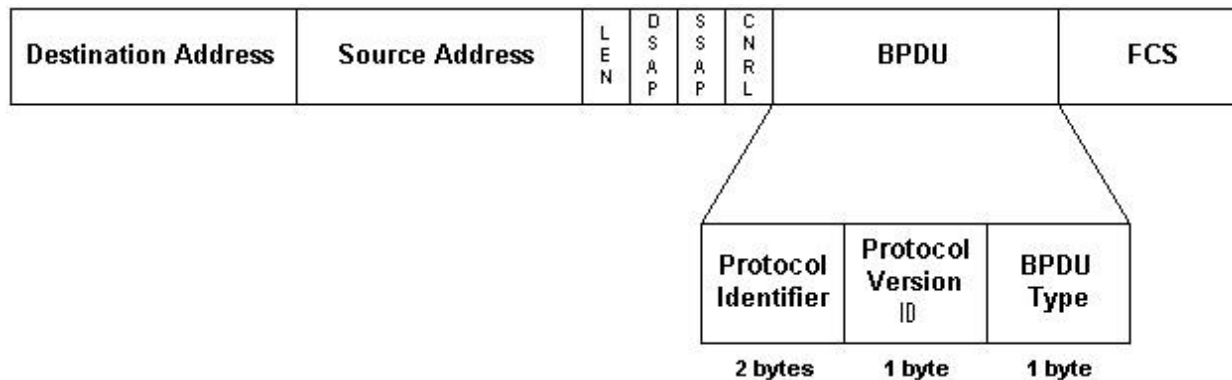
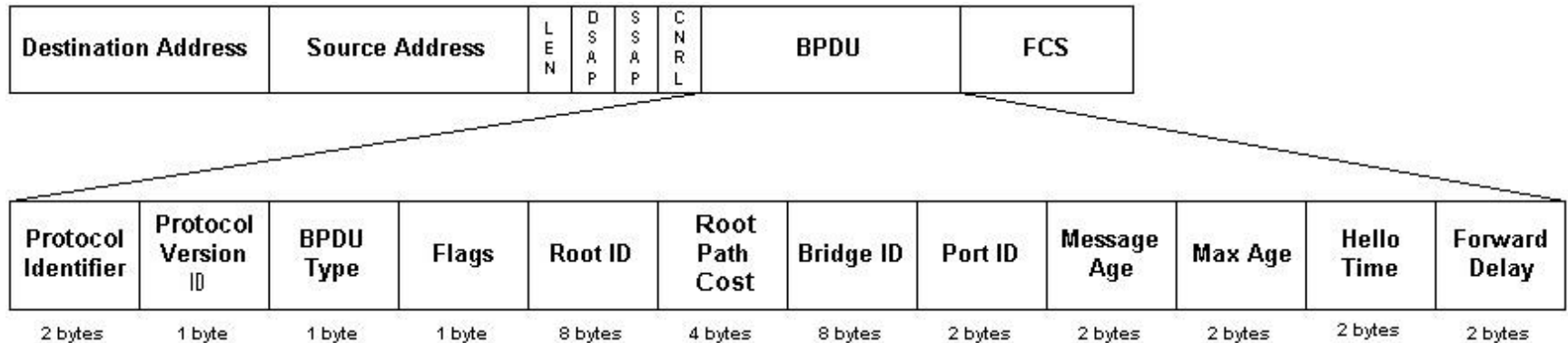
- Az interfész amelyen a Root Bridge-t éri el, a *Root Port* lesz
- Az a bridge, amely egy LAN-t szolgál ki a root fele, a LAN *designated bridge*-e lesz
 - A designated bridge portja *forwarding state*-be kerül
- Az összes többi, topológiában részt nem vevő interfész blokkolni fog (*blocking state*)
- Létezik egy adminisztratív kikapcsolt állapot is, a *disabled state*

- Az STP-t úgy tervezték, hogy a topológiában bárhol lehetnek HUB-ok
 - Pl. egy HUB 2 bridge-hez kötve
- Figyelembe kellett venni a médium hozzáférés szempontjából is
 - Full duplex linknél ezt nem kell nézni

Bridge BPDU-k



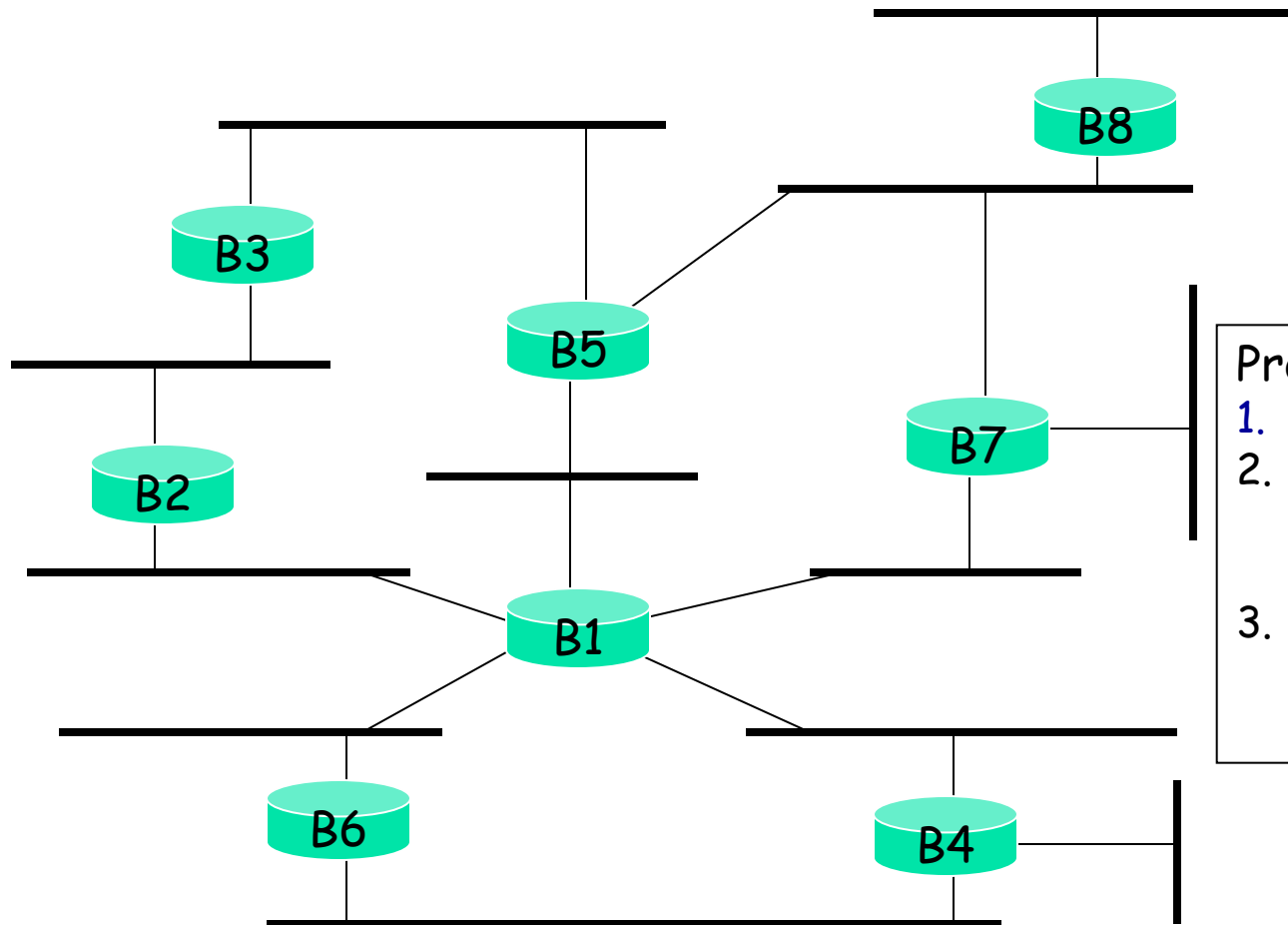
BME-TMIT



Példa – Fizikai topológia



BME-TMIT



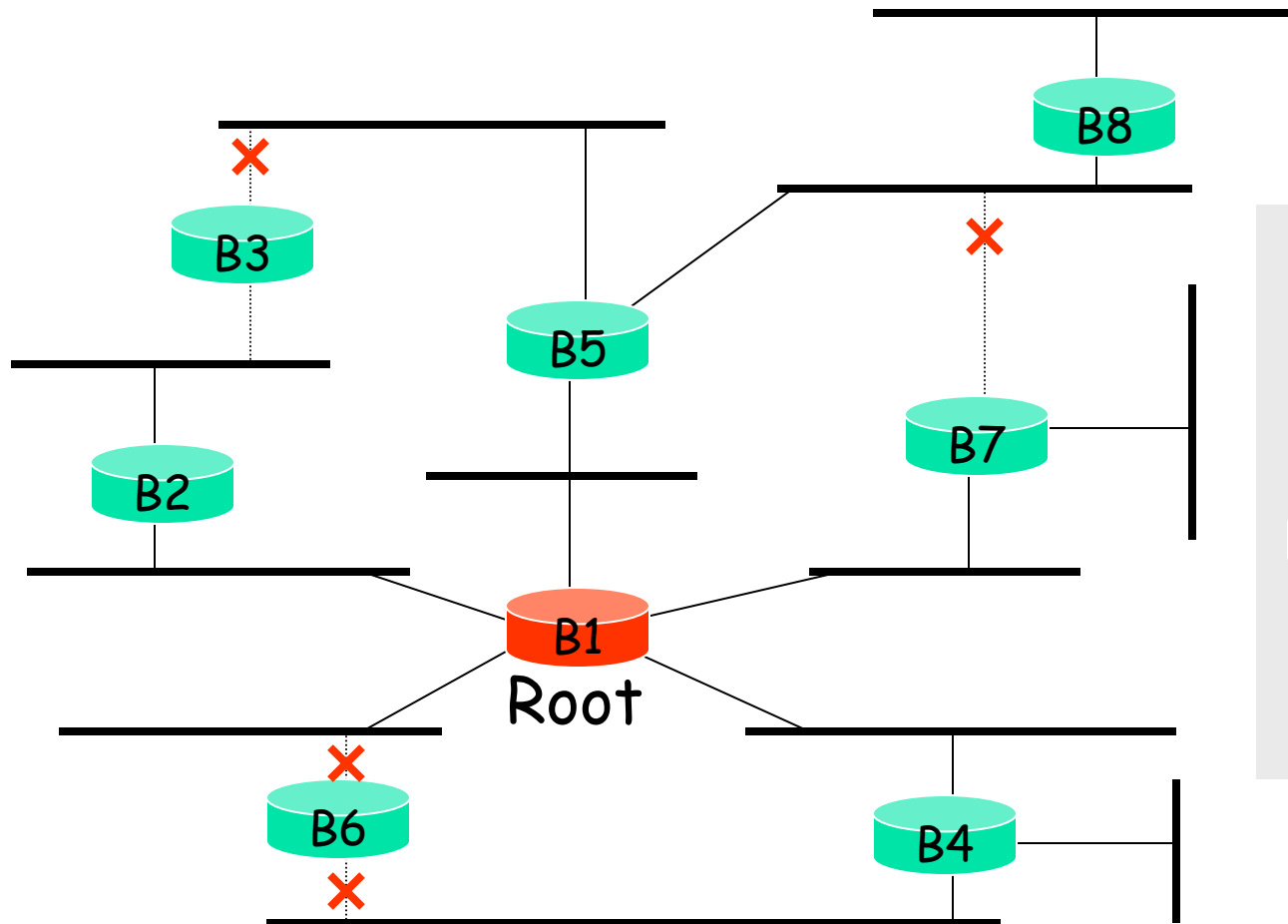
Protocol működés:

1. **Root** kiválasztás
2. minden LAN-ra kiválasztja a **designated** bridge-et, a legközelebbit a root-hoz.
3. Minden bridge a **root** fele a **designated** bridge-en keresztül küld.

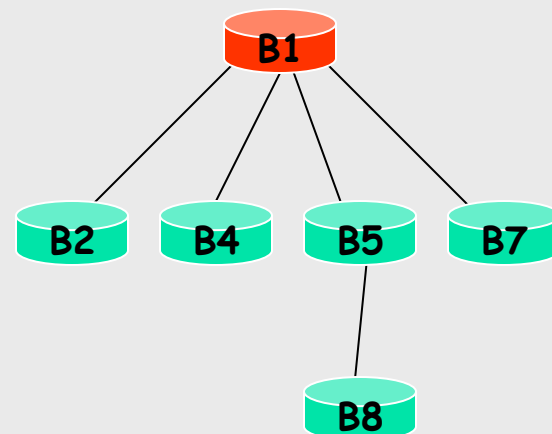
Példa – STP Topológia



BME-TMIT



Spanning Tree:

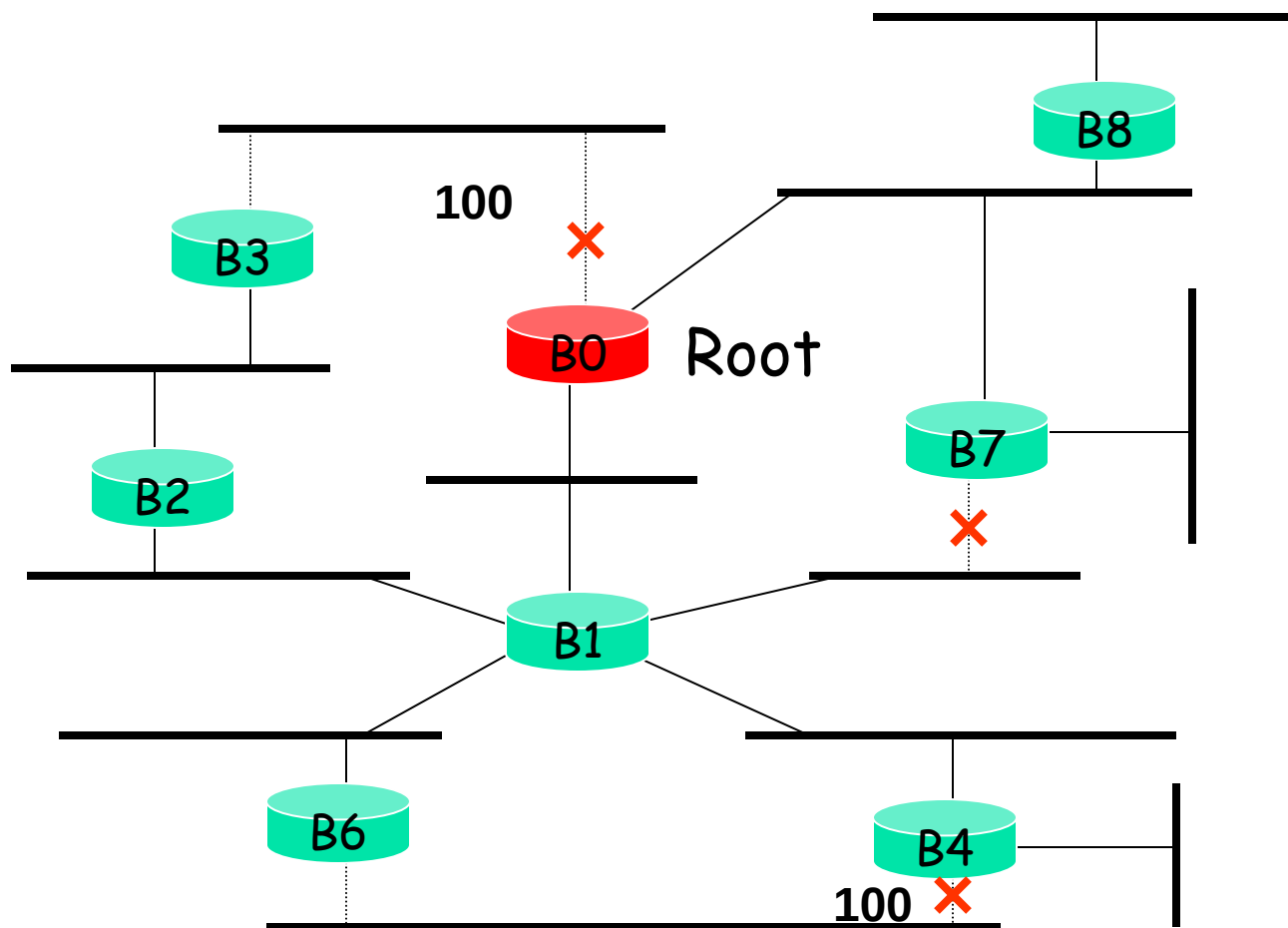


- A bridge ID manuális beállításával megválaszthatjuk a root-ot
 - Általában célszerű megválasztani
- A Port Cost-ok meghatározásával megváltoztathatjuk a kialakuló fát
 - Csak indokolt esetben érdemes átállítani
 - Hiba esetén a rossz beállítás szuboptimális topológiához vezethet

Példa – Módosított STP Topológia



BME-TMIT



- Minden porton periodikusan HELLO üzeneteket küldenek
- 2 HELLO üzenet elmaradása hibát jelent
 - A bridge-ek újraszámolják a topológiát
 - Ha van blokkolt port akkor azt fogja használni
- Az új topológia kialakítására van 15 sec
- Következik a MAC címek megtanulása a portokon
- 30 másodpercen belül újra működőképes a rendszer

- A STP nagyon lassan reagál hiba esetén
 - Timer alapú működés
- Nagyszámú kapcsoló esetén nem alkalmazható
 - max bridge diameter = 7
- Nem lehetséges a terhelés megosztása

802.1w - Rapid STP alapok

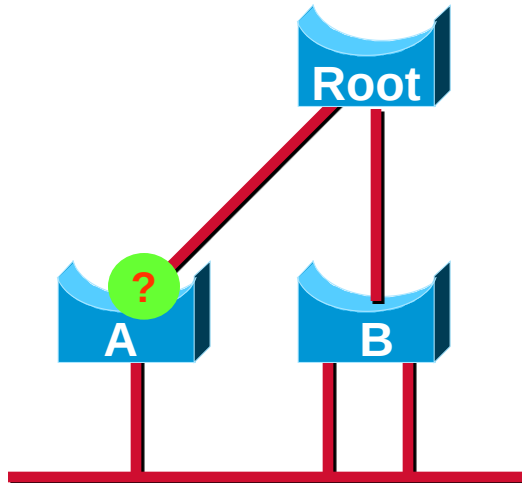


- Újabb IEEE protocol (802.1w)
Kompatibilis a 802.1D-1998-al
- Gyorsabb konvergencia & Timer független
 - CSAK point-to-point FDX Link esetén
 - CSAK ha az edge portok helyesek
 - CSAK ha 802.1D együttműködés nem kell
- Gyártók egyedi megoldásait tartalmazza
- Újítások:
 - új Port Role
 - Más BPDU
 - BPDU kezelés
 - Gyors port állapotváltás
 - Új topologia kialakítás
 - PVST+/802.1D kompatibilitás

RSTP Port Állapotok



BME-TMIT



STP 802.1D	RSTP 802.1w	Active ?	Learning ?
Disabled	Discarding	No	No
Blocking	Discarding	No	No
Listening	Discarding	No	No
Learning	Learning	No	Yes
Forwarding	Forwarding	Yes	Yes

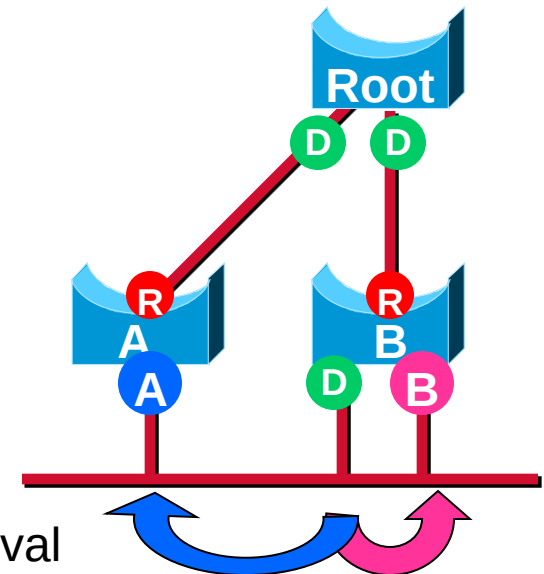
- RSTP szétválasztja a Port Szerepét és Állapotát
- RSTP csak 3 Port állapotot tartalmaz:
 - DISCARDING
 - LEARNING
 - FORWARDING

RSTP Port Roles

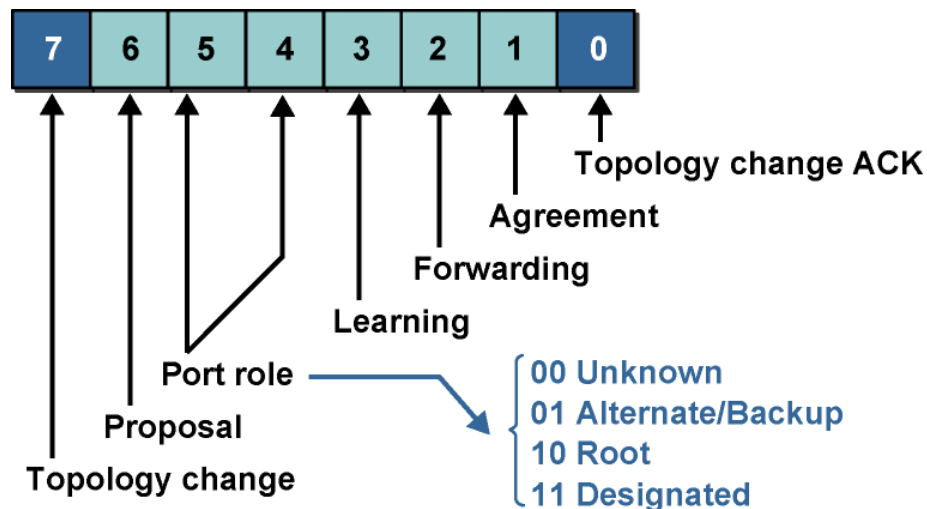


BME-TMIT

- R** Root Port (Fwd):
Az a port amely a legjobb BPDU-t kapja
– shortest path a Root irányába a cost függvényében
- D** Designated Port (Fwd):
A port amely a legjobb BPDU-t küldi egy segmensen
- A** Alternate Port (Disc):
Más bridge BPDU-k által blokkolt port
– redundáns út a Roothoz
- B** Backup Port (Disc):
Port blokkolva ugyanazon bridge által küldött BPDU-val
– redundáns út egy segmenshez



RSTP BPDUs



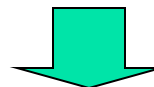
IEEE 802.1d/w BPDUs have the following layout:

```

protocol id: 0000 IEEE 802.1d
version id: 00 -> 02
bpdn type: 00 config bpdn, 02 RSTP bpdn
            80 tcn bpdn
bit field: 1 byte
 1 : topology change flag
 2 : unused 0 Agreement
 3 : unused 0 Forwarding
 4 : unused 0 Learning
 5 : unused 0 Port Role
 6 : unused 0 Port Role
 7 : unused 0 Proposal
 8 : topology change ack
root priority 2 bytes
root id: 6 bytes
root path cost: 4 bytes
bridge priority: 2 bytes
bridge id: 6 bytes
port id: 2 bytes
message age: 2 bytes in 1/256 secs
max age: 2 bytes in 1/256 secs
hello time: 2 bytes in 1/256 secs
forward delay: 2 bytes in 1/256 secs
    
```

- Minden bridge hello_time indőnként BPDn-t küld
- Port állapot érvénytelenítődik 3 x hello_time max idő alatt
 - i.e. 3 BPDn elvesztése

802.1D eldobják a 802.1W BPDn-kat

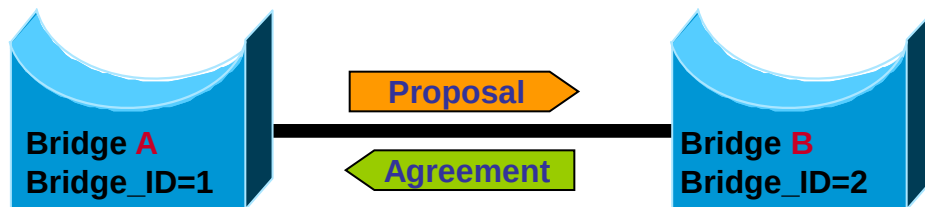


Együttműködés + RSTP előnyök elvesztése

Gyors átmenet Forwarding-ba

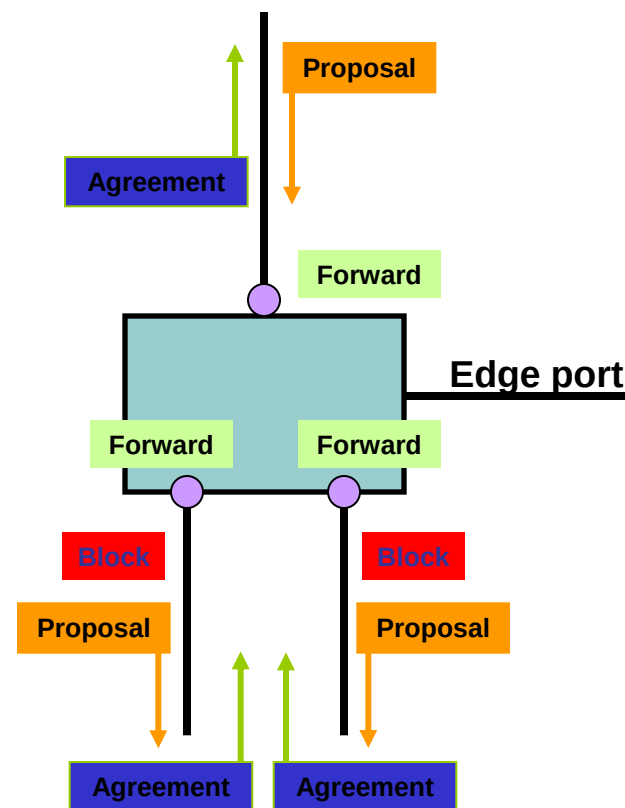


BME-TMIT



- Amikor egy port bekapcsol,
 - A bridge egy BPDU-t küld **proposal** flag-el hogy designated legyen a segmensen
- A válasz egy
 - BPDU **agreement** flag-el ha a távoli bridge elfogadja
- Amint megérkezik az **agreement** - forwarding

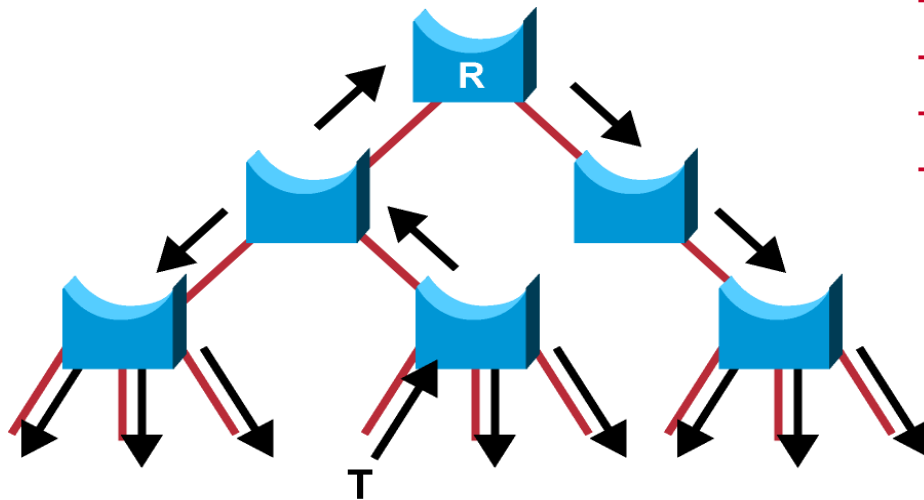
- Proposal érkezés
 - Blokkol minden nem edge port
- Visszaküld egy agreementet
 - Az új port forwarding állapotba kerül
- A többi porton proposal küldés
- Agreementek fogadása
 - Portok forwarding módba



RSTP Topology Change



BME-TMIT



- + Port on which the TCN was received is not flushed
- + Edge ports are not flushed
- +/- Flooding but connectivity restored immediately
- + No need for proxy multicasts

- TC csak azon a linken amely **forwarding-be megy**
- A kezdeményező küldi (nem a Root) és a szomszédok továbbítják az **aktív topológia mentén**
- TC bit 2 x hello time időre
- **Azonnal törli a CAM táblát**

- a reakcióidő lecsökkentésére találták ki
- a STP protokoll továbbfejlesztett változata
- működése nem időzítőkön alapul
- új port állapotokat vezet be a gyorsabb átkapcsolás érdekében
- a HELLO üzenetek sűrűségét is megnövelték
- 802.1w – később 802.1D-2003-ként már minden eszköznek kötelező

- Proposal-agreement módszer
 - a jobb útvonallal rendelkező kapcsoló felajánlja az útvonalat
 - a legjobb ajánlatot fogja elfogadni
- A topológia helyreállása nem függ időzítőtől
- Több kapcsolón keresztül is működik

- A root, forwarding, blocking állapotok mellett a bridge még a következő állapotokat ismeri:
- Alternate port
 - Alternatív útvonal hiba esetén a root felé
- Backup port
 - Egy párhuzamos útvonal egy olyan LAN felé, melynek designated bridge-e
- Új topológia kialakulásakor ezek a portok kapcsolnak be

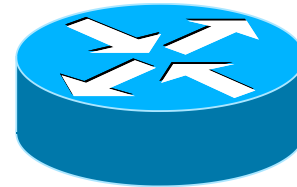
- Az új topológia kialakulása nem timer függő, de topológia függő
 - Nincs korlát – csak az időzítő (STP-ből)
 - Nagy topológia esetén lassú lehet
- Terhelés megosztás továbbra sem lehetséges
- A tanulási fázis nem csökkent le
 - Bár a táblák ürítésére megoldást ad

- Layer 2 továbbítás – MAC címek alapján
- Megtanulja a MAC címeket akár a bridge
- Erőssége a store-and-forward működés amely egyidejű továbbítást végezhet különböző portok között
 - Nincs ütközés
 - Nagy sebességű *backplane*
- Nagyszámú interfész
 - Különböző sebességek/médiumok

- A nagyobb teljesítményű kapcsolók egyben bridge-ek is
 - Támogatják az STP protokollt
- Típusok
 - Nem menedzselhető:
 - irodai célokra, HUB-ok összekapcsolására kiváló
 - Nem támogatja az STP protokollt, sem a VLAN-okat (jövő óra)
 - Menedzselhető
 - VLAN és STP támogatás
 - Menedzsment interfész
 - L2/L3

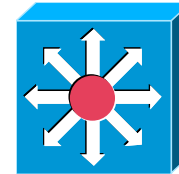
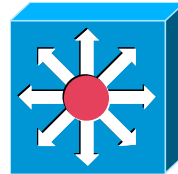
- Egyes kapcsolók képesek pont-pont módban működni
 - Nincs ütközés
 - Full duplex kapcsolat
 - Nagyobb elérhető sávszélesség
- Beállítható
 - Kapcsolók között
 - Kapcsoló és munkaállomások között

- Hierarchikusan



Router

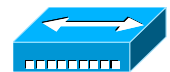
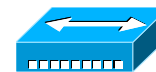
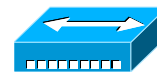
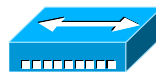
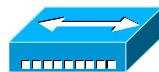
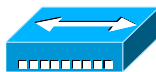
Nagy teljesítményű,
multiservice kapcsoló



SWITCH, Bridge



HUB



- O'Reilly ***Ethernet: The Definitive Guide***
 - by Charles E. Spurgeon
- IEEE 802.3 szabvány
- Cisco – Understanding STP, RSTP

Folytatása következik



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**

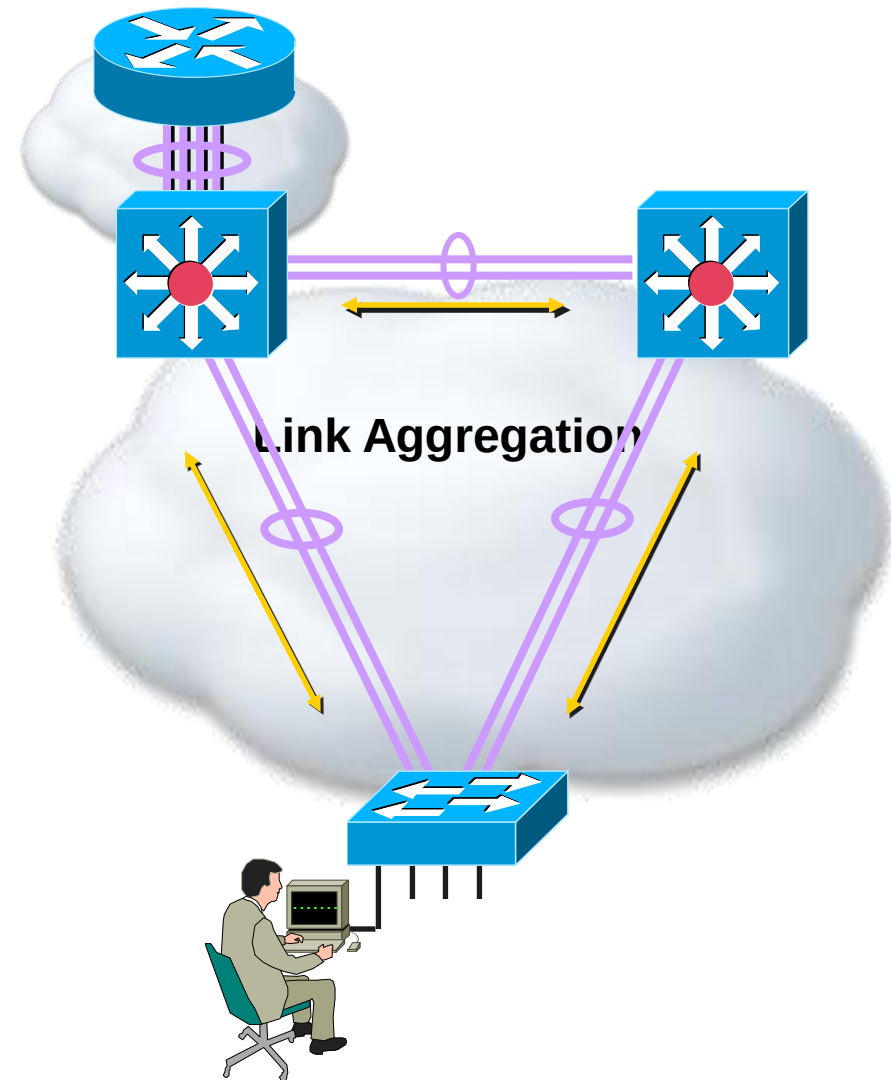


Link Aggregáció



BME-TMIT

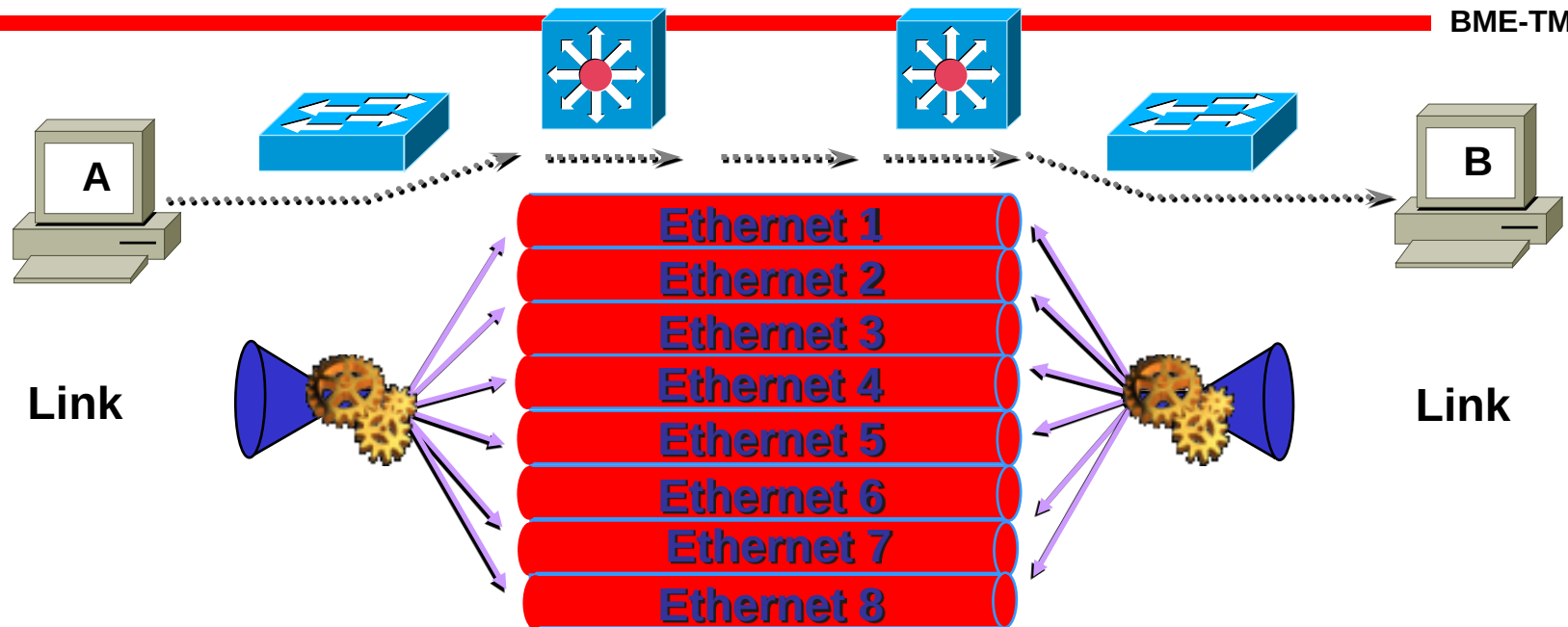
- Link Aggregáció
 - Hasonló linkek összefogása (összesen 8-ig)
10/100/1000/10GE ports
- Használható switchek, routerek és néhány típusú NIC között
- Aggregált Link
 - mindig point-to-point



Link Aggregáció



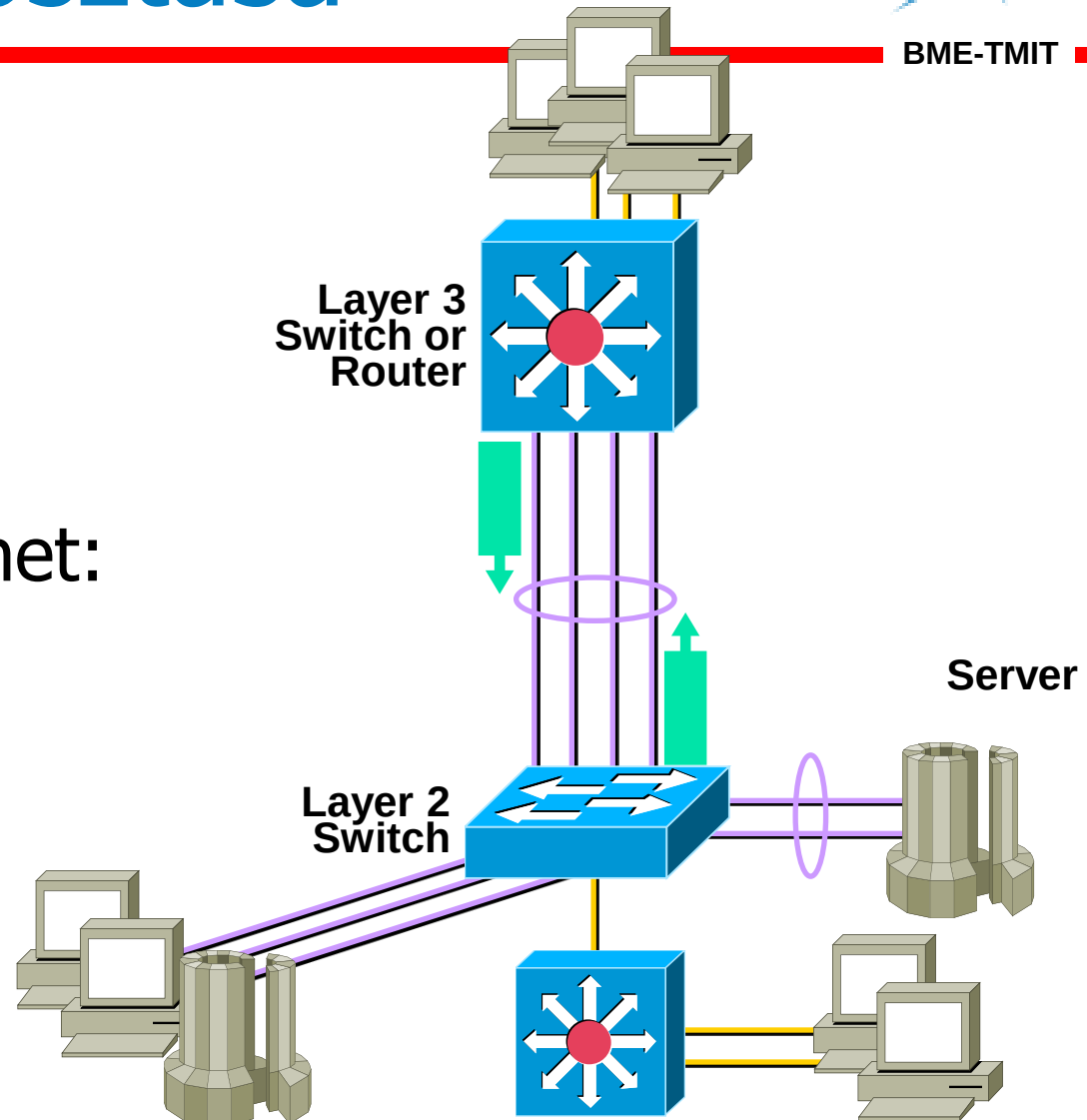
BME-TMIT



- Load sharing és Layer-1 redundanciát nyújt
 - Hash függvény a következő alapokon:
 - MAC, IP addresses, Layer-4 ports (session)
 - Az egész csatorna egyetlen logikai port
 - Vezérlő Protokoll: IEEE 802.3ad LACP

Terhelés megosztása

- Függ a beállított konfigurációtól és a hash-függvénytől
- A hash függvény lehet:
 - MAC-cím alapú
 - IP cím alapú
 - Layer-4 alapú (session)



- Adminisztrátor által beállítható állapotok - LACP
 - ON
 - Csatorna akarok lenni, nem érdekel mit gondolsz!
(nem generál LACP üzeneteket)
 - OFF
 - Nem akarok csatorna lenni, és nem érdekel mit gondolsz!
(nem fogadja az LACP üzeneteket)
 - Active
 - Csatorna szeretnék lenni. Érdekel a dolog?
(Used when you are interested in being a channel)
 - Passive
 - Várom az ajánlatokat, nekem mindegy!
(a “plug-and-play” működésnél használatos,
DE a másik oldalt jól kell konfigurálni)

**Aggregált
LINK megvalósul:**

- **Active-active,**
- **Active-passive,**
- **On-on**

Ethernet – első labor

Moldován István



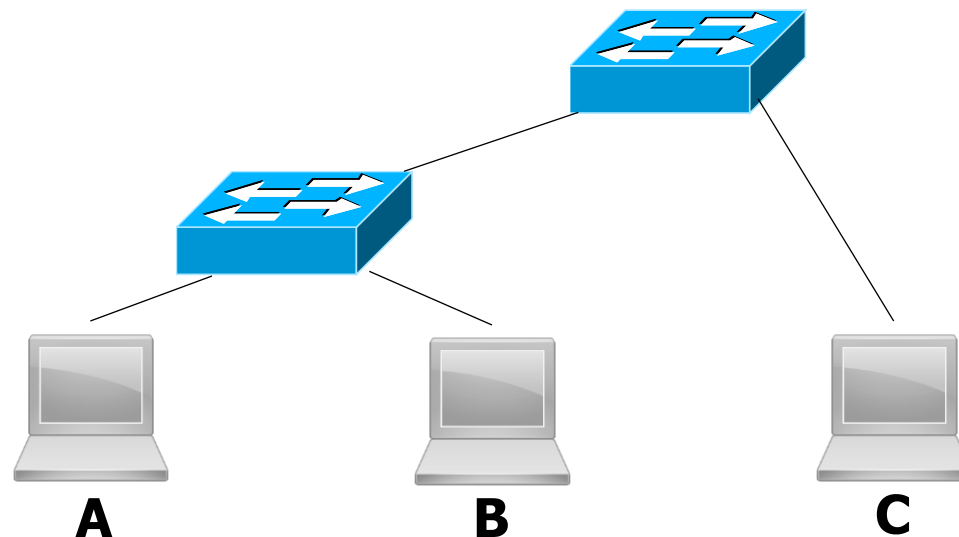
Budapest University of Technology and Economics



**Department of
Telecommunications and Media Informatics**

- MAC címek
 - MM:MM:MM:SS:SS:SS
 - MM-MM-MM-SS-SS-SS
- An OUI {Organizationally Unique Identifier}
 - 00:00:0A -- this is owned by Omron
 - 00-0D-4B -- this is owned by Roku, LLC
- <https://www.wireshark.org/tools/oui-lookup.html>
- Broadcast: FF:FF:FF:FF:FF:FF
- Multicast

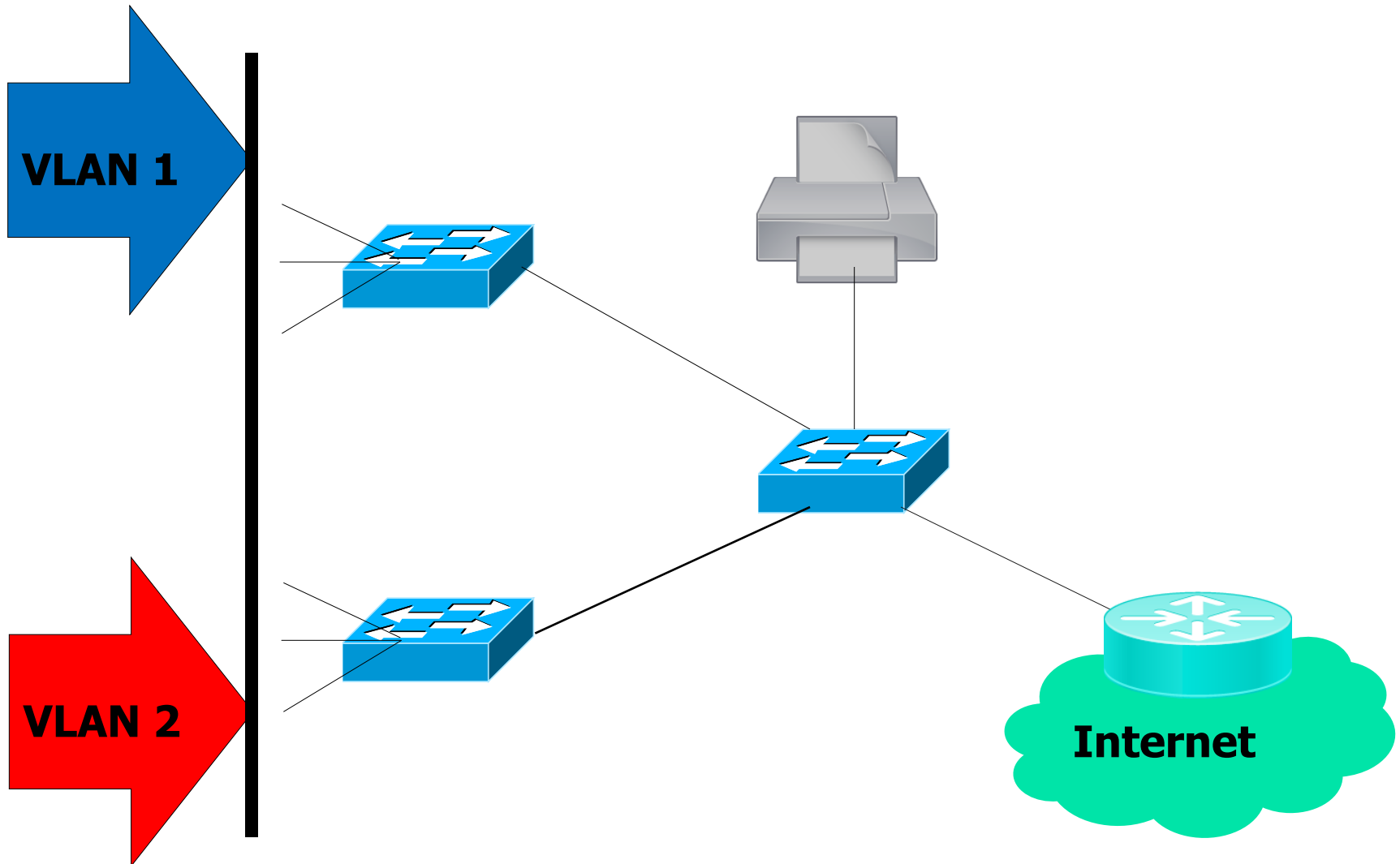
- Ethernet II vs 802.3 – mikor melyik?
 - Type/length
 - packETH / wireshark
- Tanulás
 - Példa



VLAN



BME-TMIT



Spanning Tree



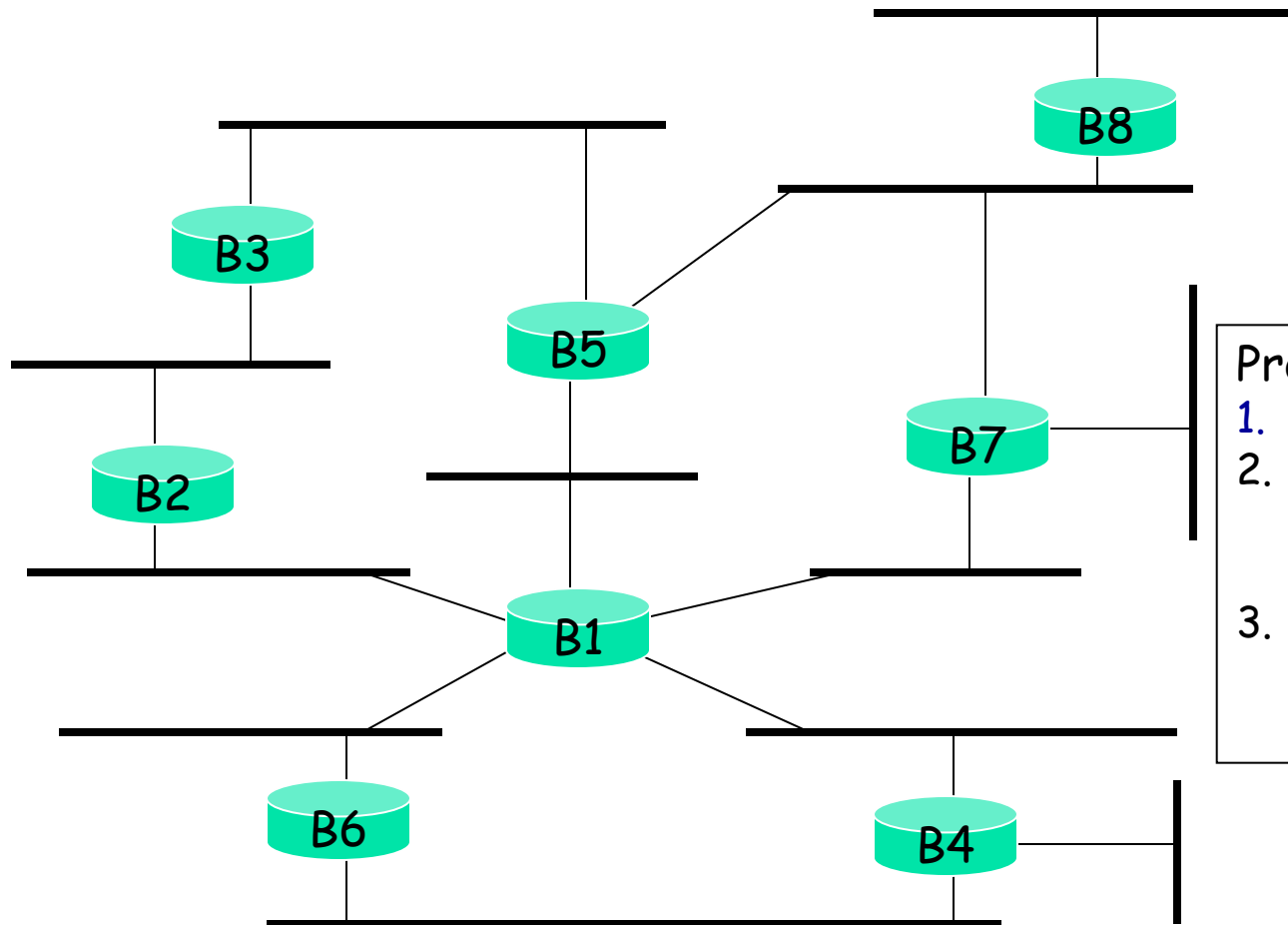
BME-TMIT

- RSTP fa kialakítása
 - BPDU-k
- Paraméterek:
 - Link speed
 - Prioritás
 - ID
- Bridge ID: Prioritás + MAC

Példa – Fizikai topológia



BME-TMIT



Protocol működés:

1. **Root** kiválasztás
2. minden LAN-ra kiválasztja a **designated** bridge-et, a legközelebbit a root-hoz.
3. Minden bridge a **root** fele a **designated** bridge-en keresztül küld.

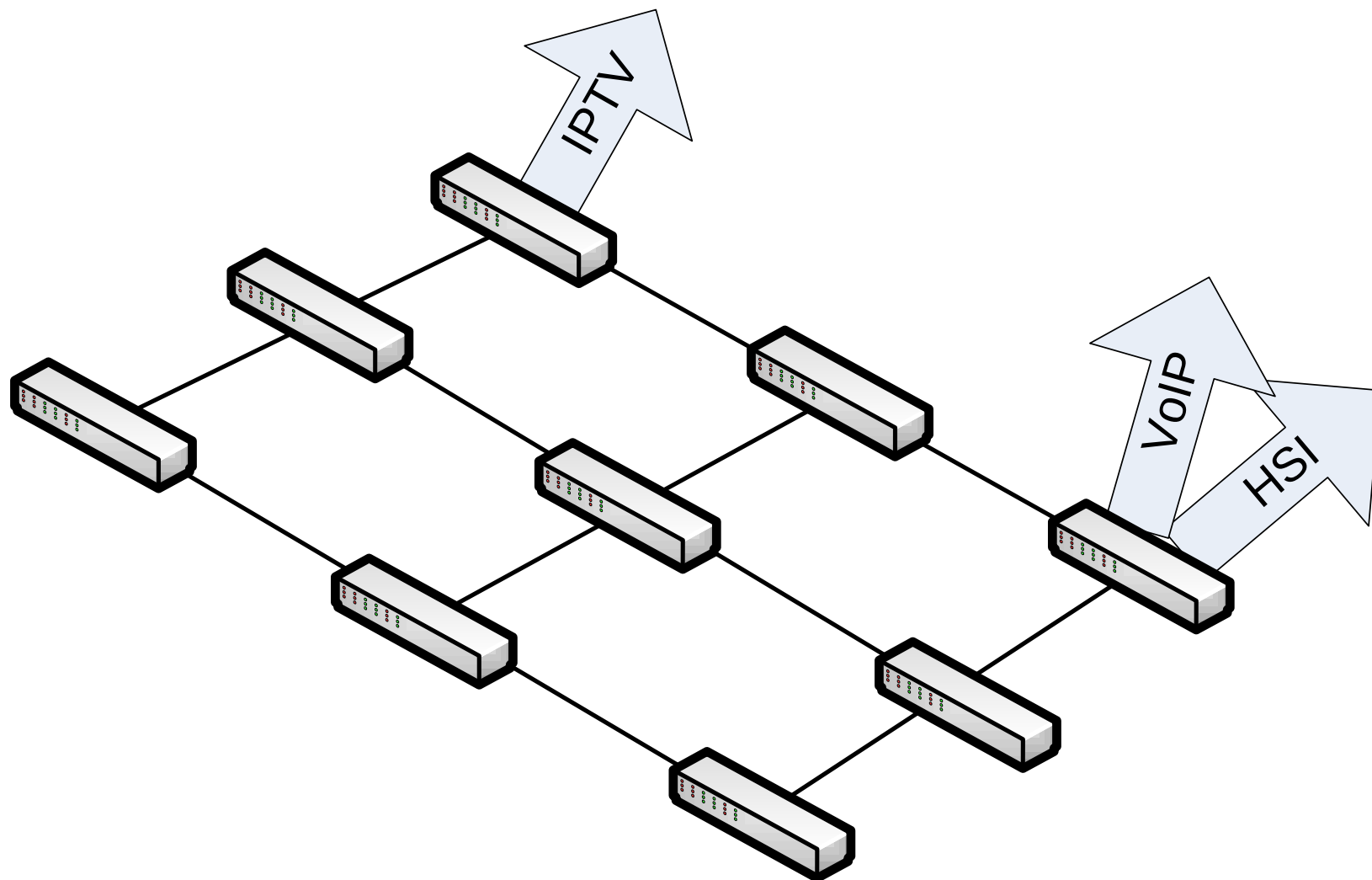

```

graph TD
    B1((B1)) --- B2((B2))
    B1 --- B4((B4))
    B1 --- B5((B5))
    B1 --- B7((B7))
    B5 --- B8((B8))
  
```

RSTP optimális beállítás



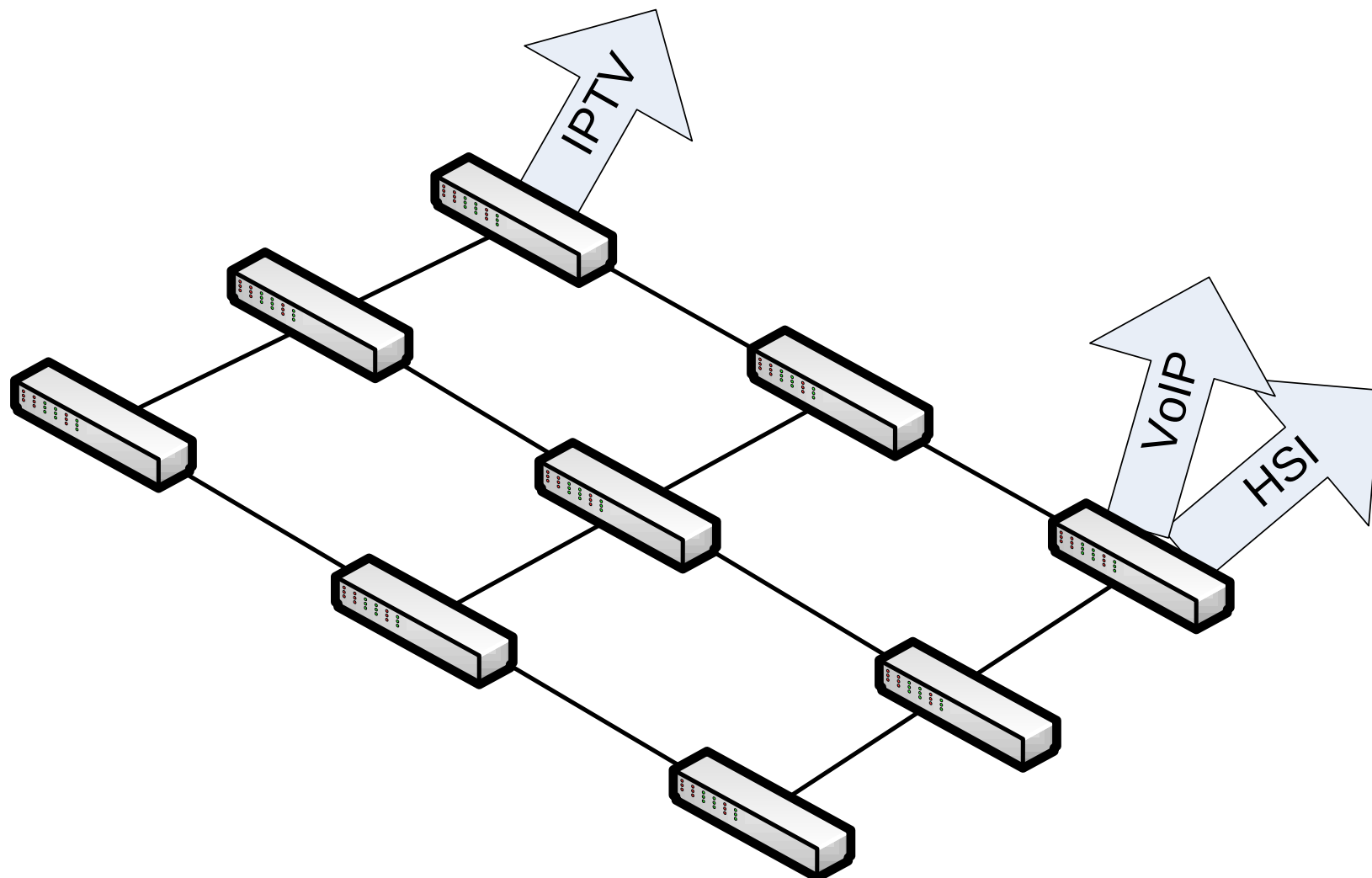
BME-TMIT



MSTP optimális beállítás



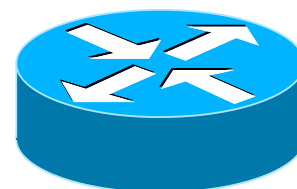
BME-TMIT



- Layer 2 továbbítás – MAC címek alapján
- Megtanulja a MAC címeket akár a bridge
- Erőssége a store-and-forward működés amely egyidejű továbbítást végezhet különböző portok között
 - Nincs ütközés
 - Nagy sebességű *backplane*
- Nagyszámú interfész
 - Különböző sebességek/médiumok

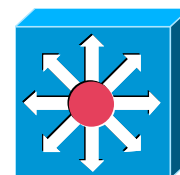
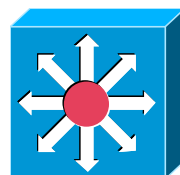
- A nagyobb teljesítményű kapcsolók egyben bridge-ek is
 - Támogatják az STP protokollt
- Típusok
 - Nem menedzselhető:
 - irodai célokra, HUB-ok összekapcsolására kiváló
 - Nem támogatja az STP protokollt, sem a VLAN-okat (jövő óra)
 - Menedzselhető
 - VLAN és STP támogatás
 - Menedzsment interfész
 - L2/L3

- Hierarchikusan



Router

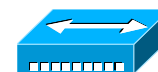
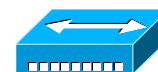
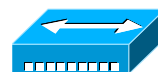
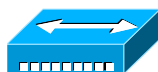
Nagy teljesítményű,
multiservice kapcsoló



SWITCH, Bridge



HUB



Kérdések?



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**



Ethernet – második rész

Moldován István



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**



- LAN (Local Area Network): broadcast tartomány
 - a tartományon belül mindenki megkapja a szórt üzenetet
 - kívül senki
 - a LAN határait a kábelezés szabja meg
 - kifelé router kell a kommunikációhoz
 - másik gép megtalálásához is szórt adás kell (ARP)
- VLAN (Virtual LAN): adminisztratív úton létrehozott broadcast tartomány
 - a rendszer adminisztrátora határozza meg, ki van benne
 - a határok itt nem fizikaiak, csak virtuálisak
 - a különböző VLAN-ok nem látják egymás forgalmát

A VLAN által nyújtott előnyök és hátrányok



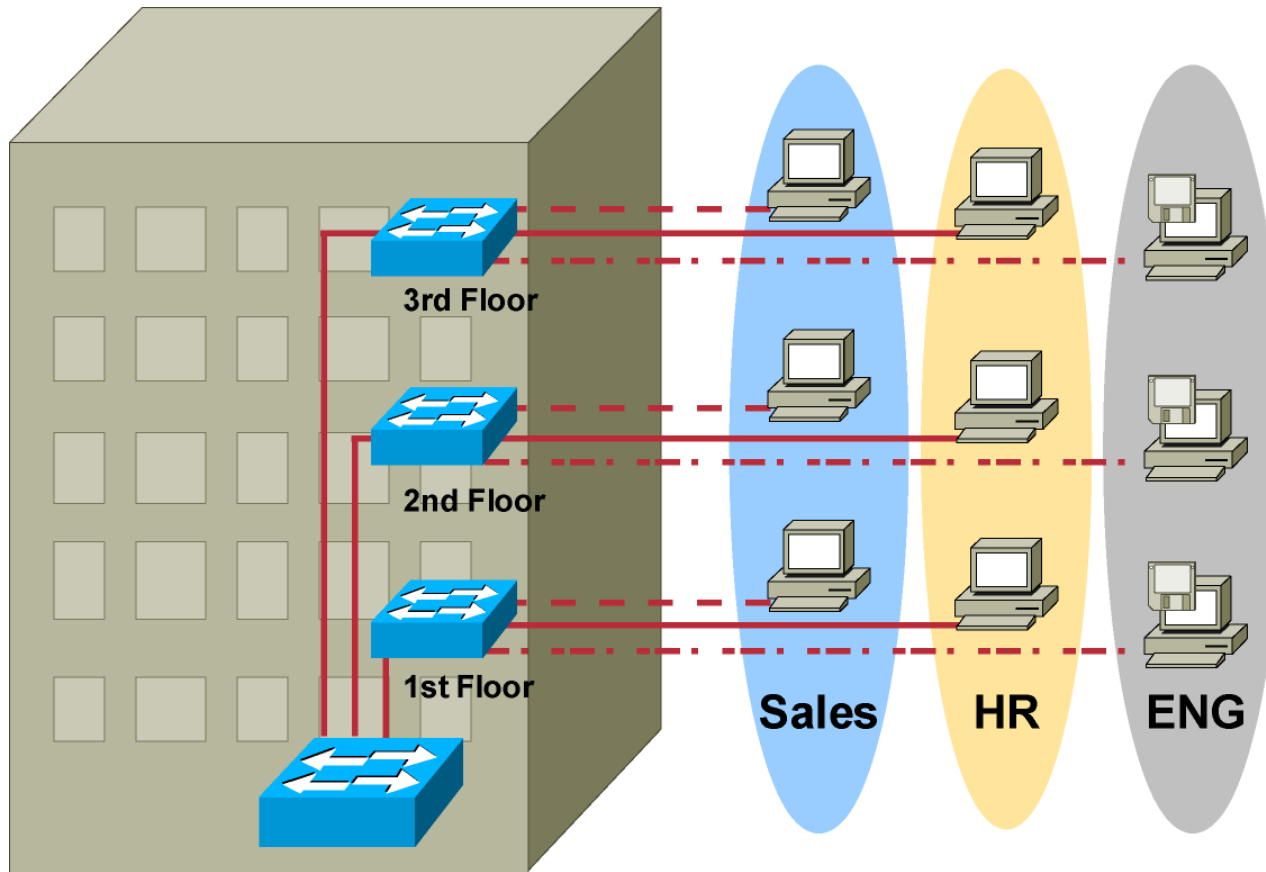
BME-TMIT

- Előnyök:
 - könnyebb kezelni a felhasználók helyválttatását (mobilitás)
 - szórt adás kiterjedésének csökkentése
 - felhasználók virtuális elkülönítése
 - hatékonyabb topológia (esetleg átlapolódás is a VLAN-ok között)
 - biztonság nő
 - virtuális munkacsoportok
- Hátrányok:
 - a virtuális munkacsoportoknál túl gyakorivá válhat a tagság változása
 - Adminisztrációs problémák
 - WAN-ra kiterjedő VLAN-nál a szórt adás már főbb vonalakat is terhelhet

VLAN Áttekintés



BME-TMIT



- Layer 2 connectivity
- Logikai elrendezési flexibilitás
- Egyetlen broadcast domain
- Management
- Biztonság

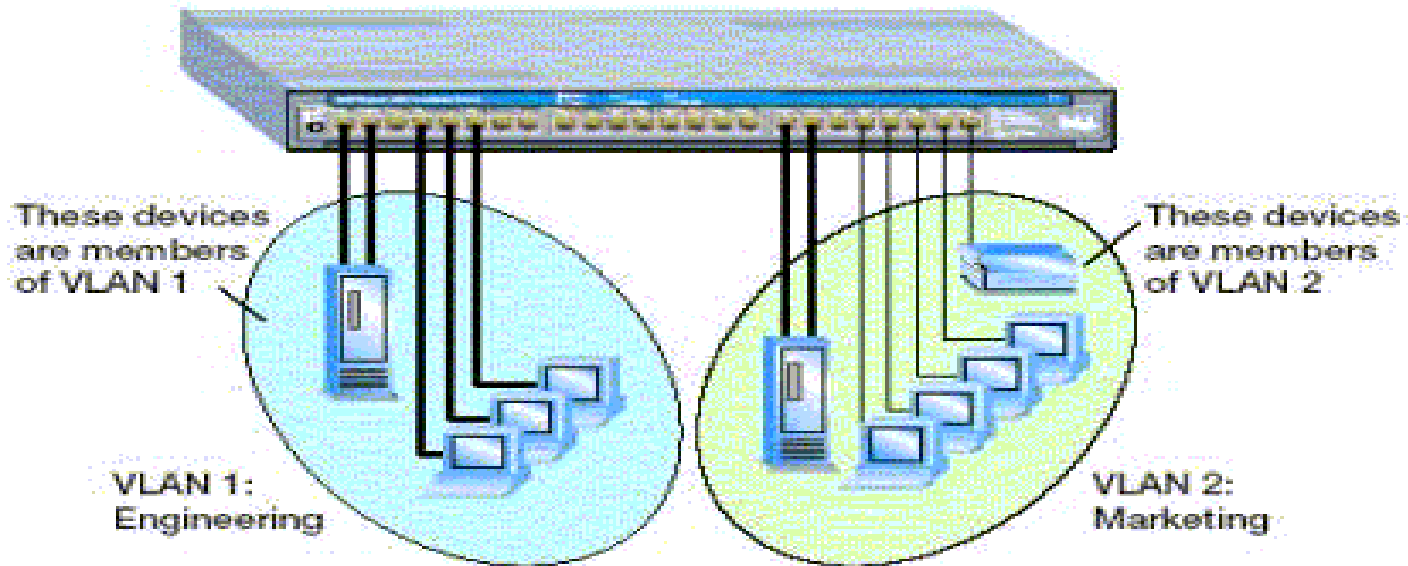
1 VLAN = 1 Broadcast Domain = 1 Logikai hálózat (Subnet)

- Megosztott kapcsolók
 - Csak kapcsolón belül értelmezett
 - Lehet port alapú, MAC alapú
- IEEE 802.1Q szabvány
 - Az Ethernet szabvány kiterjesztése
 - Új mező az Ethernet fejlécben
 - Egy LAN-on belül értelmezett

Megosztott kapcsoló példa



BME-TMIT



1. VLAN: 2., 4., 6., 9. port

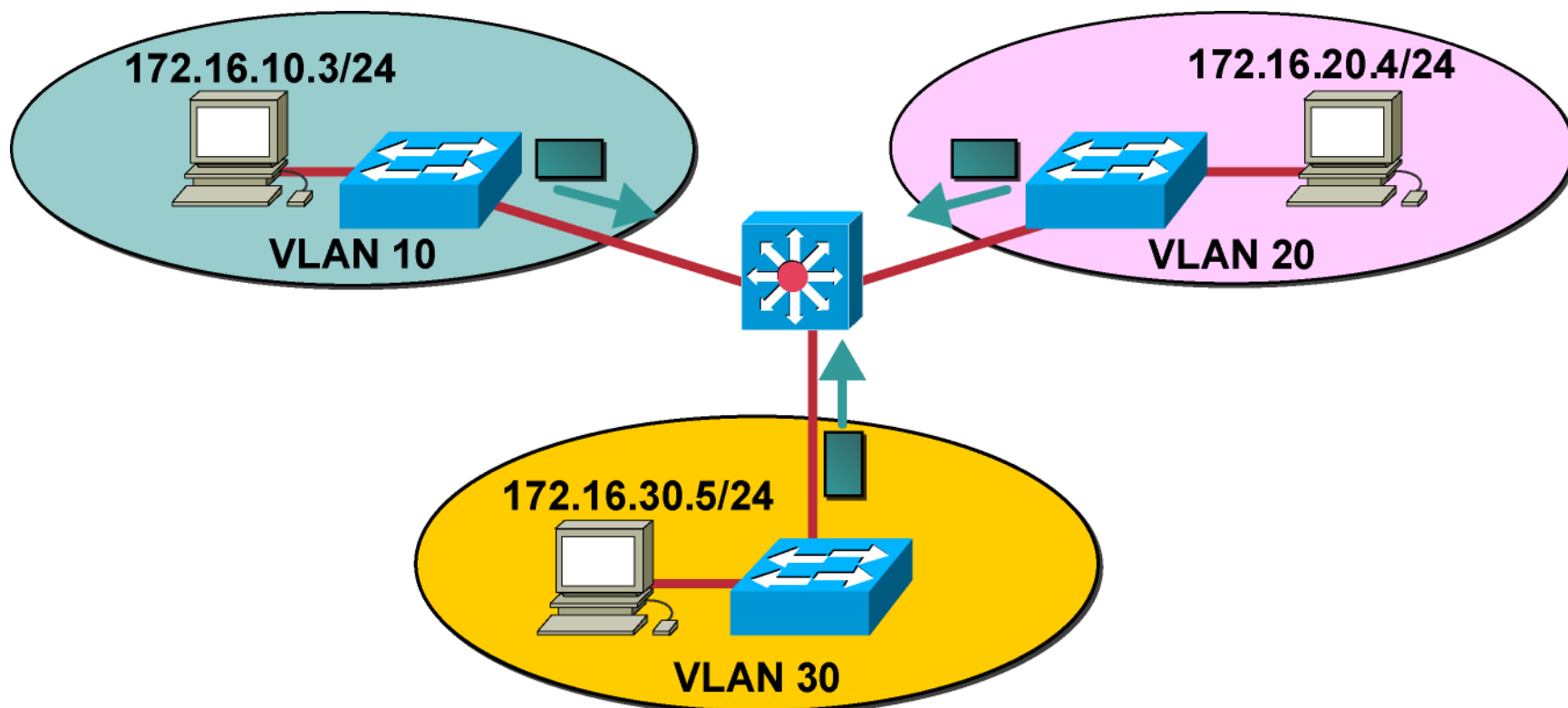
2. VLAN: 17., 19., 21., 23. port

Egymás forgalmát nem látják!

Átjárás VLAN-ok között



BME-TMIT



- Probléma: elkülönített Broadcast Domain
 - Természetüktől fogva a VLAN-ok tiltják a VLAN-ok közti kommunikációt
 - A VLAN-ok közti átjárás 3. szintű eszközt igényel: L3 SW vagy Router

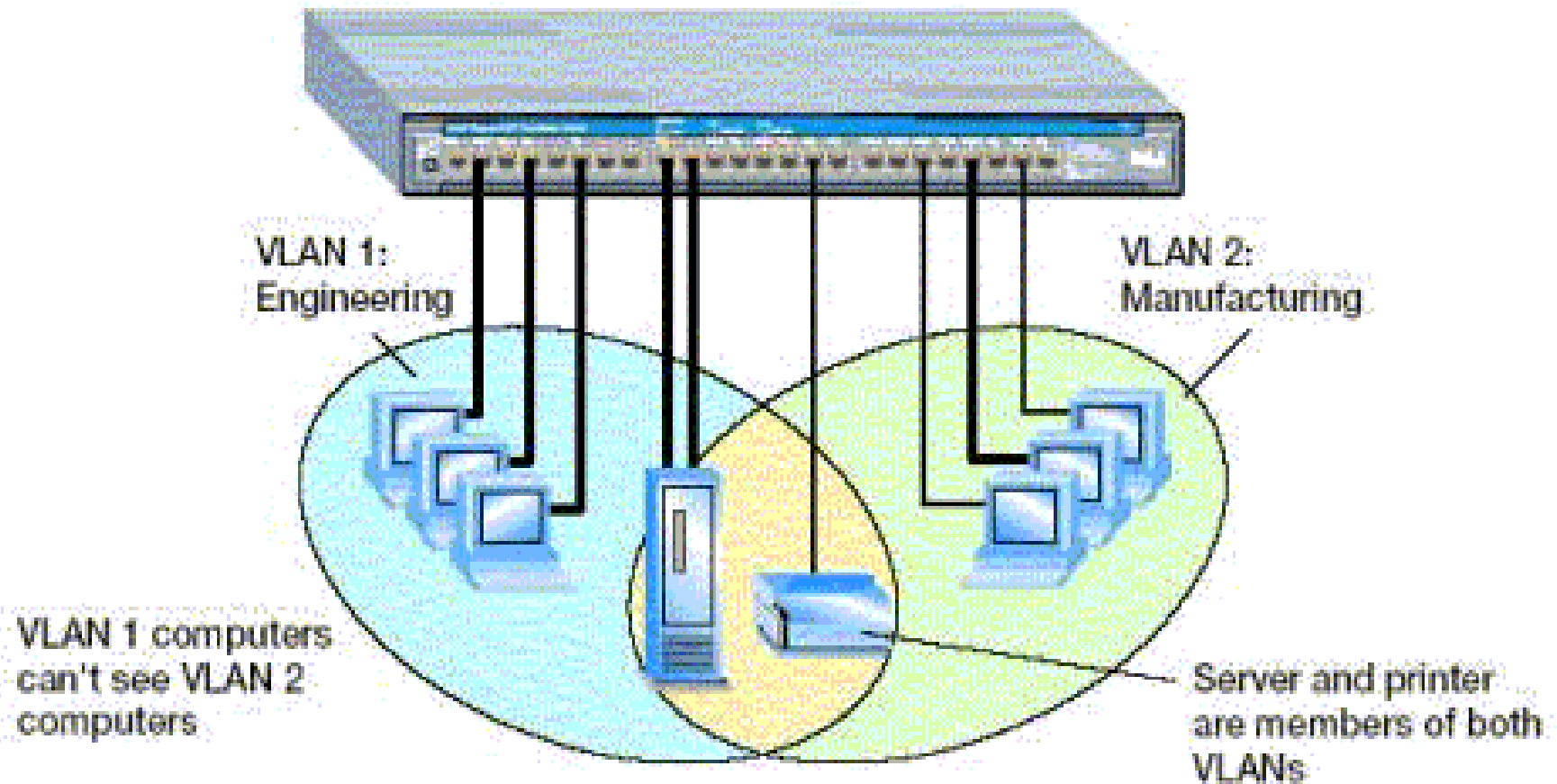
- minden IEEE 802 LAN MAC protokoll felett használható
- könnyű adminisztráció (létrehozás, tag hozzáadása, stb.)
- két különböző VLAN között nincs forgalom (adatkapcsolati szinten)
- egy LAN-on belül több csoport is elkülöníthető
- jól megválasztott csoportok esetén forgalomcsökkenés érhető el

- minden VLAN-nak van egy azonosítója: VID
- a switch nyilvántartja, melyik porthoz melyik VLAN-ok tartoznak
- egy felhasználó több VLAN-nak is tagja lehet
- egy port több VLAN-hoz is tartozhat
- a VLAN több switchre is kiterjedhet
- csökken a szórt adás által felemésztett kapacitás, biztonság nő

VLAN-ok átfedése



BME-TMIT

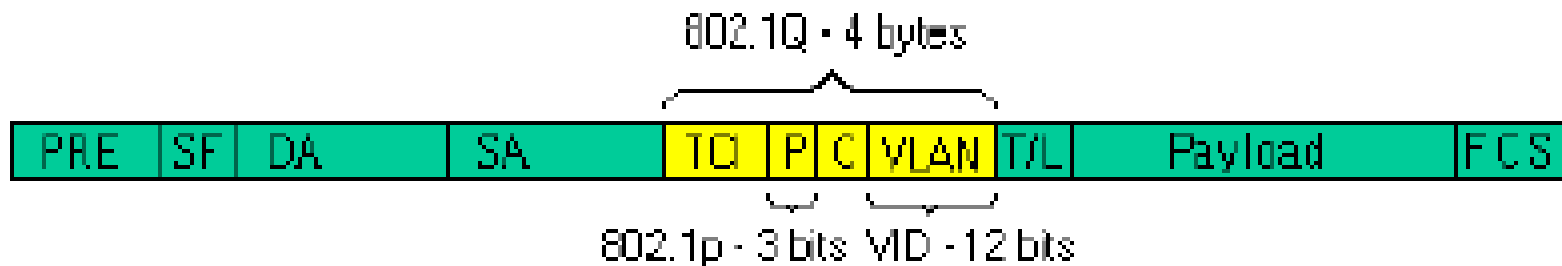


.1Q VLAN keretformátum



BME-TMIT

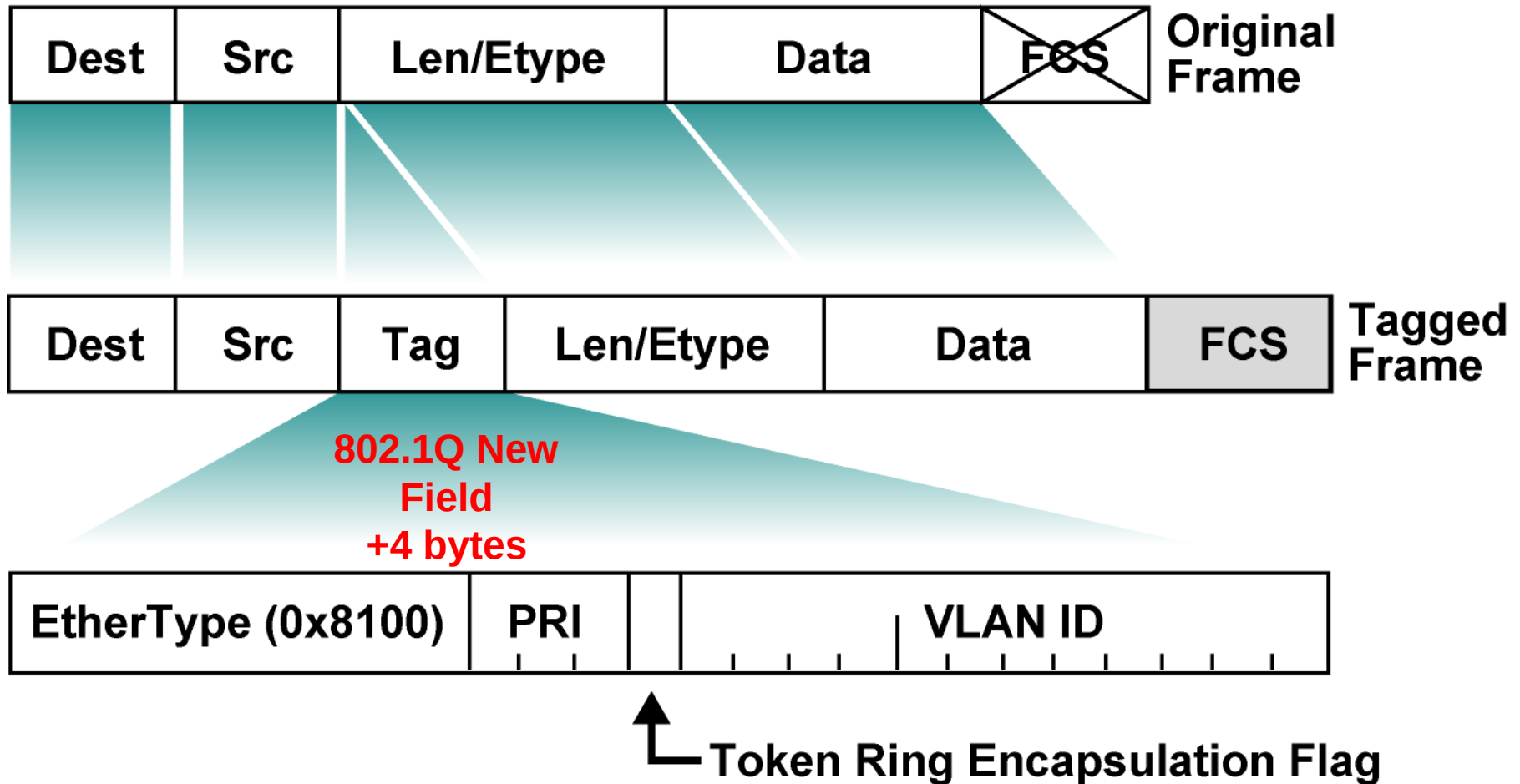
- A VLAN tag tárolásához új mezők az Ethernet-keretben
 - 4 bájtal nő az Ethernet keret
- Helye közvetlenül a forrás MAC cím után
- Lehetővé teszi a minőségbiztosítást is
 - 802.1p prioritás bitek



802.1Q Tag Formátum



BME-TMIT

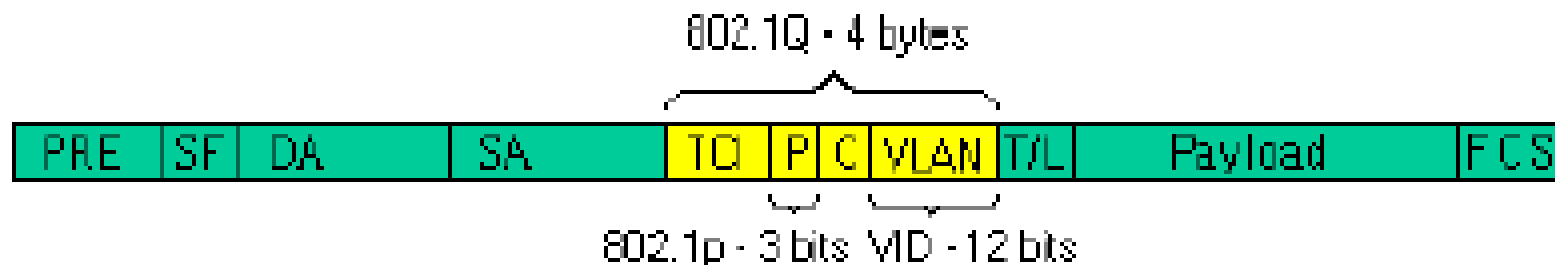


Új keretformátum – 2.



BME-TMIT

- TCI (Tag Control Info): 8100-as értéke mutatja a 802.1p és Q használatát
- P: prioritás (0..7)
- C (Canonical Indicator): megmutatja, hogy kanonikus formátumban vannak-e a MAC-címek
- VLAN: a VID-t tárolja (0..4095)



- Régi berendezések:
 - A TCI mező értéke 0x8100, melyet típusként értelmeznek
 - Ha nem támogatja, nem tudja értelmezni tehát eldobja a keretet
- megoldás: az új, tag-el rendelkező kereteket támogató eszközök csatlakoztatásánál a régi típusúak felé csatlakozó portokon a tag leválasztása

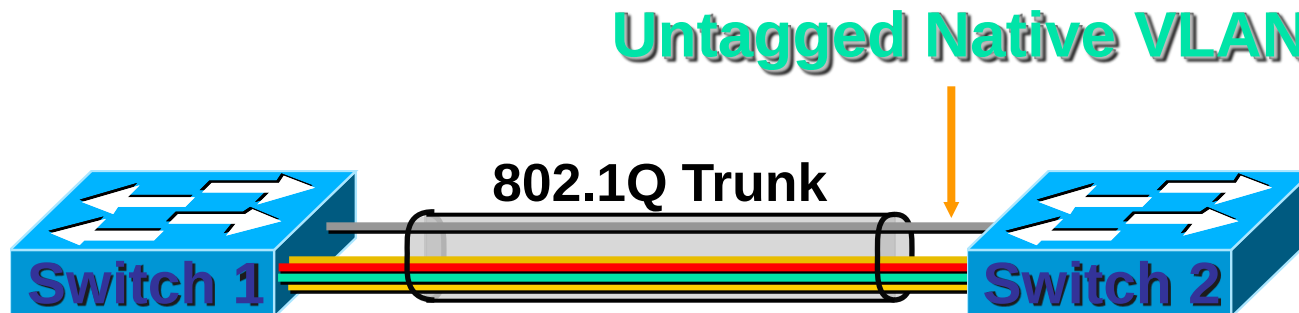
- VLAN Identifier
- 0-4095 tartomány
- 0: alapérték
 - ha a keret eredetileg tag nélküli volt, ezt kapja meg
- 4095: fenntartva, az ilyen VID-jű kereteket eldobjuk
- lehetőség van rá, hogy a tag nélküli keret 0-tól különböző értéket kapjon (port-based VLAN, Port VID)
 - Port VID: a switch minden portjának van, azt mutatja, hogy a tag nélküli keretet melyik VLAN-ba továbbítsák

A Natív VLAN és a 802.1Q



BME-TMIT

- Egyetlen nem-csomagolt VLAN – nincs Tag
- Általánosan elterjedt menedzsment célokra
- Meghatározva és kötelező a 802.1Q -ban
- Bármely VLAN lehet, nem feltétlen a VLAN 1
- Egy hálózati tartományban egységes



- megnézzük a fejléct
- ha nincs tag:
 - VID=0 értékkel teszünk bele
 - ha port alapú VLAN-t használunk, a port PVID-je lesz a VID
- ha van tag, szűrés:
 - megnézzük, kik a tagjai a VID által mutatott VLAN-nak
 - ha a fogadó port nincs a VLAN-ban, eldobjuk a keretet
 - egyébként megnézzük, a cél is tagja-e a VLAN-nak
 - ha igen, továbbítjuk

- megnézzük, a másik oldal támogatja-e a tagelt kereteket
 - ha nem, levesszük a taget
 - ha igen, továbbküldjük
- Átjárás VLAN-ok között csak routeren keresztül
 - Ha egy porton több VLAN van, a router virtuális interfészekként kezeli
 - Átjárás csak IP szinten!

VLAN-ok hozzárendelése



BME-TMIT

- Port-based VLANs: fizikai interfészenként
- MAC-based VLANs: a kapcsoló rendelkezik egy MAC-VLAN listával
- Protocol-based VLANs: a kapcsolóban be van állítva hogy milyen protokoll milyen VLAN-hoz tartozik
- IP subnet alapú (nem elterjedt)

- a legegyszerűbb megoldás
- a switch egyes portjaihoz hozzárendeljük a megfelelő VLAN azonosítóját
- egy porthoz legfeljebb egy azonosító tartozhat
- a VLAN változtatásakor csak a switch-hez kell nyúlni – a felhasználó számára átlátszó

- a switchben egy lista van, ez tartalmazza az egyes VLAN-okhoz tartozó eszközök MAC-címét
- előny:
 - ha egy felhasználó portot változtat a switch-en belül, az nem igényel semmilyen beavatkozást
- hátrányok:
 - általában nehezebb az adminisztráció (mindenki MAC-címét ismerni kell)

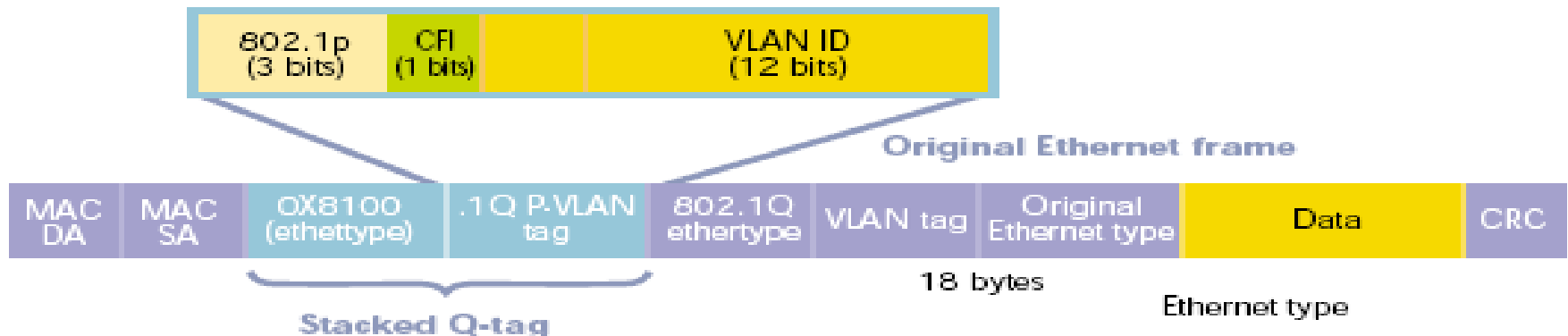
- Különböző protokoll típusokhoz más-más VLAN tag lesz rendelve
- Feltételezi, hogy a kapcsoló megnézi az IP fejléceket
- Nem felhasználót azonosít, hanem szolgáltatást

Q-in-Q (VLAN trönk)



BME-TMIT

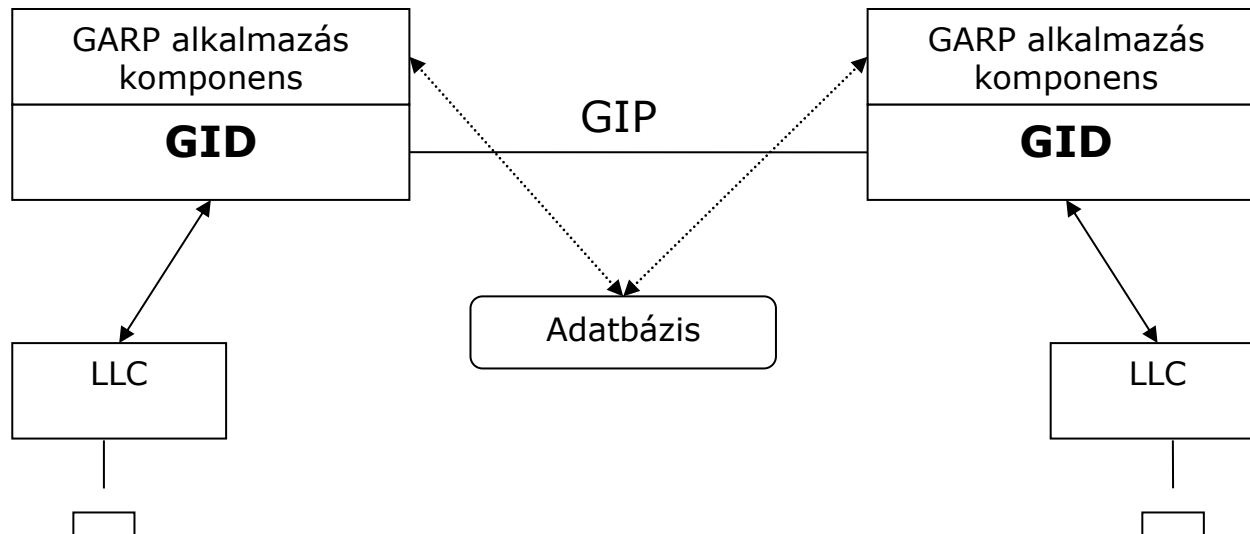
- 4096 VLAN nem elég egy nagyobb hálózatban (>4095 felhasználó)
 - A felhasználók azonosítására használva: C-VLAN
- Bevezetnek egy szolgáltatói VLAN tag-et
 - S-VLAN



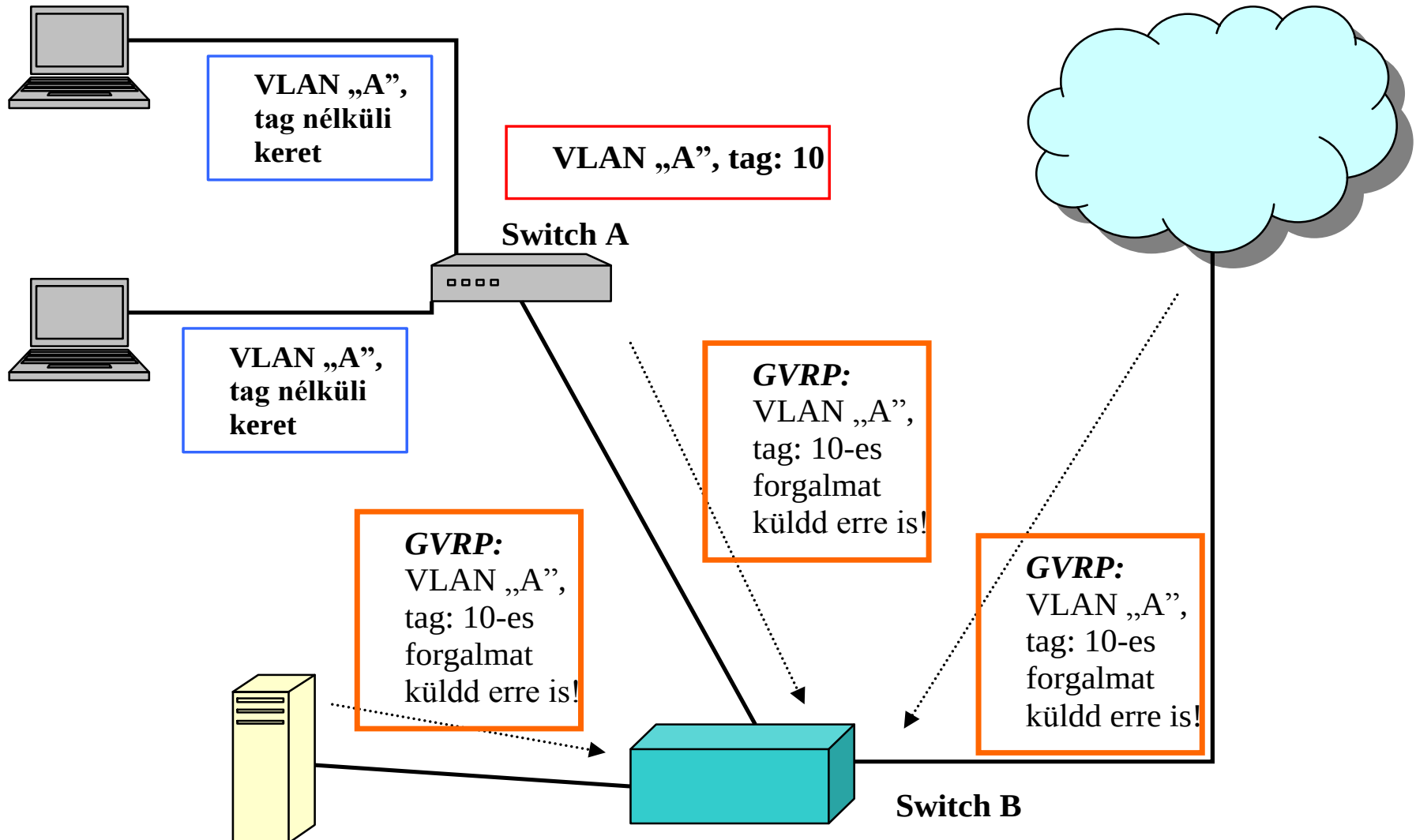
- Előnye hogy a VLAN –ok száma megnő
- Léteznek más megoldások is:
 - MAC-in-MAC
 - V-MAN tag bevezetése
 - MPLS tunnel
- A IEEE 802.1ad provider bridge szabvány a QiQ-t támogatja

- *Statikus:*
a VLAN tagság információk manuálisan állíthatóak, a dinamikus terjesztés protokolljai tiltva vannak.
- *Dinamikus:*
GVRP (GARP VLAN Registration Protocol) használatával az információ dinamikusan terjed, ha valami változás történik.
- *Vegyes:*
egyes nyilvántartott VLAN-ok információi csak statikusan, másoké csak dinamikusan változtathatóak.

- Alapja a GARP (Generic Attribute Registration Protocol).
 - GARP alkalmazás komponens
 - GARP Information Declaration (GID)
 - GARP Information Propagation (GIP)

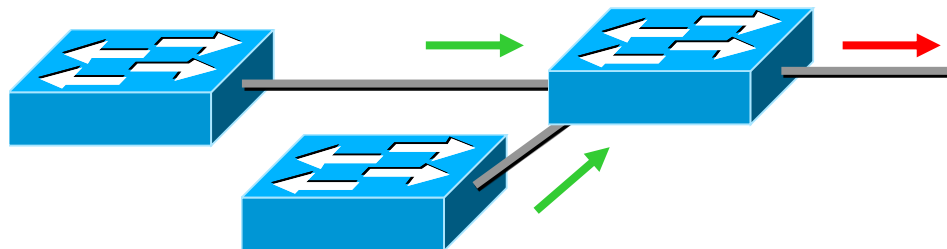


GVRP – GARP VLAN Registration Protocol



- Prioritás kezelés a 802.1Q VLAN tag alapján
 - 3 prioritás bit = 8 osztály
 - Ez azt jelenti hogy QoS-hez 802.1Q VLAN-okat kell használni
- A prioritás bitek értelmezését a 802.1P szabvány adja meg
 - A VLAN prioritás és az IP TOS hasonló
 - Jelen Ethernet kapcsolók nem mind támogatják a 8 osztályt

- IP QoS/ToS Mechanizmusok
 - Transzparens módon használhatók a TAG-al ellátott vagy anélküli linkeken
 - Ethernet szintű torlódás esetén az IP QoS nem garantált
 - Ethernet szintű torlódás:
- Mapping functionalitás szükséges az Ethernet CoS és IP ToS együttműködéséhez
- Az QoS/CoS aktuális megvalósítása QoS/CoS a kapcsoló architektúrájától függ és változó lehet
 - Nem egyértelműen szabványosított

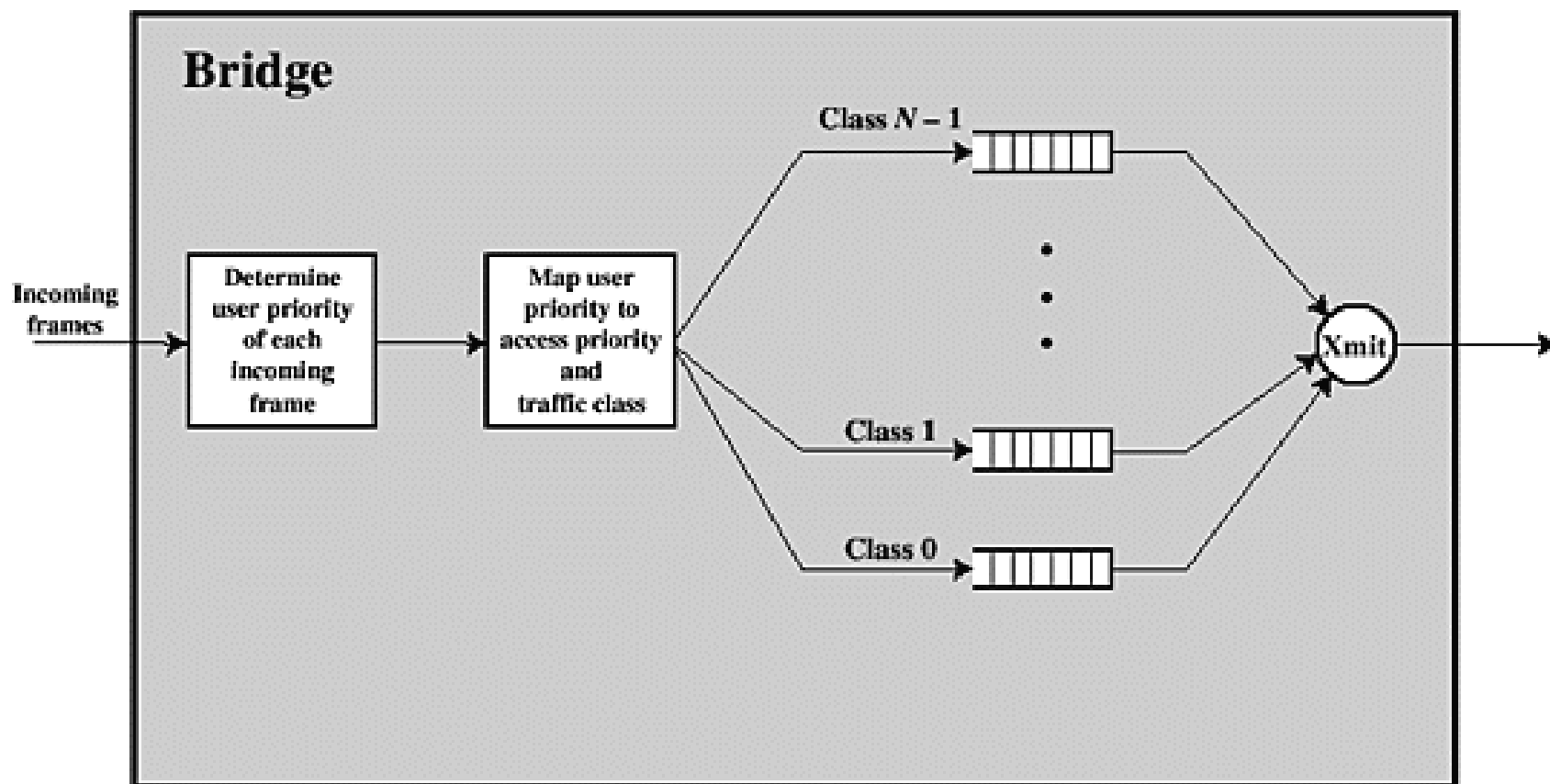


- **Network Control:** legnagyobb prioritás
- **Voice:** kisebb mint 10 ms késleltetés
- **Video:** kisebb mint 100 ms késleltetés
- **Controlled Load:** fontosabb alkalmazások
- **Excellent Effort:** fontosabb előfizetők BE forgalma
- **Best Effort:** alap prioritás
- **Background:** letöltések, játékok, stb

IEEE 802.1D forgalmi osztályok kezelése

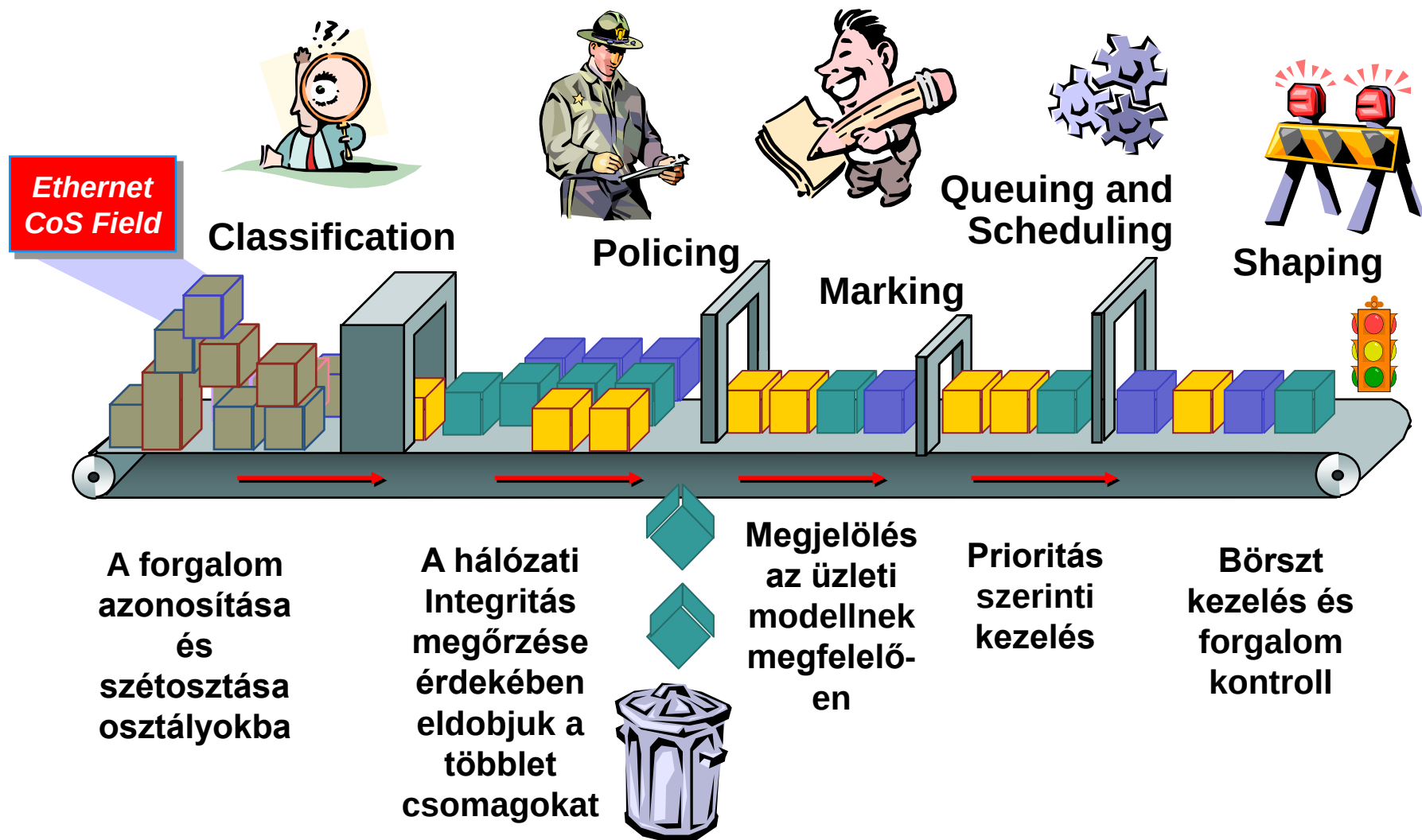


BME-TMIT



QoS Architecture Components

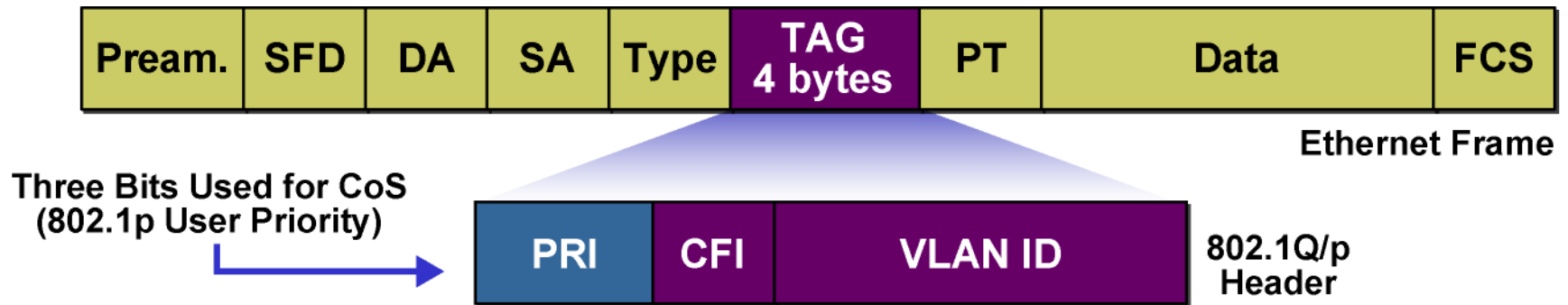
BME-TMIT



Layer 2 jelölés: 802.1p, CoS



BME-TMIT



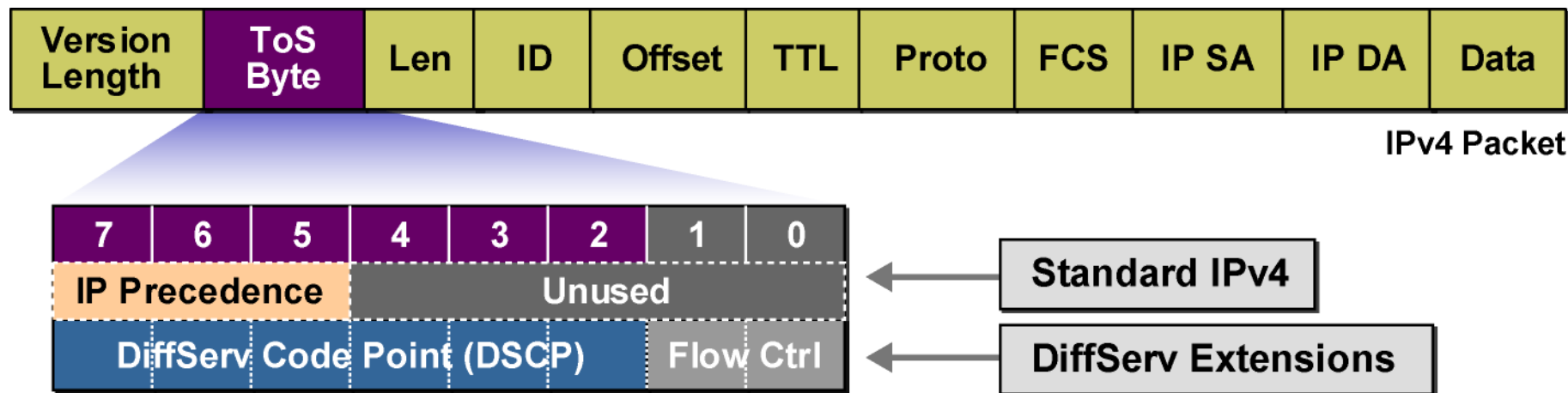
- 802.1p User Prioritást szokás Class of Service (CoS)-nek nevezni
- Különböző forgalmi típusok más-más CoS értéknek felelnek meg
- **Semmilyen CoS nincs a TAG nélküli keretek esetében !!**
 - VLAN ID = 0: alap prioritás

CoS	Typical Application
7	Reserved
6	Reserved
5	Voice Bearer
4	Video Conferencing
3	Call Signaling
2	High Priority Data
1	Medium Priority Data
0	Best Effort Data

Layer 3 jelölés: IP Precedencia



BME-TMIT



- IPv4
 - A ToS byte felső 3 bitje az IP Precedencia
 - A többi bit nem használt (delay, throughput, reliability, cost, unused)
- DiffServ
 - A ToS byte felső 6 bitje a DiffServ Code Point (DSCP)
 - DSCP visszafele kompatibilis a ToS-al
 - A megmaradt bitek - flow control

802.1P QoS protokoll (Microsoft)



BME-TMIT

- Prioritációs rendszer
- Nincs CAC – túlterhelhető a hálózat
- Nincs határ szabva az applikáció forgalma számára – implementáció kérdése
- QoS együttműködés a Subnet Bandwidth Manager (SBM) alapján

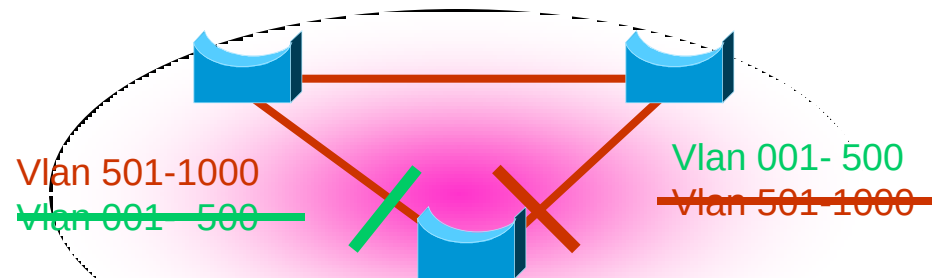
- E2E QoS biztosítása Ethernet és más technológia között
 - RSVP+802.1P
- SBM egy jelzési protokoll az RSVP alapú CAC-hoz IEEE 802 hálózatok (pl. Ethernet) számára
 - Leírja az RSVP-képes eszközök és hálózati szintű eszközök (switchek) működését hogy támogassák az erőforrásfoglalást
 - Összehangolja a 802.1p prioritások kezelését a kapcsolókban
 - Megfelelteti a szolgáltatási osztályokat az RSVP kliensek és az RSVP-t támogató hálózatok között

- Az RSTP hátránya: rossz hálózati kihasználás
- Cisco: PVST (Per-VLAN feszítőfa)
 - Minden VLAN: egy RSTP
 - Sok VLAN – nem skálázható, fölösleges
- IEEE: MSTP
 - Lehetővé tesz több feszítőfát
 - A VLAN-ok a feszítőfákhoz vannak rendelve

PVST vs. 802.1Q vs. 802.1s

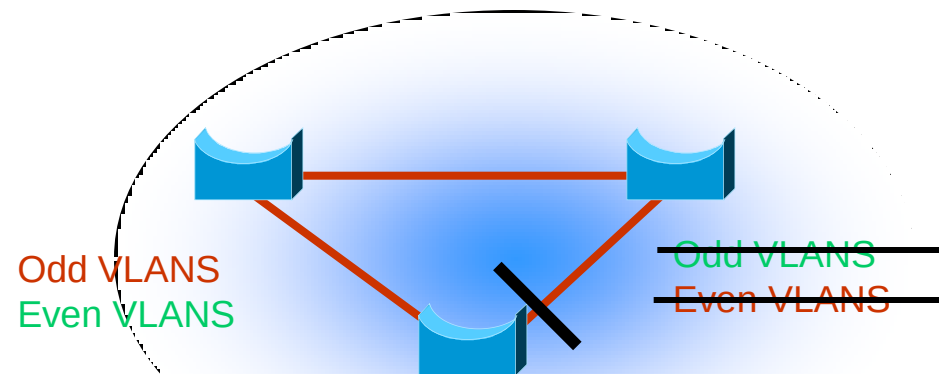


BME-TMIT



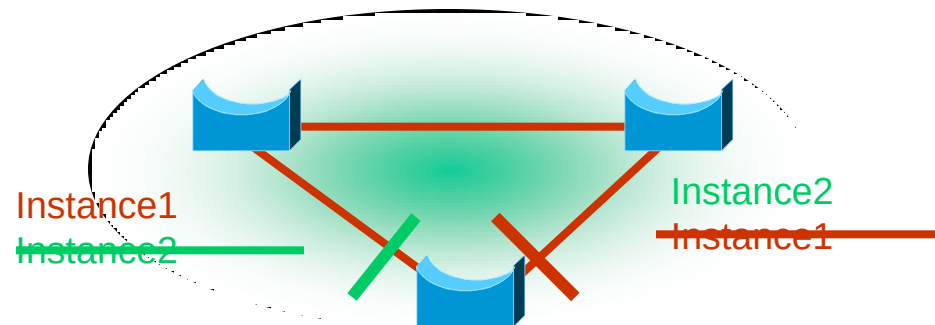
Prop. PVST egy STP per VLAN

Skálázhatóság 1000 nagyságrendű Vlan-ra (SPT!) nagy kihívás



802.1Q egyetlen SPT

Load sharing nem lehetséges



802.1s több SPT

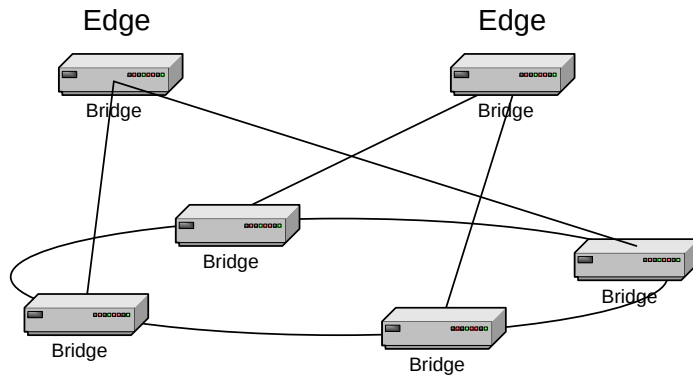
HATÉKONY Load sharing lehetséges

- RSTP alapú, a szabvány továbbfejlesztése
- Max. 64 fa (MST instance)
- Minden fának beállíthatjuk
 - A gyökerét
 - A link cost-okat
 - A hozzá tartozó VLAN-okat
- Egy VLAN csak 1 fához tartozhat!

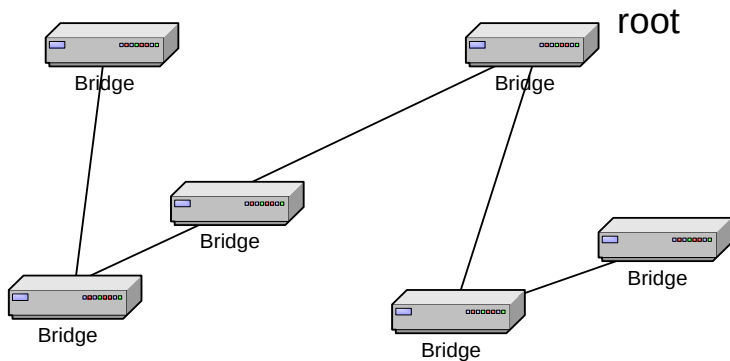
MSTP előnyei



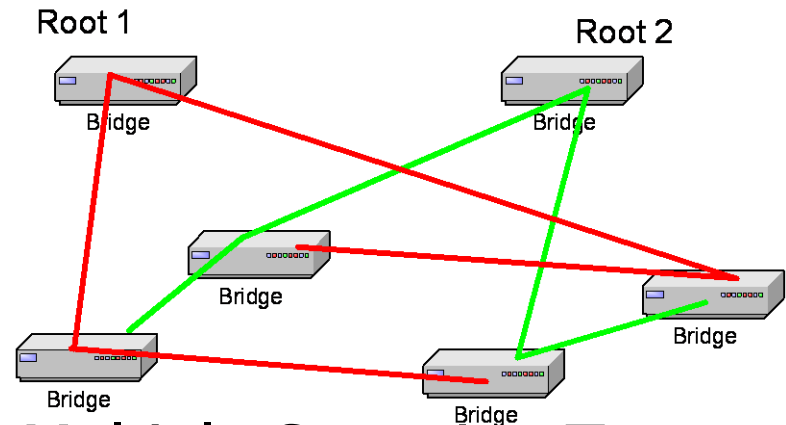
BME-TMIT



- Hálózati topológia: 2 kijárat
- A gyűrű redundanciát jelent
 - Nagyobb megbízhatóság



- STP: Egy feszítőfa



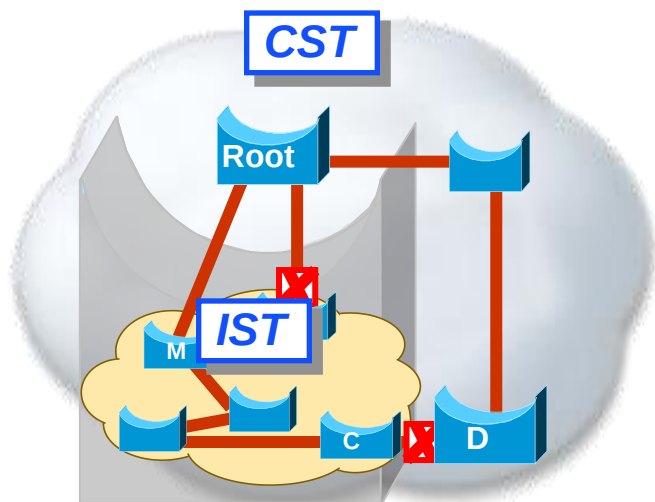
- Multiple Spanning Tree
 - 2 feszítőfa

802.1s: CST, IST, MST – Sok fa

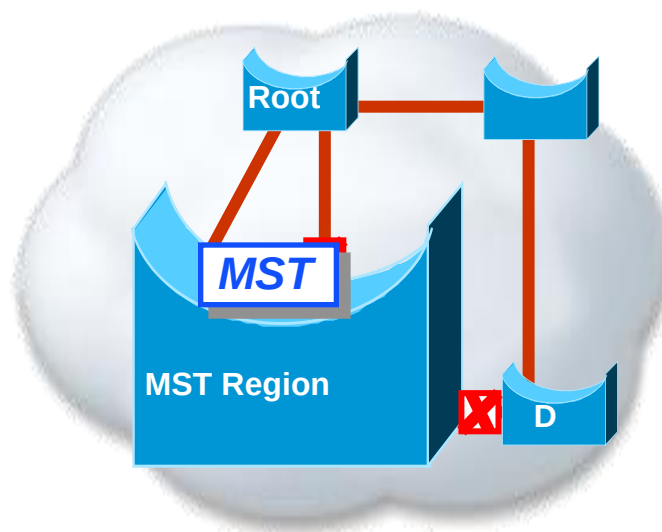


BME-TMIT

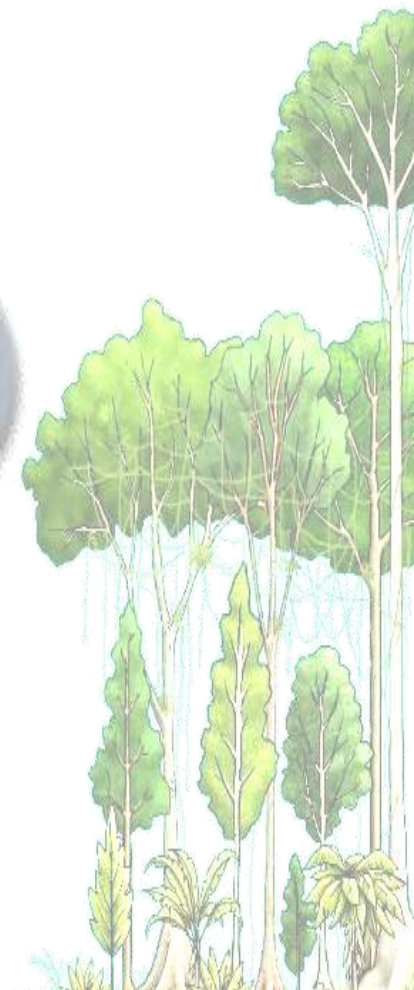
Belső nézet



Külső nézet



- **CST 802.1Q Common SPT** => Egyetlen fa
- **IST 802.1s Internal SPT** => a külső világ számára az MST-t egyetlen CST kapcsolónak mutatja
- **MST 802.1s Multiple SPT** => több VLAN egyetlen MST Instance-ba való összefogása



- Az MSTP lehetőséget nyújt régiók kialakítására
- A régiókat egymástól adminisztratív módon választhatjuk el
 - RG mező
- Előnyök:
 - egy régió belüli hiba nem zavarja a többi régió működését
 - Negy hálózat felbontása régiókra javítja a skálázhatóságot
 - A VLAN-ok lokális jelentőséget kapnak

MST régiók - 2



BME-TMIT

- Miért használjunk régiót?
 - Különböző adminisztratív vezérlés az L2 hálózat különböző részei között
 - Nem minden switch támogat/futtat MSTP-t – különböző STP-k felosztják a hálózatot STP régiókra
 - Az MST előnyei régióon belül érvényesülnek azon kívül egyetlen példány (topológia) minden VLAN számára
- Az MST régió egy csoport összekapcsolt MST switch amely ugyanolyan MST konfigurációt használ
- Régióon belül: több példány
 - IST – Internal Spanning Tree (instance 0), mindig jelen van minden porton
 - MSTI - Multiple Spanning Tree Instance
- Régióon kívül: egy példány

IST Master



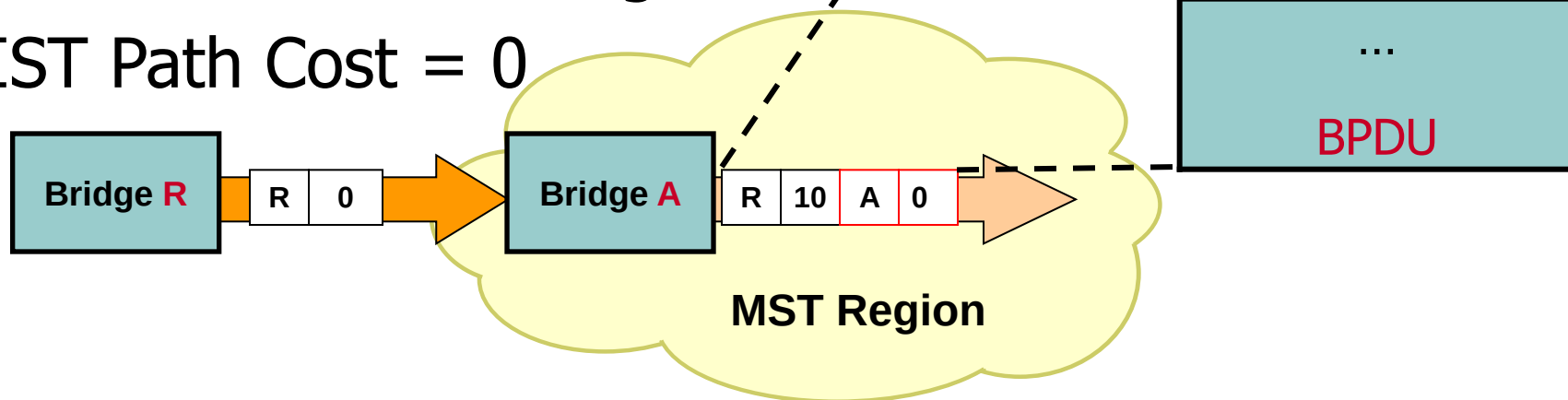
BME-TMIT

- A határnál az MST bridge hozzáadja:

- IST Master ID
- IST Master Path Cost

- Alapértelmezés:

- IST Master ID = Bridge ID
- IST Path Cost = 0

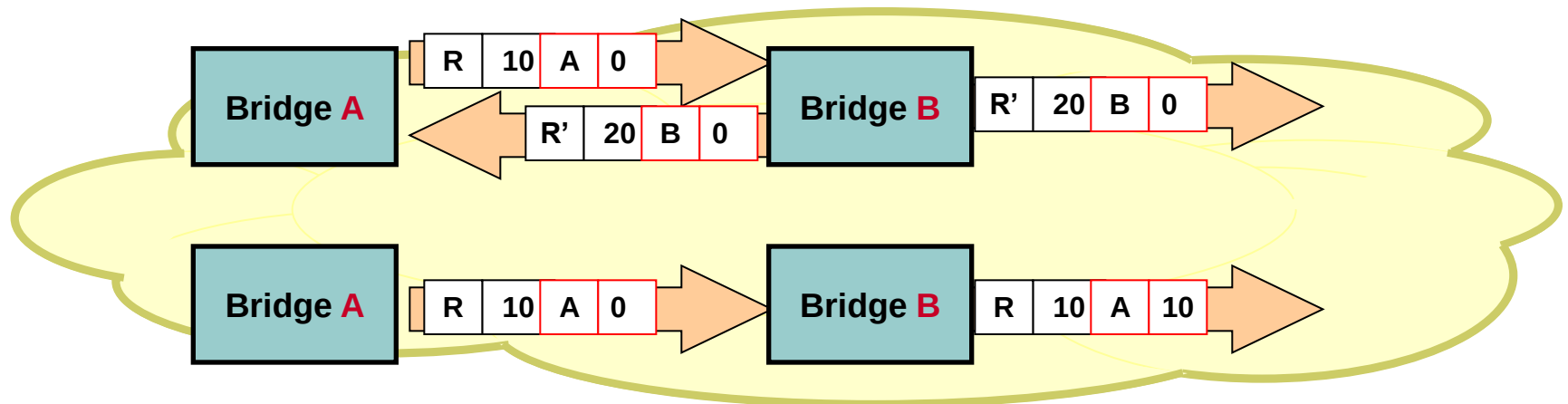


IST Master választás

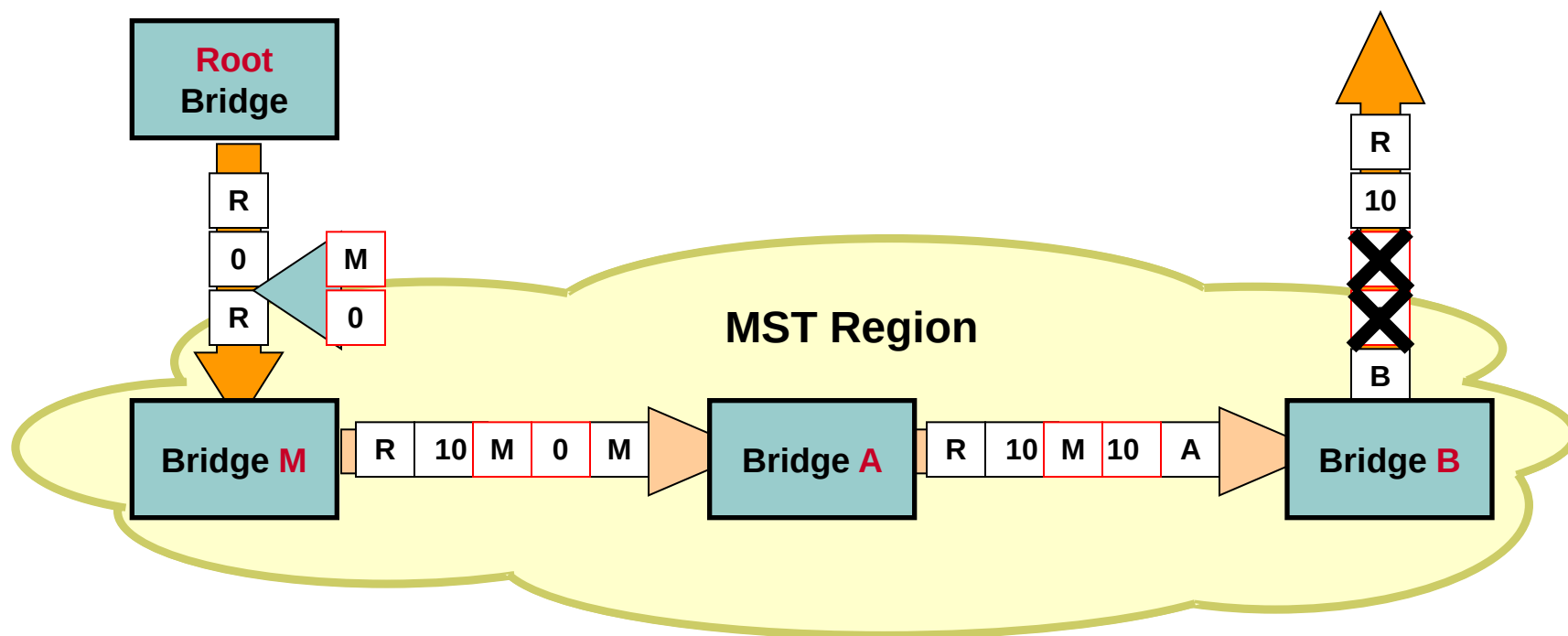


BME-TMIT

- Hasonló a Root választáshoz
 - Kezdetben minden bridge a régióban IST Master-nek tartja magát
 - Az a bridge amely jobb BPDU-t kap, a kapott IST Master-t küldi tovább
 - Az IST Root lesz a master – ha ugyanaz a switch – vagy a Root-hoz legközelebbi



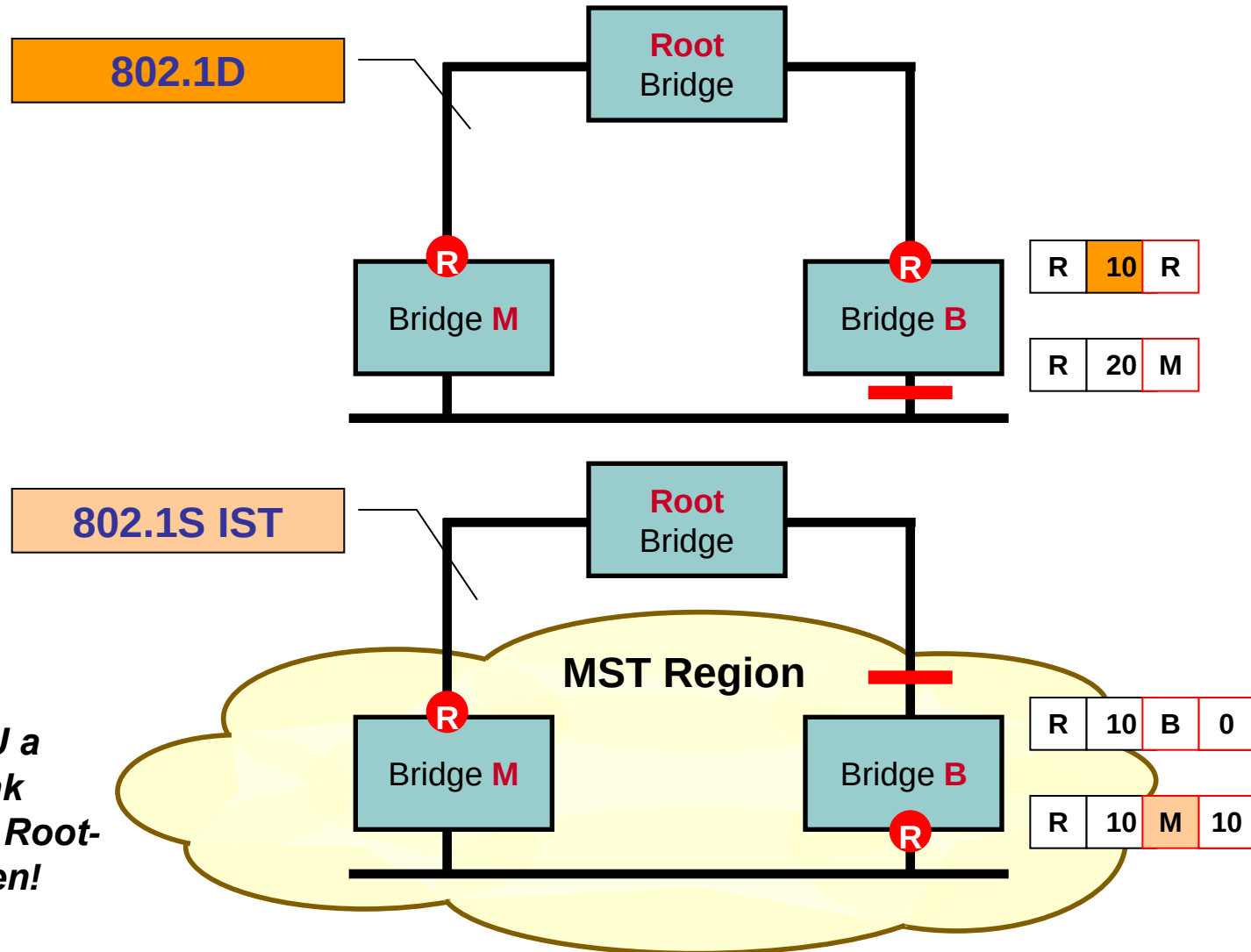
- A Root path cost mindig 1-el nő, mintha 1 bridgen haladt volna át
- Az IST csak az IST master ID-t és master Path Cost-ot használja



802.1D és 802.1S IST konvergencia



BME-TMIT

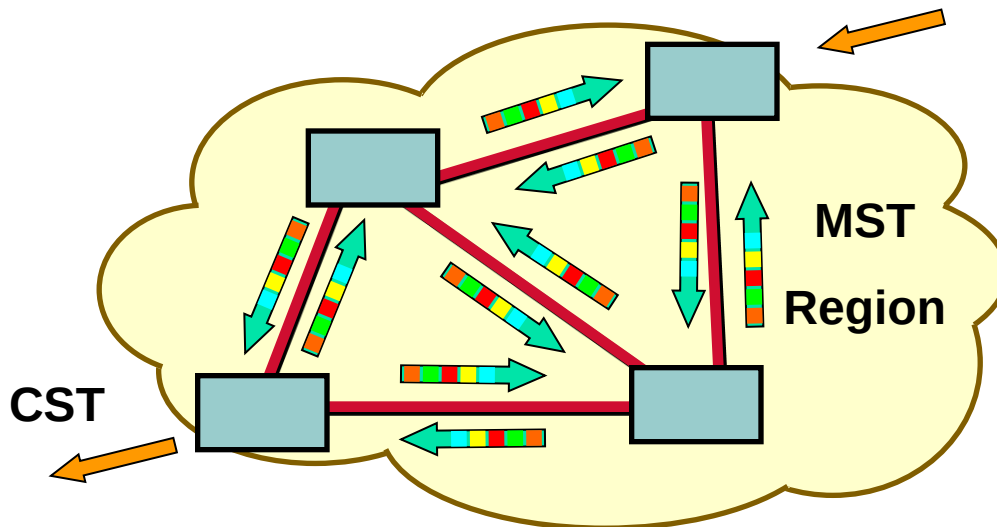


MST instance-ok



BME-TMIT

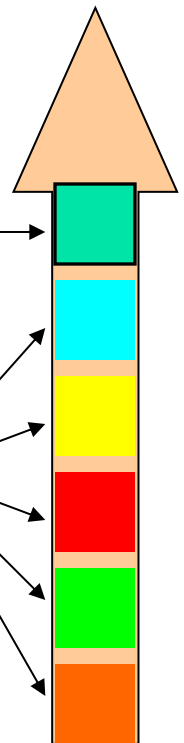
- Az MSTI-k STP példányok amelyek **csak a régió belül** értelmezettek
- Az MSTI-k nincsenek kapcsolatban a régió kívüli eszközökkel
- Az MST egyetlen BPDU-t küld az összes példánynak egy M-rekorddal példányonként
- Egyetlen példánynak van timer-alapú paramétere (az IST instance)
- Az MST BPDU-k **minden porton kiküldődnek**
- **A BPDU-k mindkét irányban küldődnek** ellentétben a 802.1D-vel, ahol csak a designated bridge küld



Protokoll információ az IST számára

Protokoll információ az MST példányoknak

MST BPDU

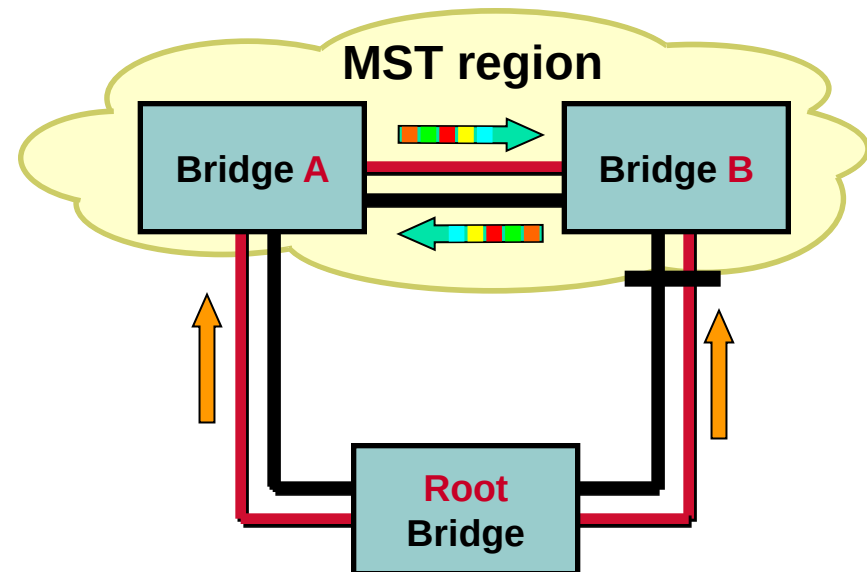


MSTI-k a határnál



BME-TMIT

- Csak az IST kapcsolódik a külső STP-hez (CST)
- Az MSTI-k nem küldenek BPDU-kat a határ portokon
- Egy határ port az MSTI-ben mindig követi az IST állapotát
- Példa: B esetében a piros példány blokkol mert az IST is blokkol

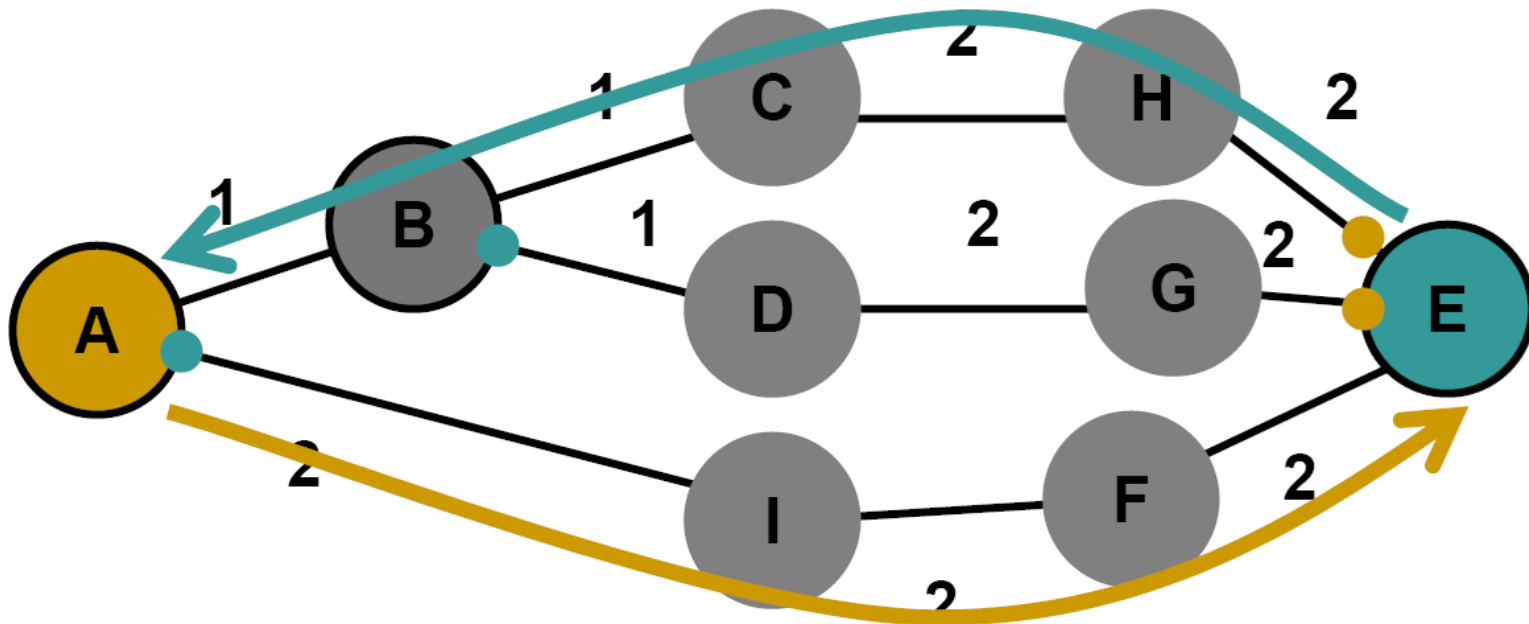


Shortest Path Bridging



BME-TMIT

- IEEE 802.1aq
- Több fa, gyökerek a bridge-ekben
 - Mindegyik a legrövidebb utat használja
- Megoldandó
 - Szimmetrikus útvonalak...



A lényegesebb L2 Protokollok ...



BME-TMIT

Destination MAC Address	Name	Description
01-80-C2-00-00-2X	GARP*	IEEE 802.1D, Generic Attribute Registration Protocol. The “carrier” protocol over which GMRP and GVRP are implemented.
	GMRP*	IEEE 802.1D, GARP Multicast Registration Protocol. Prunes delivery of multicast MAC addresses back from ports that don't need to see them. L2 equivalent of IGMP.
	GVRP	IEEE 802.1Q, GARP VLAN Registration Protocol. Prunes each VLAN's broadcasts, multicasts, and unicast floods back from ports they don't need to go to.
01-80-C2-00-00-10	All Bridges*	IEEE 802.1D. Defined as an ordinary multicast address to be used to reach all bridges in a bridged LAN.
01-80-C2-00-00-04 - 01-80-C2-00-00-0F	Undef. 802.1 bridge addrs.	Reserved for use by 802.1. IEEE 802.1D states that a bridge will never forward a frame with one of these addresses.

A lényegesebb L2 Protokollok ...



BME-TMIT

Destination MAC Address	Name	Description
01-80-C2-00-00-00	STP	IEEE 802.1D, Standard Spanning Tree Protocol. Protocol packets called, “Bridge Protocol Data Units”, or BPDUs.
	RSTP	IEEE 802.1W, Rapid Spanning Tree protocol (RSTP). Same function as STP, but converges (typically) in tens of milliseconds, rather than tens of seconds.
	MSTP	IEEE 802.1S, Multiple Spanning Tree Protocol. Carries multiple STP instances on top of a single RSTP BPDU.
01-80-C2-00-00-01	Pause	IEEE 802.3 Clause 31, Point-to-point Pause function. Used to implement L2 flow control on a whole physical link. Handled by hardware.
01-80-C2-00-00-02	LACP	IEEE 802.3 Clause 43, Link Aggregation Control Protocol. Protocol to automatically establish groups of point-to-point links between two devices for load sharing.
	OAM	IEEE 802.3ah EFM Draft 1.3, Operations, Administration, and Maintenance.
	LLDP	IEEE 802.1ab Draft, Link Layer Discovery Protocol. Allows stations to exchange chassis and port information.
	Slow Protocols	Future IEEE 802 standard protocols which expect no more than about 1 packet per second are expected use this MAC address.

A lényegesebb L2 Protokollok ...



BME-TMIT

Destination MAC Address	Name	Description
01-80-C2-00-00-03	802.1X	IEEE 802.1X, Port-Based Network Access Control. Port-level secure authentication, usually using a RADIUS server.
01-00-5E-XX-XX-XX	IGMP*	IETF RFCs 1112 and 2236, Internet Group Management Protocol. Layer 2.5 Multicast subscription protocol which runs between hosts and routers. Snooped by switches to control distribution of L2 multicast MAC addresses.
00-00-5E-00-00-XX (Unicast address)	VRRP*	IETF RFC 2338, Virtual Router Redundancy Protocol. This unicast MAC address may move around. It may be used by two different MACs in two different locations on a bridged network, on different VLANs.
* This protocol's packets may be tagged with a VLAN ID.		

33 speciális Layer 2 multicast MAC cím van: 16 a BPDU blokk-ban, 16 a GARP blokk-ban, és egy "All Bridges" cím.

Néhány protokoll pont-pont kapcsolat alapú és nem szabad multipoint-to-multipoint fölött használni

- Speciális Multicast MAC címek
- Kevesebb mint IPv4 multicast
 - Mapping kell
- GMRP – Multicast regisztráció
 - GMRP és IGMP együttműködés szükséges

Multicast – IGMP snooping



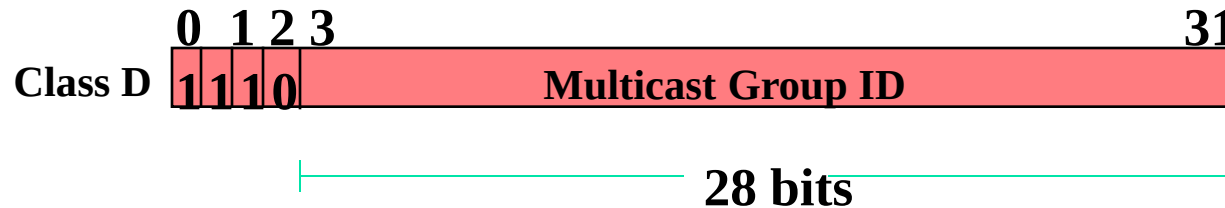
BME-TMIT

- A kapcsoló figyeli az IGMPv2 üzeneteket
- Minden IGMP join üzenetre hozzáad egy bejegyzést a bridge által megtanult címekhez
 - Így a következő multicast csomag csak azon a porton fog megjelenni
- Az IGMP leave üzenet törli a bejegyzést
- Megszűnik a multicast broadcast jellegű továbbítása
- IGMPv3 – még nem széles körben támogatott!

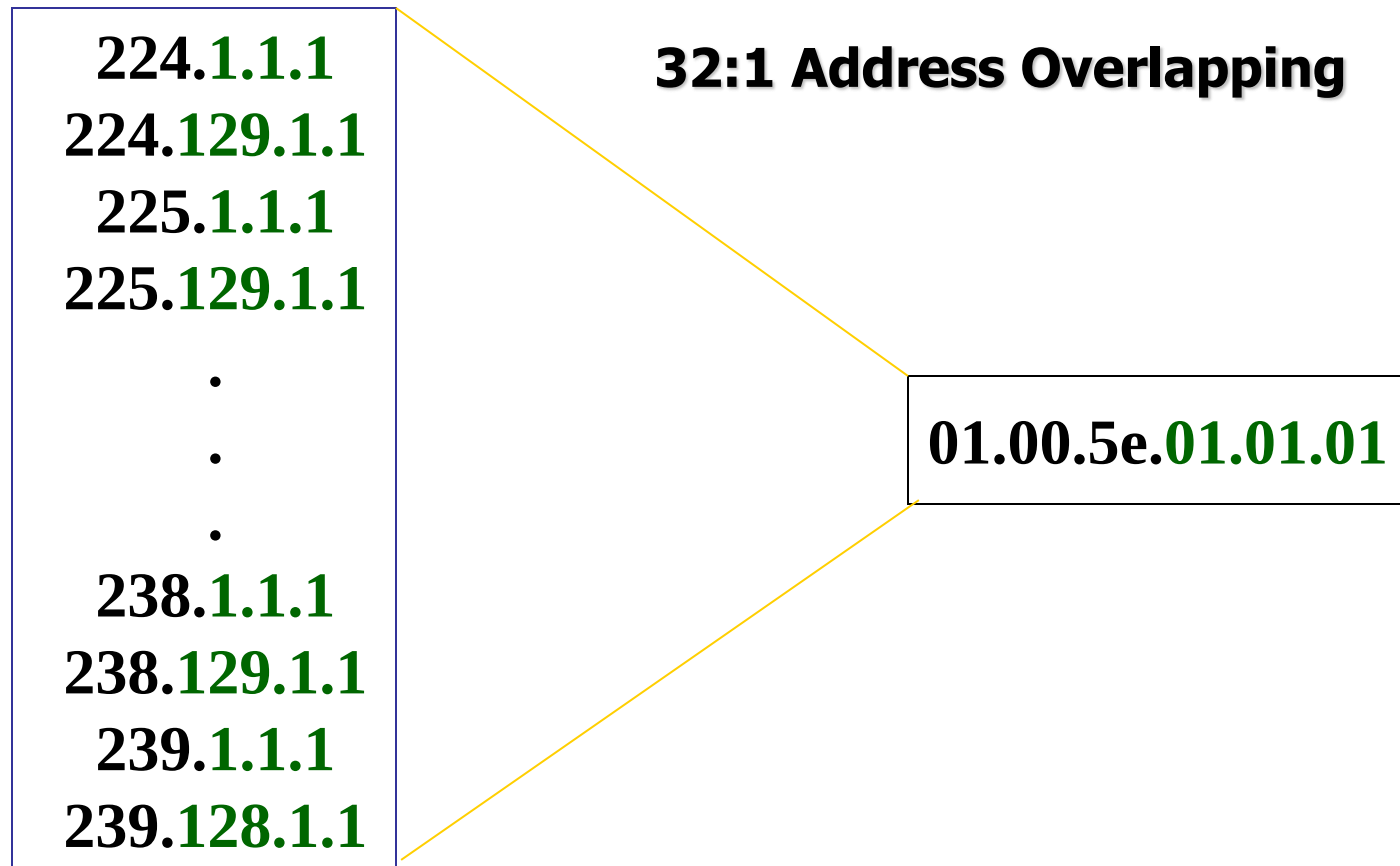
Multicast címzés



BME-TMIT



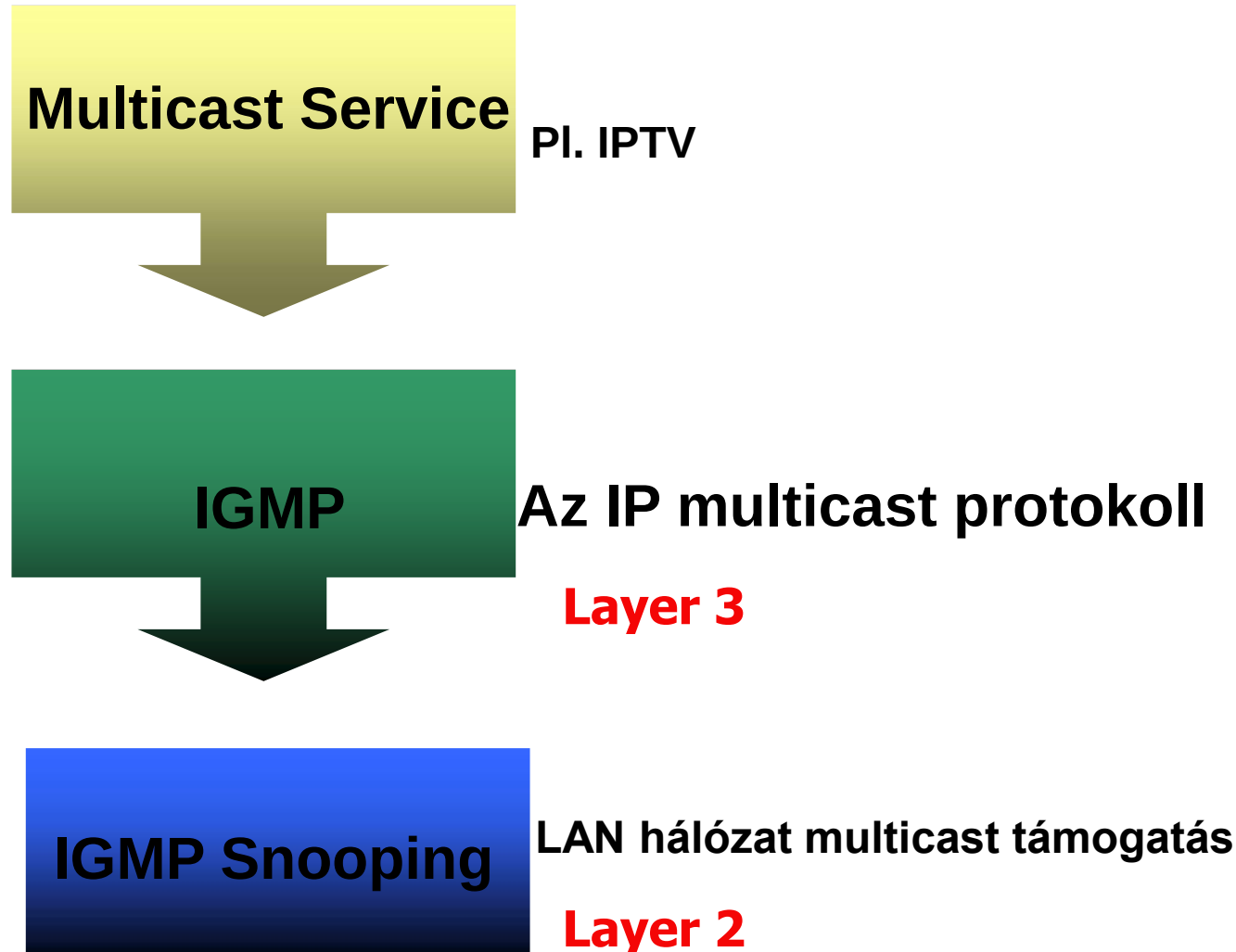
- Internet Assigned Numbers Authority (IANA)
- 224.0.0.1-224.0.0.255-->Reserved
- **224.0.1.0-238.255.255.255-->Multicast Group**
- 224.0.0.1: All multicast-capable hosts group
- 224.0.0.2: All multicast routers group
- 224.0.0.4: All DVMRP routers



Multicast Service



BME-TMIT



- Internet Group Management Protocol
- Membership management
- Membership establishment
- IGMPv1 [RFC1112](#)
- IGMPv2 [RFC2236](#) (aktuális verzió)
- IGMPv3 [RFC3376](#)

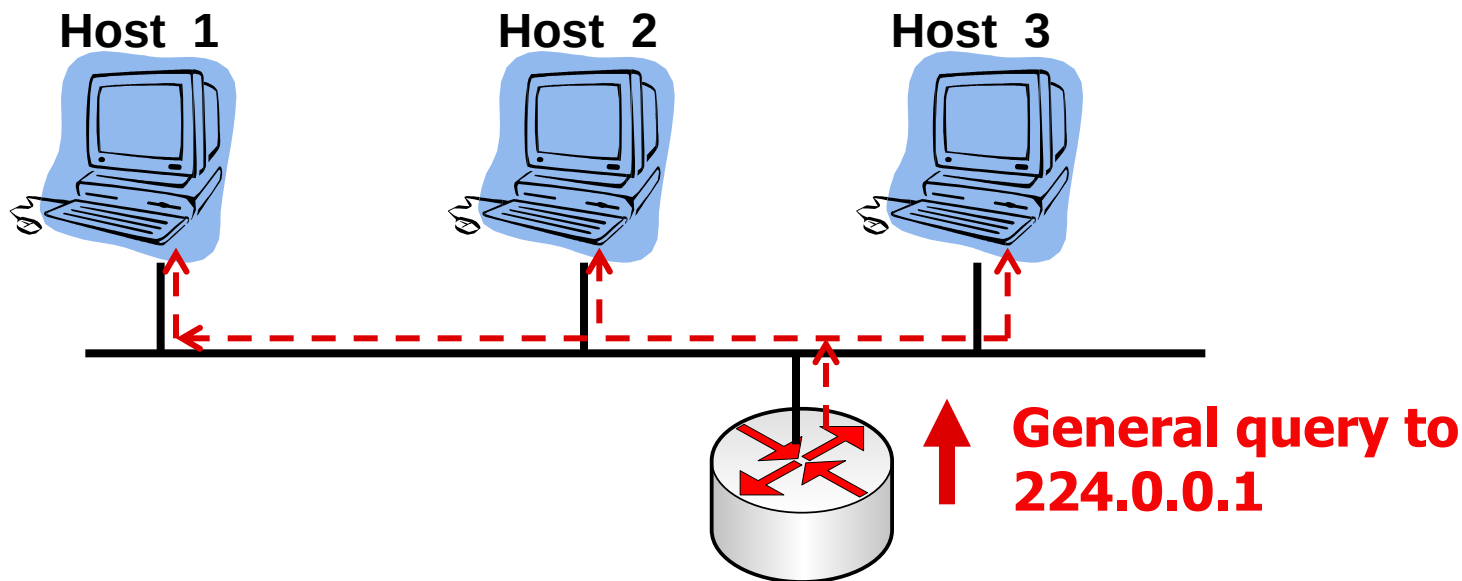
- General Query
 - Periodikus lekérdezés, tagsági információ lekérdezésére.
- Join Report
 - Host csatlakozni akar egy csoporthoz
- Leave Message
 - Host elhagyni készül egy csoportot

General Queries



BME-TMIT

- Példa



Periodikus lekérdezés a router által

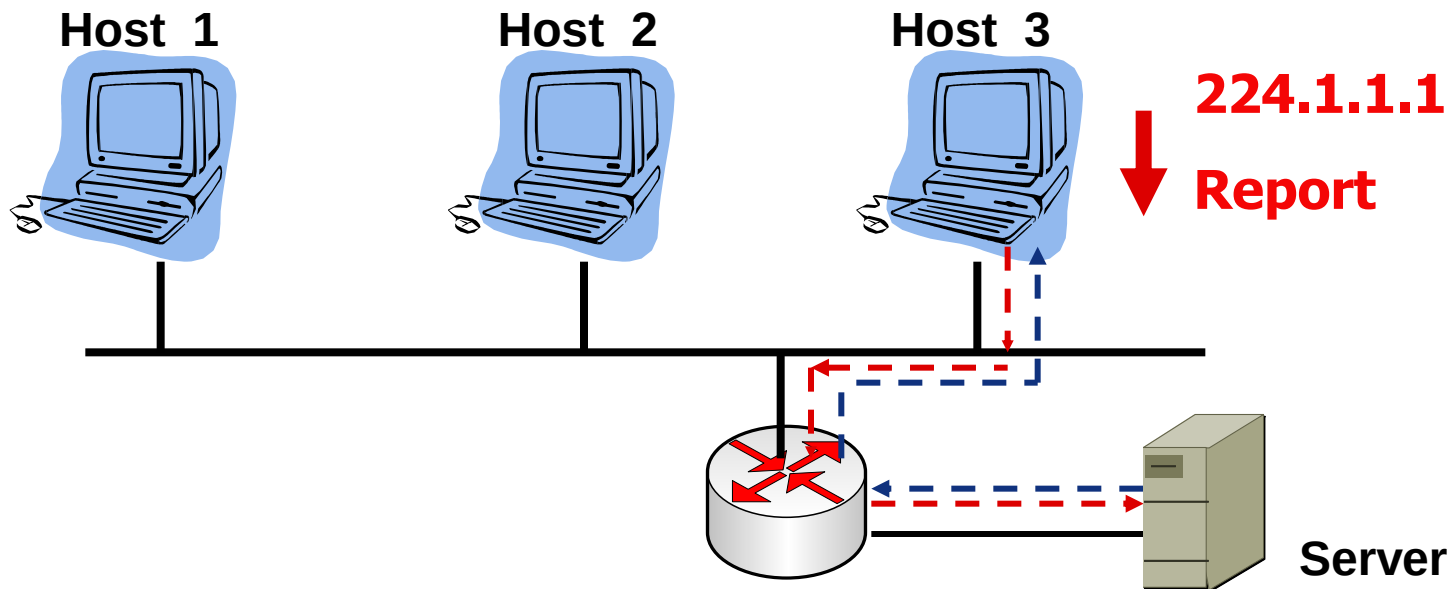
224.0.0.1: All multicast-capable hosts group

Join – csatlakozás csoporthoz



BME-TMIT

- Példa



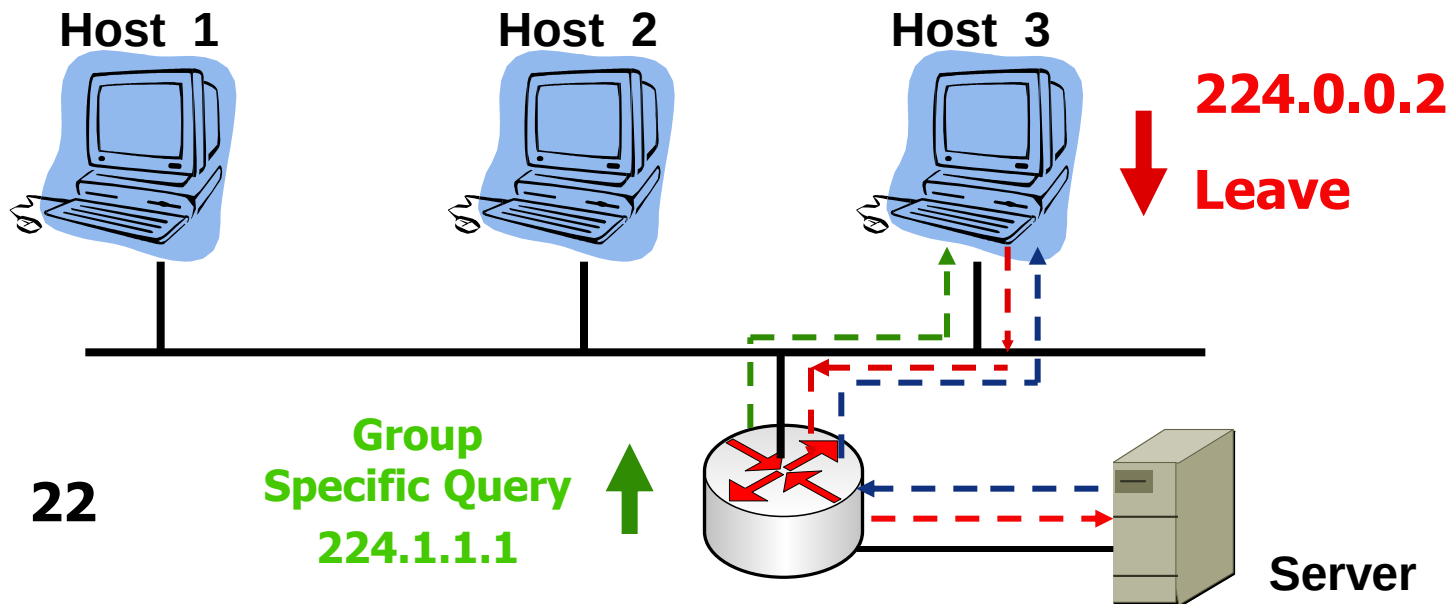
**Csatlakozás – a hoszt küld egy „report”
üzenetet a 224.1.1.1 címre**

Leave - csoport elhagyása



BME-TMIT

- Példa



Leave üzenet küldése a 224.0.0.2 címre a csoport elhagyásához

224.0.0.2 – The multicast address for all routers on subnet

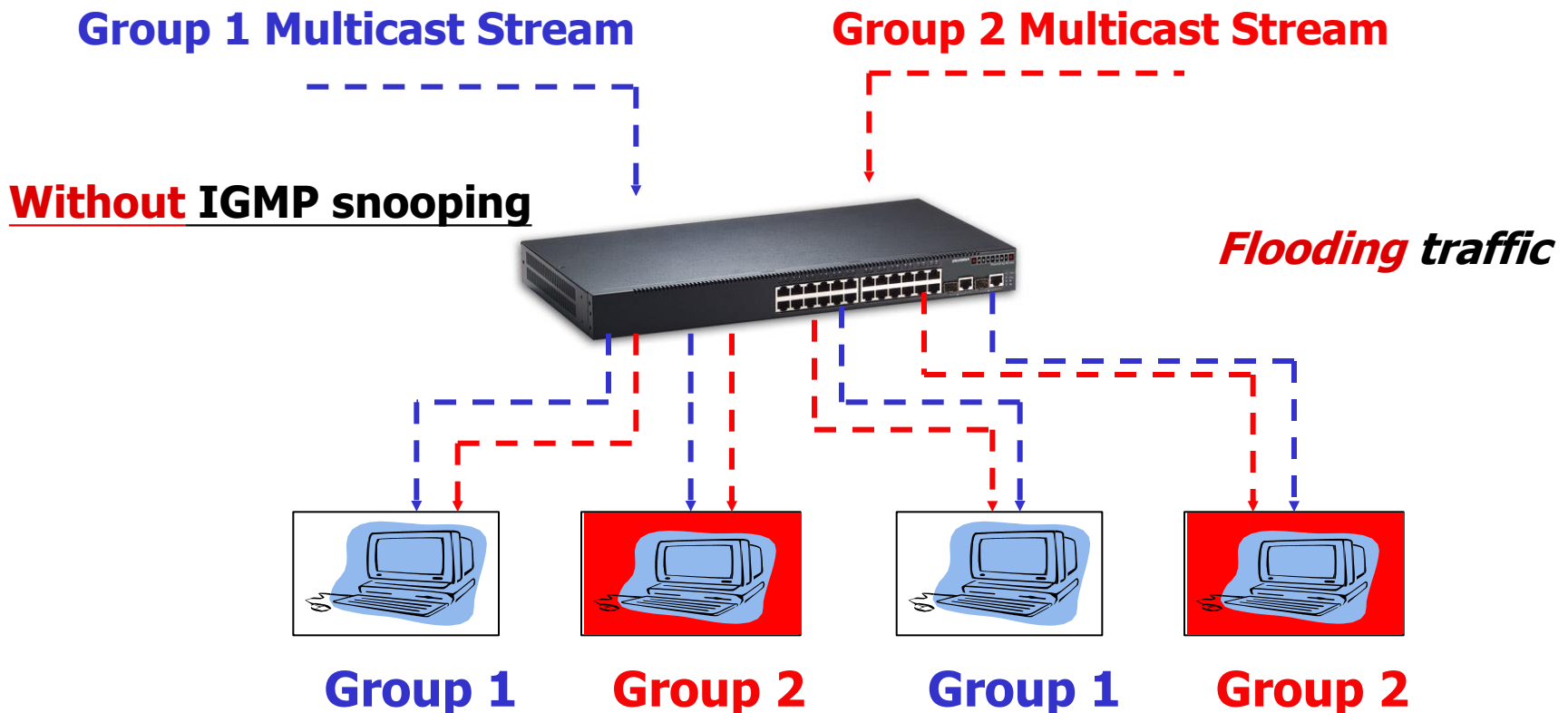
Az IGMP Snooping



BME-TMIT

IGMP Snooping - hatékony multicast Etherneten

All hosts need to handle the traffic whether they need it or not.



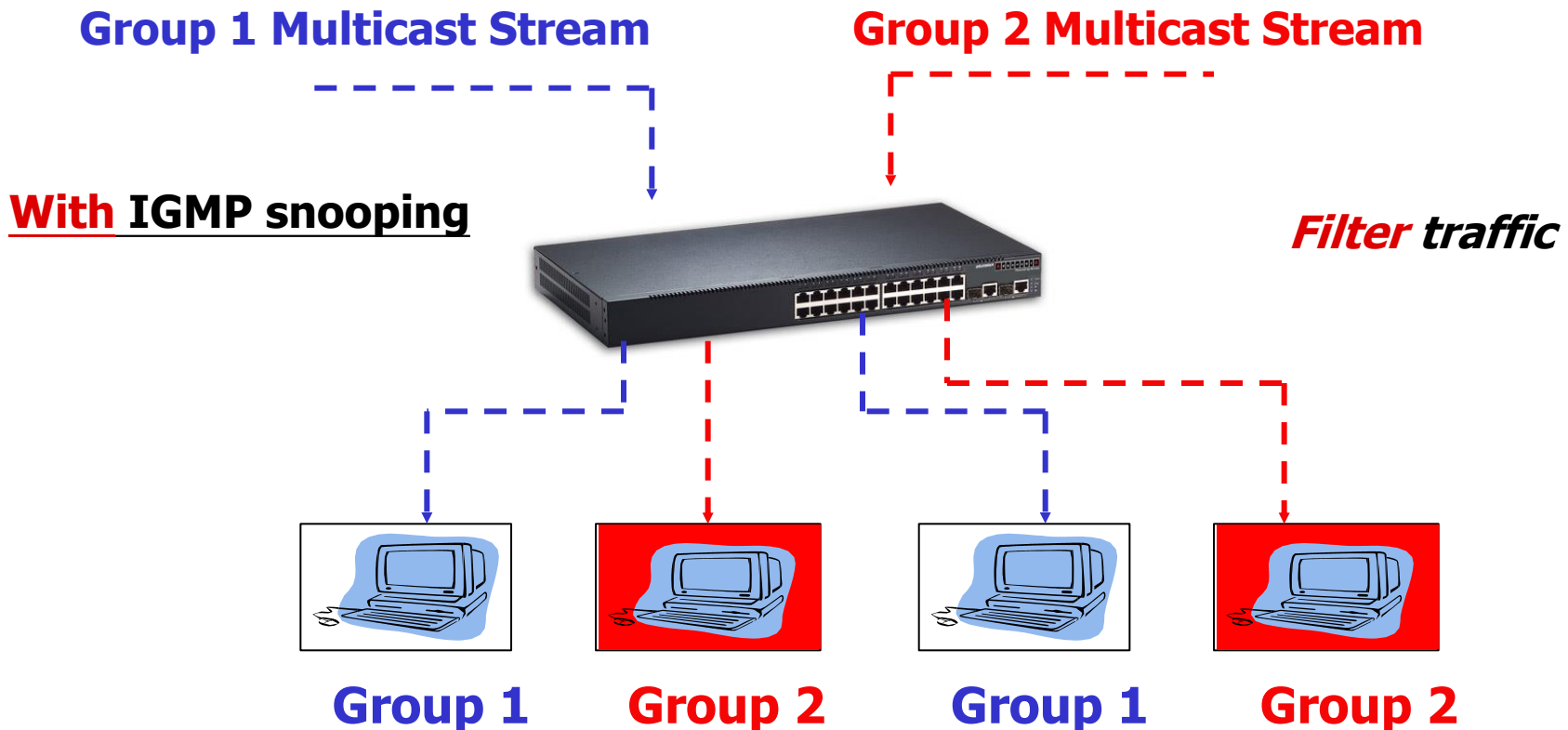
Az IGMP Snooping



BME-TMIT

IGMP Snooping - hatékony multicast Etherneten

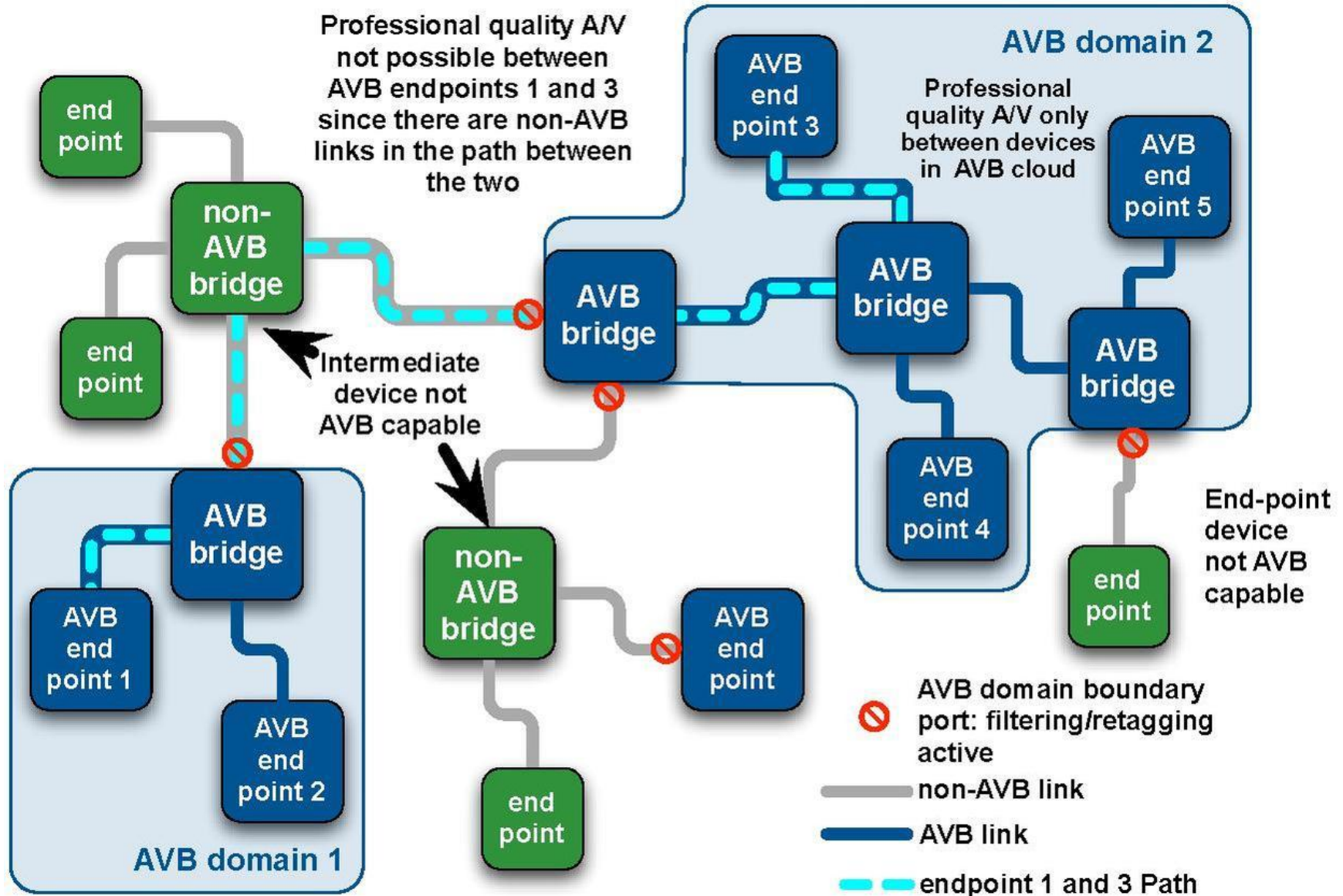
Hosts only receive dedicated traffic belonging to the same group



- A broadcast jelleg miatt nagy a veszély
- Támadható pontok/területek
 - SPT DOS
 - ARP
 - MAC címek kicserélése
- A C-VLAN használata leszűkíti támadó lehetőségeit

- IPoE specifikus
 - DHCP snooping használata
 - Dinamikus ARP figyelés
 - ARP proxy
- Védekezni kell
 - BPDU szűrés – STP támadás ellen
 - PVLAN használata a trónk portokon
 - Traffic policing – a broadcast stormok ellen
 - A MAC címek limitálása egy porton

AVB – Audio Video Bridging



- 802.1AS – idő szinkron
- 802.1Qav – továbbítás
- 802.1Qat – Erőforrás foglалás
- 1722 - beágyazás

AVBTP network stack.

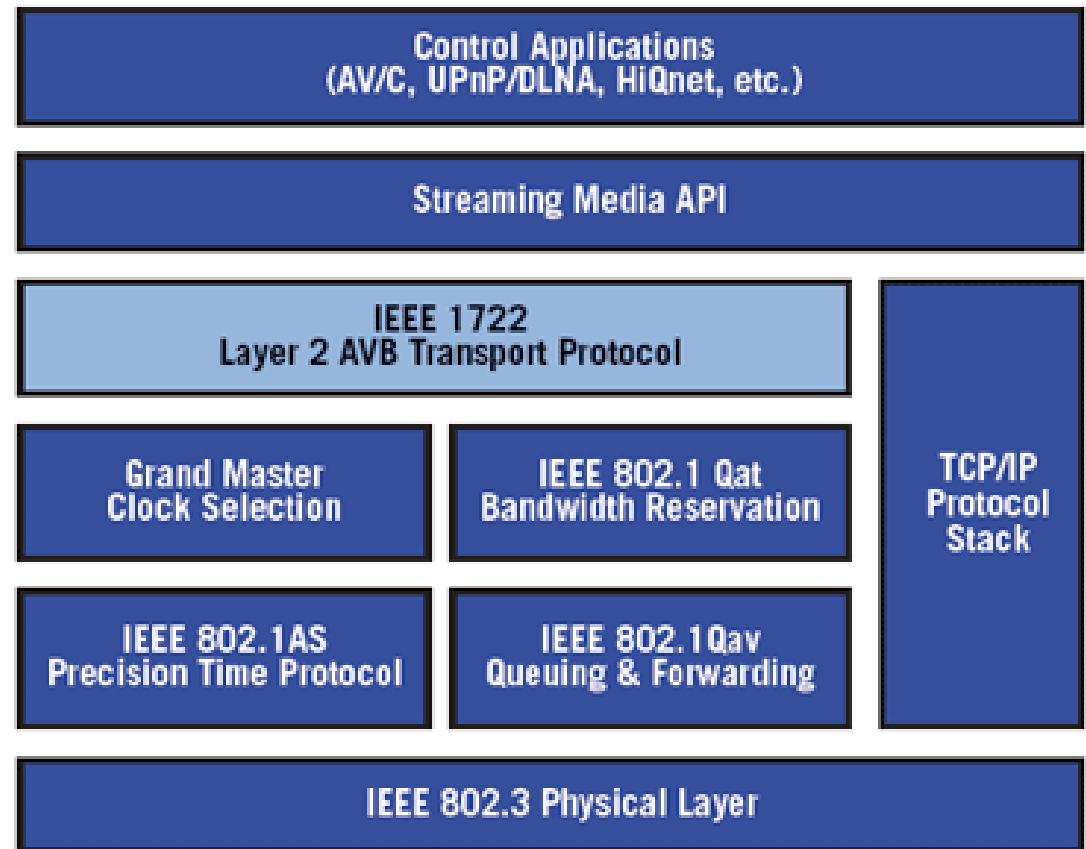


Figure 1

- Az AVB „utódja” – átnevezés
- Tágabb felhasználási terület
 - Ipar
 - Autók
 - Stb.
- A szabványosítás jelenleg is folyik
 - Új szabványok

- › P802.1Qbu – Frame Preemption – ready
 - › P802.1Qbv – Enhancements for Scheduled Traffic – ready
 - › P802.1Qcc – Stream Reservation Protocol (SRP) Enhancements and Performance Improvements
 - › P802.1Qci – Per-Stream Filtering and Policing
 - › P802.1Qch – Cyclic Queuing and Forwarding
 - › 802.1Qcj – Auto-attach to PBB services
 - › P802.1AS-Rev – Timing and Synchronization – Revision
 - › P802.1CB – Frame Replication and Elimination for Reliability
 - › P802.1CM – Time-Sensitive Networking for Fronthaul
- related
- related to each other

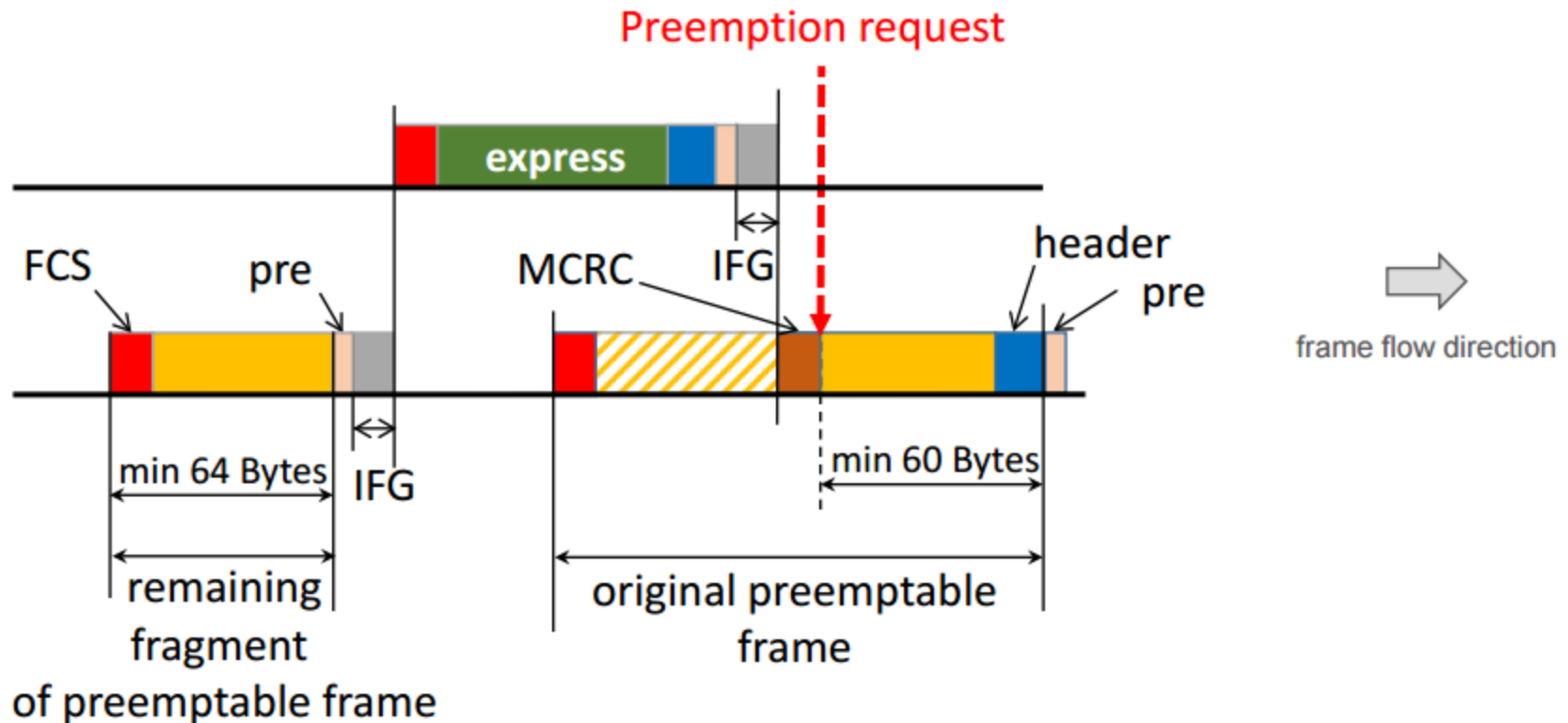
Frame Preemption



BME-TMIT

2. 802.1Qbu – Frame Preemption

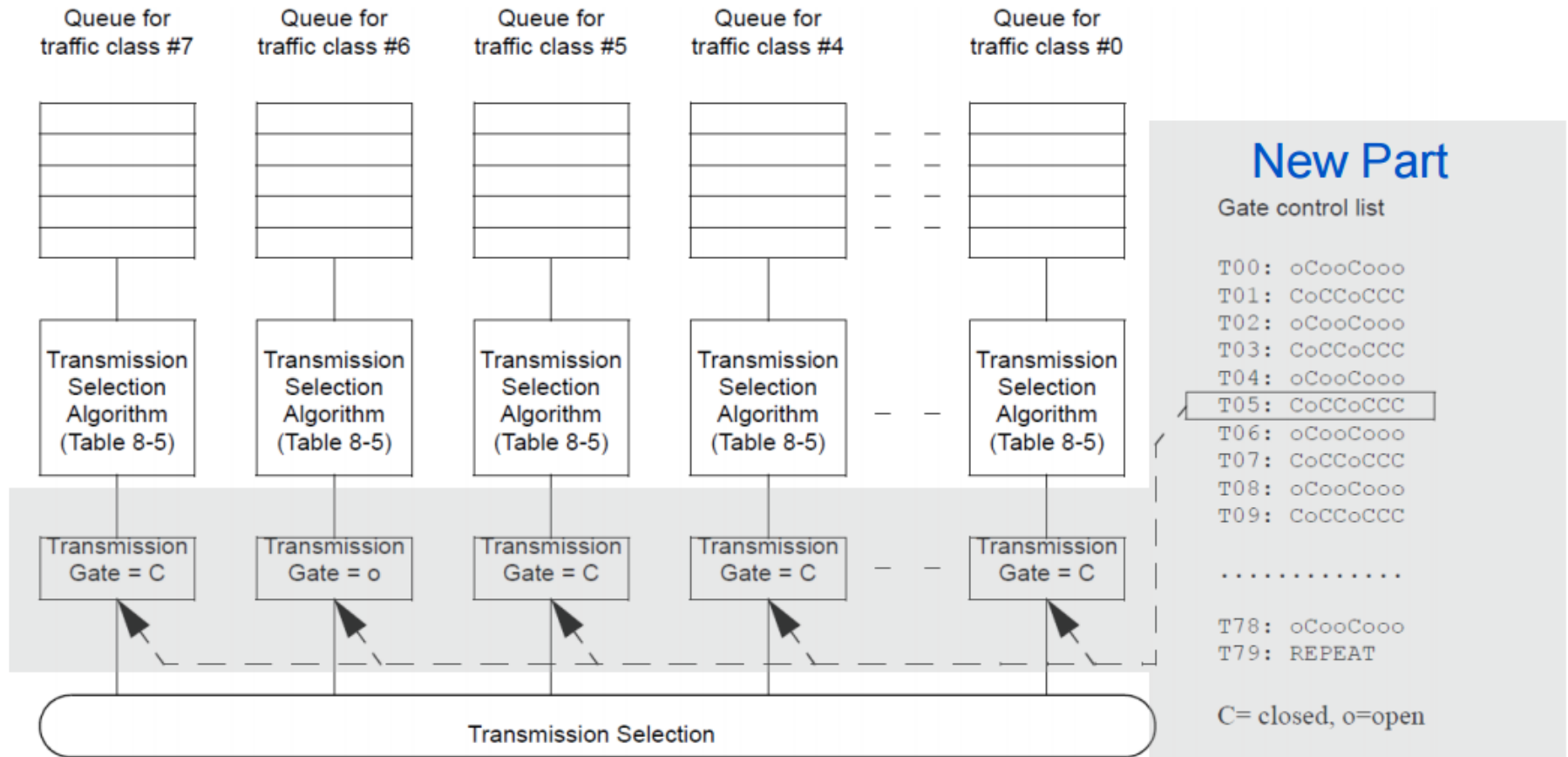
1. 802.3br – Interspersing Express Traffic (IET)



802.1Qbv – Time Gated Queuing



BME-TMIT

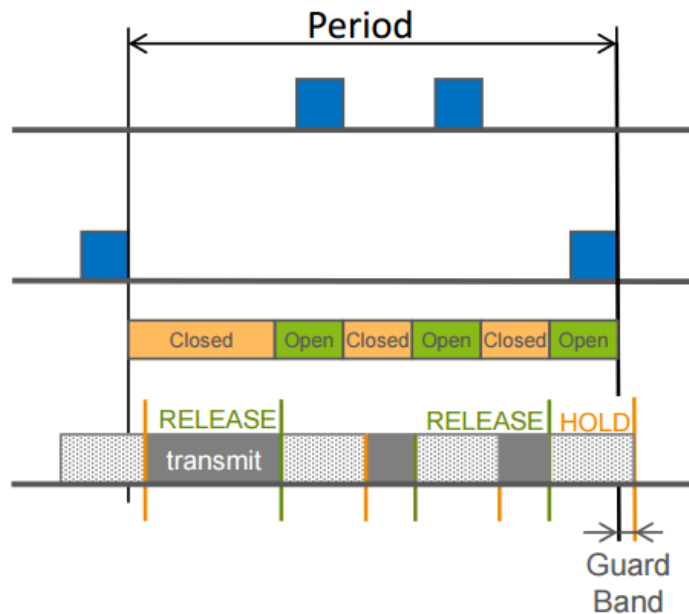


Time Gated Queuing – Működés



BME-TMIT

Example 1



Legend

Express From Port 1

Express From Port 2

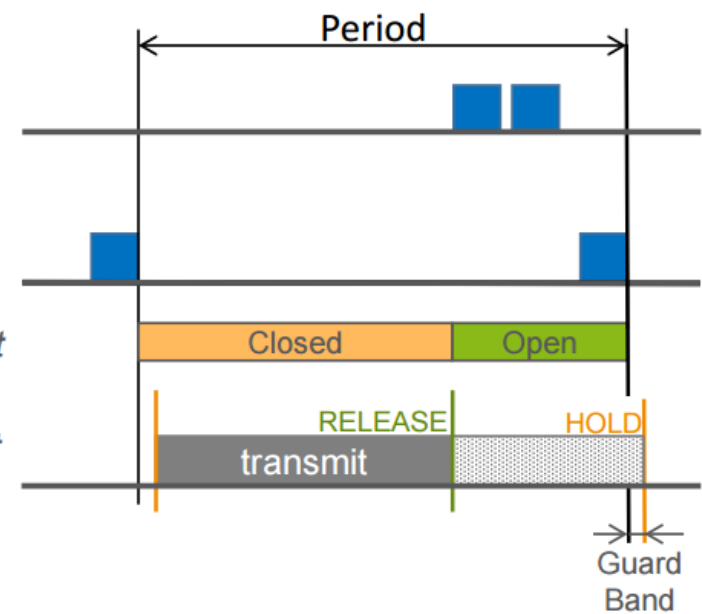
Express Gate at egress port

Preemptable at egress port



frame flow direction

Example 2



Köszönöm a figyelmet

- Vége -



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**



Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. október 5.



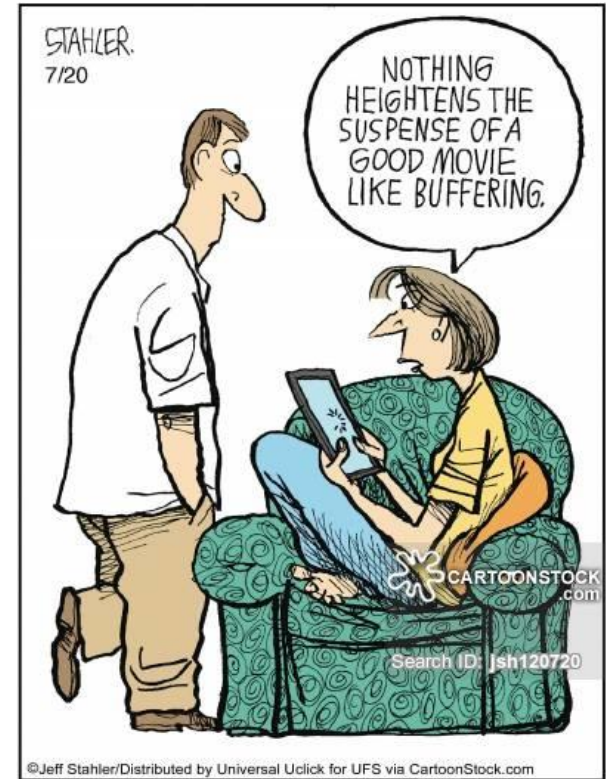
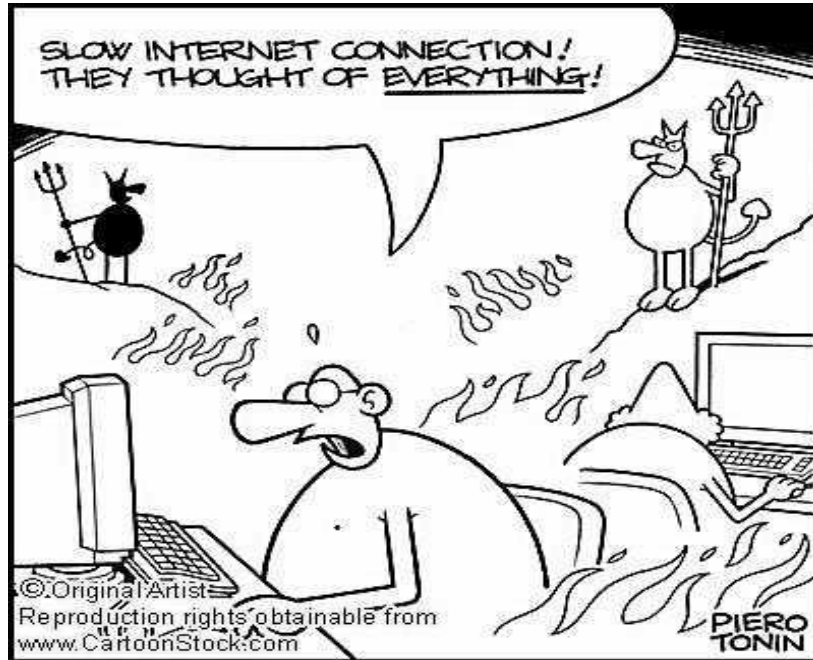
Miért kellene mégis optikai hálózatok?

- Ma már nem a webezés, hanem a multimédia a fontos
 - MPEG-1 – ISO/IEC szabvány
 - Moving Pictures Experts Group
 - 50:1 – 100:1 video tömörítés
 - 1.5 Mbps, VHS minőségű kép
 - MPEG-2
 - DVD minőségű kép
 - Nagy felbontás, nagy színmélység, sok mozgás (pl. sportközvetítés) – 4-8 Mbps
 - HDTV – 14 Mbps, **8K UHD TV – 50 Mbps** (7680 x 4320, 60 fps)
- Az xDSL sávszélessége messze nem elegendő ehhez
 - Csak nagyon rövid helyi hurkok esetén

Miért kellene mégis optikai hálózatok?

- HFC (Hybrid Fiber Coax)
 - Az eredeti 300-550 MHz-es kábeleket 850 MHz-s koax kábelek váltják föl
 - Plusz 300 MHz → 50 db új 6 MHz-es csatorna
 - QAM-256-al 40 Mbps egy csatornán → 2 Gbps új sávszél
 - 500 ház egy kábelén → mindenkinek jut 4 Mbps downstream, ami elég egy MPEG-2 filmhez
 - Szépen hangzik, de...
 - Minden kábelt le kell cserélni 850 MHz-es koaxra
 - Új fejállomások, új fényvezető csomópontok (fiber node), kétirányú erősítők
 - Szinte a teljes kábelhálózati rendszert le kell cserélni
- **Akkor miért ne legyen minél több fényvezető szál benne?**

A kis sebesség ma már kínzás!!



Adatátvitel fényvezető szálon

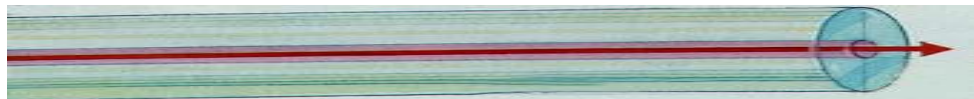
- Három fő komponens:
 - **Fényforrás**
 - LED (light emitting diode), félvezető lézer
 - **Átviteli közeg**
 - Rendkívül vékony üvegszál
 - **Fényérzékelő (detektor)**
 - fény hatására elektromos impulzusokat állít elő
- Az adatátviteli sebességet az átalakítás sebessége határozza meg
 - A gyakorlati sebesség egy szálon ma 10-50 Gbps

Fényvezető szálak



- Többmódusú szál

- A fényimpulzusok hosszanti irányban szétszóródnak a szálban
- Egyszerre több, különböző szögben visszaverődő fénysugár halad
- Olcsó megoldás, de csak kis távolságokra hatékony (500 m)

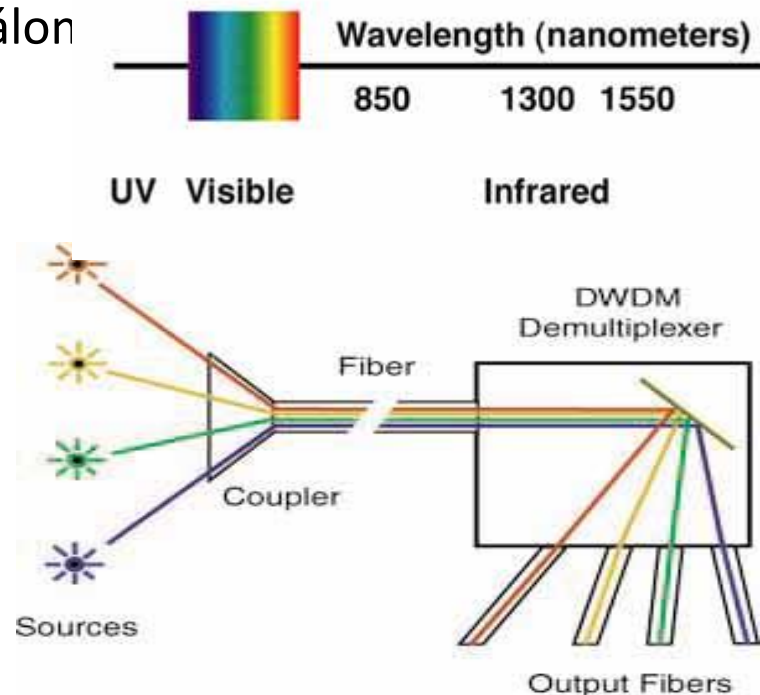


- Egymódusú szál

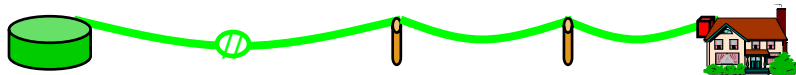
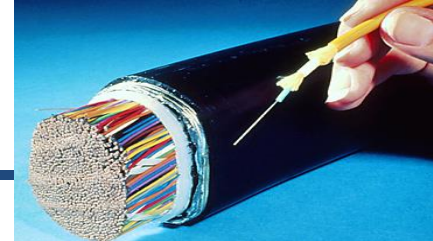
- Ha az üvegszál átmérője nagyon kicsi, a fény visszaverődés nélkül, egyenesen terjed
- Jóval drágább a szál, és nagyobb kapacitású, jobb lézereket igényel
- Nagyobb távolságok áthidalására sokkal jobb
 - 50 Gbps 100 km távolságba erősítés nélkül
 - A transzatlanti optikai kábeleknél nagyon fontos, hogy kevés erősítő legyen
- A gerinchálózatban csak egymódusú szálakat használnak

Hullámhossz osztás

- WDM – Wavelength Division Multiplexing
 - Több hullámhossz (szín) ugyanazon az üvegszálon
 - Kezdetben csak 2 szín
 - Ma már akár 160
 - 10 Gbit/s szálon elméletileg 1.6 Tbit/s



Fiber vs. Réz érpár



- Optikai kábel

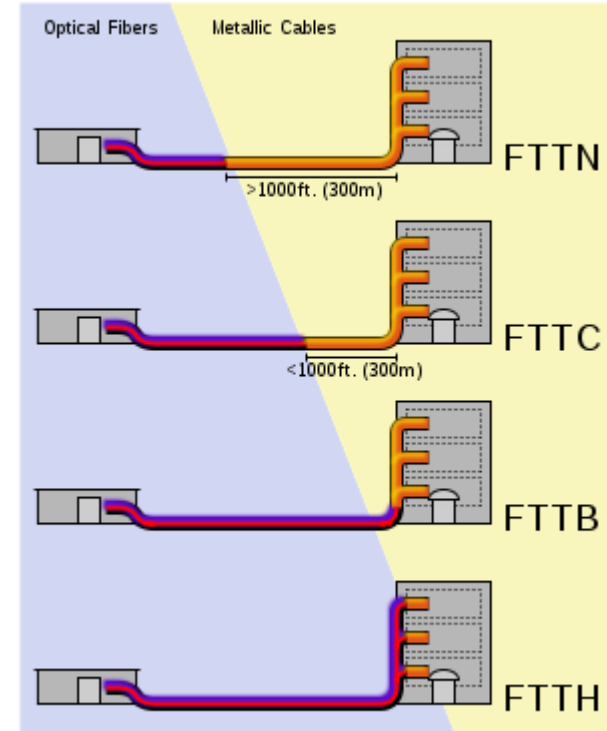
- Fényjelekkel működik
- Nem érzékeny az elektromágneses interferenciákra
- Ismétlők kb. 30 km után
- Kismértékű hőágulás
- Törékeny, viszonylag merev anyag
- Kémiaailag stabil

- Réz érpár

- Elektromos hullámok
- Érzékeny az elektromágneses interferenciákra
- Ismétlők 5 km után
- Nagymértékű hőágulás
- Hajlítható anyag
- Érzékeny a korrózióra és galvanikus reakciókra
- Újrahasznosítható
 - Jó pénzért el lehet adni a rezet

FTTx

- FTTx – Fiber To The x
 - FTTN – Fiber To The Neighborhood
 - FTTC – Fiber To The Curb
 - FTTB – Fiber To The Building
 - **FTTH – Fiber To The Home**
 - ...
 - FTTO – Fiber To The Office
 - FTTD – Fiber To The Desk
 - FTTE – Fiber To The Enclosure
 - FTTP – Fiber To The Premises
 - FTTU – Fiber To The User

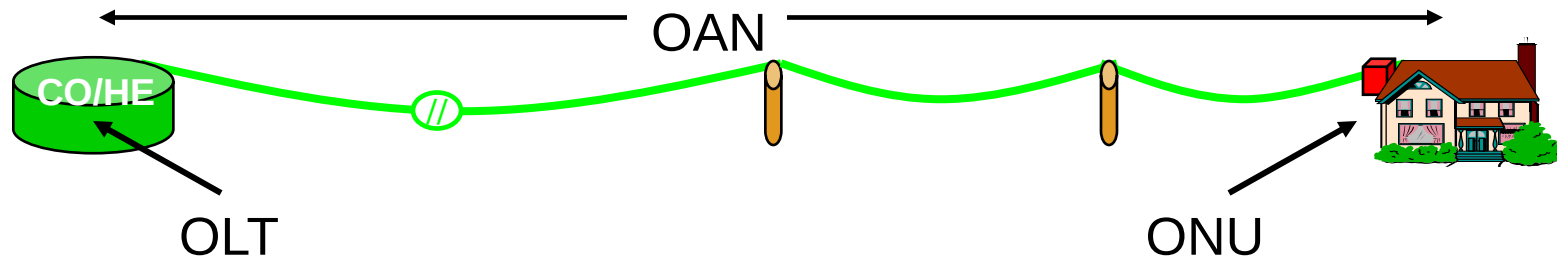


FTTC/FTTB

- **Fiber To The Curb / Building**
 - Üvegszál az elosztódobozig / épületig
- Üvegszál a helyi központból minden lakókörzetig
 - A szál egy ONU-ban végződik
 - Optical Network Unit – optikai hálózategység
 - Több helyi rézhurok, coax, Ethernet kábel csatlakozhat hozzá
 - Nagyon rövid hurkok, lehetséges szimmetrikus nagysebességű kiterjesztés
 - Pl. VDSL – Dél-kelet Ázsiában nagyon elterjedt
 - Alkalmas MPEG-2 átvitelre, videokonferenciázásra
 - Az FTTC/FTTB maga szimmetrikus átviteli sebességeket biztosít

FTTH – Fiber To The Home

- Rendszerelemek
 - OAN: Optical Access Network
 - Optikai hozzáférési hálózat
 - ONU/ONT: Optical Network Unit/Terminal
 - Az előfizető otthonában
 - OLT: Optical Line Termination
 - végződtetés a szolgáltató hálózatában

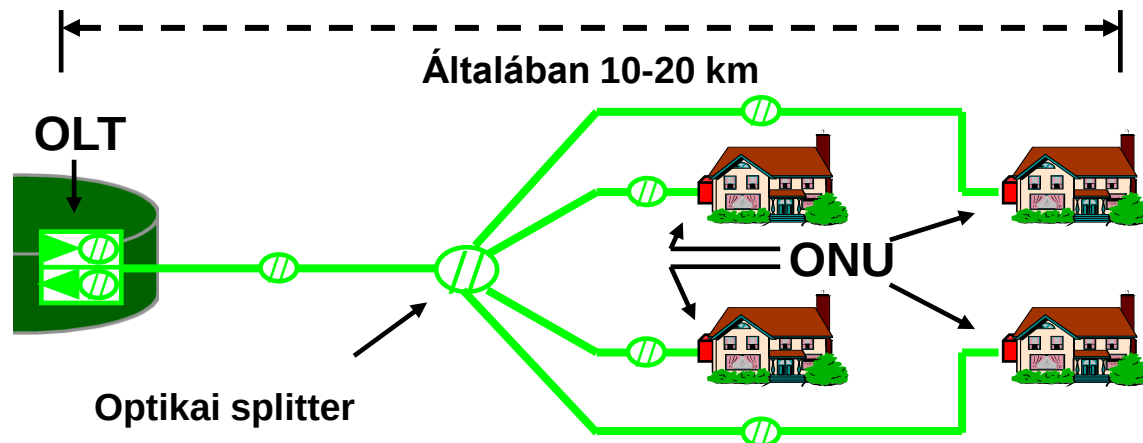


FTTH architektúrák



- **PON – Passive Optical Networks**

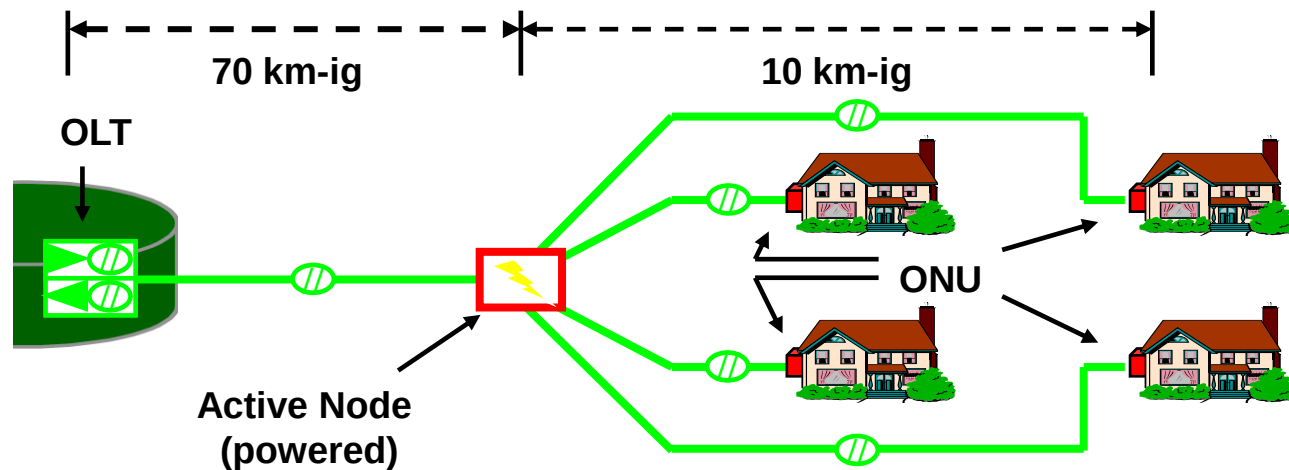
- Több felhasználó megoszt egy fényvezető szálát
- Optikai splitter-ek a jel szétválasztására és aggregálására
- Áramellátás csak a végeknél szükséges
- Osztott hálózat – Point to Multipoint (P2MP)



FTTH architektúrák

- **Active Node**

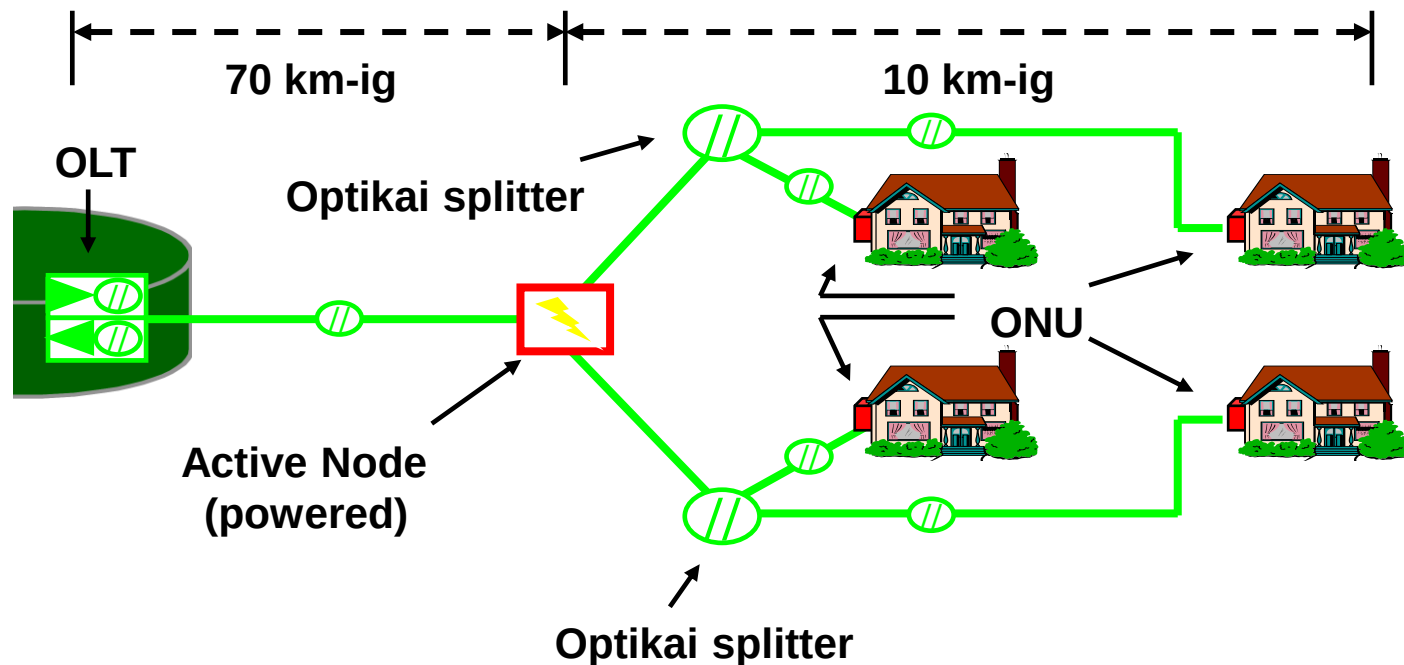
- Az előfizetőknek saját fényvezető száluk - Point to Point (P2P)
- Aktív, árammal táplált csomópontok a forgalom elosztására - Ethernet switch
- Layer2/Layer3 switching/routing



FTTH architektúrák

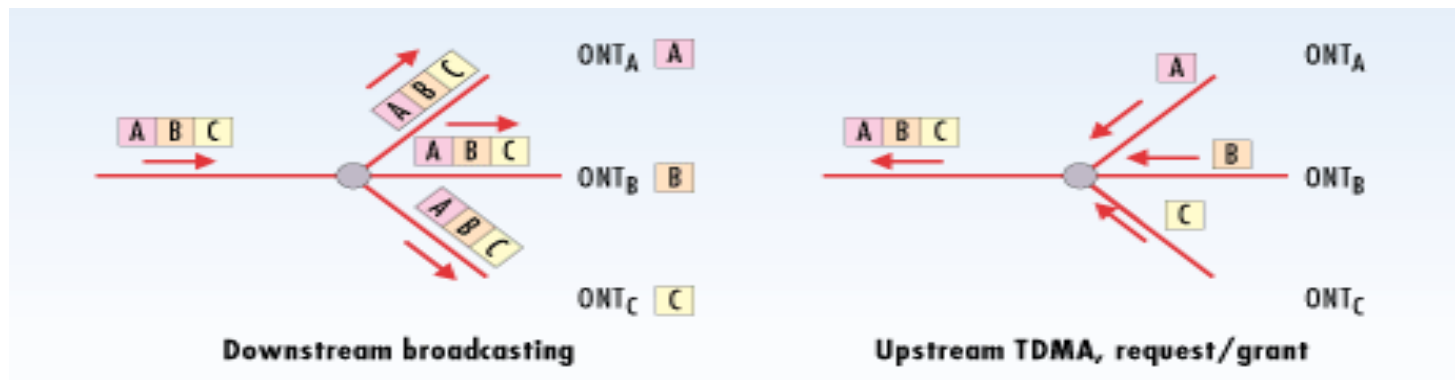
- Hybrid PON

- Az előbbi két architektúra kombinált változata



PON le- és feltöltés

- A le- és feltöltés nem egyformán működik
 - A letöltés broadcast
 - A splitter minden szálra kitesz minden csomagot
 - Az ONU csak azt a csomagot kezeli melyet neki címeztek (fejléc alapján)
 - A feltöltés TDMA-t használva történik
 - Az OLT időszeleteket oszt ki az ONU-knak
 - Szinkronizált csomagküldés
 - Időszeletek kiosztása igénylések függvényében



Ethernet vagy ATM alapú PON?

- Két külön technológia vetélkedik egymással
 - APON – ATM-based PON
 - Az első PON implementáció
 - EPON – Ethernet-based PON

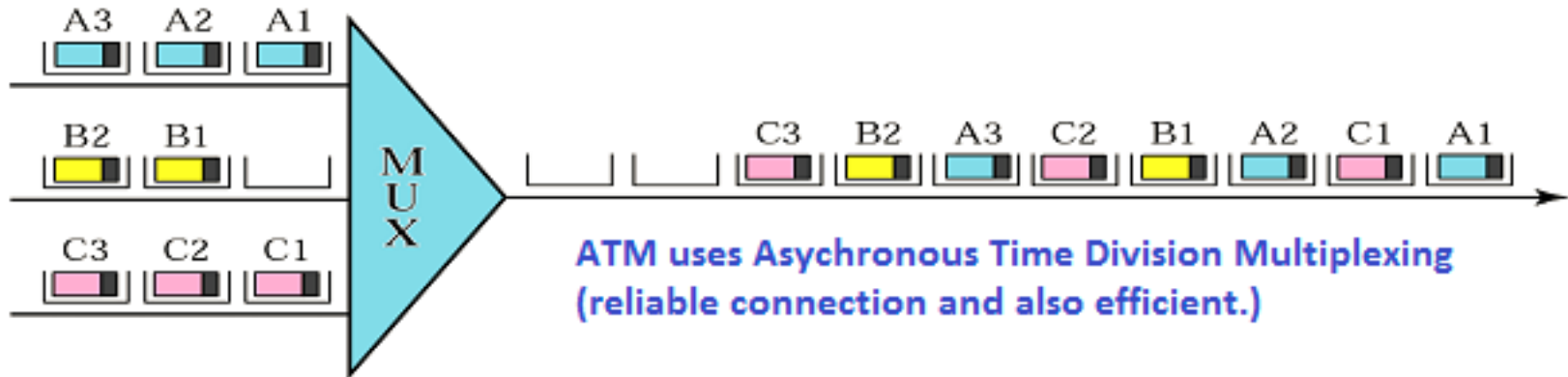
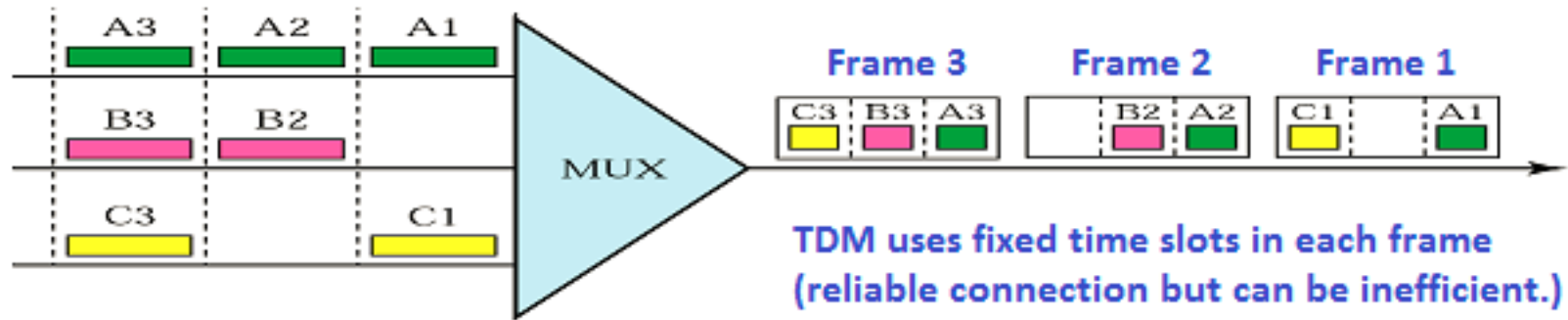
ATM (Asynchronous Transfer Mode)

- A különböző típusú forgalmak (audio, video, data) párhuzamos átvitelére találták ki
 - Az 1500 byte-os Ethernet csomagok túl nagyok
 - 1.500 byte = 12.000 bit
 - 10 Mbps-os Etherneten 0.1 μ s bit time $\rightarrow$ 1.2 ms / keret
 - Ha több forrás (gép vagy alkalmazás) áll sorban, túl nagy várakozási idők
- Az audio és video alkalmazásoknak szoros **késleltetés (delay)** és **késleltetés-ingadozás (jitter)** követelményei vannak

ATM (Asynchronous Transfer Mode)

- ATM megoldás
 - Fix méretű **ATM cellák**: 5 byte fejléc + 48 byte adat = **53 byte**
 - **Segmentation and Reassembly (SAR)**
 - Változó méretű keretek feldarabolása, majd visszaállítása a vevőnél, a fejléc alapján
 - **Asynchronous Time Division Multiplexing**

ATM (Asynchronous Transfer Mode)



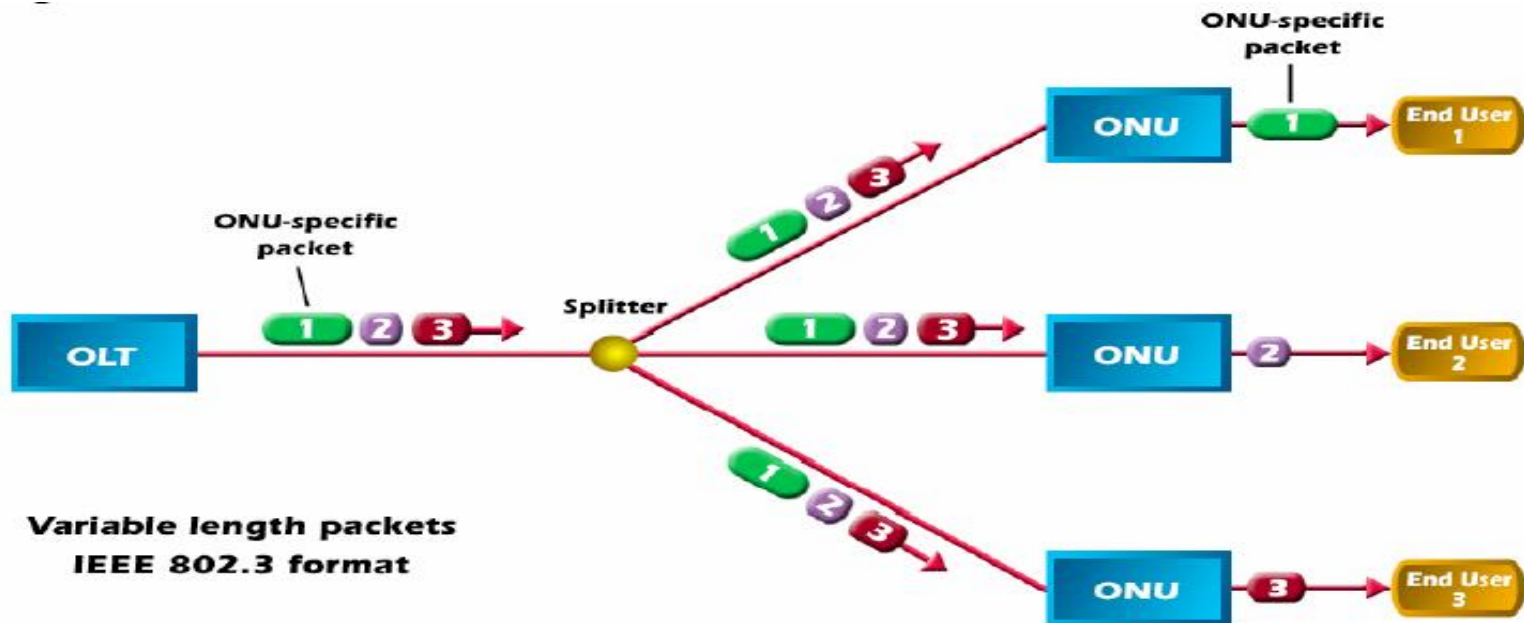
APON

- **Segmentation and Reassembly (SAR)**
 - Fix hosszúságú csomagok
 - 53 byte-os ATM cellák
 - Az adatok átmennek egy ATM Adaptation Layer-en (AAL) ahol 48 byte-os darabokra osztják őket
 - Plusz 5 byte a fejléc
 - A címzettnél az eredeti forgalmat újból összerakják
- A SAR miatt az ATM kifejezetten alkalmas video, hang és adatátvitelre
 - A kis, fix hosszúságú cellákban jól lehet késleltetésre érzékeny forgalmat szállítani
 - A procedura időigényes, az 5 byte-os fejléc pedig nem hatékony (10%-os overhead)
- A fix hosszúságú cellák jól illeszkednek a PON TDMA alapú feltöltéséhez
 - Könnyű az időszeletek kezelése

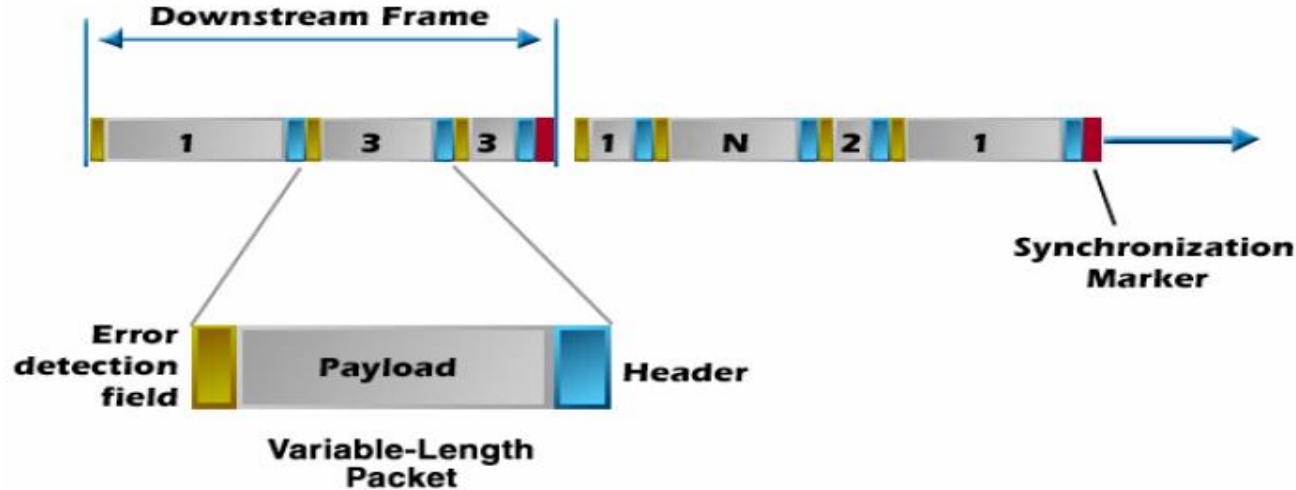
EPON

- Az adatok az IEEE 802.3 (Ethernet) formátumot használják
 - Változó hosszúságú csomagok 64 és 1518 byte között
- Hogyan oldjuk meg a TDMA alapú feltöltést?
 - Feloszthatnánk az Ethernet kereteket (frame) fix hosszúságú részekre
 - Egyszerűbb lesz a feltöltés
 - Az ár egy SAR funkció hozzáadása az EPON protokoll stack-hez
 - **Alkalmazott megoldás:** fix hosszúságú időszeletek, melyekbe több csomagot be tud rakni az ONU
 - Javít a hatékonyságon
 - Nehéz változó hosszúságú csomagokkal jól feltölteni egy fix hosszú időszeletet

EPON downstream forgalom

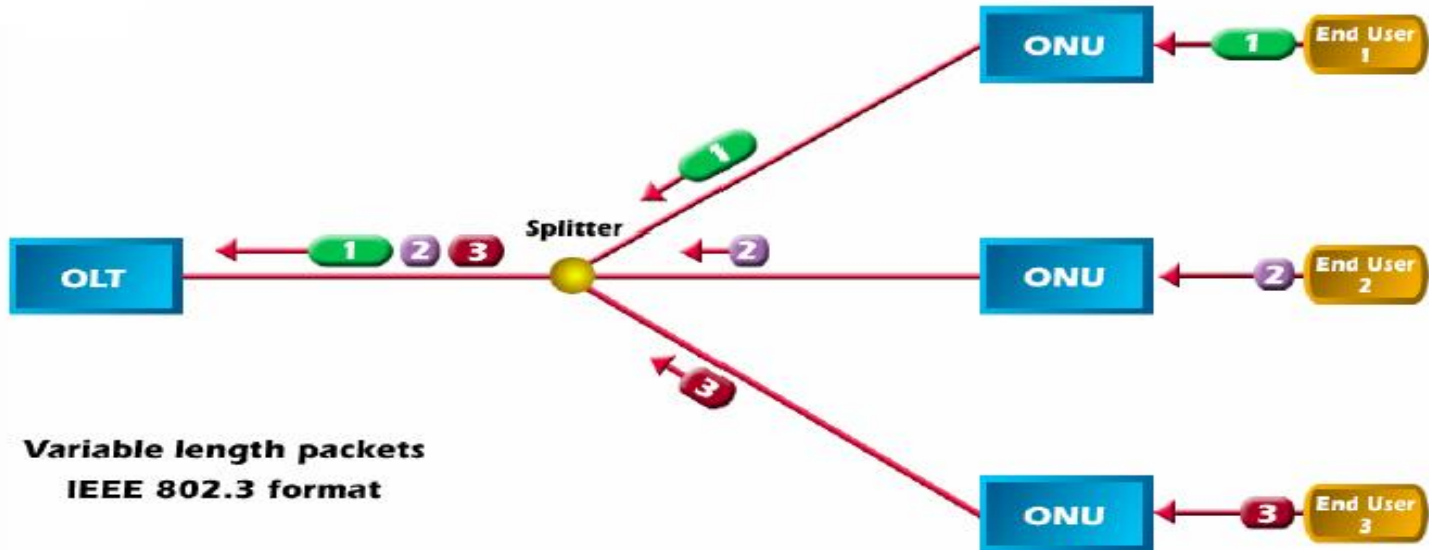


EPON downstream csomagok

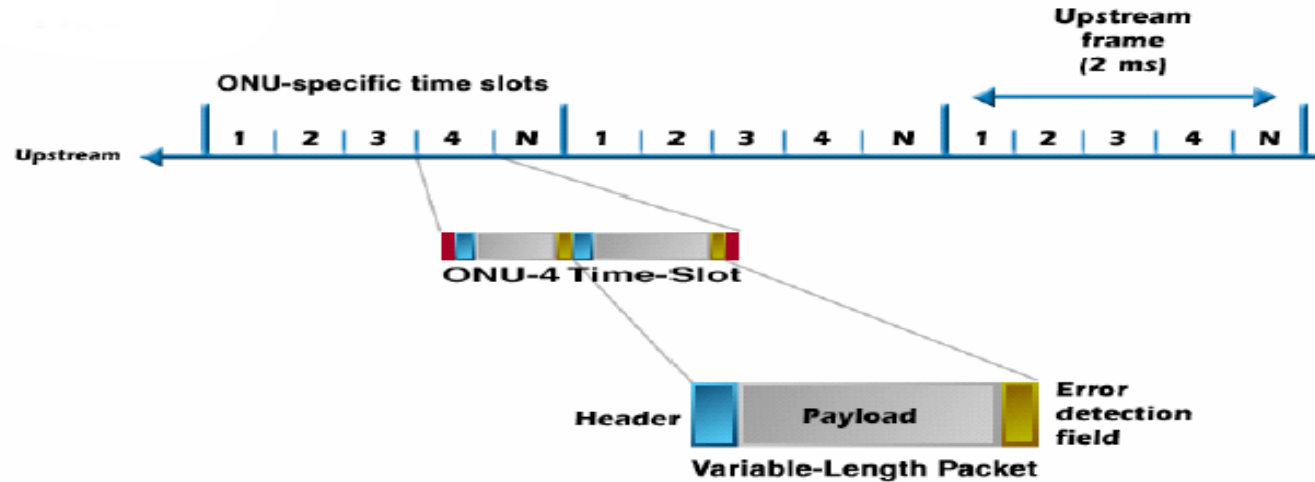


- Fix időközönként (2ms) küldött frame-ek, változó hosszúságú csomagokkal
- Szinkronizációhoz szükséges információ minden frame előtt
- Minden csomag fejléce megmondja ki a címzett
- Hibaellenőrző információ a csomag végén

EPON upstream forgalom



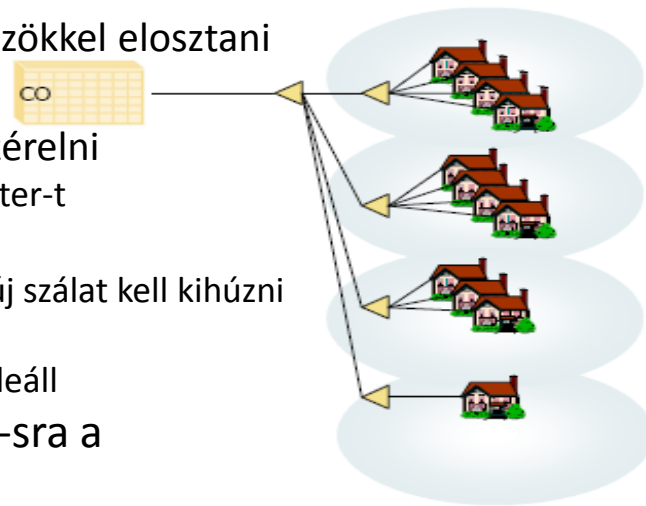
EPON upstream csomagok



- Az upstream forgalom frame-ekre (2ms) osztva
- Minden ONU-nak van egy saját időszelete, melyet változó hosszúságú csomagokkal tölthet fel

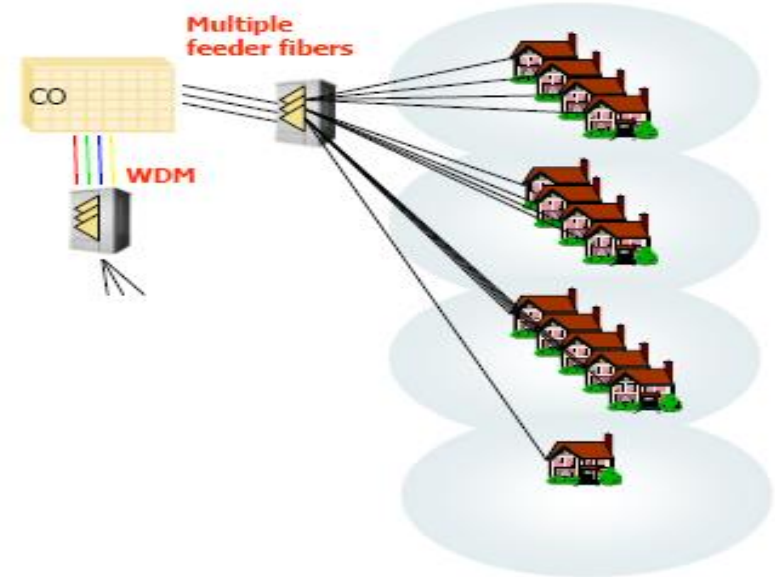
Hagyományos PON

- Az alapötlet:
 - Mindenki nem éri meg külön szálát kihúzni az OLT-től
 - Elég egy szálát közel vinni a felhasználókhoz, majd passzív eszközökkel elosztani
- Hátrányok
 - A splitter-ekben nincs intelligencia, nem tudod őket távolról vezérelni
 - Ha valami hiba van, nem könnyű egyenként megnézni minden splitter-t
 - Nem flexibilis
 - Ha egy 4-es splitter-en keresztül csatlakozol, egy 5-ik előfizetőnek új szálát kell kihúzni
 - Újratervezni a hálózatot, betenni egy nagyobb splitter-t
 - Egy splitter cseréjénél minden downstream előfizető szolgáltatása leáll
- Megoldás: ha 1x32-es splittert használsz, ne tervezd 32 ONU-sra a hálózatot, csak 16-osra vagy 24-esre
 - Van hely bővítésre
 - A maradék 16-nak többre fog kerülni a szolgáltatás



Passive Star PON

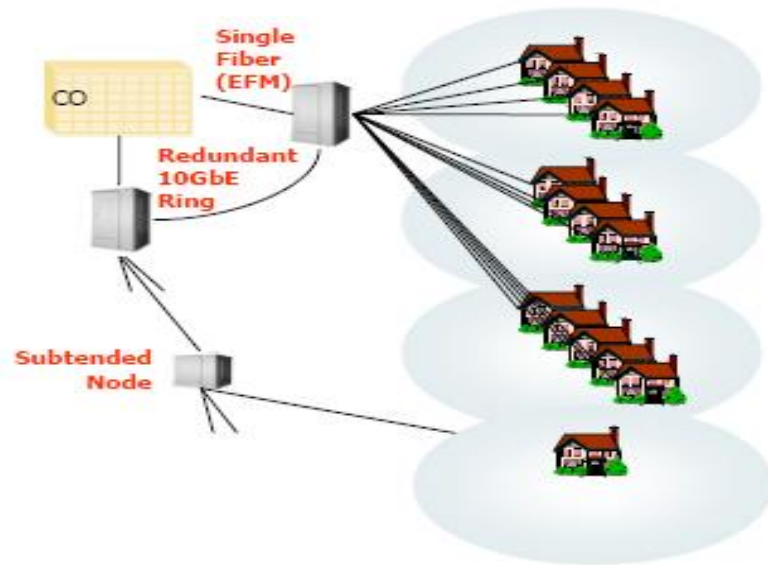
- A splitter-ek egy dobozban csoportosítva
 - Egyszerűbb a hibaelhárítás
- Továbbra is fa struktúra
 - Ha a splitter és a CO közötti szál meghibásodik, nincs backup
 - A splitterek passzívak, nem tudnak átváltani egy új útvonalra hiba esetén



Active Star

- Hátrány az aktív (árammal ellátott) node szükségessége
- Sok szempontból előnyös intelligens eszközöket használni a hálózat szélén
 - Az aktív node **IGMP*** proxy-ként működhet
 - Multicast forgalom támogatása
 - Hatékony erőforráskihasználás
 - Hibatűrő megoldás
 - Az aktív node-ok gyűrűbe kötve
 - **Ethernet Protection Switching Rings (EPSR)**
 - 50 ms alatti váltás hiba esetén
 - Video esetén pillanatnyi kockás kép
 - Egy telefon kapcsolat nem szakad meg
 - Könnyen menedzselhető, könnyű hibaelhárítás

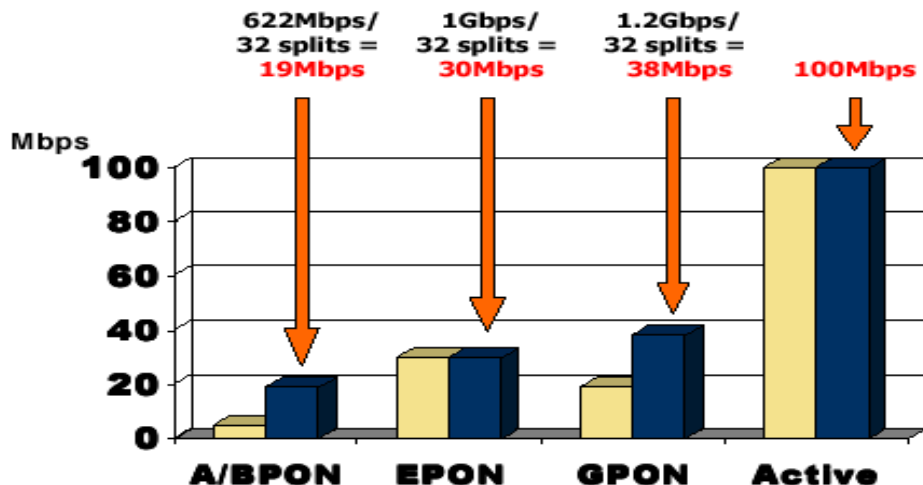
***IGMP- Internet Group Management Protocol**



Újabb verziók

- **Broadband PON**
 - 622 Mbps downstream, 622 Mbps upstream
- **Gigabit PON**
 - Több downstream/upstream változat
 - Legelterjedtebb az 2.48 Gbps downstream és 1.244 Gbps upstream
- **XGPON (10G-PON) – 2010**

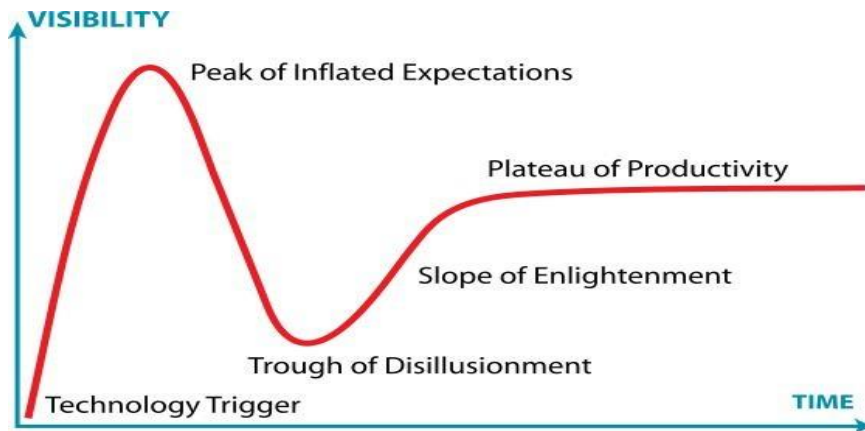
Adatátviteli sebességek összehasonlítása



- PON megoldásoknál kisebb sebességek
 - Osztott rész az OLT és az első splitter között
 - Valamivel jobb a helyzet ha nem telített a splitter
 - Nem 32-be, hanem csak 16-ba vagy 24-be kell osztani
- Active Node-nál mindenkinek saját fényvezető szála
 - Magánfelhasználóknak általában 100 Mbps mindkét irányban
 - Üzleti előfizetőknek akár több Gbps

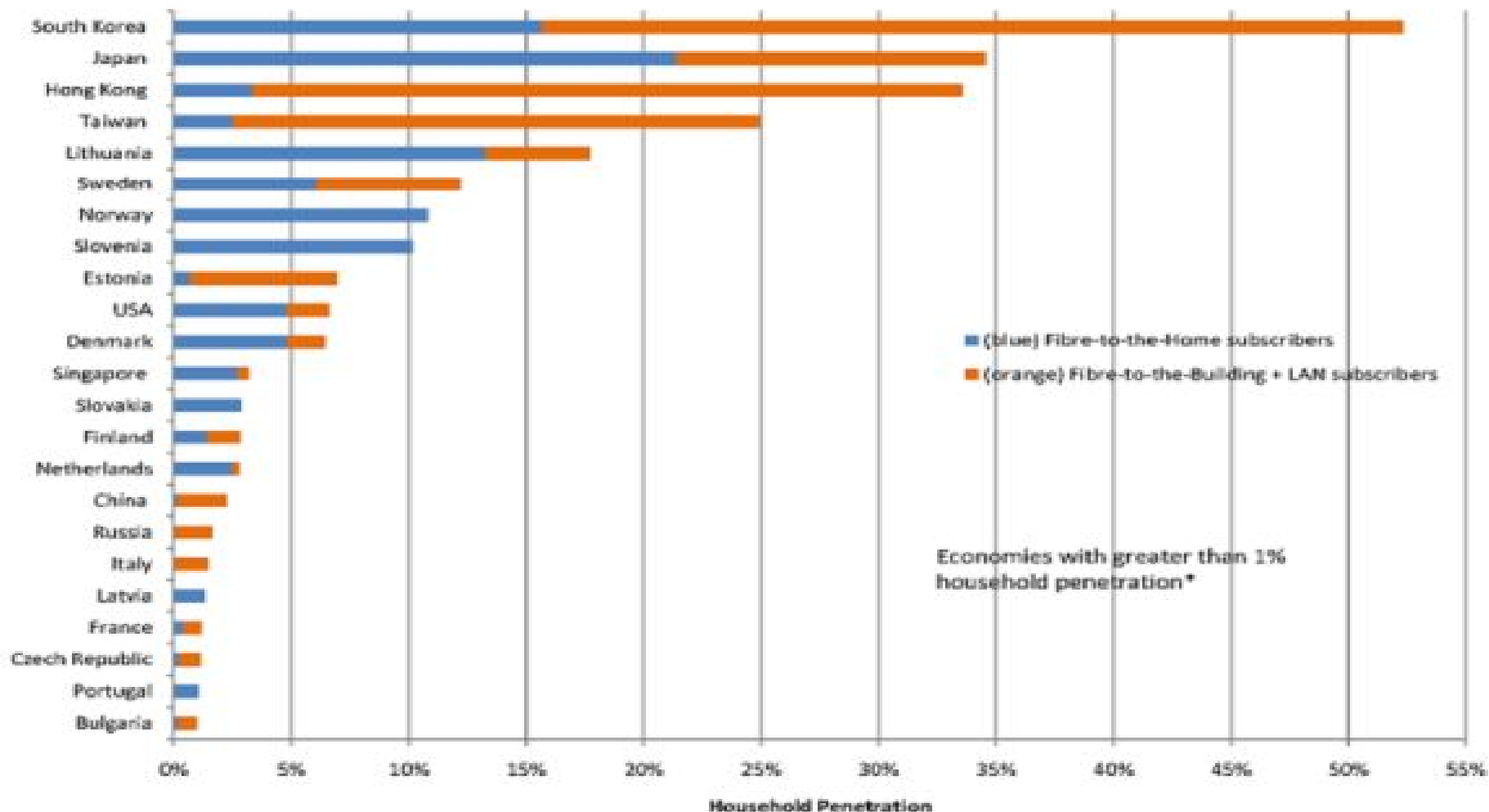
FTTH/FTTB elterjedése

- Az FTTH/FTTB még gyerekcipőben
 - Kezdeti, növekedési fázis
 - A közeljövőben továbbra is Ázsia vezeti majd a versenyt
- A nagy Telco cégek közül kevesen vezették be

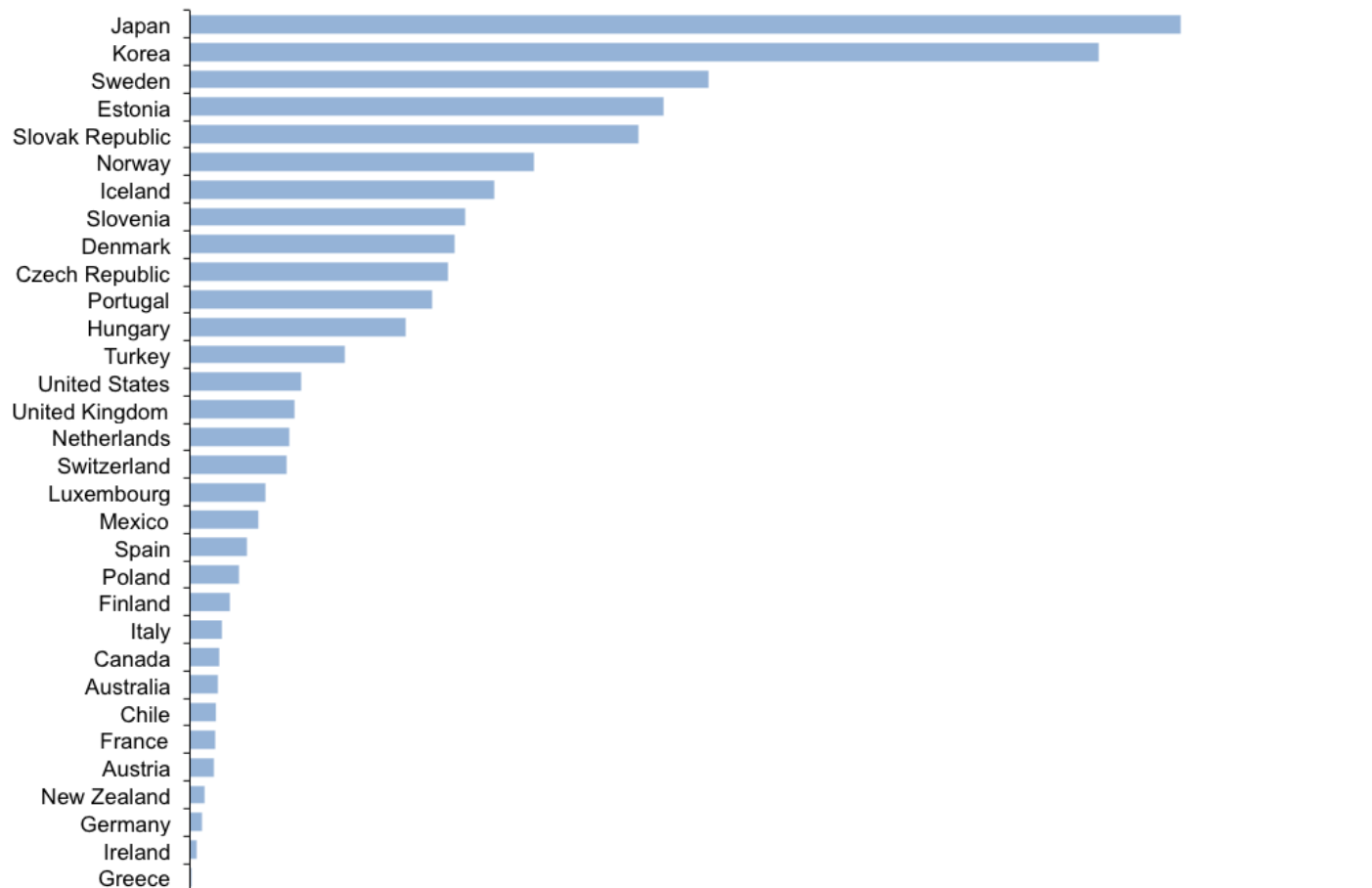


Feb. 2010

Economies with the Highest Penetration of Fibre-to-the-Home / Building + LAN



Percentage of fibre connections in total broadband subscriptions, June 2013

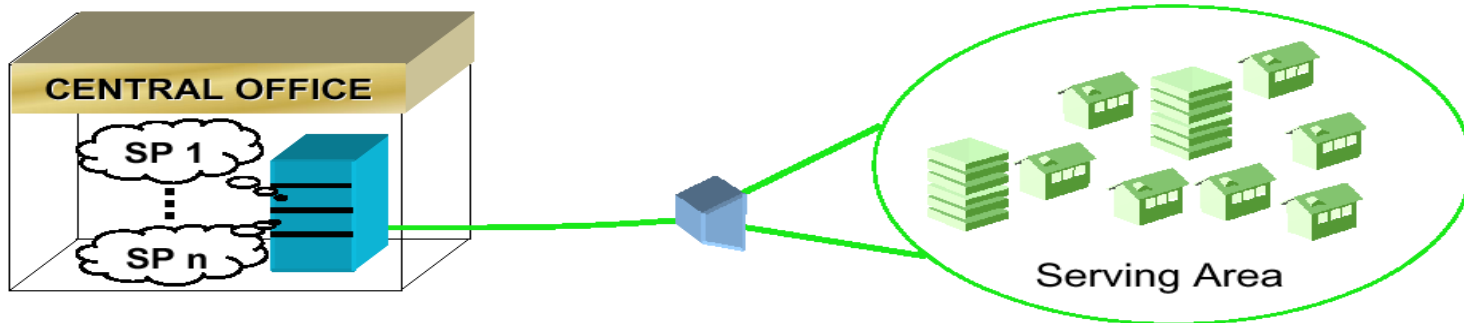


FTTx szolgáltatás

- Két szolgáltatási modell
 - Saját hálózat
 - Az FTTx szolgáltatások nagy része
 - A hálózat tulajdonosa egyenesen a felhasználóknak adja el a szolgáltatást
 - Hagyományos telefon és kábeltévé szolgáltatási modell
 - Nyílt hozzáférés
 - Több országban törvényi szabályozás miatt
 - A hálózat tulajdonosa átadja az infrastruktúrát több viszonteladó szolgáltatónak, ők szerződnek a felhasználókkal

Nyílt hozzáférés (Open Access)

- A tulajdonos egyenlő feltételek mellett adja át a hálózatát különböző szolgáltatóknak (Telco, ISP, video szolgáltató, stb)
 - Saját maga nem lép be a versenybe
- Általában önkormányzati, városi hálózatok
 - A hálózati infrastruktúra közszolgáltatásnak számít
 - Úgy mint a víz, az áram vagy az úthálózat



Open Access példák

- Sok önkormányzati Open Access hálózat Nyugat Európában és főként Skandináviában
 - Stokab (Stockholm) – az első önkormányzati FTTx hálózat (1996)
 - Vasterbotten – vidéki régió, fele akkora mint Hollandia, 260.000 lakos
 - 15 önkormányzat összekötve egy FTTx hálózaton
 - Svédországban 289 önkormányzat, több mint 200-nak saját hálózata
 - CityNet, Amsterdam – 450.000 házat bekötő hálózat
 - Több önkormányzati hálózat Dániában
- Franciaországban és Angliában új törvényjavaslatok a nyílt hozzáférésű hálózatok támogatására vagy kötelezővé tételére
- Néhány önkormányzati hálózat az USA-ban

FTTH Európában

- Sok országban jogilag szabályozva
 - Nemzeti szélessávú stratégiák
- Miért nem építenek saját optikai hálózatot az „incumbens” szolgáltatók?
 - Így is uralják a piacot, nincsenek rákényszerítve
 - A rövid előfizetői hurkok miatt viszonylag magas xDSL sebességek
 - Skandináviában olcsóbb az önkormányzatok hálózatait bérelni, mint sajátot építeni
 - A videoátvitel még nem annyira követelmény mint Ázsiában
- A helyi önkormányzatoknak az FTTH egy fontos eleme a regionális fejlesztésnek
 - Vonzóvá teszi a régiót, megéri befektetni

FTTx saját hálózaton

- Versenyhelyzetes piacok
 - Minden szolgáltatónak saját hálózata, mellyel lefedik ugyanazt a területet
 - Leginkább jellemző az USA-ban és Japanban
 - 9 japán szolgáltatónak van saját hálózata
 - Európában is van rá példa (Hollandia)
 - Nagyobb sebességek, kisebb üzemeltetési költségek (OpEx)
 - Nagyobb tőkeberuházás (CapEx)



FTTH verseny Tokió belvárosában



Ethernet

L2VPN Szolgáltatások

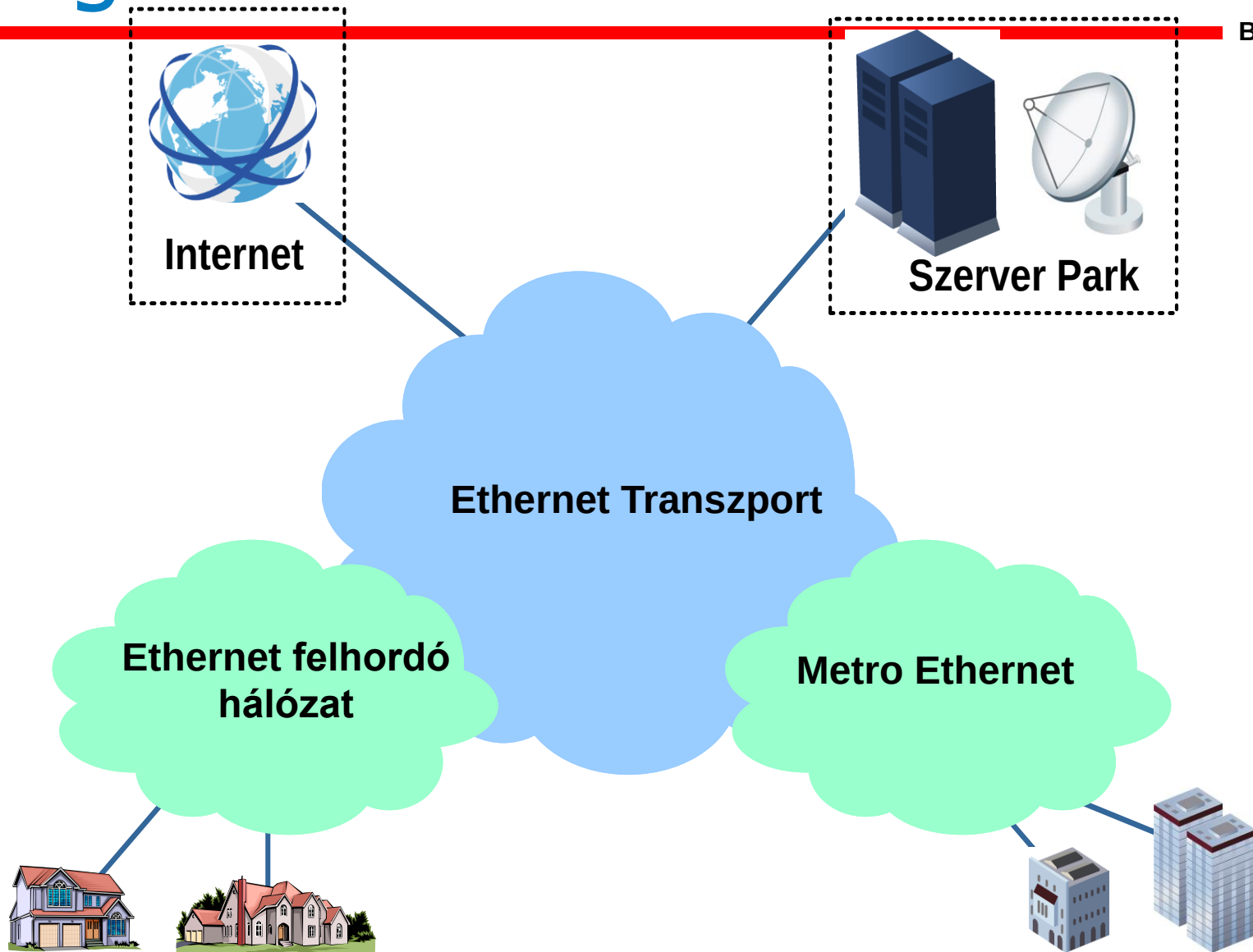
Moldován István



Szolgáltatói Ethernet



BME-TMIT



- L3 VPN – IPVPN, VPRN
 - IP szintű kapcsolatot biztosít a telephelyek között
 - A csomagok routing segítségével jutnak célba
- L2 VPN – VLL vagy PW
 - Pont-pont kapcsolat
- L2 VPN – TLS, VPLS
 - Ethernet szintű kapcsolat a telephelyek közt
 - Bridging
 - A két telephely egyetlen LAN-t lát

Szabványosítás



BME-TMIT



Provider Bridges (Q-in-Q)
Provider Backbone Bridges (Mac-in-Mac)
Provider Backbone Transport (PBB-TE)



UNI specifikáció
Szolgáltatás Definíciók
Szolgáltatás specifikációk



I E T F®

Pseudowire
Virtual Private LAN Service
Hierarchical VPLS



UNI specifikáció– szolgáltató oldal
Szolgáltatás Definíciók
T-MPLS

Ethernet szolgáltatások

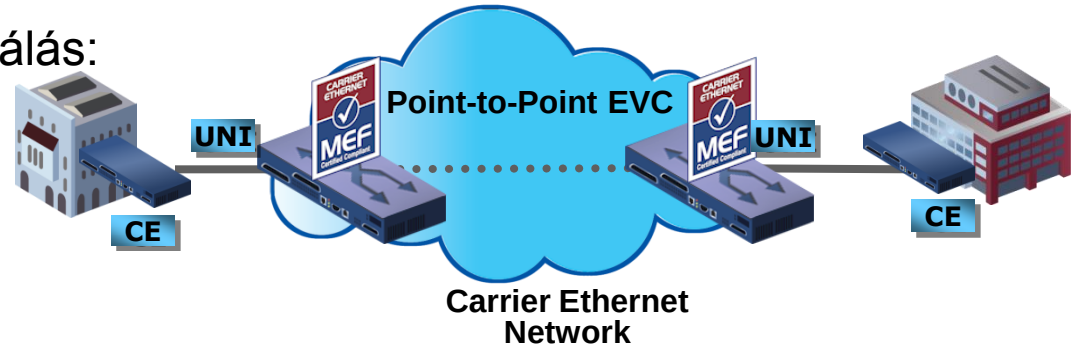


BME-TMIT

E-Line szolgáltatás típus

- E-Line Szolgáltatás – felhasználás:

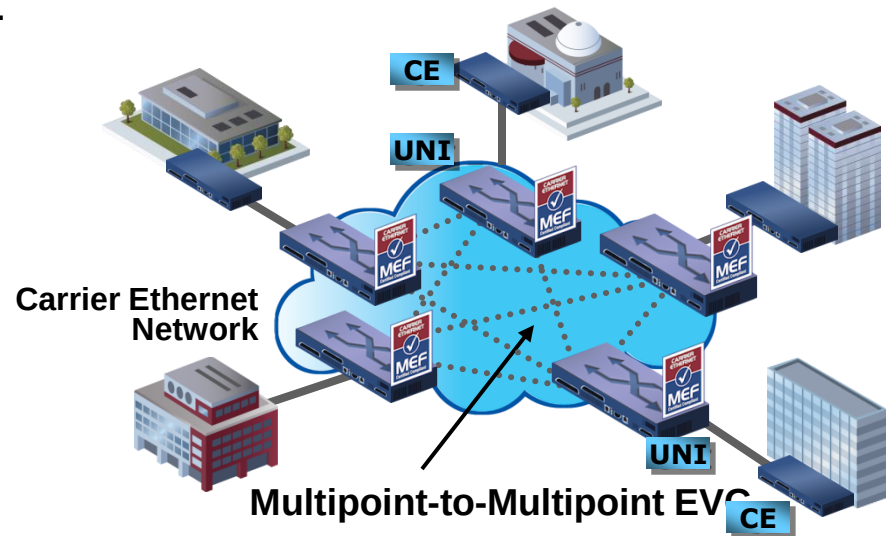
- Ethernet Private Line
- Virtual Private Line
- Ethernet Internet Access



E-LAN szolgáltatás típus

- E-LAN Szolgáltatás – felhasználás:

- Multipont L2 VPN-ek
- Transzparens LAN Szolgáltatás
- Alap az IPTV és Multicast szolgáltatásokhoz stb.



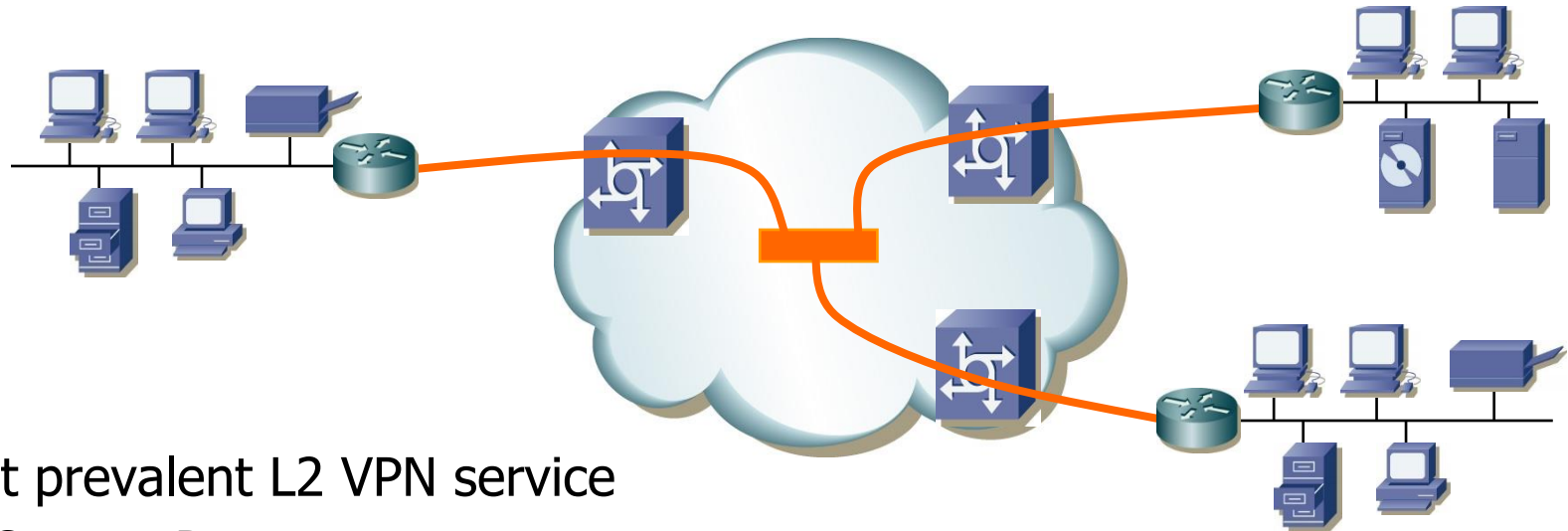
MEF által hitelesített Carrier Ethernet termékek

UNI: User Network Interface, CE: Customer Equipment

Router Inter-connect Characteristics



BME-TMIT

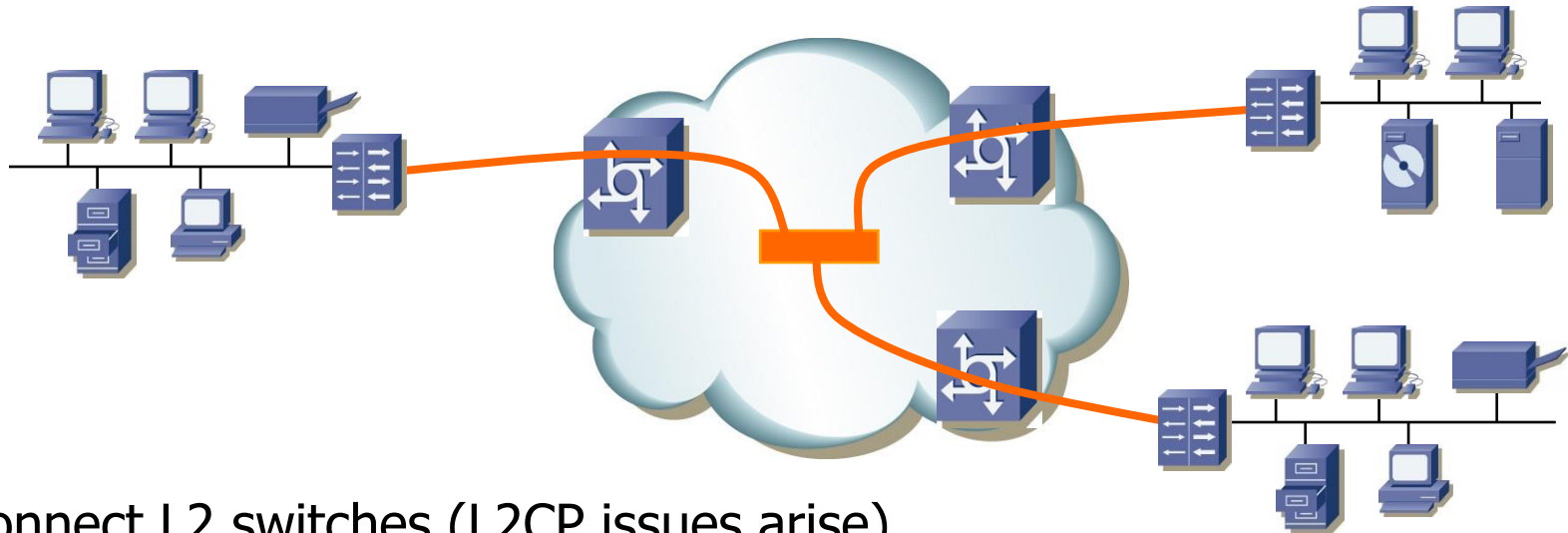


- Most prevalent L2 VPN service
 - Connect Routers
 - Routers provide clear demarcation
 - Possibility of unnecessary BC/MC eliminated
 - L2CP issues eliminated
 - One MAC address per site
- Most customers use routers to connect multiple sites together
 - Well-known design
 - Considerably reduces management overhead (single connection per site)

Switch Inter-connect Characteristics

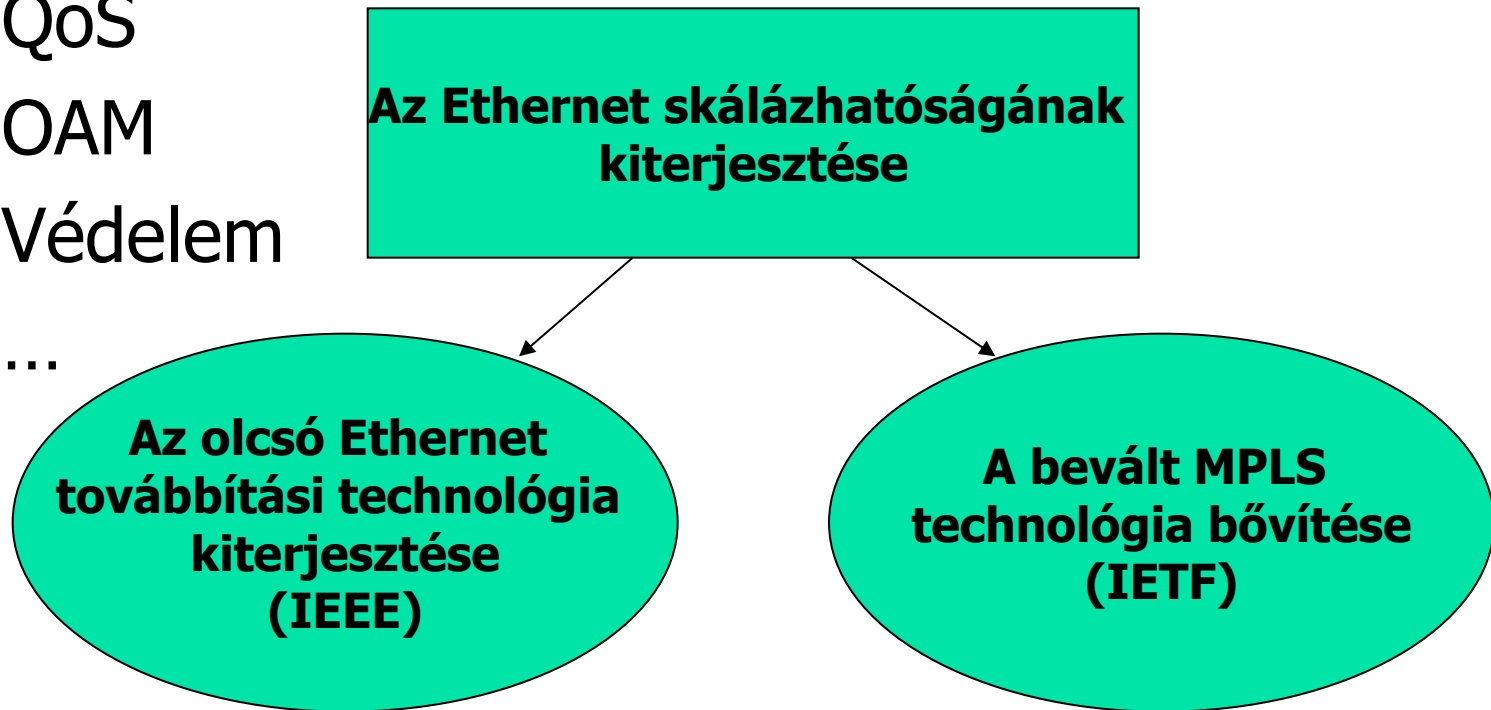


BME-TMIT



- Connect L2 switches (L2CP issues arise)
- Multiple MAC address per site
 - Potential scalability issues
 - Malfunctioning L2 switches can also flood the provider network with BC/MC traffic
 - Requires controls of FIB size (# MAC per site) and BC/MC rate-limiting
- Should be sold as a premium service
 - Charge customers per site per block of MAC addresses

- Szolgáltatói tulajdonságok
 - Skálázhatóság
 - QoS
 - OAM
 - Védelem
 - ...



Ethernet alapú átvitel



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**

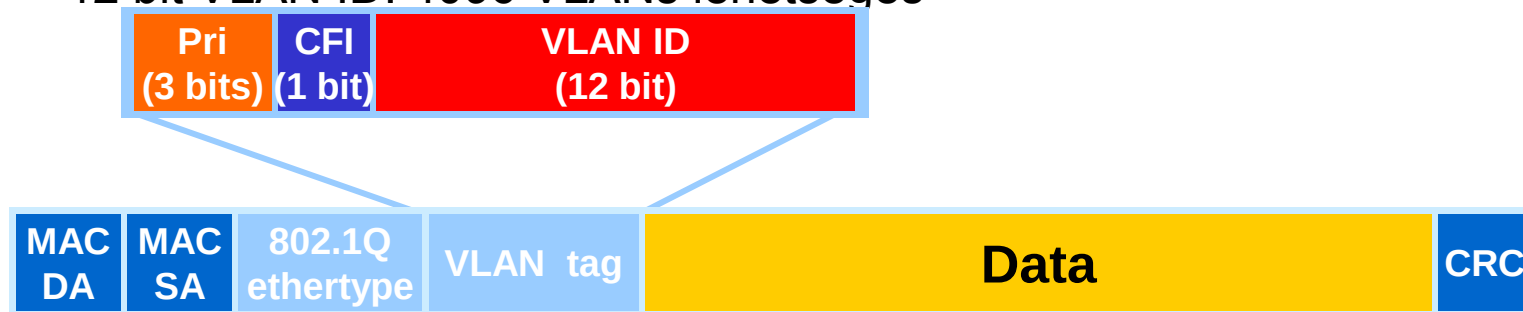


IEEE 802.1Q - VLAN



BME-TMIT

- VLAN tag
 - QoS: prioritás bitek
 - 12 bit VLAN ID: 4096 VLANs lehetséges



- Felhasználási módok:
 - Felhasználó azonosítás
 - Szolgáltatás azonosítás
- Mindkét esetben korlát a 4096 – **szolgáltatói környezetben kevés!**
- De: a legelterjedtebb UNI
 - Fel kell készülni hogy transzparensen át kell vinni a VLAN csomagot

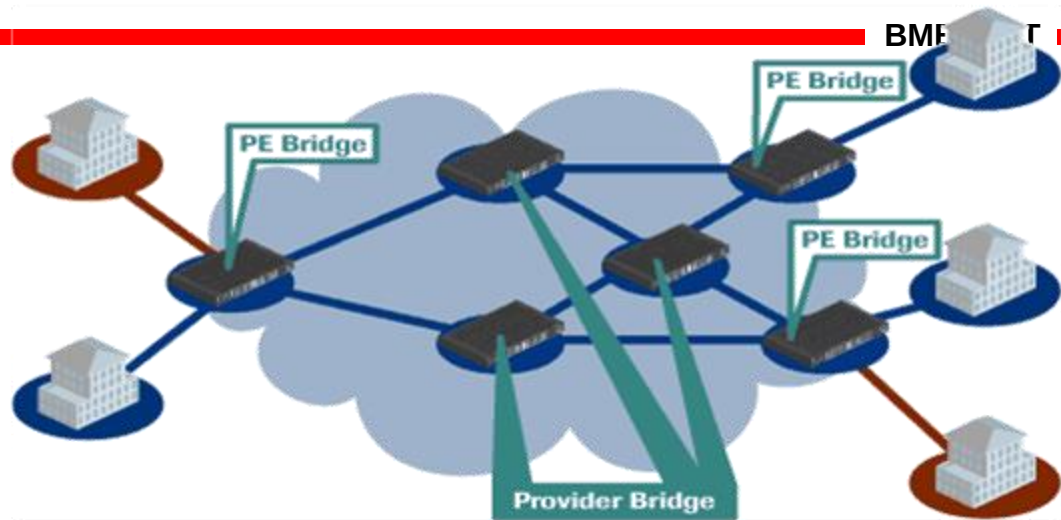
Provider Bridges (IEEE 802.1ad)

BME T

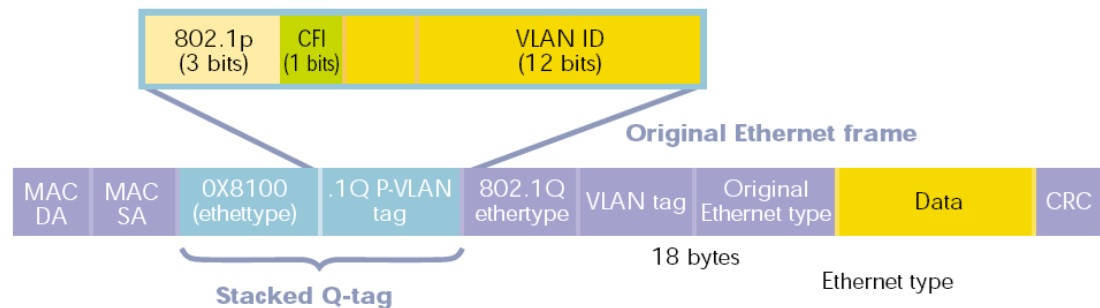
- Q-in-Q néven ismert
- Széles körben elterjedt

- A legelterjedtebb felhasználói interfész
- 4K Szolgáltatás (12-bits)
- Egyedi szolgáltatás ID
 - (S-VID)

- A csomagtovábbítás a megszokott L2 tanuló bridge alapú a MAC DA/SA alapján S-VID szinten és xSTP a hurkok elkerülésére



- Skálázhatóság
 - 4K szolgáltatás

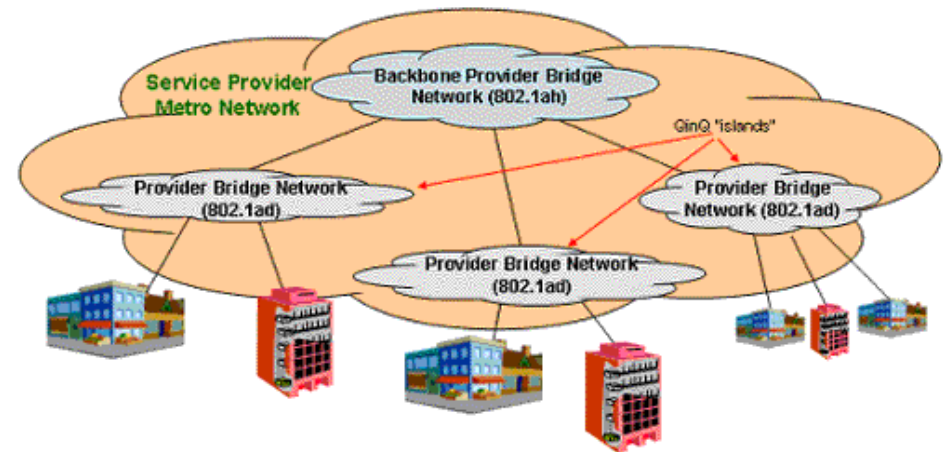
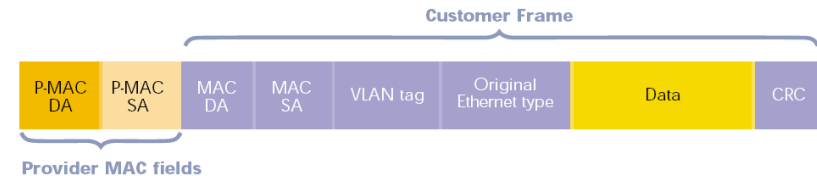


Provider Backbone Bridges



BME-TMIT

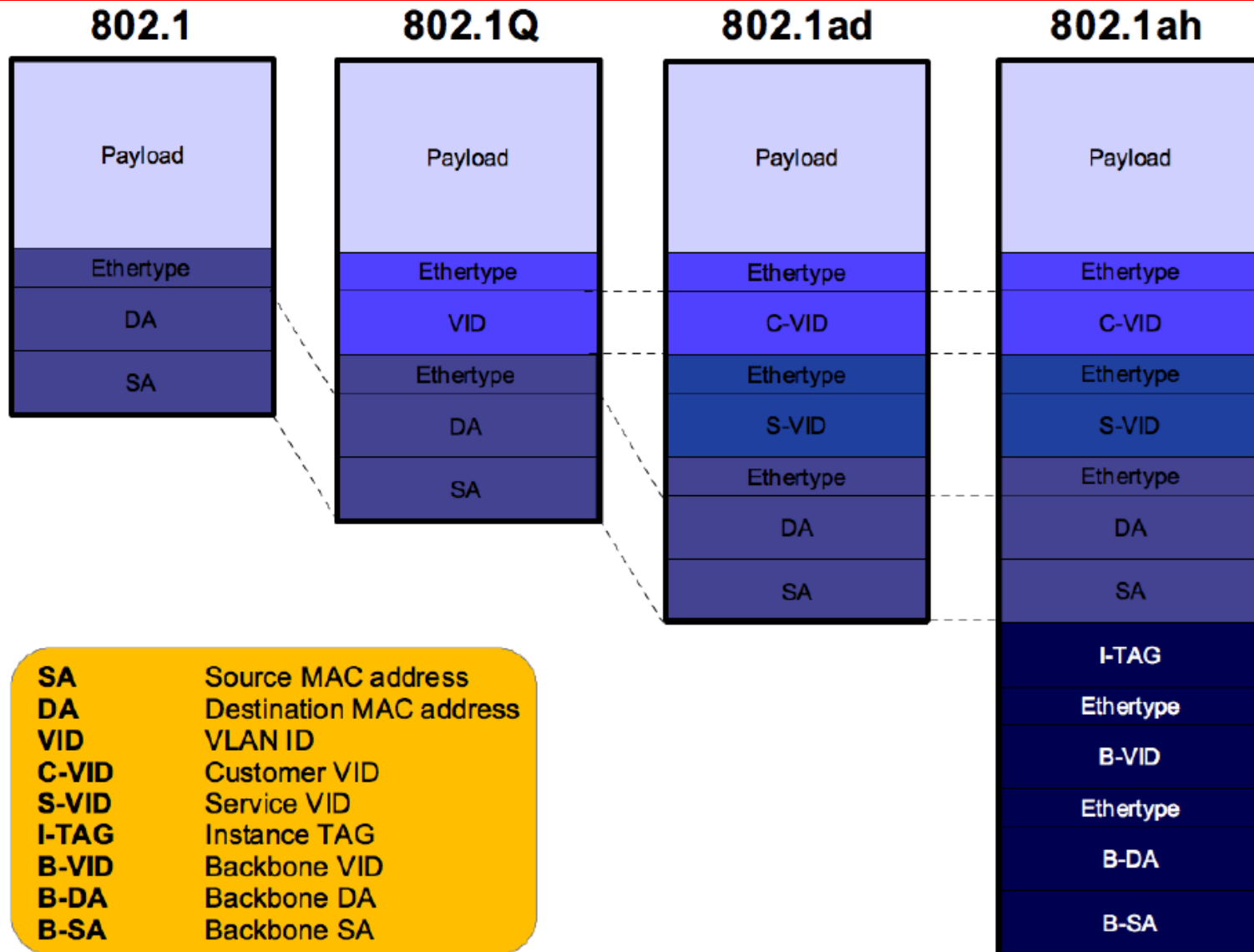
- 4K összekötött többesadós hálózat
- Egyedi Szolgáltatás ID
 - (LAN = I-SID)
- A csomagtovábbítás a megszokott L2 tanuló bridge alapú a MAC DA/SA alapján B-VID szinten és xSTP a hurkok elkerülésére
- A szolgáltatások engedélyezése egyszerű mert az I-SID társítás hasonlít az S-VID beállításhoz
- Skálázhatóság
 - Maszív szolgáltatás skálázhatóság (24-bit)
 - Nincs szükség hogy minden csomópontban beállítsunk egy I-SID-et egy P2P mesh hálózat kialakításához
 - A C-MAC címeket megtanulja és B-VID/I-SID alapján társítja



Összehasonlítás – hozzáadott fejlécek



BME-TMIT



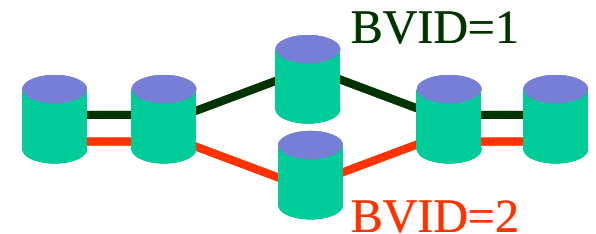
PB/PBB tények



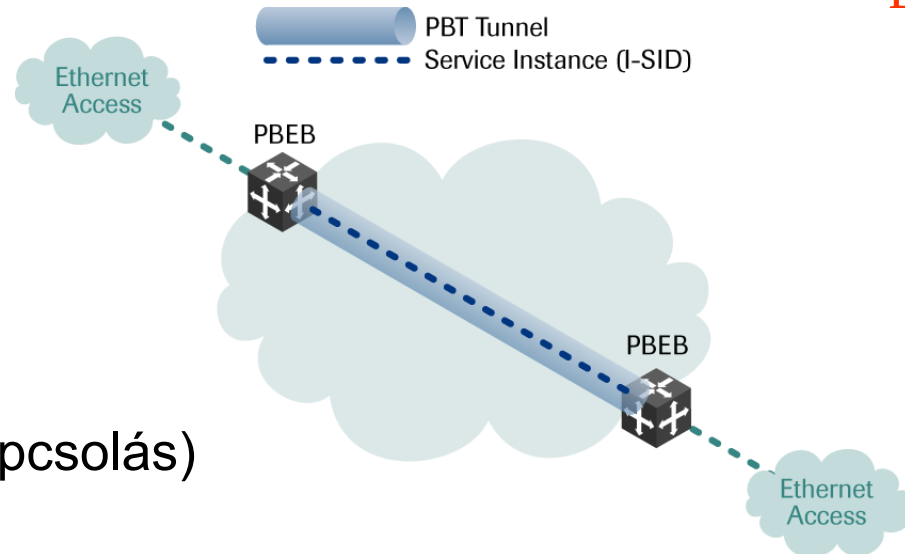
BME-TMIT

- Skálázhatóság megoldva 
 - Marad az olcsó Ethernet kapcsolás 
-
- Továbbra sincs Traffic Engineering 
 - Védelem/helyreállítás STP alapú 
 - Nehézkes a menedzsment 
 - Több szintű VLAN-ok sokasága
 - Nincs hiba- és performancia menedzsment
 - Még mindig nem megfelelő a maghálózatban...

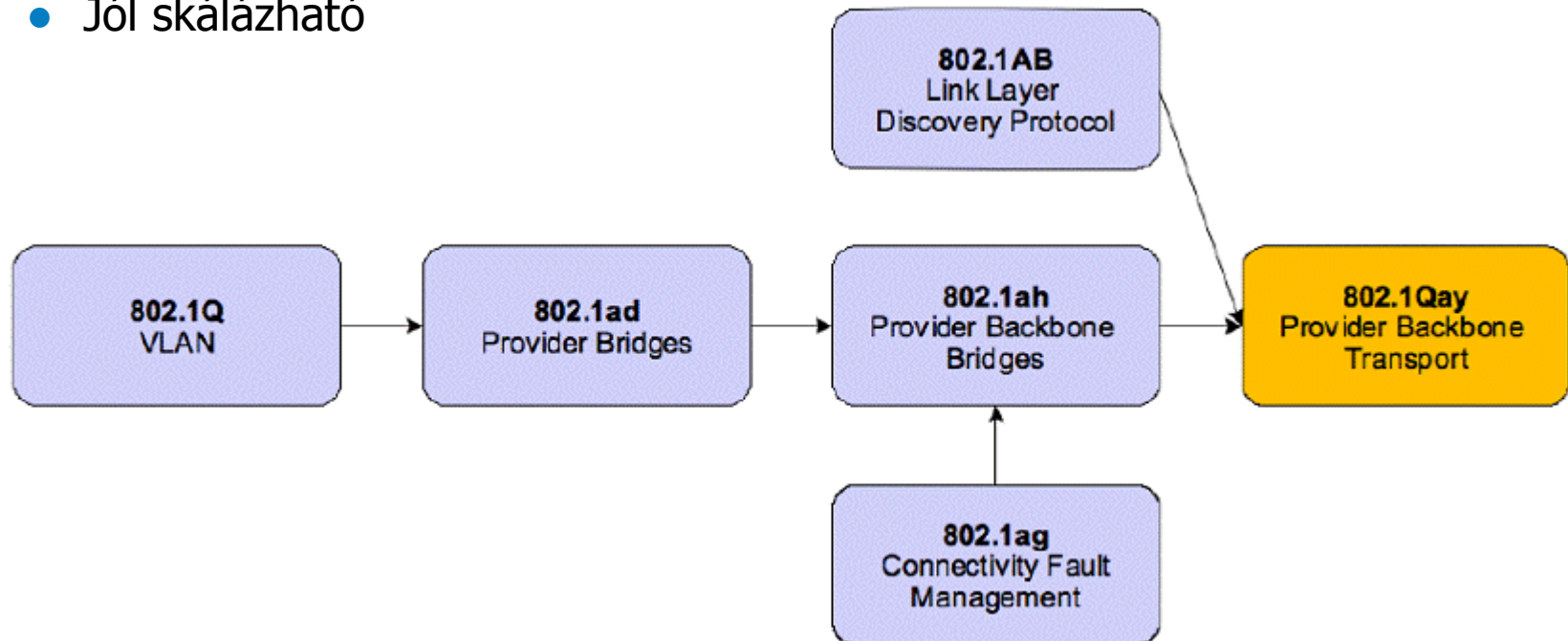
- A cél:
 - marad az Ethernet csomagtovábbítási mechanizmus
 - Lecseréljük a vezérlési síkot (kikapcsoljuk az STP-t és tanulást)
 - Az útvonalakat a hálózatban előre beállítjuk
= Traffic Engineering - Ethernet



- Amit kapunk:
 - Pont-pont tunnelek
 - Traffic Engineering
 - Védelem (védelmi kapcsolás)



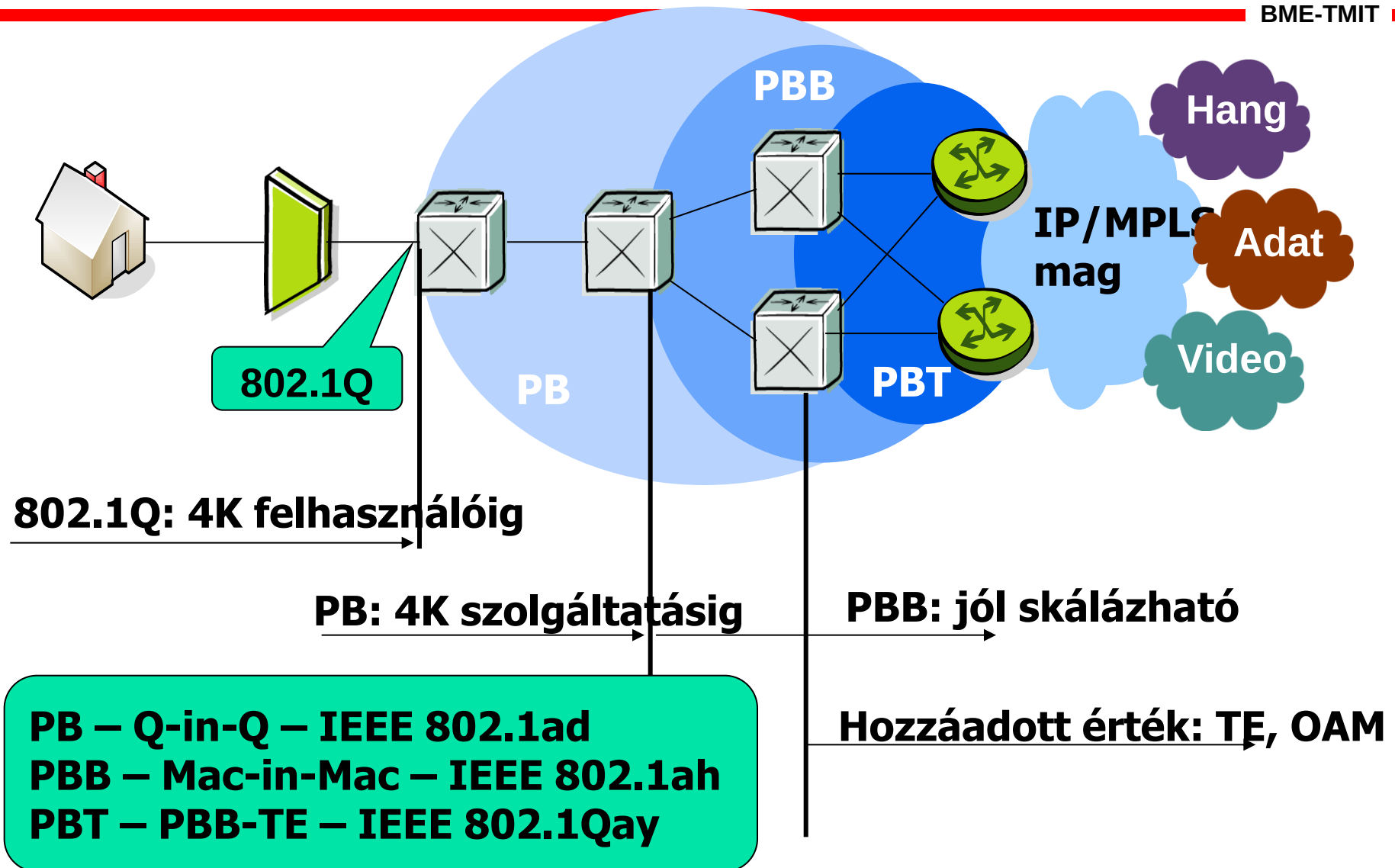
- Provider Backbone Transport – IEEE 802.1Qay
 - Nortel kezdeményezte
 - A PBB-re épül
- Újrahasznosítja a meglévő technológiákat
 - Determinisztikus QoS-t nyújtó szolgáltatás a cél
 - Jól skálázható



Ethernet Transzport technológiák alkalmazása



BME-TMIT



MPLS alapú átvitel



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**

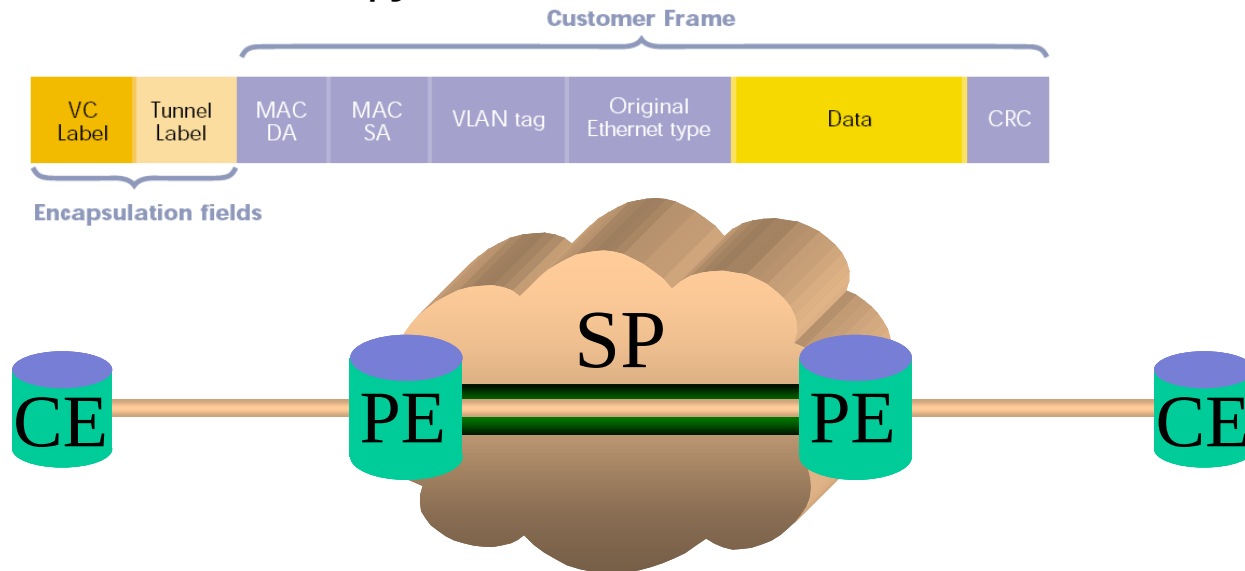


MPLS Pseudowire - VPWS



BME-TMIT

- Ethernet p2p kapcsolatot tesz lehetővé
 - Az IETF pwe3 csoport dolgozta ki, a draft neve alapján Martini – enkapszulációnak is nevezik
 - Az MPLS címke is beágyazott, egy UNI-n belül több virtuális kapcsolatot megkülönböztetve (VC)
 - A tunnel címke alapján továbbítódik a hálózatban

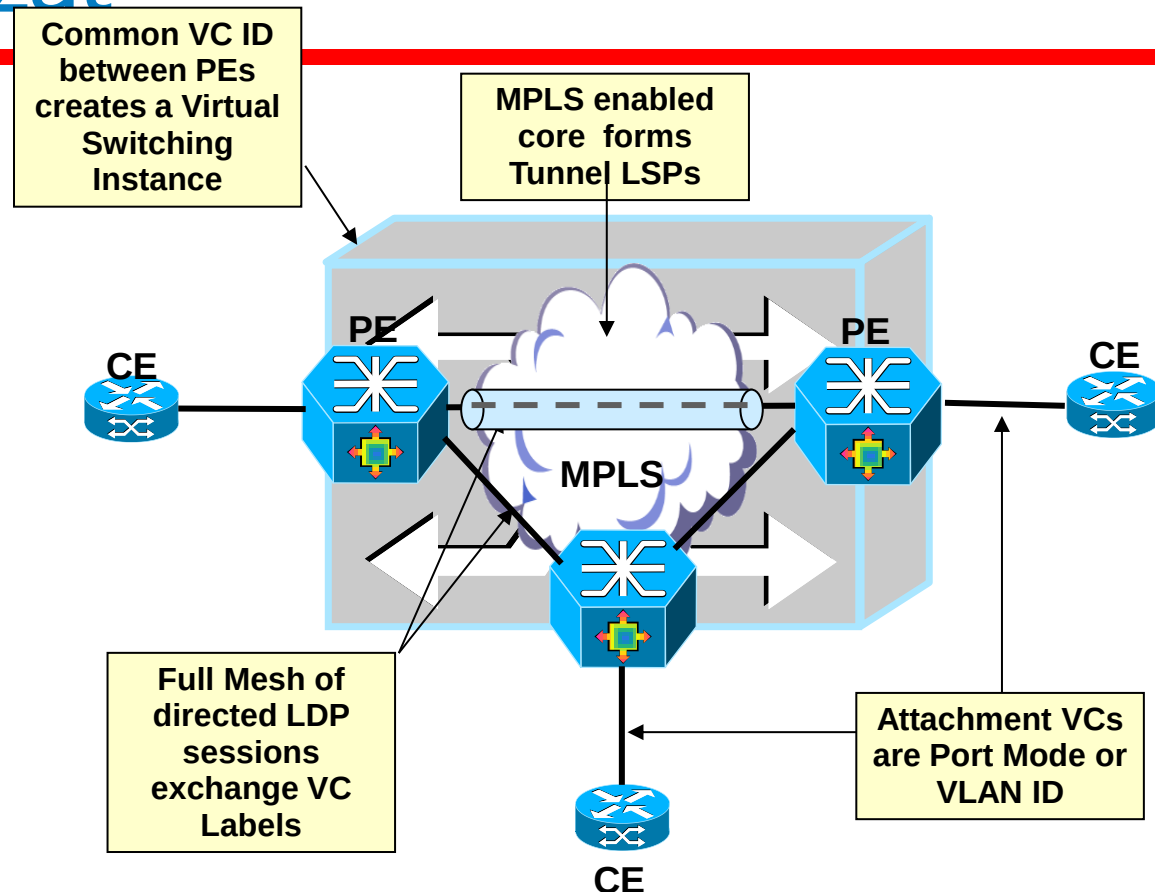


- A megoldás örökli az MPLS összes jó tulajdonságát:
 - Traffic Engineering, Védelem, OAM

VPLS hálózat



BME-TMIT



Customer Edges (CE): kliens oldali eszköz, tipikusan Etherneten csatlakozik

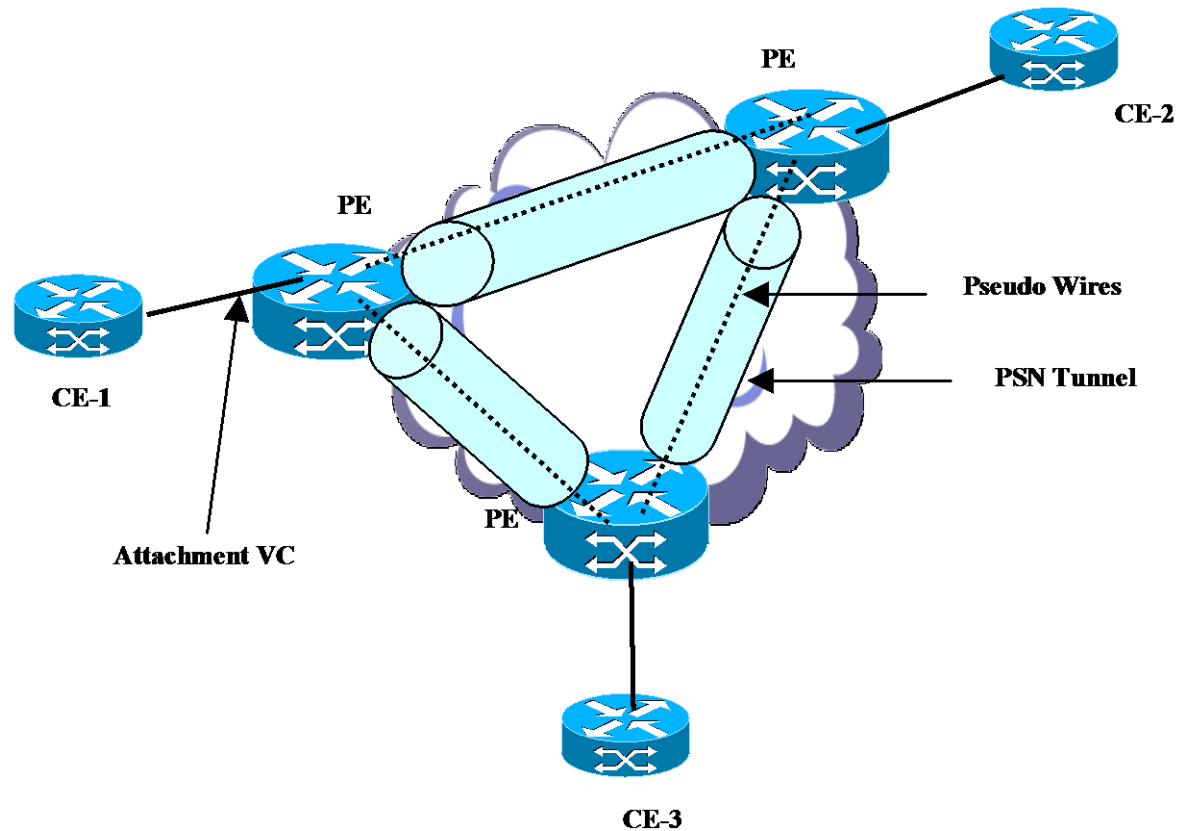
Provider Edges (PE): itt található a VPLS intelligencia, kezdő/végső pont

Core: csak a továbbításban vesz részt

VPLS rendszer példa



BME-TMIT



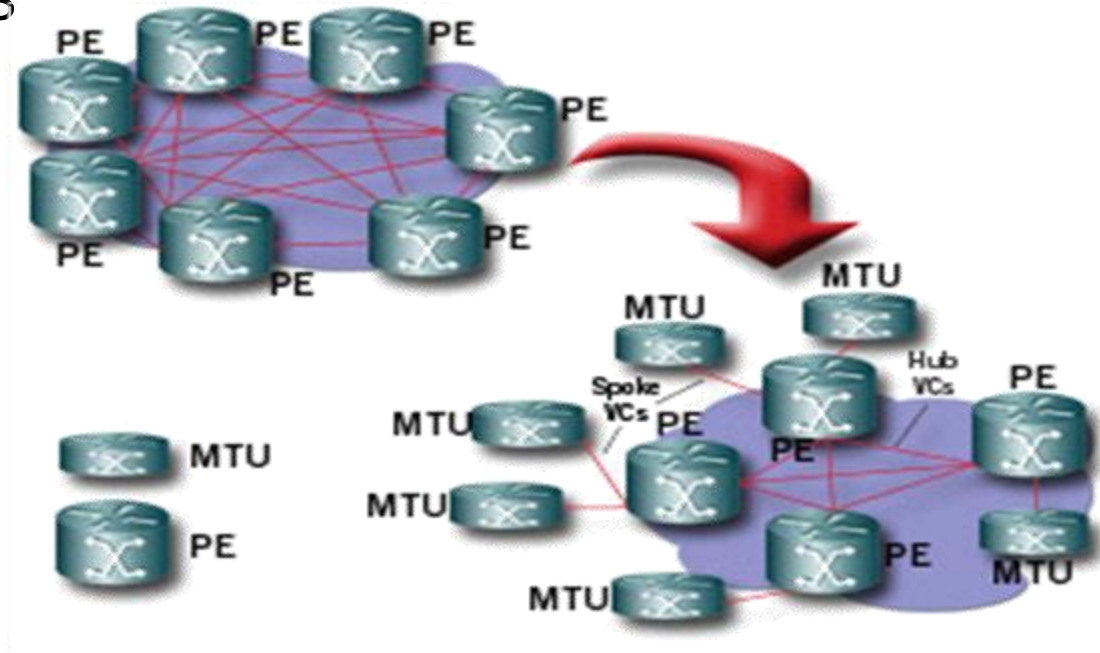
Full Mesh alagutak a PE-k között
- nem feltétlen fizikai, sőt!
PE-k egy virtuális bridge-t mutatnak a CE-k felé

- VPLS instance : Service–identifier (Svc-id)
- Full mesh tunnelek kialakítása
 - Célzott LDP üzenetekkel
- Csomag továbbítás: tanuló bridge
 - Ismeretlen cél: broadcast az összes tunnelre
 - Split-horizon: soha nem továbbít oda, ahonnan érkezett – a hurkok ellen

VPLS skálázhatóság - hierarchia

- MTU - Multi-Tenant Unit: több felhasználó által bérelt eszköz, bridge
- MTU-ig kiterjeszthető a VPLS

- MAC/VLAN skálázhatóság megnő
- Komplexebb MTU

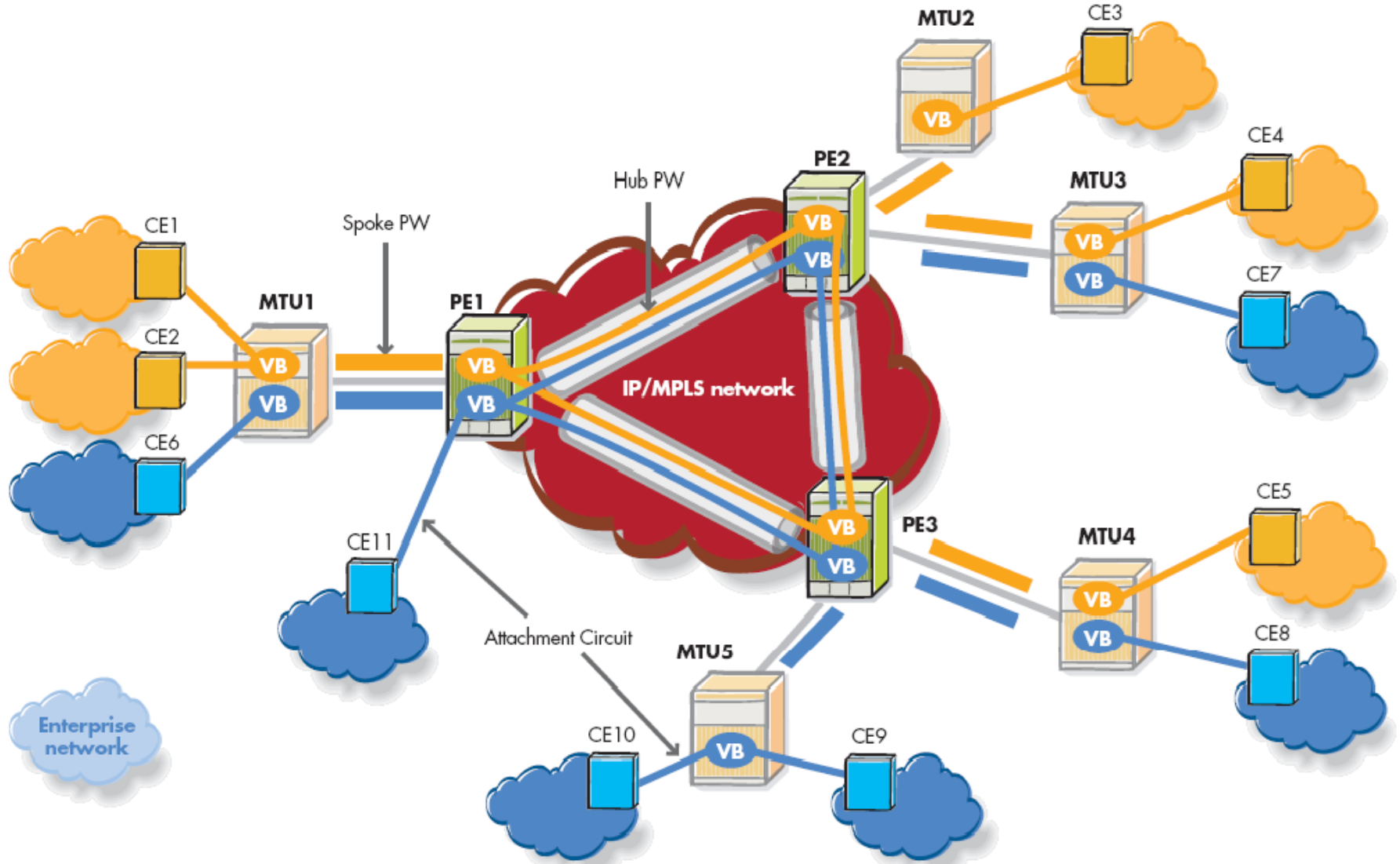


- Hierarchikus VPLS
 - PE-k között „HUB” pseudowire-k (hub PW)
 - MTU-PE között „spoke” PW
 - Spoke PW lehet QiQ, MPLS, ...

Hierarchikus VPLS



BME-TMIT



VPLS jelzés és automatikus felderítés



ME-TMIT

VPN
felderítés

Központosított
DNS, LDAP, Radius, Directory Services

Elosztott
BGP

Jelzés

Címke kiosztási
Protokoll (LDP, BGP)

- VPLS megköveteli a full-mesh LSP-k kihúzását a PE-k között egy adott VPLS instance számára:
 - Manuális eljárás (statikus beállítás)
 - Menedzsment rendszerek (NMS/OSS alapú beállítás)
 - Jelzési protokollok:
 - LDP ("Lasserre-V. Kompella" draft, a legtöbb gyártó által támogatva)
 - BGP ("Kompella" draft, kevesen támogatják)
 - other (Radius, DNS, etc.)

VPLS javaslat	Automatikus felderítés	Jelzés / címke kiosztás
Draft Kompella VPLS	BGP	BGP
Draft Lasserre-Vkompella VPLS	Nincs (több lehetséges opció)	LDP

VPLS és H-VPLS skálázhatóság



BME-TMIT

- PW kapcsolatok az MPLS vezérlő sík által kihúzva
- A Hub-and-Spoke csökkenti a full-mesh PW igényét
- A szolgáltatások számának növekedésével a MAC tábla mérete és a PW-k száma közt egyensúlyt kell teremteni

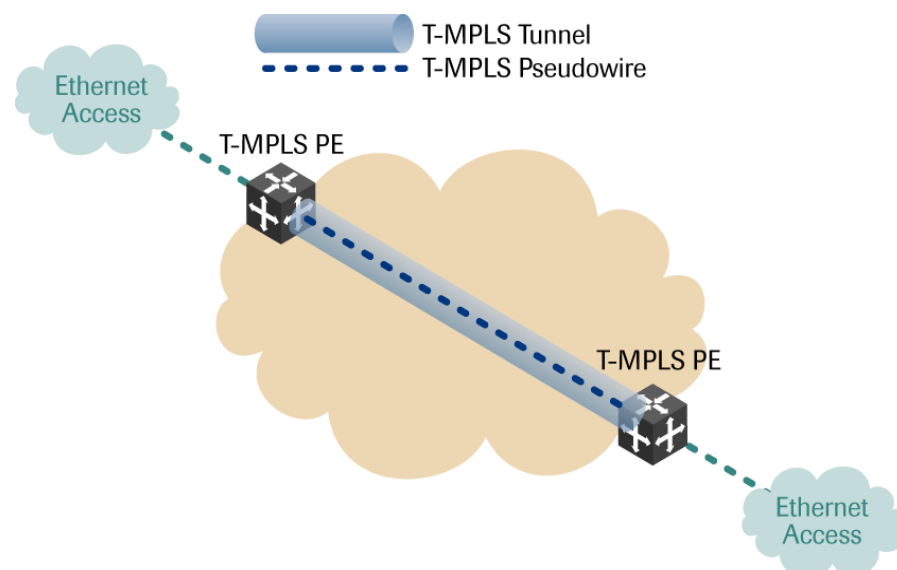
	Services	PEs	MTU-s	PW mesh	MACs (1K per service)
VPLS	512	64	-	1M	512K
H-VPLS	512	16	256	60K	512K
	8,192	16	256	1M	8M

Vissza az áramkör kapcsoláshoz: T-MPLS



BME-TMIT

- Transport-MPLS
- **Kimondottan kapcsolat orientált alcsoportja az MPLS-nek**
- Az ITU-T szabványosítja
 - Az MPLS egy része, amely a p2p kapcsolatok kezeléséhez szükséges
- arra helyezi a hangsúlyt, hogy az IP menedzsmentet eliminálva az MPLS megoldások menedzsmentjét leegyszerűsítse
- Fő szempontok
 - Védelem: 50ms
 - Traffic Engineering
 - Garantált SLA
 - OAM: hiba és performancia menedzsment



"Carrier Ethernet" szolgáltatások - összehasonlítás

BME TMT

	VPLS	H-VPLS	T-MPLS	PBB	PBT	PBT-LAN
Ethernet Services	P2P, LAN	P2P, LAN	P2P	P2P, LAN	P2P	P2P, LAN
Tunnel Scale (N=PE nodes)	N up to $N! / 2*(N-2)!$	N up to $N! / 2*(N-2)!$	Service count	4K Flood Domains	2 * Service Count	$N! / 2*(N-2)!$
Pseudowire Scale	$N! / 2(N-2)! *$ Services	$N! / 2(N-2)! *$ Services	Service count	N/A		
Service Scale	100s	1000s	1M	16M		
Protection	FRR	FRR	Y.1720 Y.mrps	xSTP	In-Tunnel CCM	In-Tunnel CCM
Dual Homing	Yes			No	Yes	Yes
Control Plane	OSPF/IS-IS w/ RSVP-TE and LDP		GMPLS	xSTP only for Loop Prevention	<i>Future</i> OSPF/IS-IS w/ RSVP-TE	<i>Future</i> OSPF/IS-IS w/ RSVP-TE
Split Horizon Forwarding	Yes		No	No		Yes
OAM	Virtual Ping Virtual Traceroute BFD			Virtual Ping Virtual Traceroute CFM		

- Pont-Pont kapcsolat
- Generic Framing Procedure (GFP)
 - Szabvány az első és második rétegbeli protokollok Sonet/SDH keretezésére
- Virtual Concatenation (VCAT)
 - Az SDH sáv szélesség felosztása virtuálisan
 - Link Capacity Adjustment Scheme (LCAS)
 - Elősegíti a dinamikus sáv szélesség kiosztást a VCAT segítségével

- Az Ethernet alapú szolgáltatások gyorsan fejlődnek
 - Szabványosítás
 - Alkalmazás
- Több, egymással versenyző technológia van az Ethernet szolgáltatások átvitelére
 - MPLS és Ethernet alapú megközelítések
 - Mindkettőnek megvan az előnye-hátránya
- A különböző technológiák arányát/határait a hálózatokban a szolgáltatók fogják eldönteni

OAM

Operations
Administration
Maintenance



Budapest University of Technology and Economics

Department of
Telecommunications and Media Informatics



- Jelenleg korlátozódik az SNMP alapú lekérdezésekre
- A szolgáltatások megvalósításához szükséges!
 - Ethernet szintű „ping”, „traceroute”
- Szabványosítás folyamatban
 - IEEE 802.1ag – CFM, ITU - Y.17ethoam
 - IEEE 802.3ah – EFM

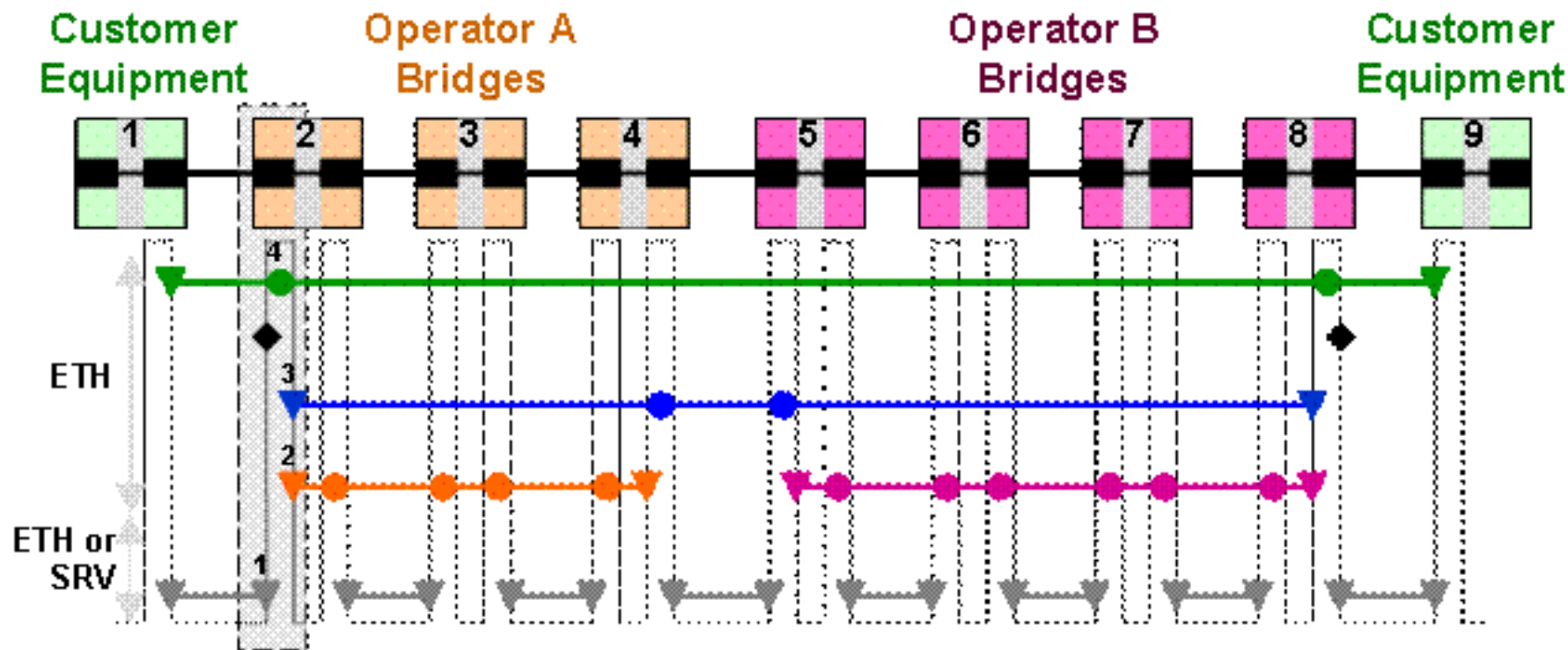
- Link monitorozás
- EVC kapcsolatok folytonosságának ellenőrzése
- Jelentés a hálózati menedzsment felé hiba észlelése esetén
- Alsóbb szintekről érkező redundáns üzenetek szűrése
- Az alapvető eszköztár nyújtása az operátor felé
- STP hiba esetén is működjön!

OAM Hierarchia



BME-TMIT

- Maintenance Entities (ME)
 - MEP – mgmt. endpoint, MIP – intermediate
 - Különböző szintek a menedzsmentben



- Hiba menedzsment funkciók
 - A kapcsolati hibák felderítése, ellenőrzése és lokalizálása
- Performancia menedzsment
 - Minőségi hibák felderítése, ellenőrzése és lokalizálása
- Kommunikáció: OAM keretek segítségével
 - Out-of-Band: különálló OAM keretek

- Hiba menedzsment funkciók (CFM)
 - Continuity Check – folyamatos ellenőrzés
 - Az aktuális szolgáltatói szinten
 - SNMP trap– hiba jelentés
 - Loopback
 - Ellenőrzés, IP megfelelő: ping
 - Link Trace
 - lokalizálás, IP megfelelő: traceroute

- Performancia menedzsment
 - Különböző metrikák mérését teszik lehetővé
- Performancia mérések:
 - Frame Loss (FL),
 - Frame Delay (FD),
 - Frame Delay Variation (FDV),
 - Availability
- Folytonos mérés: Frame Loss

Köszönöm a figyelmet



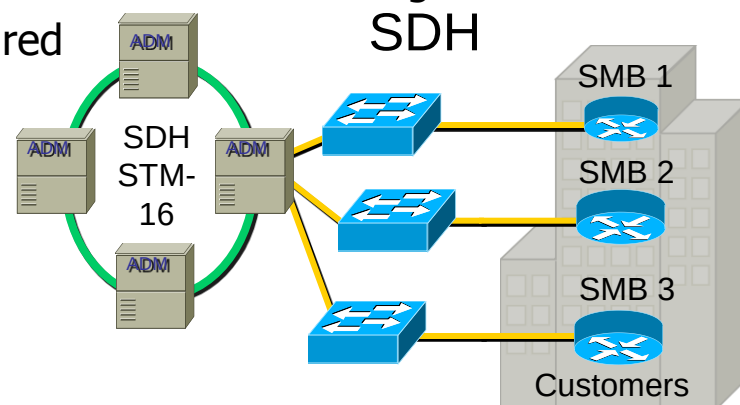
Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**



- “Some” development was required
 - G.7041: GFP (Generic Framing Procedure) = protocol agnostic frame container
 - two mapping modes: frame mapped GFP (one MAC PDU in one GFP frame), transparent GFP (into fixed length GFP frame)
 - G.707: VCAT (Virtual Concatenation) = better bandwidth granularity
 - bond any number of noncontiguous VCs as a single flow (Virtual Concatenation Group: VCG)
 - Similar to the concept of IMA (Inverse Multiplexing for ATM)
 - G.7042: LCAS (Link Capacity Adjustment Scheme) = dynamic bandwidth settings
 - signaling method for dynamically add and remove members of a VCG
 - also useful for fault tolerance and protection
 - Similar to the concept of IMA (Inverse Multiplexing for ATM)
 - Many open points in the standard, rarely implemented, interworking problematic
- Ethernet frames are mapped into VCs (VC-12, VC-3, VC4)

- SDH protections mechanisms can be used (ring topology, APS, etc.)
- P2P by nature of SDH
 - A tag (e.g. VLAN or MPLS) can be used to overcome this limitation and share bandwidth between customers
 - MP2MP
 - Currently very limited implementations
 - VPHS (Virtual Private Hub Service) = VPLS without MAC learning
 - VPLS like mechanisms and solutions required
- Using MPLS results in same solution as VPWS/VPLS



Protected SDH Ring w/ Ethernet interfaces

Hálózati réteg - áttekintés

Moldován István



Rétegződés

**OSI
Reference Model**

Application
Presentation
Session
Transport
Network
Link
Physical

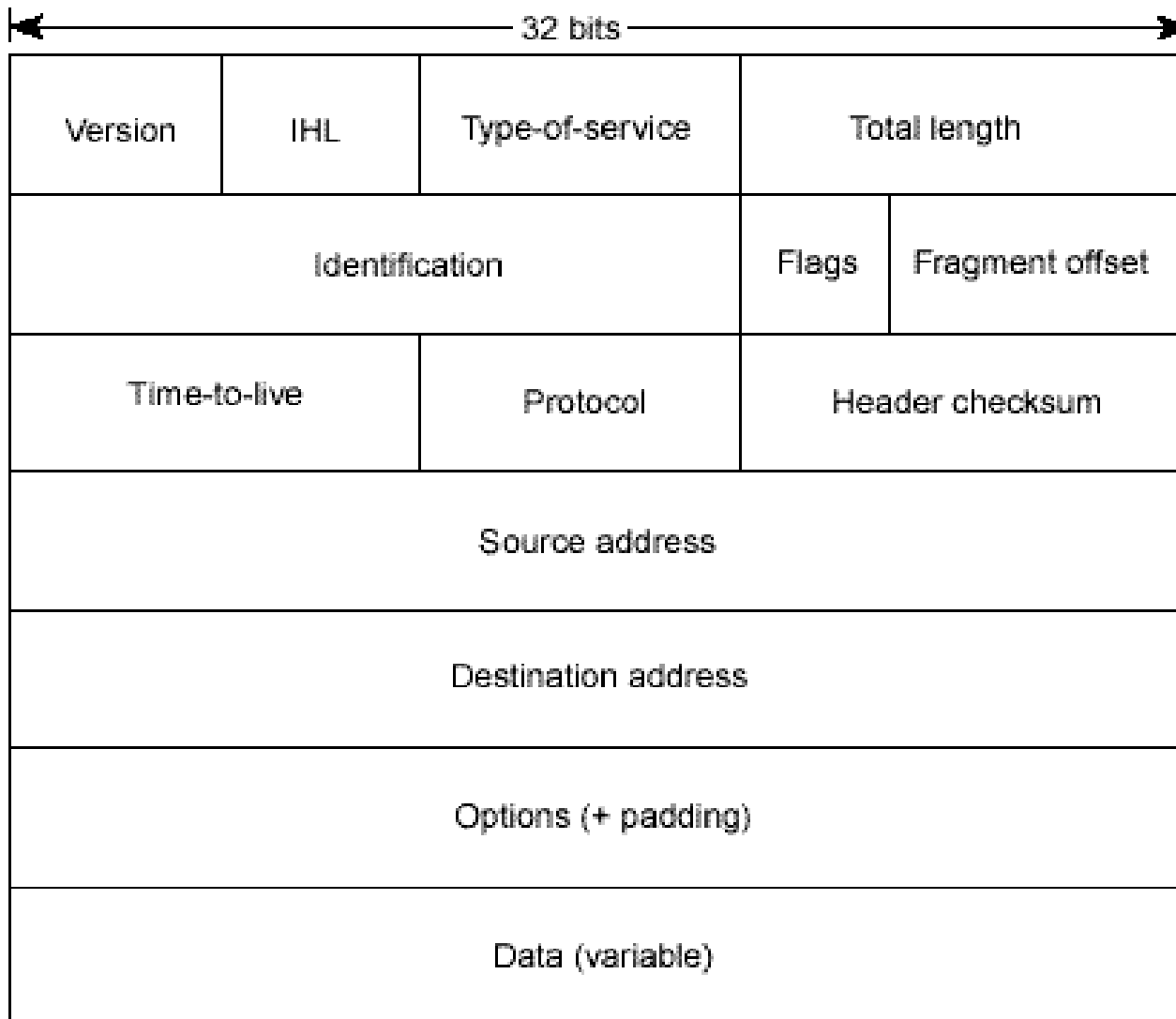
Internet Protocol Suite

FTP, Telnet, SMTP, SNMP	NFS
	XDR
	RPC
TCP, UDP	
Routing Protocols	IP ICMP
ARP, RARP	
Not Specified	

Az IP

- Lehetővé teszi hogy bármely két Internetre kötött gép kommunikáljon egymással
- Feladata a csomag eljuttatása a célállomáshoz – semmi garancia (best effort)
- A csomag több átjárón, útvonalválasztón haladhat át
 - Útvonalválasztás
 - Hurok detektálás

Az IPv4 fejléc



IP fragmentálás

- Továbbítás során a csomag több hálózaton haladhat át
 - kisebb MTU -> darabolás
 - Az IP fejléc tartalmazza a darab számát
 - A darabok összeállítását is az IP végzi
- A darabolás elkerülhető
 - “Path MTU discovery”

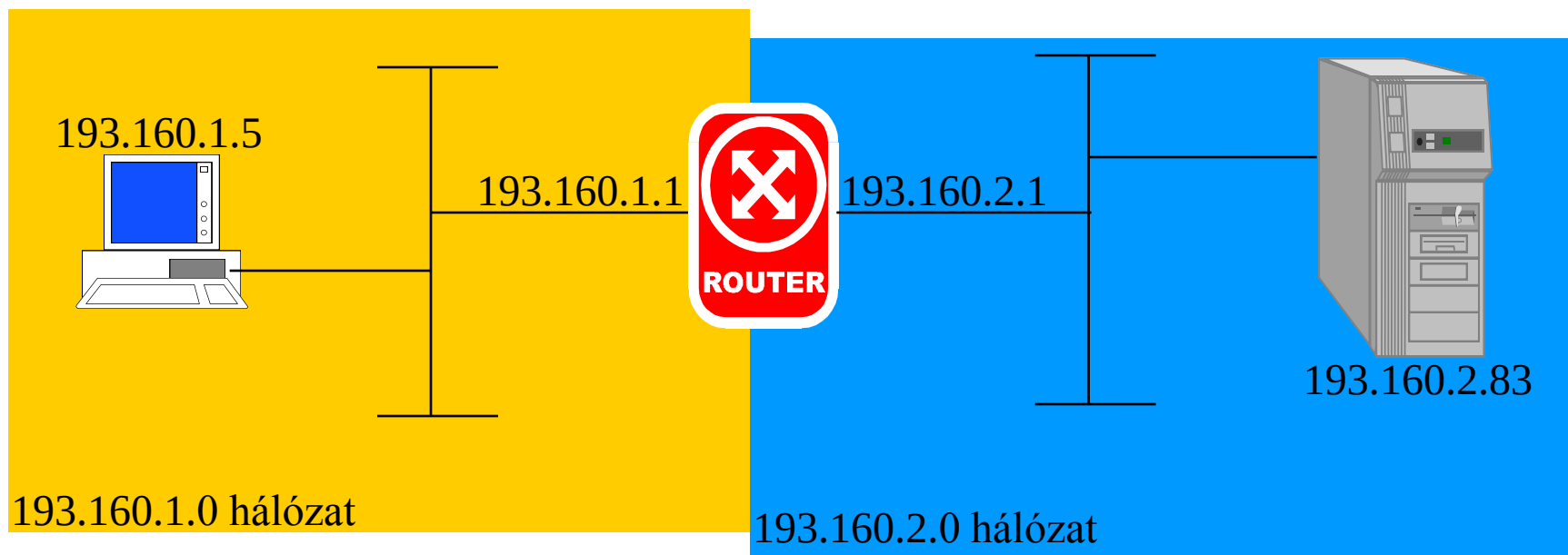
IP címek részei

- 2 részből épülnek fel, pl.:



- Hosszuk változó

IP címek részei - példa



- 193.160.1.0 hálózat:
 - Hosztok: 193.160.1.1 ... 193.160.1.254-ig
 - Az első 24 bit: hálózat azonosító

Bináris formátum	11000001 10100000 00000001 00000000
IP cím	193.160.1.0

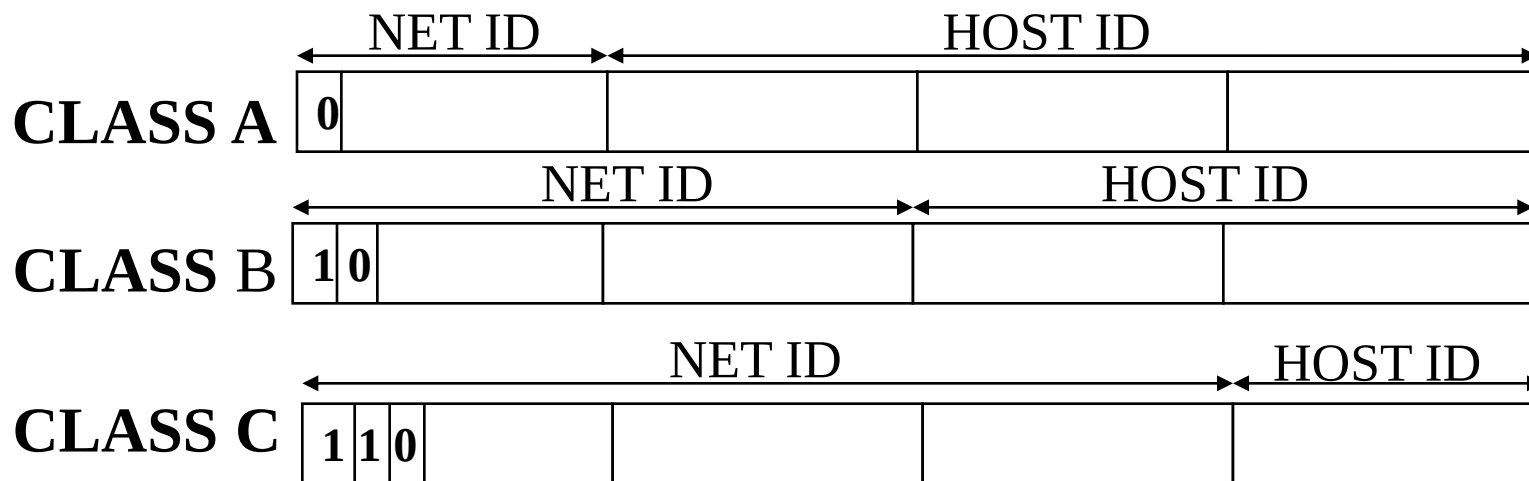
Tradicionális IP címosztályok

- Az IP címek csoportokra osztottak:
 - 5 osztály (A,B,C,D,E)
 - Hálózat/hoszt azonosító hossza változik
- Általános célú címek: A, B, C osztályok
- Előző példa:
 - C osztályú cím:
 - Címosztály azonosítója



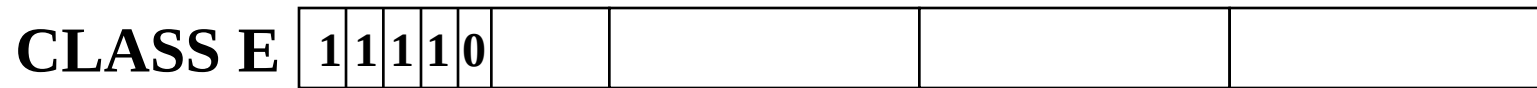
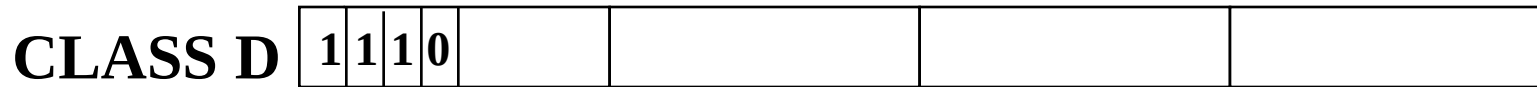
Bináris formátum	11000001 10100000 00000001 00000000
IP cím	193.160.1.0

Tradicionális IP címosztályok



	Hálózatok száma	Egy hálózaton hosztok maximális száma	Első octet értéke
Class A	126	16,777,214	1 – 126
Class B	16,384	65,534	128 – 191
Class C	2,097,152	254	192 - 223

Tradicionális IP címosztályok



- Class D
 - Multicast csoportok címezésére
 - Első oktet értéke: 224..239
- Class E
 - Foglalt, „jövőbeni” használatra
 - Első 5 bit: 11110

Címhasználati szabályok

- *A Hálózati azonosító (NET ID)* nem lehet 127
 - 127 foglalt a loop-back interfésznek
- *A Hoszt azonosító* nem lehet 255 (minden bit 1-es)
 - 255 un. broadcast cím
- *A Hoszt azonosító* nem lehet 0 (minden bit 0)
 - 0 jelentése: „az adott hálózat”
- *A Hoszt azonosítónak* egyedinek kell lenni az adott hálózaton

Címhasználati szabályok - 2

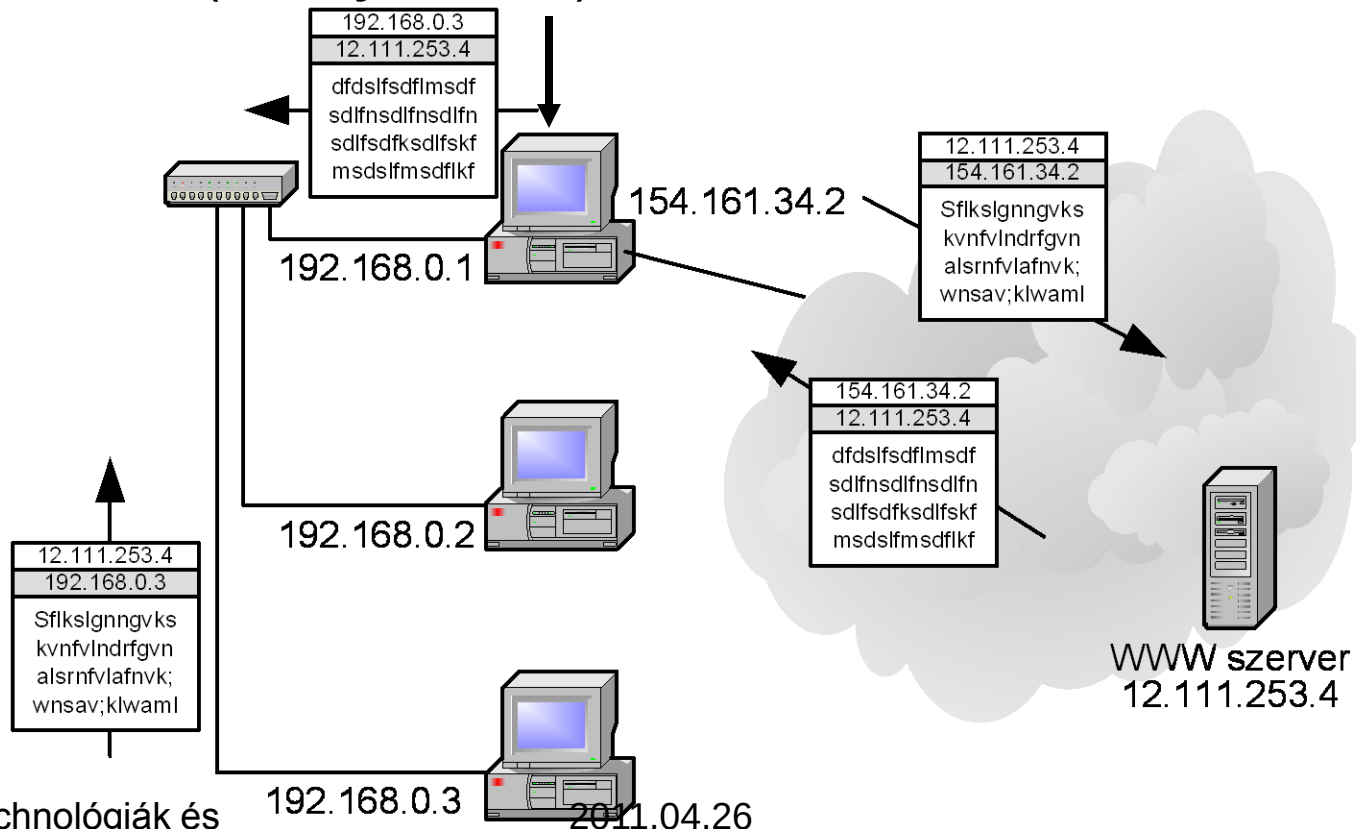
- Broadcast – Üzenetszórás
 - Az alhálózat utolsó címe: pl. 152.66.244.255
 - A szabvány engedi a 255.255.255.255 használatát
 - Az alhálózat minden gépének szól
- Hálózati cím: 152.66.244.0
 - Subneten belüli kommunikációhoz
 - Ha egy subneten – ARP
 - Különben Gateway kell

IP címek kimerülése

- 4 294 967 296 (2^{32}) elvi kiadható címmennyiség
- Csökkenti:
 - Címzési osztályok
 - Címhasználati szabályok
 - Class B címosztály „népszerűsége”...
- IP címek kimerülésének megakadályozása
 - Privát IP címek, ezek többszörös felhasználása
 - Network Address Translation (NAT), címfordító használatával
 - Kevésbé népszerű címosztályokban alhálózatok kialakítása
 - Subnetting
 - Classless InterDomain Routing (CIDR)
 - A tradicionális címosztályok feloldása
 - IPv6
 - nagyobb címtér: 32 bit helyett, 128 bit

Magán IP címek többszörös használata

- A magán IP címek az Internet felől nem „látszanak”
- Címfordítás (az átjáróban)



NAT problémák

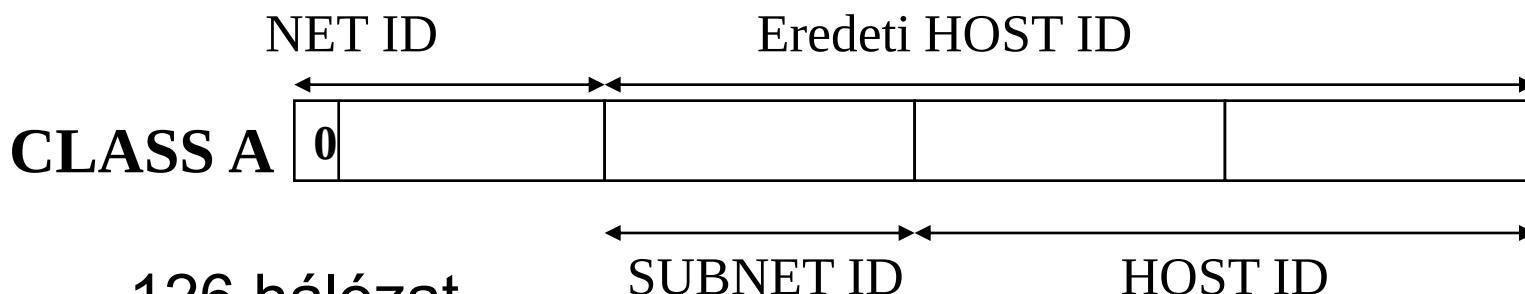
- Átmeneti megoldás
 - Kívülről nem lehet kapcsolatot teremteni egy NAT-olt géppel
 - Egyre több alkalmazás ahol globálisan routolható IP cím kell
 - VoIP, videokonferencia, hálózati játékok
 - Sok protokoll nem működik NAT-olt hálózaton

Magán IP cím típusok

- 1 db „Class A” hálózat:
 - 10.0.0.0 - 10.255.255.255
- 16 db „Class B” hálózat:
 - 172.16.0.0 - 172.31.255.255
- 256 db „Class C” hálózat:
 - 192.168.0.0 - 192.168.255.255
- 224.0.0.0 – 240.255.255.255 (D oszt.)
 - Multicast címek

Alhálózatok - Subnet

- A címosztályok adta hálózatokon belül
 - Alhálózatok kialakítása, kevesebb hoszttal
- Pl. „Class A” címek népszerűsítése:
 - (Túl sok hoszt egy hálózaton)



- 126 hálózat
- 254 alhálózat/hálózat
- 65534 hoszt/alhálózat

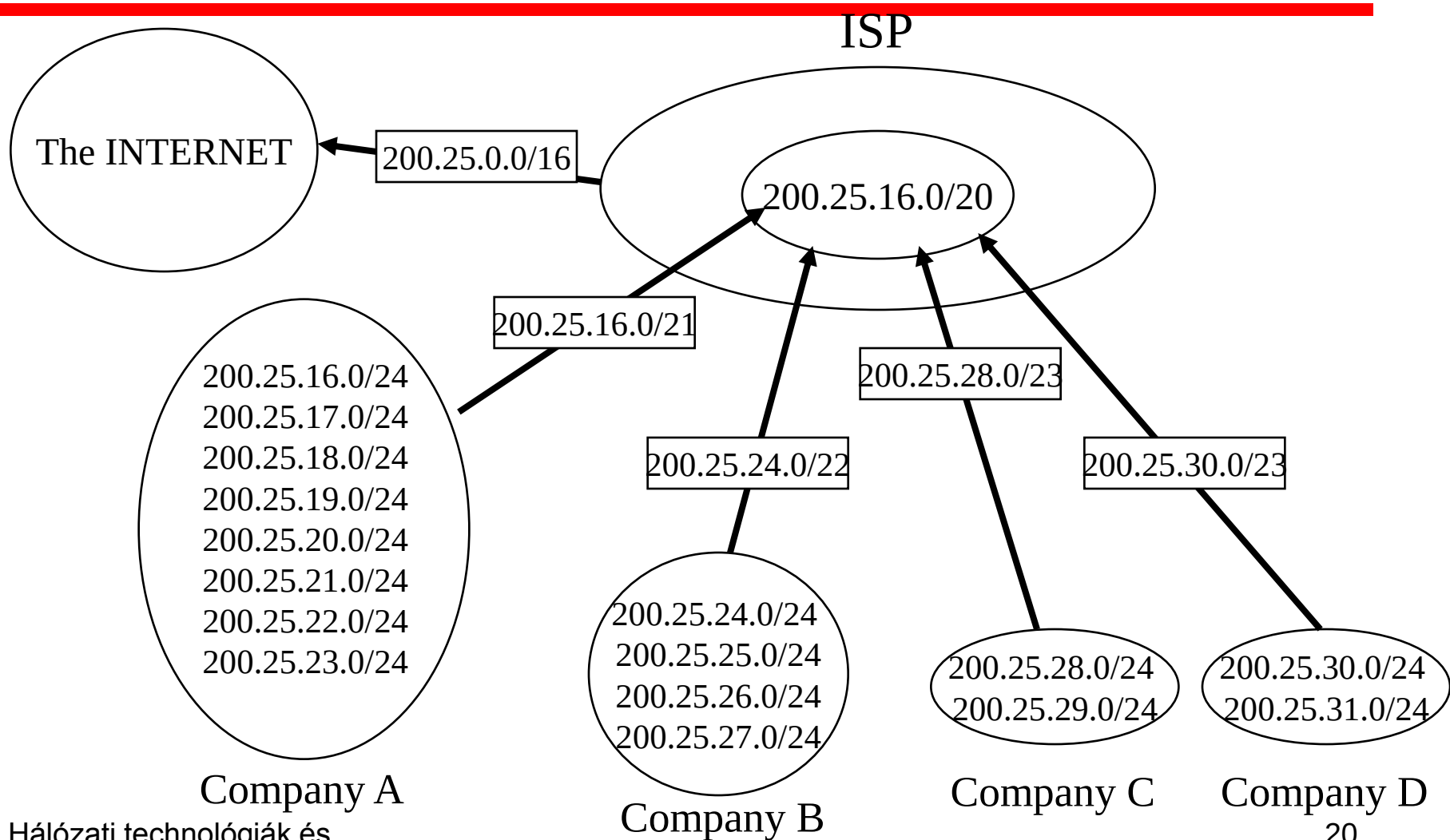
CIDR - Classless InterDomain Routing

- Hatékonyabb kihasználása a címtartománynak
- Szabványosítás – RFC 1518, RFC 1519 (1993)
 - Aktuális verzió – RFC 4632 (2006)
- A hagyományos („classful”) modellben csak 8, 16, vagy 24 bites NET ID
 - Rossz kihasználtság
 - Sok kis Class-C hálózat, melyek magukat hirdetik
 - A topológia különböző pontjain, nehezen aggregálható routing bejegyzések
- **CIDR ötlet:** változó hosszúságú hálózati prefixek (Net ID)
 - CIDR címezés: *A.B.C.D/N*
 - N a prefix hossza bitekben

CIDR címkiosztás

- **IANA – Internet Assigned Numbers Authority**
 - Az RIR-eknek oszt ki rövid prefixes CIDR blokkokat
 - Regional Internet Registries
 - Pl. 62.0.0.0/8 a RIPE NCC-nek
 - **Réseaux IP Européens Network Coordination Centre**
 - Az európai RIR
- Részekre bontva továbbosztják a blokkokat
 - Nagy ISP-k nagyobb szeleteket kapnak, amit továbboszthatnak
 - Tetszőleges méretűek a továbbosztások
 - **Supernet**: összefüggő IP címtartományok egy útválasztási csoportba sorolhatók
 - Sokkal kevesebb routing bejegyzés az aggregáció miatt

CIDR példa



IP címzési példa

- Az IP címek kiosztása: hierarchikusan
- Az alhálózatok méretét a netmask adja
- Példa:
 - BME hálózat
 - IP címtartomány: 152.66.x.x : 255.255.0.0
 - TMIT egyik alhálózata:
 - IP címtartomány: 152.66.244.x : 255.255.255.0
- 255.255.255.0 → 24 bites netmask : C osztály

IP címzés 26-os netmask

- 26 hálózatok – 26 1-es a netmaskban
 - Netmask: 255.255.255.192
 - 4 alhálózat:
 - x.x.x.0-63
 - x.x.x.64-127
 - x.x.x.128-191
 - x.x.x.192-255
- A netmask és az IP cím megadja hogy melyik hálózatban van a gép

IPv4 címkiosztás

MAP OF THE INTERNET
THE IPV4 SPACE, 2006



THIS CHART SHOWS THE IP ADDRESS SPACE ON A PLANE USING A FRACTAL MAPPING WHICH PRESERVES GROUPING--ANY CONSECUTIVE STRING OF IP'S WILL TRANSLATE TO A SINGLE COMPACT, CONTIGUOUS REGION ON THE MAP. EACH OF THE 256 NUMBERED BLOCKS REPRESENTS ONE /8 SUBNET (CONTAINING ALL IP'S THAT START WITH THAT NUMBER). THE UPPER LEFT SECTION SHOWS THE BLOCKS SOLD DIRECTLY TO CORPORATIONS AND GOVERNMENTS IN THE 1990'S BEFORE THE RIR'S TOOK OVER ALLOCATION.

0 1 14 15 16 19 →
3 2 13 12 17 18
4 7 8 11
5 6 9 10

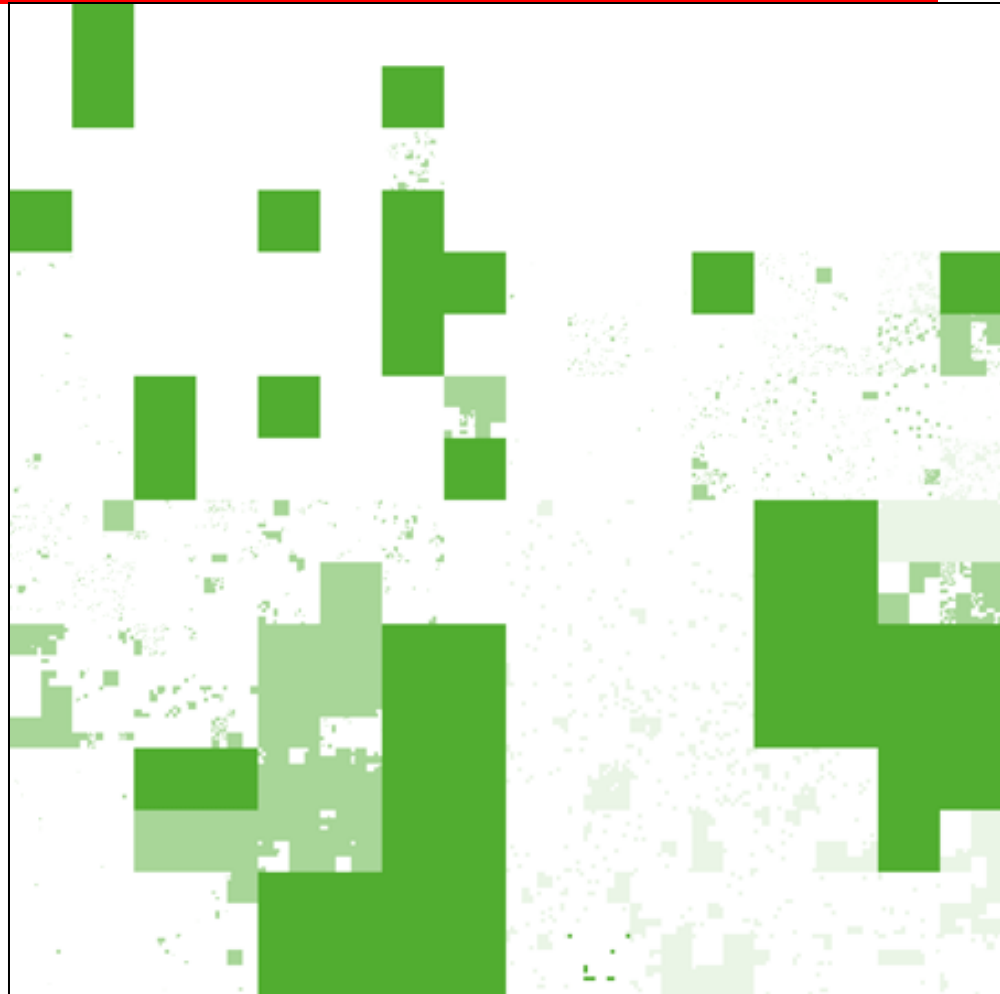


 = UNALLOCATED BLOCK

- Kezdetben intézményeknek
 - Bal felső sarokban
 - HP, Apple, MIT, IBM, Ford, stb.
- Később megjelennek az RIR-ek
 - Regional Internet Registrar

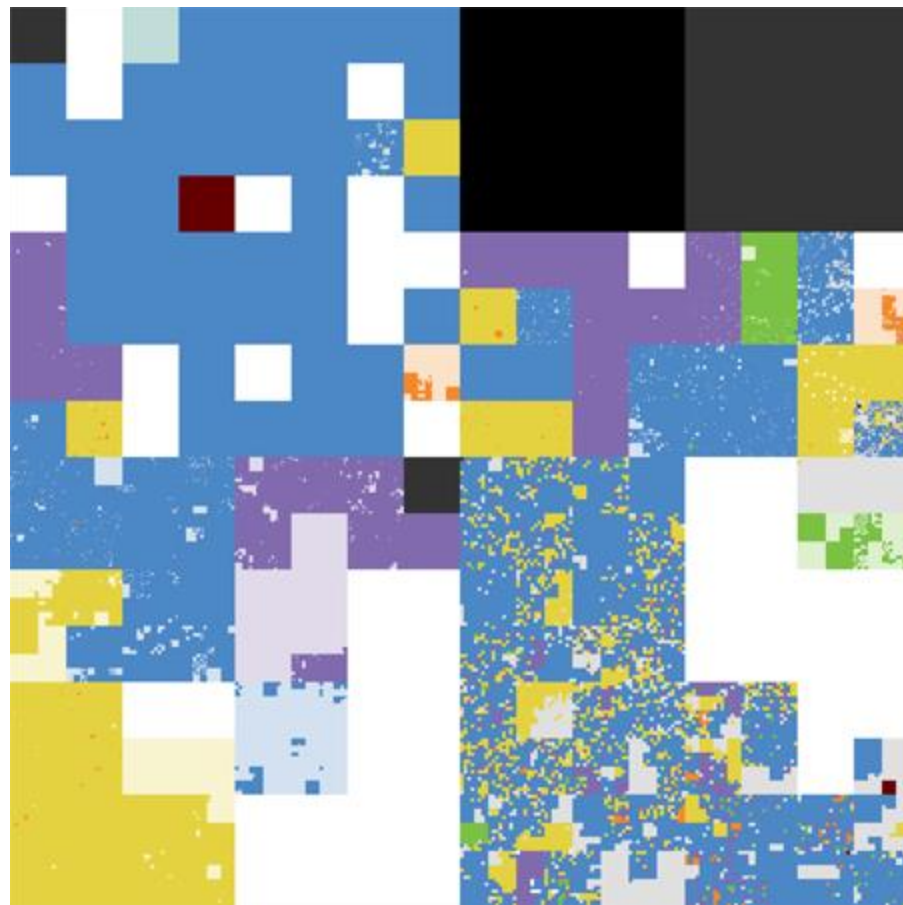
IPv4 címkiosztás (2006)

- Zöld
 - szabadon kiosztható
- Világos zöld
 - RIR-eknek kiosztott, de ők nem osztották tovább
- Fehér
 - elhasznált IP címek
 - kiosztott vagy speciális



IPv4 címkiosztás (2006)

- Kék: ARIN – Észak Amerika
- Sárga: RIPE NCC – Európa
- Lila: APNIC – Asia-Pacific
- Zöld: LACNIC – Latin-Amerika
- Narancs: AfriNIC – Afrika
- Fekete: Multicast
- Szürke: Speciális címek
 - Loopback, privát, class E, stb.
- Fehér: szabad

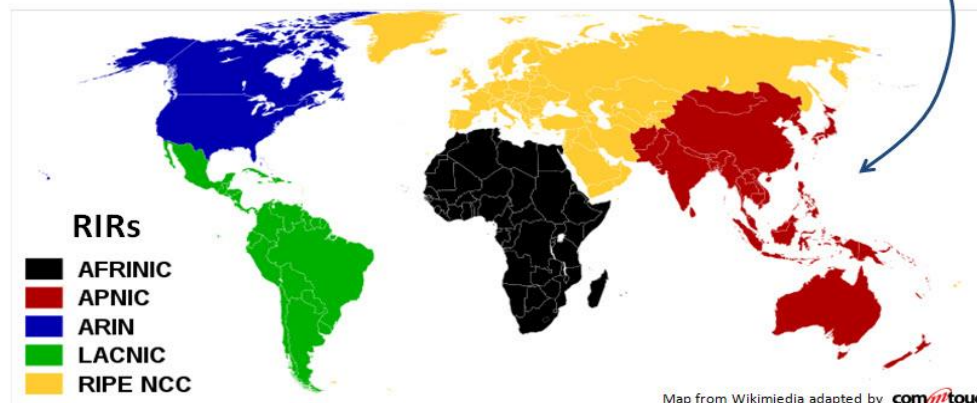
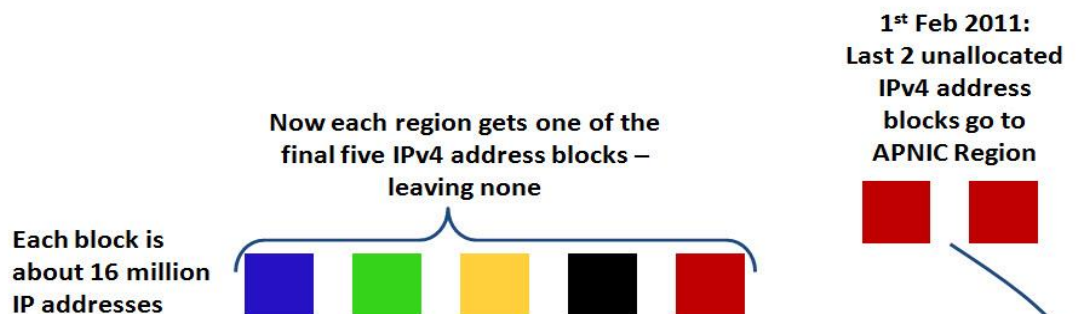


Elfogynak az IPv4-es címek?

- Amerikában nincs gond
 - „Internet Heaven”
- Mindenhol máshol komoly gond
 - Kínában kértek címeket 60.000 iskola bekötésére, kaptak egy Class B címet (65.534 cím)
 - Sok európai és afrikai országnak Class C címe (254 cím)
- Fejlődő Internet Észak-Amerikán kívül
 - Ázsia (2.5 milliárd ember), Kelet-Európa (250 millió), Afrika (800 millió), Dél- és Közép-Amerika (500 millió)
- Új kommunikációs eszközök melyeknek IP cím kell
 - Mobil telefonok, PDA-k, szenzorok, hűtő szekrények, stb.
- Mindig a jelenlegi év végére jósolják, hogy elfogynak a címek

„Betelt” az Internet?

- 2011. február 1.-én kiosztották az utolsó /8-as IPv4-es címblokkokat az RIR-eknek
- Az RIR-ek várhatóan 2011. végén osztják ki az utolsó címeiket



ARP

- Minden Ethernet kártyának saját MAC címe van
- A címzéshez meg kell tudni a gép MAC címét - az ARP végzi ezt
- Az ARP üzenetváltás:
 - Req: Who-has 152.66.244.102? Tell 2f:34:35:67:67:8d
 - Ans: 152.66.244.102 is 37:66:f3:d4:2b:8e
- Ethernet szintű broadcastot használ
- ARP cache

Útvonalválasztás

- Statikus
 - Default gateway
 - Statikus útvonalak
- Dinamikus
 - RIP
 - OSPF
 - ISIS
 - BGP4
- Útvonalválasztó tábla
 - Cím/netmask, next hop, cost

Dinamikus útvonalválasztás

- A legkisebb költségű útvonalat keresik
 - SPF: a legrövidebb útvonal
- Szempontok:
 - Csomópontok száma
 - Link cost (beállítható)
- RIP: legegyszerűbb, periodikus útvonaltábla csere
- OSPF: több area kezelése
 - Link state adatbázis, frissítő üzenetek

ICMP

- Üzenetek hálózati hibák, nem várt eseményekről való értesítésre
- Több típus, pl:
 - ICMP Echo
 - Destination Unreachable
 - Network unreachable
 - Host unreachable
 - Port unreachable
 - Protocol unreachable

IP beállítások

- Alapvető beállítások egy gépen
 - IP cím/netmask
 - Default gateway
 - DNS szerver
- A beállítások kiegészíthetők
 - Alapértelmezett domain név megadása
 - Több DNS szerver

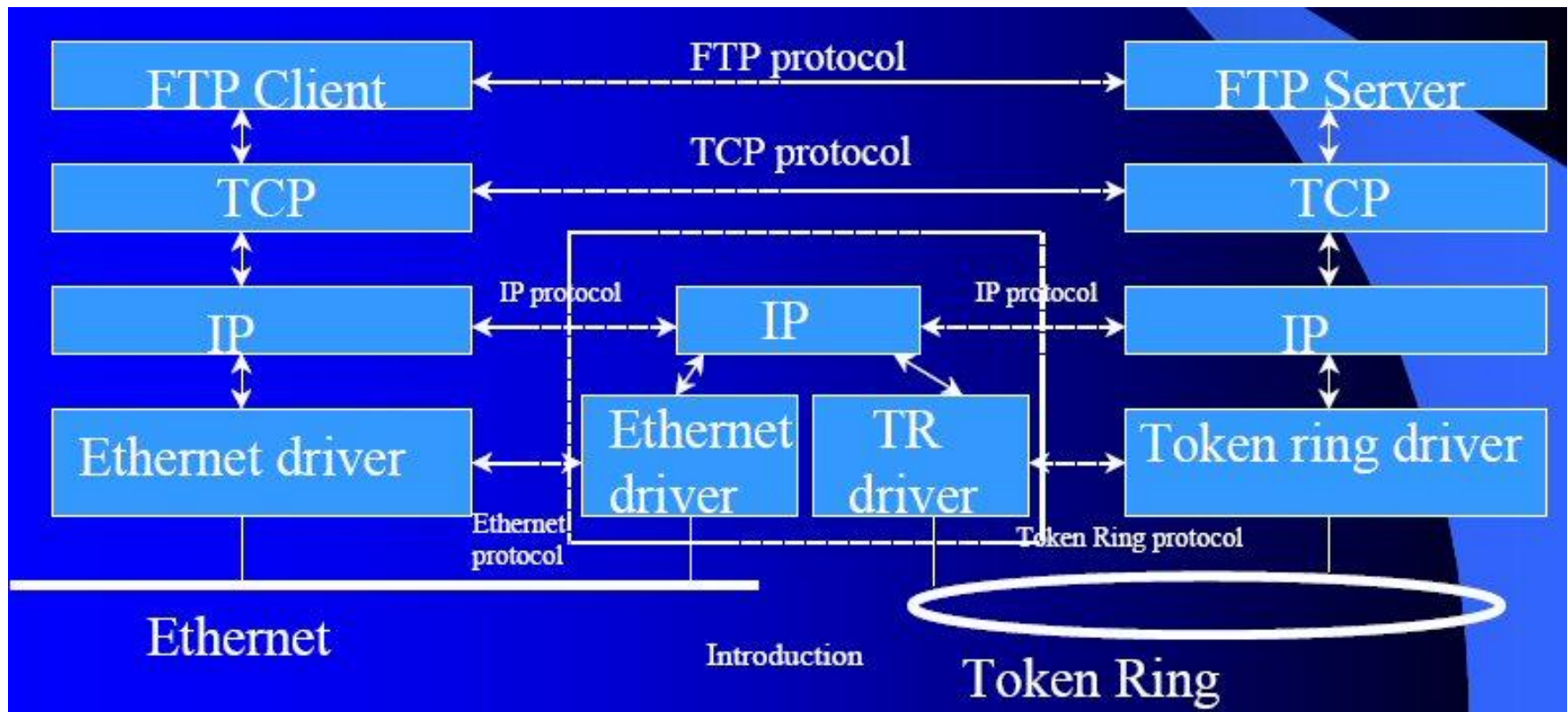
DHCP

- Dynamic Host Control Protocol
- Lehetővé teszi hogy egy gép IP címet kérjen a hálózattól
- Megadhatja a többi hálózati paramétert is:
 - Átjáró, DNS szerver
- Az IP nincs géphez rendelve, változhat minden kéreşor
 - MAC címhez rendelhető azonban

További segédanyagok

- TCP/IP illustrated

Útvonalválasztó



IP címzés - gyakorlat

Moldován István

moldovan@tmit.bme.hu



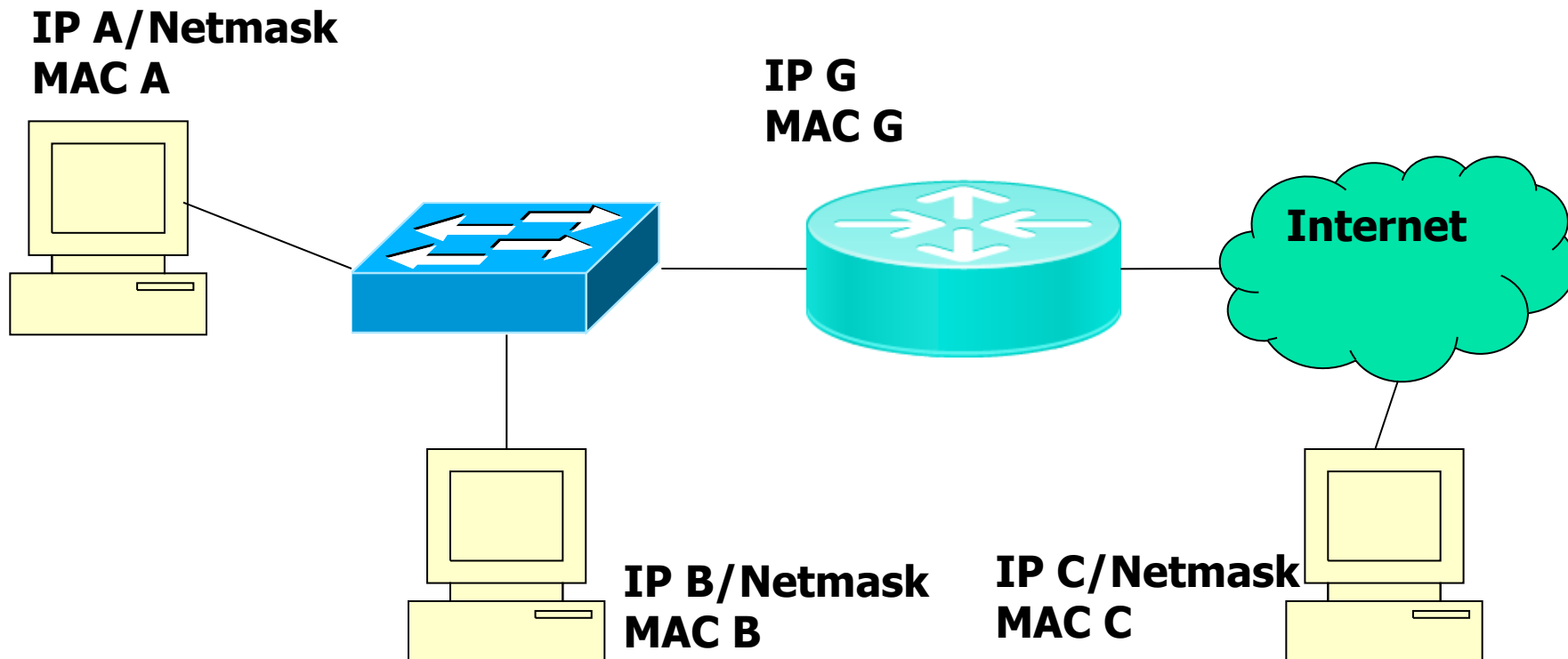
BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
TÁVKÖZLÉSI ÉS MÉDIAINFORMATIKAI TANSZÉK

- IP címzés (subnet)
 - 2 subnet 1 if?
 - Vlan interfesz routing
- L2 és L3 VPN
- QoS
 - Linux PC

IP+Ethernet működés



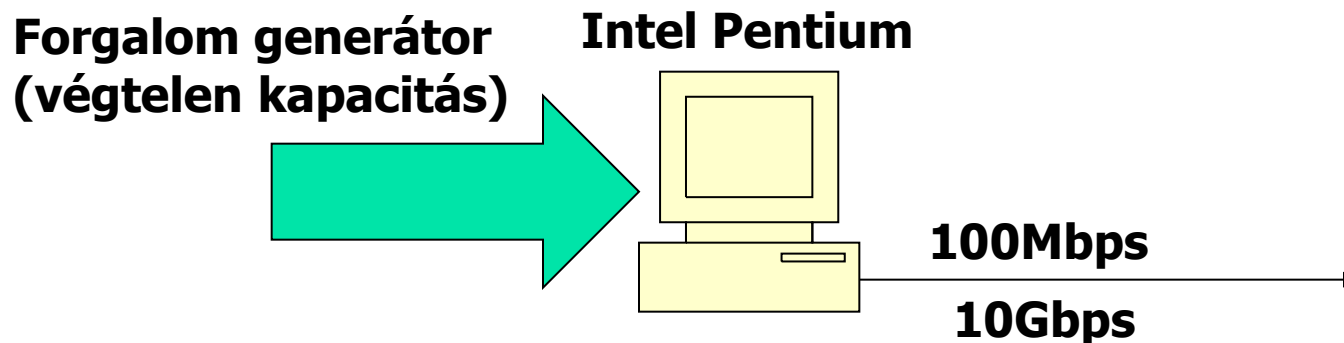
BME-TMIT



1. A->B – ugyanaz a subnet, cél MAC: B
2. A->C – nem lokális háló, GW-en keresztül – cél MAC: G

MAC cím feloldása: ARP

- Szűk keresztmetszet kialakulása
 - Ethernet interfész
 - Processzor

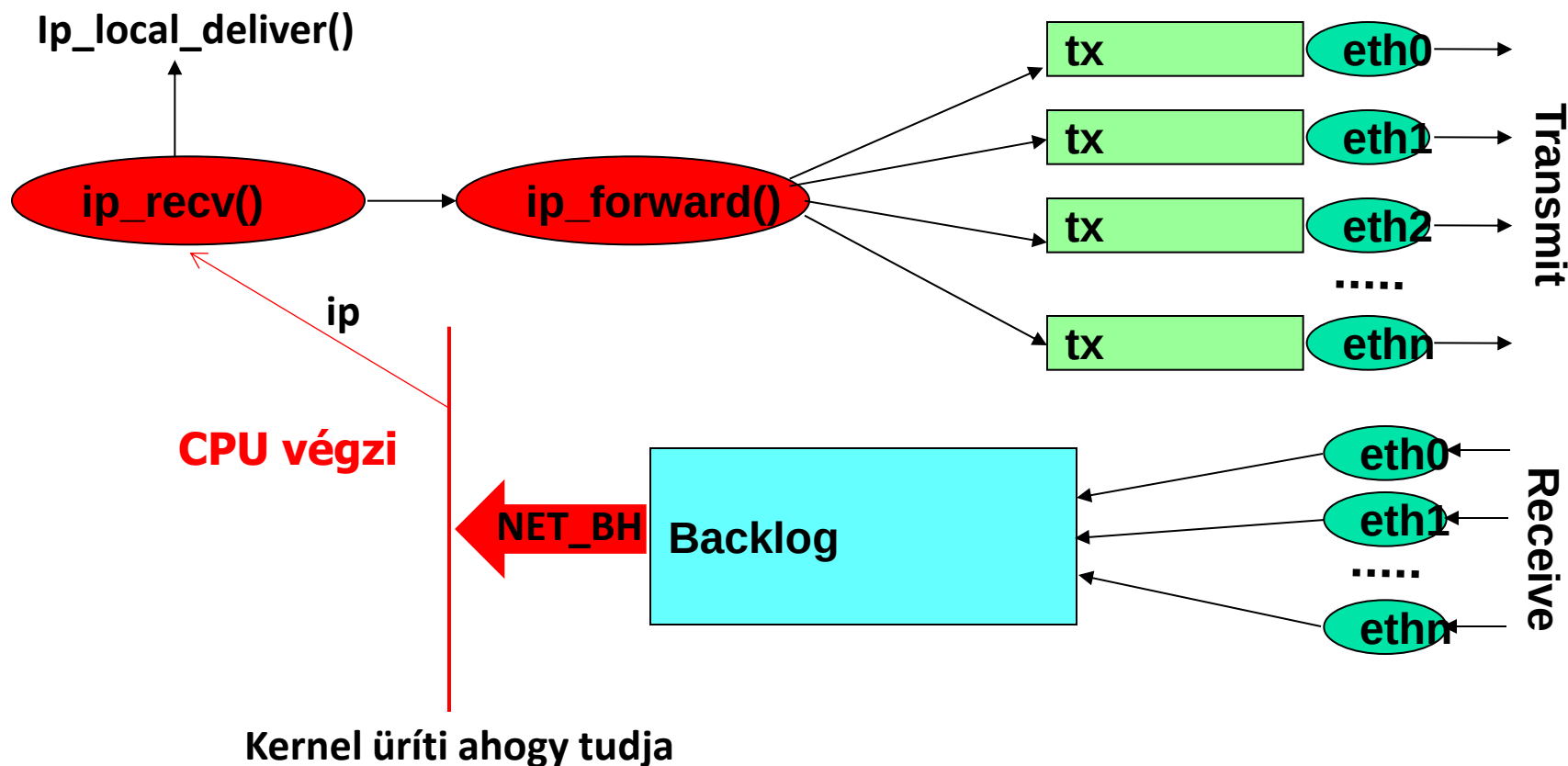


- 100Mbps – a PC bírja, a korlát az interfész
- > 10Gbps – a PC nem képes meghajtani az interfészt maximális sebességgel

Hol alakul ki szűk keresztmetszet - Linux



BME-TMIT

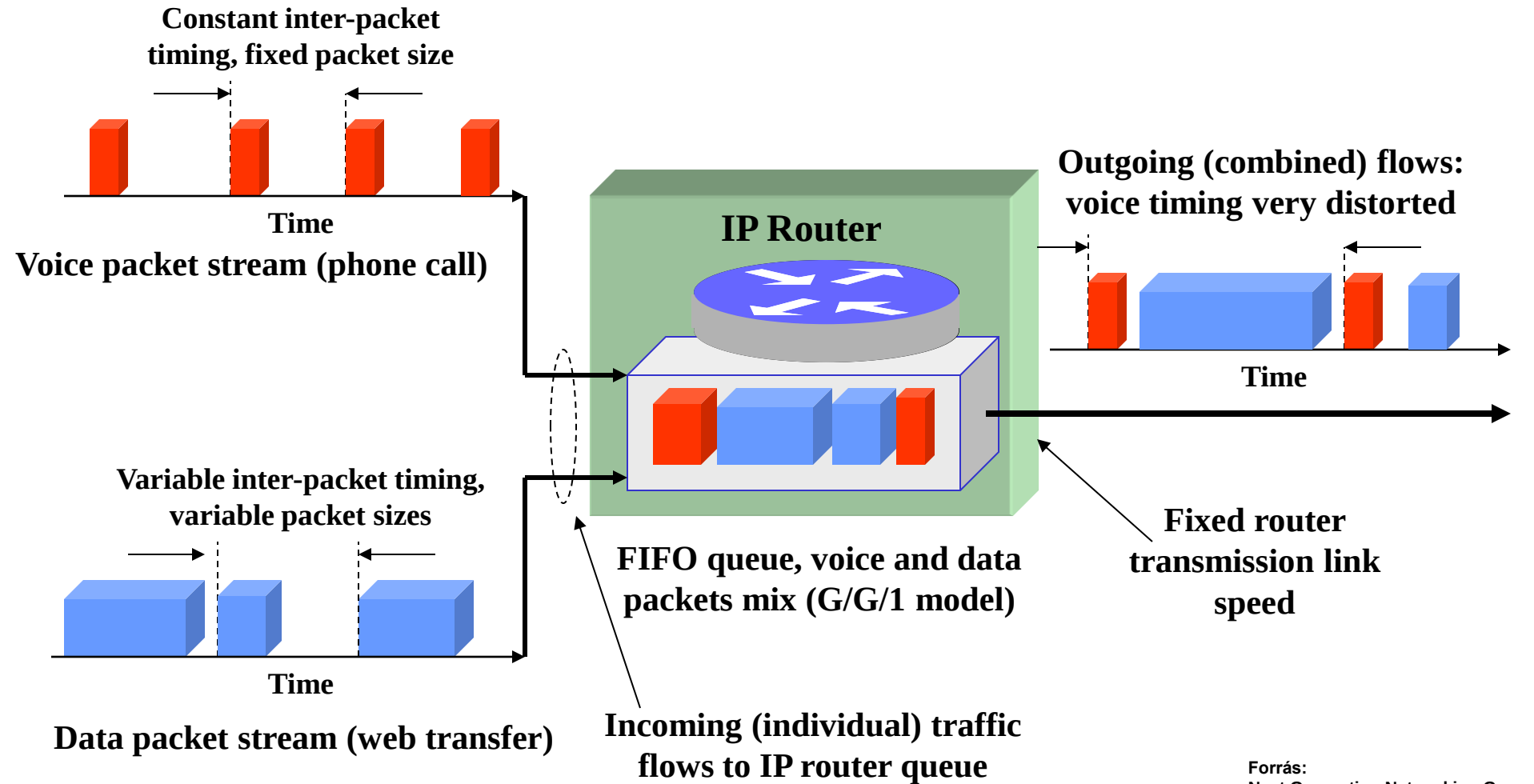


Szűk keresztmetszet:

Interfésznél: tx sorok

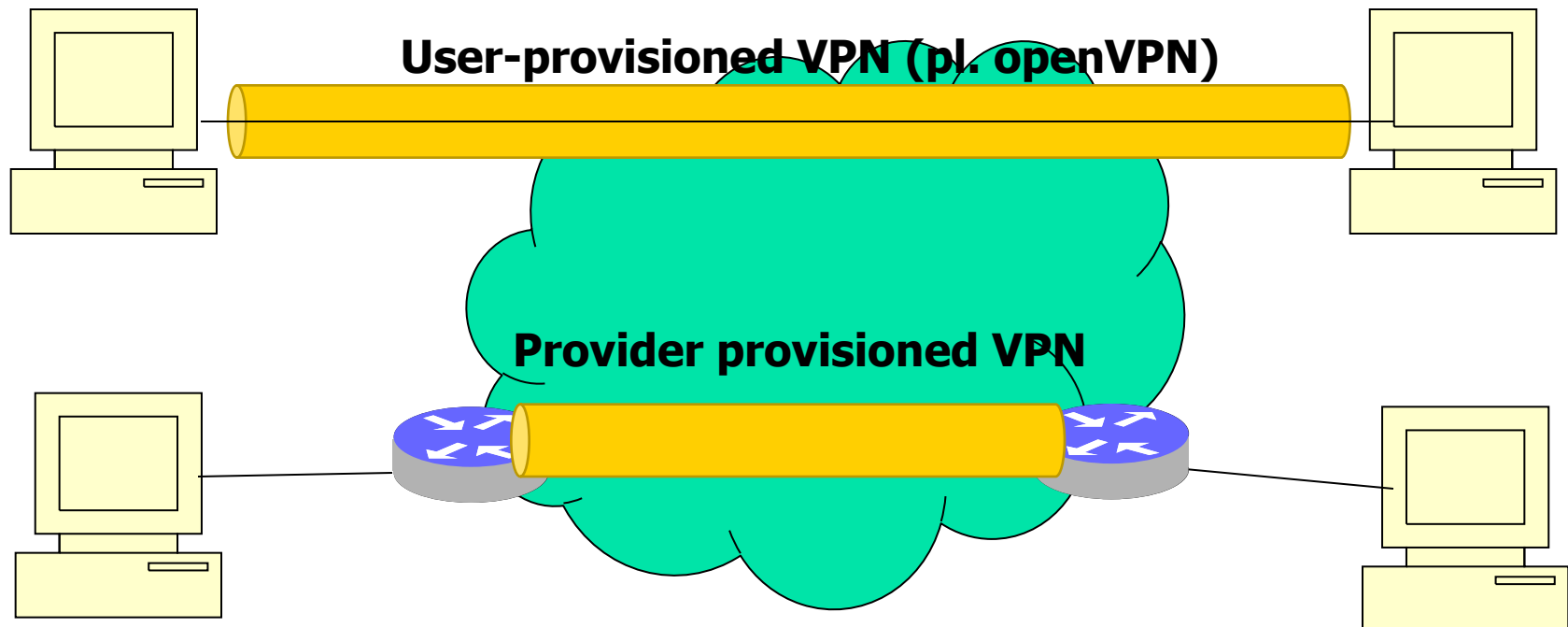
CPU-nál: backlog

„Best Effort” - aggregációs mechanizmus



- Virtual Private Network (VPN)
- Két alapvető típus
 - User-space VPN
 - Provider Provisioned VPN (ppvpn)
- Mindkettő a hálózat erőforrásainak költséghatékony kihasználását célozza

- L2 és L3 VPN példa



Köszönöm a figyelmet!

Hálózati Technológiák és Alkalmazások

Vida Rolland
Moldován István
BME TMIT

2016. október 21.



Routing - Router



- **Routing (útválasztás)**
 - Folyamat, mely során a hálózati protollok csomagjai a célállomáshoz jutnak
 - A routing tábla és a megvalósított protollok szerint a routerek meghatározzák a beérkező csomagok útvonalát
- **Router (útválasztó)**
 - Útválasztást végző csomópont
 - Egymással kommunikálnak
 - A szomszédoktól szerzett információkat gyűjtik és tárolják
 - **Útválasztó táblákat** hoznak létre és tartanak karban
 - Tartalmuk: <célcím, kimenő interfész> párok



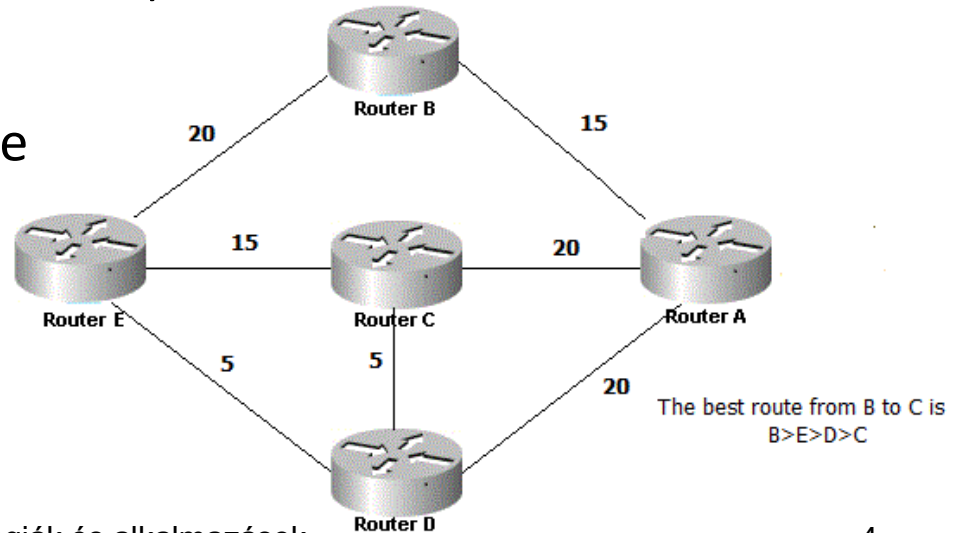
Router

- Router lehet
 - operációs rendszer routing modulja
 - dedikált eszköz (nem csak szoftver, hanem hardver támogatottsággal is rendelkezik) - gyorsabb
 - Cisco, Juniper, Alcatel-Lucent, HP, Huawei, NEC, etc.
- Router kapacitása
 - hány csomagot képes továbbítani időegység alatt (packet/s - PPS)
 - Pl. HP 8800 router – 864 Mpps (2012)



Routerek feladata

- Az optimális útvonal kiválasztása az adott csomag számára
- Alábbi szempontok (metrikák) szerint:
 - az út hossza (hány linken vezet át)
 - költség
 - az adott útvonal terheltsége
 - sávszélesség
 - megbízhatóság
 - Késleltetés



IP Routing – útvonal keresés

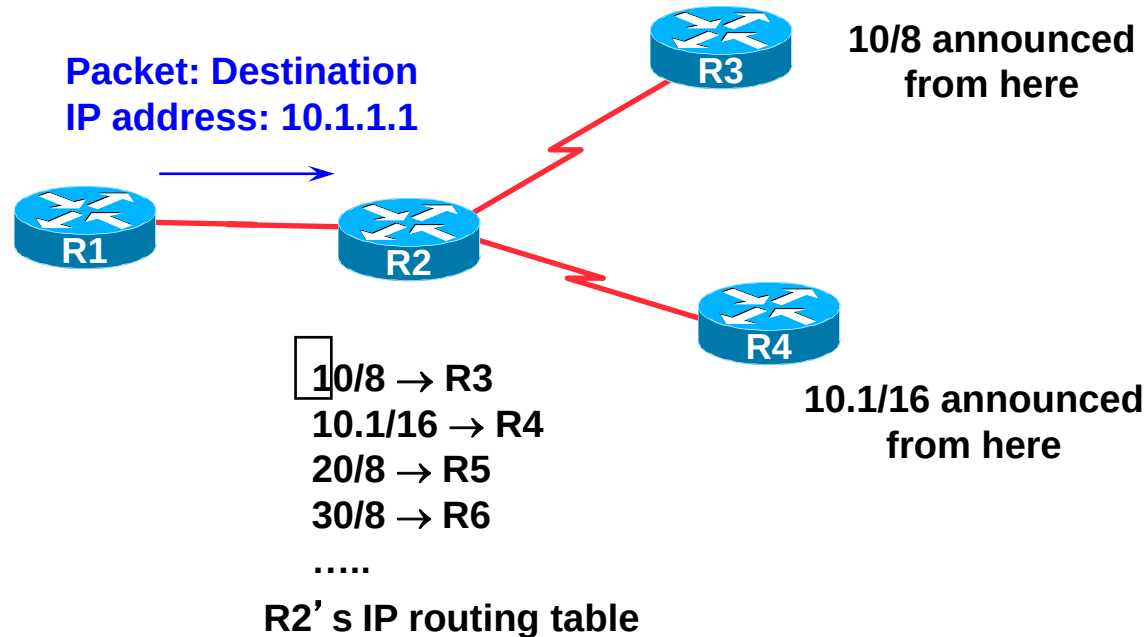
- Routing protokoll által gyűjtött információk alapján
- Több alternatív útvonal lehet
 - A legjobb útvonal tárolva a **forwarding** táblában
- A döntéseket periodikusan frissíti, vagy ha változik a topológia (event driven)
- Döntések:
 - topológia, szabályok és metrikák (hop count, filtering, delay, bandwidth, etc.)

IP route „lookup”-a legjobb útvonal kiválasztása

- Cél IP cím alapján történik
- “longest match” routing
 - A pontosabb (hosszabb prefix-el rendelkező) útvonalat választja
 - **Példa:** ha a cél IP 10.1.1.1/32, akkor inkább annak a routernek küldjük aki 10.1/16 prefix-el jelentette mint annak aki 10/8-al.

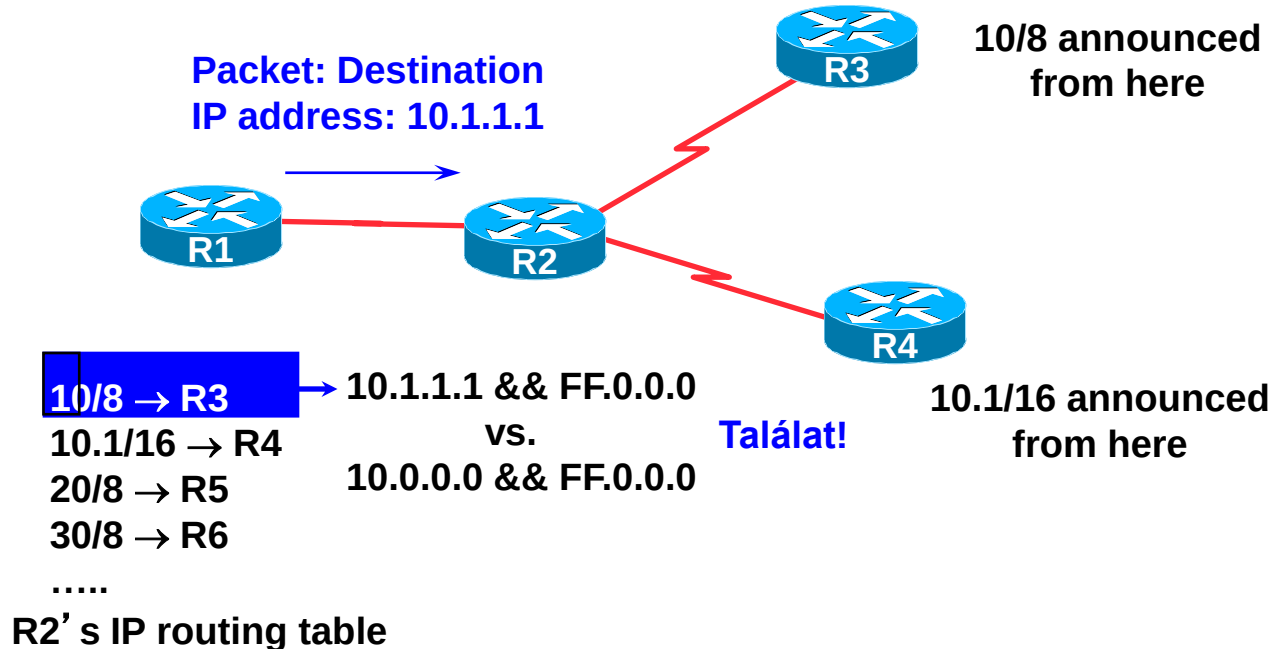
IP route lookup

- Cél IP cím alapján



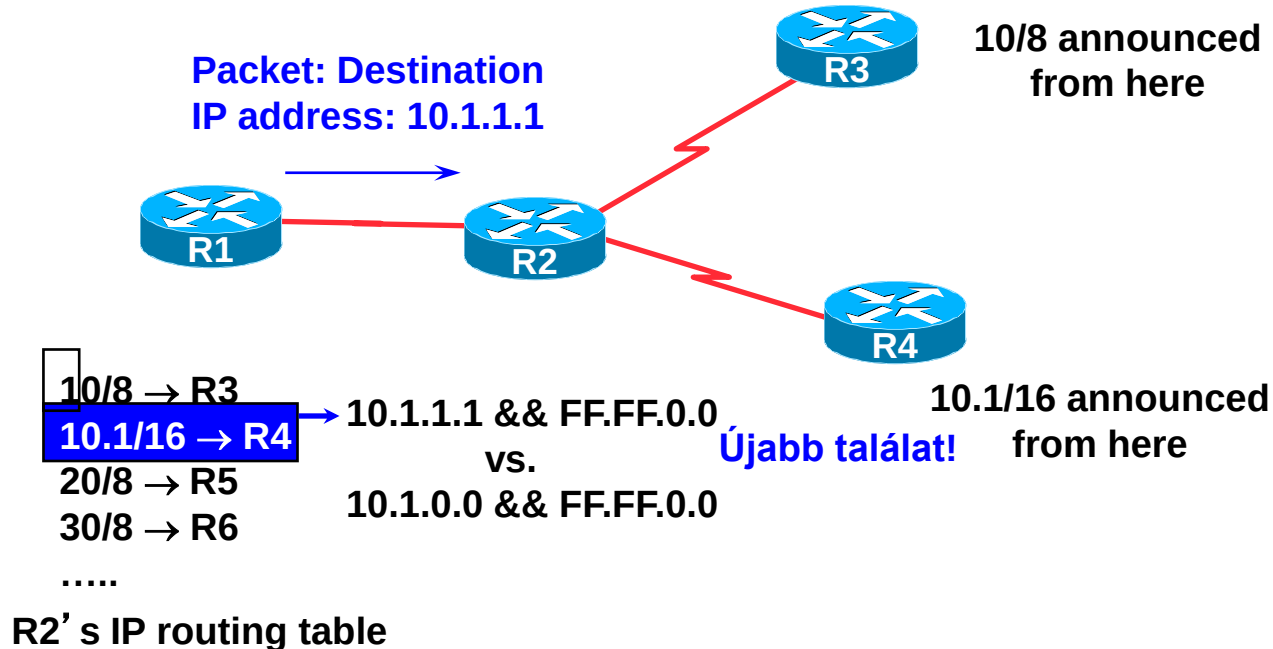
„leghosszabb találat” routing

- Cél IP cím alapján



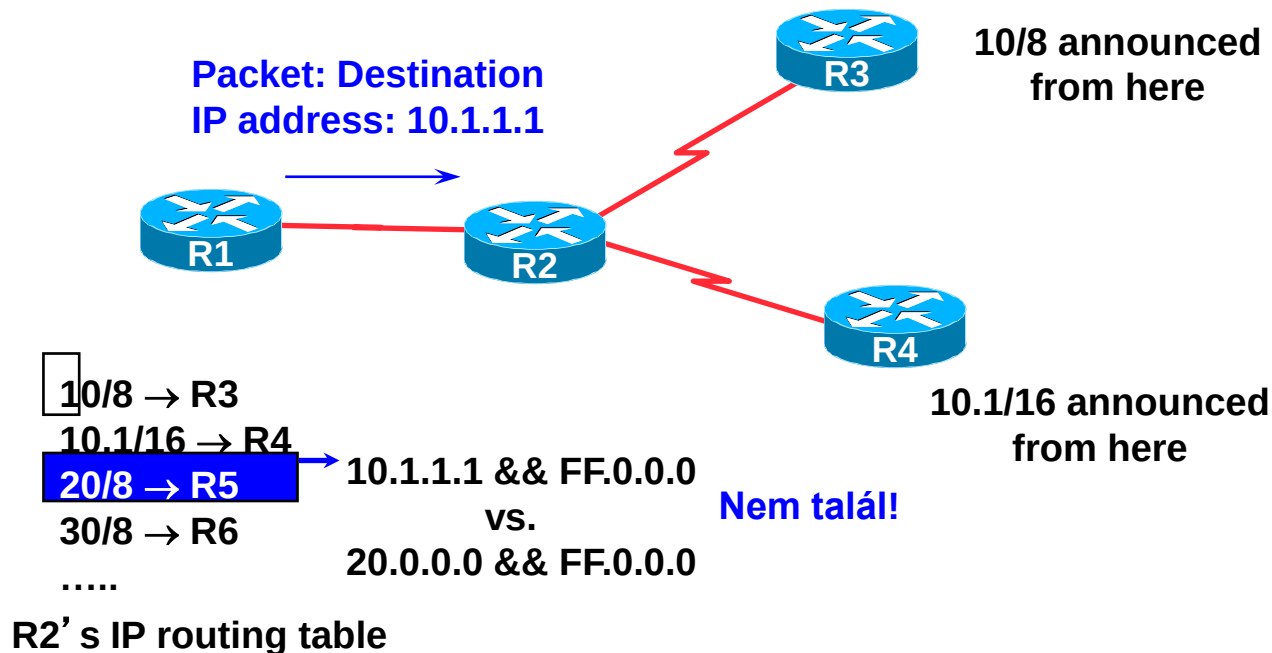
„leghosszabb találat” routing

- Cél IP cím alapján



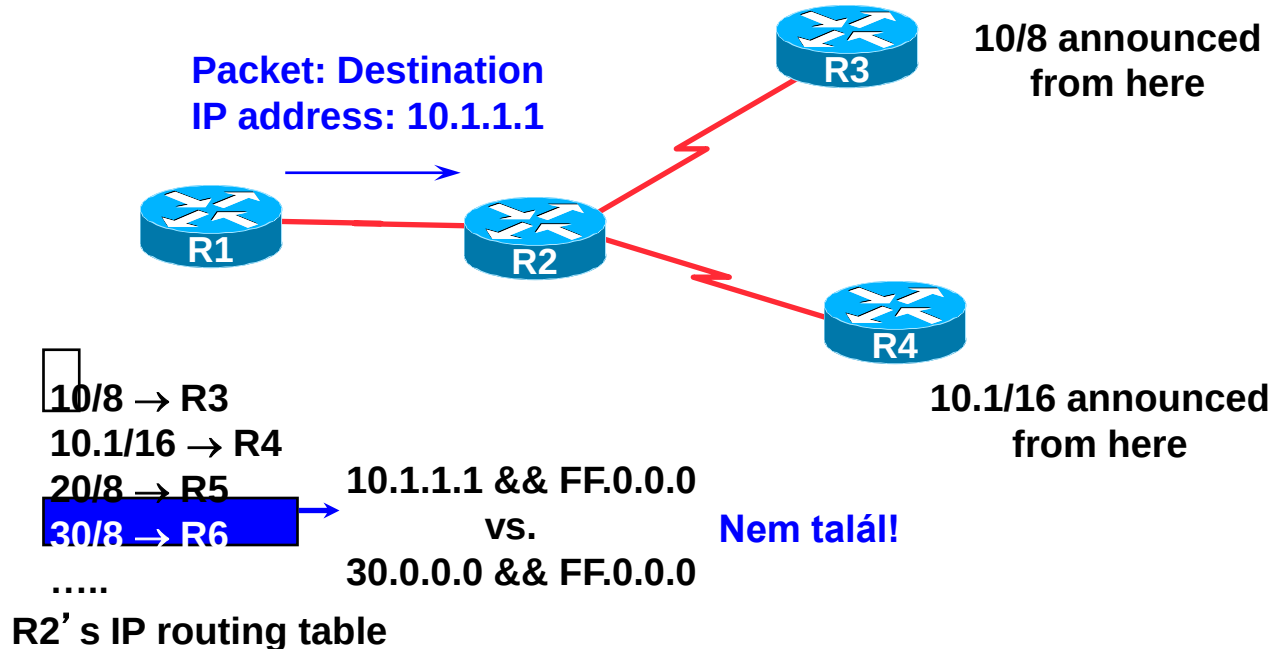
„leghosszabb találat” routing

- Cél IP cím alapján



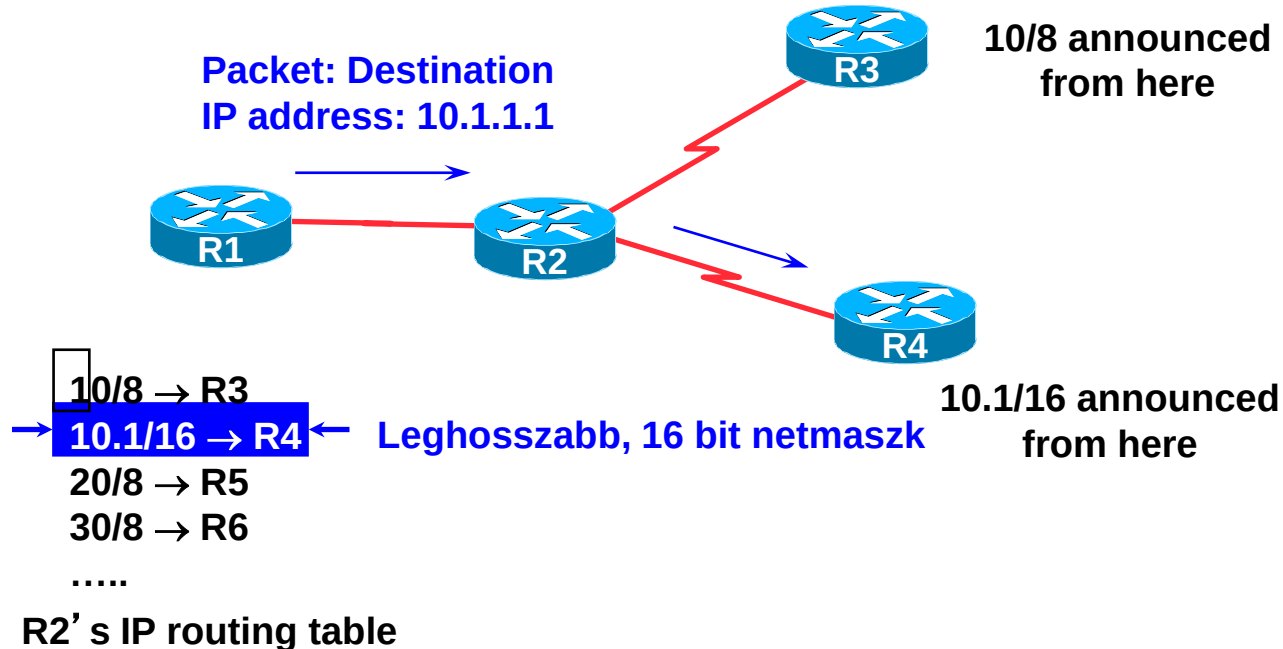
„leghosszabb találat” routing

- Cél IP cím alapján

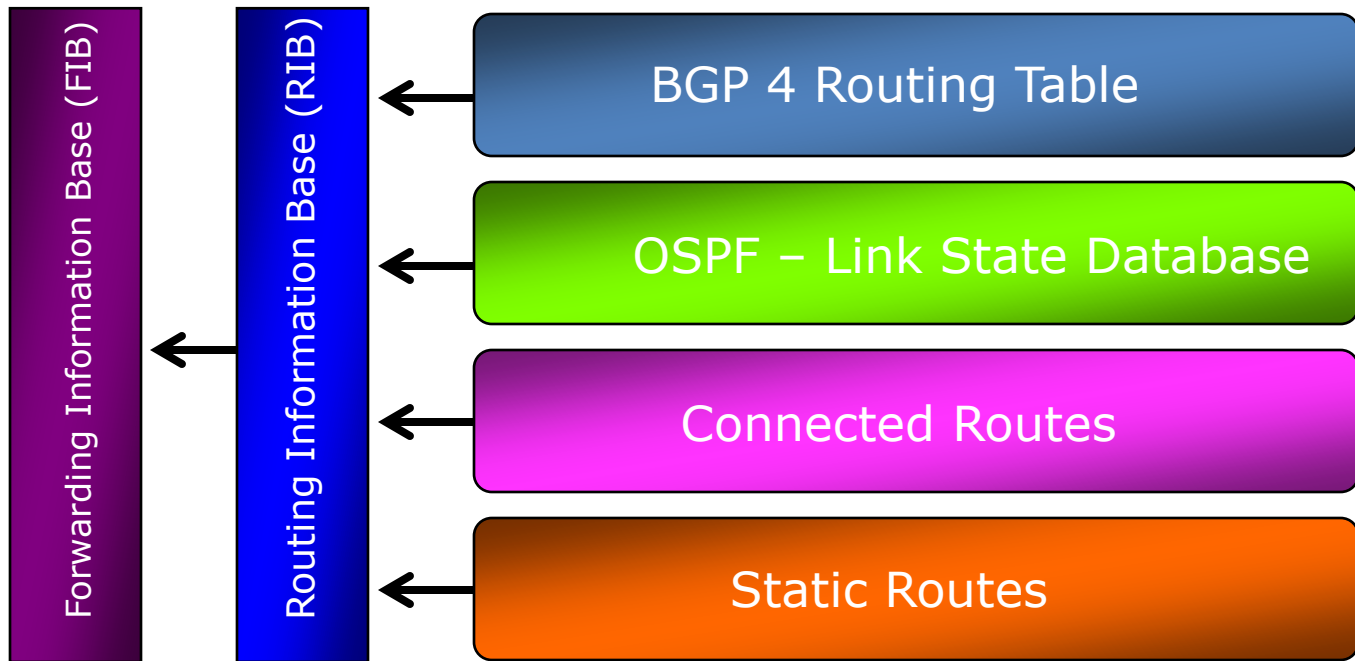


„leghosszabb találat” routing

- Cél IP cím alapján



Routing Táblák alapján tölti fel a továbbító táblákat



RIB és FIB

- FIB = Forwarding Table
 - Interfészek, lokális hálózatok
 - Segít a routernek eldönteni hová küldje a csomagot
 - Néha nem különböztetik meg a routing táblától
 - Gyors (gyakran hardveres) keresés
- RIB = Routing Table
 - Cél alhálózat címek és a hozzájuk tartozó next hop bejegyzések – egyéb információkkal
 - Egy cél – több next hop is lehet
 - Csak a legjobb kerül a FIB-be

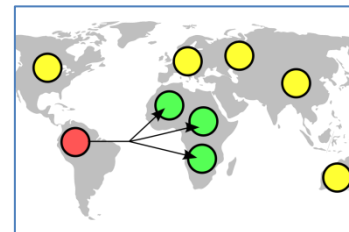
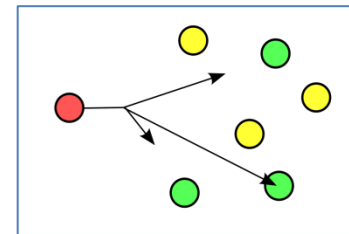
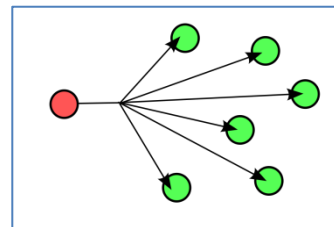
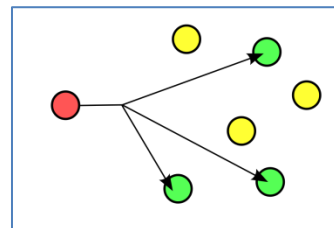
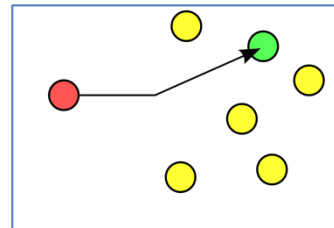
Egy host routing táblája

Network Address	Netmask	Gateway Address Gateway	Interface IP Address	Metric	Purpose
0.0.0.0	0.0.0.0	157.55.16.1	157.55.27.90	1	Default Route
127.0.0.0	255.0.0.0	127.0.0.1	127.0.0.1	1	Loopback Network
157.55.16.0	255.255.240.0	157.55.27.90	157.55.27.90	1	Directly Attached Network
157.55.27.90	255.255.255.255	127.0.0.1	127.0.0.1	1	Local Host
157.55.255.255	255.255.255.255	157.55.27.90	157.55.27.90	1	Network Broadcast
224.0.0.0	224.0.0.0	157.55.27.90	157.55.27.90	1	Multicast Address
255.255.255.255	255.255.255.255	157.55.27.90	157.55.27.90	1	Limited Broadcast

Subnet Mask

Útválasztási szemantika

- **Unicast** – csomag küldése egy adott célcsomópontnak
- **Anycast** – csomag küldése bárkinek (pl. a legközelebbi csomópontnak) egy adott csoportból
- **Multicast** – csomag küldése egy csoportnak
- **Geocast** – csomag küldése egy adott földrajzi területre
- **Broadcast** – csomag küldése minden csomópontnak a hálózatban



Az Internet felépítése

- ISP-k, nagyobb hálózati entitások
 - Autonomus System-ek
- Összekapcsolásuk:
 - „Carrier-ek” – nagybani szolgáltatók: globális gerinc
 - Pl. Telia Carrier
 - „Peering” – ISP-k között
 - Példa: BIX
- Szabályok – Policy-k
 - Peering forgalomra

Routing működés

- Az ISP-k lehetnének közvetlen kapcsolatban
 - Nem skálázható
- Inkább az elérhetőségüket tudatják a környező routerekkel
 - A lokális IP registry-től kapott IP cím tartomány információkat
 - Aggregáltan
- Ezeket az információkat tovább terjesztik a routerek – „route hirdetés”

Routing Protokollok

- A Routers “routing protokollokat” használnak az információcserére
 - IGP –t használnak a saját hálózatukon belül
 - EGP-t használnak a különböző ISP-k között
 - AS-ek között

Mi is az IGP?

- Interior Gateway Protocol
- egy Autonomous System-en belül
- Belső infrastruktúráról terjeszt információkat
- Gyakrabban használt IGP-k:
 - OSPF
 - ISIS
 - RIP (-R.I.P.)

Minek az IGP? Nem lenne elég csak 1 szint?

- Az ISP-k közötti hálózat skálázhatóságát javítja
 - Hierarchia
 - A hálózati hibák lokálisak maradnak
 - Csak az ISP-k **infrastrukturájának** adatait osztja meg, felhasználói információkat nem
 - **Minimizálja** a hálózati prefixeket hogy növelje a skálázhatóságot (és gyorsabb is lesz)

Mi az EGP?

- Exterior Gateway Protocol
- Autonomous System-ek közötti információcserére szolgál
- Független az IGP-től
 - De felhasználhatja az összegyűjtött információkat
- Jelenleg az EGP szerepét a BGP protokoll játssza

Routing protokollok osztályozása

- **Statikus:**
 - a routing tábla manuális kitöltése
 - automatikusan soha nem frissítődik
- **Dinamikus:**
 - a routerek egymás között kommunikálva a hálózat topológiájának megfelelően állítják elő az útvonalválasztó táblát
- **Egyutas:**
 - minden célpont felé csak egy utat tárol
- **Többutas:**
 - minden célpont felé több (esetleg minden) utat tárol.
 - Ezek a protokollok képesek load balancing-ra (terhelés megosztás)

Routing protokollok osztályozása

- **Lapos (flat):**
 - minden router minden célpontról tud
 - Régebben (kisebb hálózatok)
- **Hierarchikus:**
 - a router-ek nem minden célpont felé ismerik az utat
 - egy ismeretlen címzettnek szánt csomagot egy előre meghatározott irányba (default route) küldenek
 - ez routing információk egy szélesebb körével rendelkezik
 - routing táblák mérete kezelhető marad
- **Intra-domain**
 - valamely területen (domain) belüli útvonalválasztásért felelős
- **Inter-domain**
 - a területek (domain) közötti útvonalválasztásért felelős

Routing protokollok osztályozása

- **Hop-by-hop:**
 - minden router autonóm módon határozza meg a továbbítás irányát
 - ezen elven működő routerek csak olyan utakat hirdetnek (szomszédjaiknak), melyeket maguk is használnak
- **Source routing:**
 - a feladó határozza meg az útvonalat (pl. IP fejléc)
 - a routerek csupán az elérhetőségi információkat terjesztik
 - magukat a csomagokat a csomagba beleírt útvonal szerint kapcsolják
- A két megoldás között léteznek átmenetek

Routing protokollok osztályozása

- **Távolság vektor (distance vector) protokollok**
 - csak a szomszédos routerek kommunikálnak
 - minden router elmondja összes szomszédjának:
 - mekkora költségű utat ismer egy adott célponthoz
 - arról nem szól, hogy az út merre vezet
 - a routerek begyűjtik szomszédaiktól ezeket a hirdetéseket és kiválasztják, hogy ki hirdette a legolcsóbb utat az adott célpontokhoz
 - a megfelelő csomagokat a legkedvezőbb irányba továbbítják
 - saját költségüket a legkedvezőbbekhez hozzáadva ők is hirdetik az adott célponthoz vezető utat

Routing protokollok osztályozása

- **Kapcsolat állapot (link state) protokollok**
 1. feltérképezik a hálózat topológiai gráfját
 2. ebben a gráfban keresik a legrövidebb utat.
 - A routerek egymás között csak saját interfészeik állapotát beszélik meg
 - Kik a szomszédjaim, milyen költségük van a linkeknek közöttünk
 - Ezeket az információkat minden, a hálózatban lévő routerrel kicserélik
 - Ebből építi fel mindenki a saját (de egymással megegyező) topológiai gráfját

Distance-Vector Protokollok

Bellman-Ford protokollok

Klasszikus Bellman-Ford algoritmus

$d_{ij} := i$ - j link költsége (végtelen, ha nincs link)

- Tényleges ár, késleltetés, csomagvesztési ráta, stb.

Következmény: **additivitás**

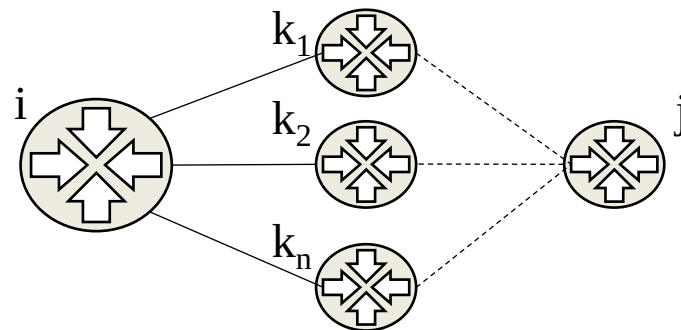
- egy útvonal költsége az azt alkotó linkek költségének összege

$D_{ij} :=$ minimum költség i és j között

Bellman egyenlet:

$D_{ii} = 0$, minden i -re

$D_{ij} = \min_k \{d_{ik} + D_{kj}\}$



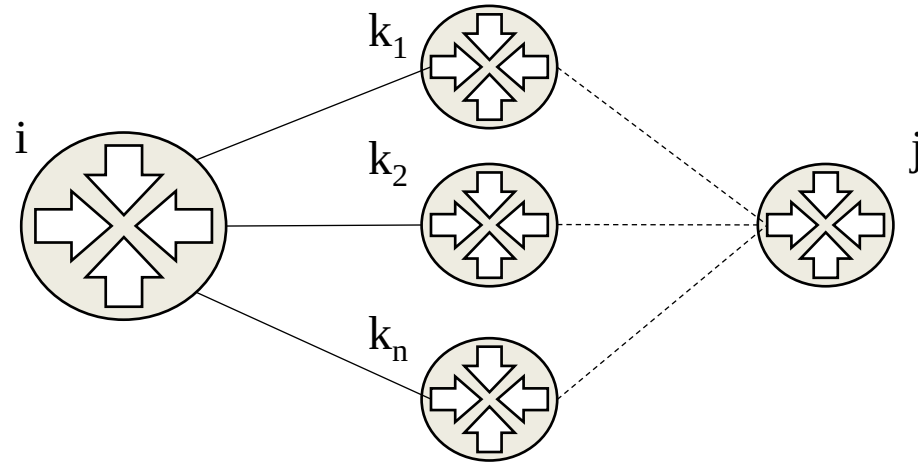
Elosztott Bellman-Ford Algoritmus

$D_{kj}^i(t)$ = minimális távolság k -tól j -ig, melyet i router lát a t időpillanatban

$D_{ii} = 0$, minden i -re

$D_{ij}(t) = \min_k \{d_{ik} + D_{kj}^i(t)\}$

- önállóan működhet az algoritmus a routerekben



Distance-vector protokollok

- RIPv1 (RFC 1058, '88)
 - Routing Information Protocol
 - Rest In Pieces 😊
- RIPv2 (RFC 2453, '98)
- RIPv6 (RFC 2080, '97)
 - IPv6-os verzió
- EIGRP
 - Enhanced Interior Gateway Routing Protocol
 - Cisco proprietary szabvány

Distance Vector protokollok

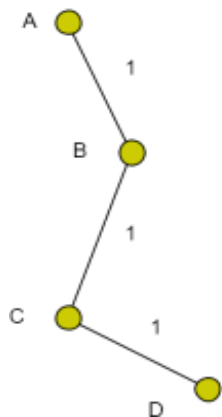
- Távolságvektorokat tárol az útvonalakról
 - Adathármasok:
 - Cél (mi a célállomás)
 - Költség
 - Következő csomópont (merre küldje)
 - Rendszeresen frissítik adataikat a közvetlen szomszédok
 - Frissítő üzenet (2 részből áll):
 - Cél, költség
 - Ha a router ilyenkor jobb utat talál egy célhoz, frissíti tábláját 2 ok miatt:
 - Kisebb költségű utat talál
 - Szomszéd költsége megváltozik

Jellemzők

- Egyszerű, de nem tökéletes:
 - A kapcsolatok ára változhat
 - Kapcsolatok meg is szakadhatnak
 - Egy megszakadt kapcsolat ára végtelen
 - Egy olyan egész érték, mely nagyobb bármilyen lehetséges valós értéknél (RIP-nél jellemzően 16)
 - A routerek topológia változás esetén nem egyszerre frissítik táblázataikat
 - Periodikus időközönként (pl. 30 s) frissítő üzenet
 - Ha 6 frissítés elmarad, az ár végtelen
 - A szomszédok is frissítik a bejegyzéseiket
 - Konvergál, de lassan
 - Csak kis hálózatokban használható

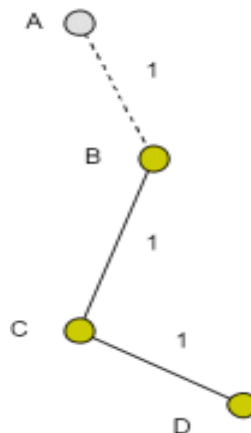
Végtelenig számolás

- A routerek a célcím költségek hirdetésekor végtelenig inkrementálhatnak



B	C	D
1	2	3

Távolság A felé



B	C	D
3	2	3

B	C	D
3	4	3

B	C	D
5	4	5

Megakadályozás

- **Split horizon módszer**
 - ha C B-től megismer egy utat, akkor az azt kiterjesztő utat B-vel már nem közli
- **Poisoned Reverse módszer**
 - Végtelen elérhetőség hirdetése az adott linken elérhető csomópontokhoz
 - ha C B-től megismer egy utat, akkor az azt kiterjesztő utat B-nek végtelen költségű útként jelzi

Egy kis animáció...

- <https://www.youtube.com/watch?v=TWDUllypHq5o>

Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. október 26.

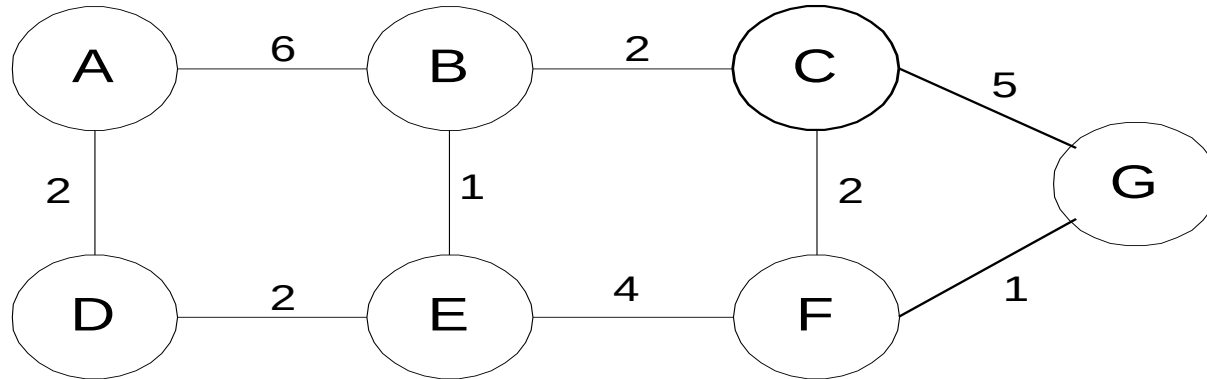


Link-state Protokollok

Link-state protokollok működése

- A link-state protokollok működése 2 részből áll:
 1. minden állomás felderíti a hálózat topológiáját
 - Hálózati topológia leírása a link állapot leíró rekordokban található
 - Link állapot leíró rekordokat kell terjeszteni
 2. a kapott gráfban megkeresi a legrövidebb útvonalat és az ahhoz tartozó első állomást
- **Fontos!**
 - A routerekben lévő topológia megegyezzen
 - Az optimális út kiválasztása ugyanúgy történjen
 - ha A router B felé számolja az optimális utat, B meg A router felé – hurok!

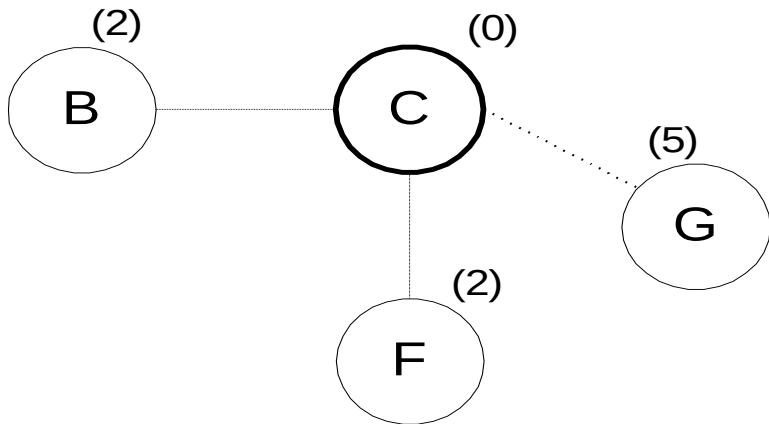
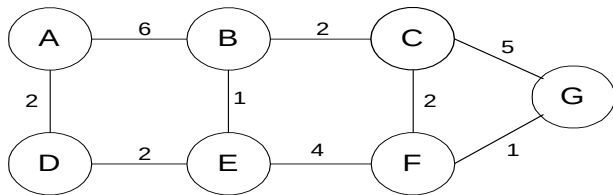
Link State Database



Link state Database						
A	B	C	D	E	F	G
B/6	A/6	B/2	A/2	B/1	C/2	C/5
D/2	C/2	F/2	E/2	D/2	E/4	F/1
	E/1	G/5		F/4	G/1	

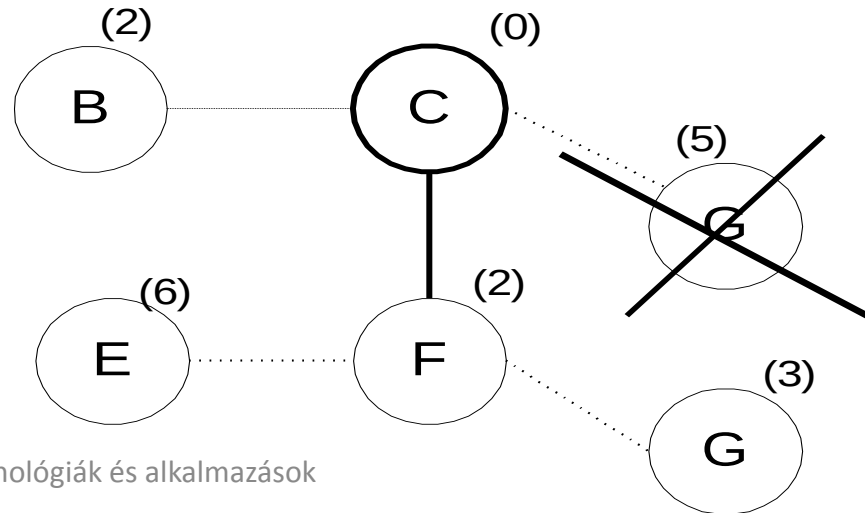
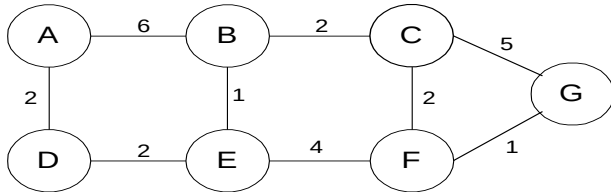
Dijkstra algoritmus

- Az útvonalválasztás Dijkstra algoritmus alapján
 - Legyen C a gyökér
 - Számoljuk ki a szomszédokhoz vezető utak költségét



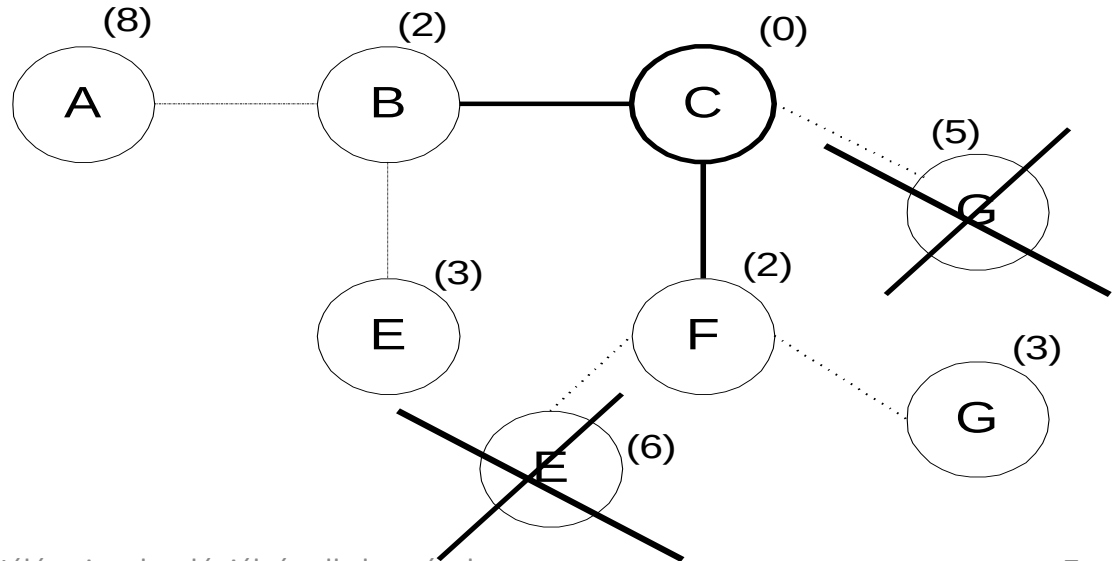
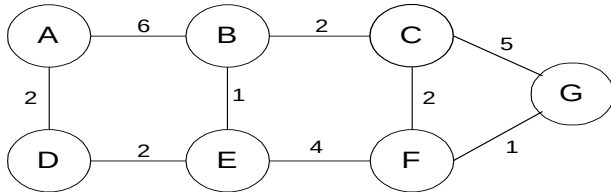
Dijkstra algoritmus 2

- Vegyük be F-et (legkisebb költségű, még nem elemzett csomópont) és számoljuk ki F szomszédjaihoz vezető utak költségét.
- Rövidebb út G-hez F-en keresztül, E megjelenik



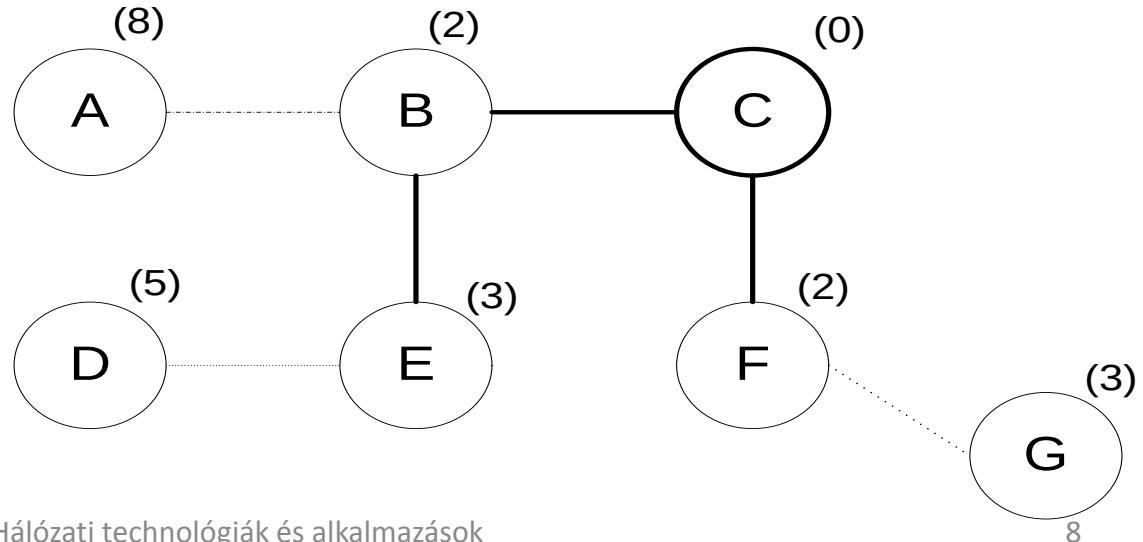
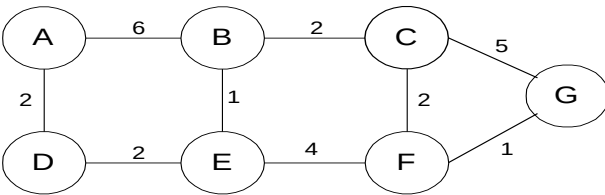
Dijkstra algoritmus 3

- Vegyük be B-t és számoljuk ki B szomszédainak költségét
- Rövidebb út E-hez B-n keresztül, A megjelenik



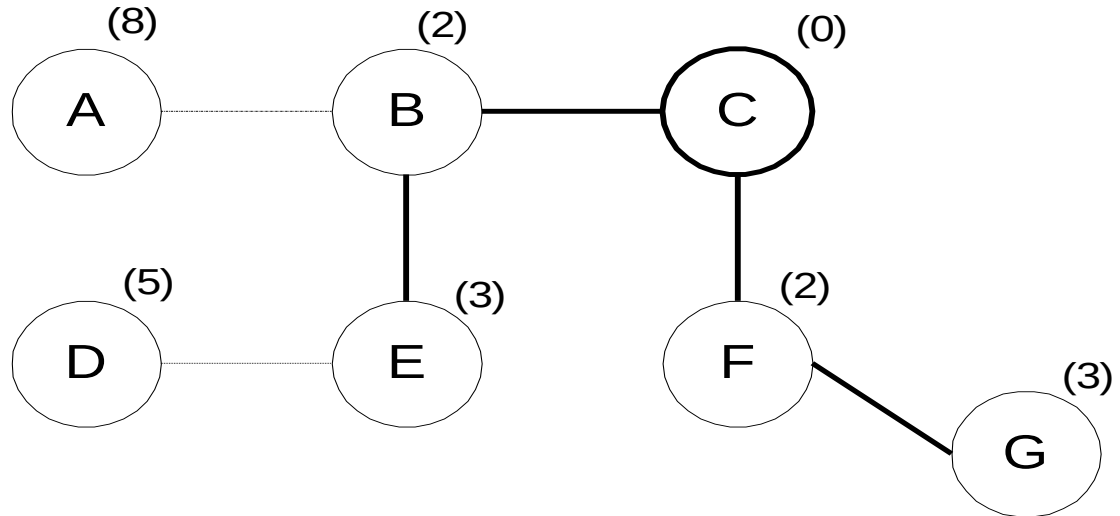
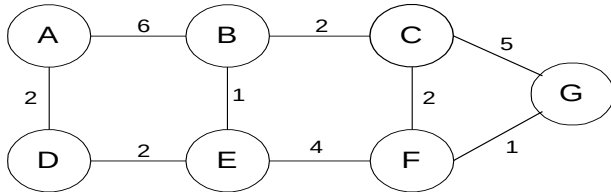
Dijkstra algoritmus 4

- Vegyük be E-t, és számoljuk ki E szomszédainak költségét
- Nincs változás, D megjelenik



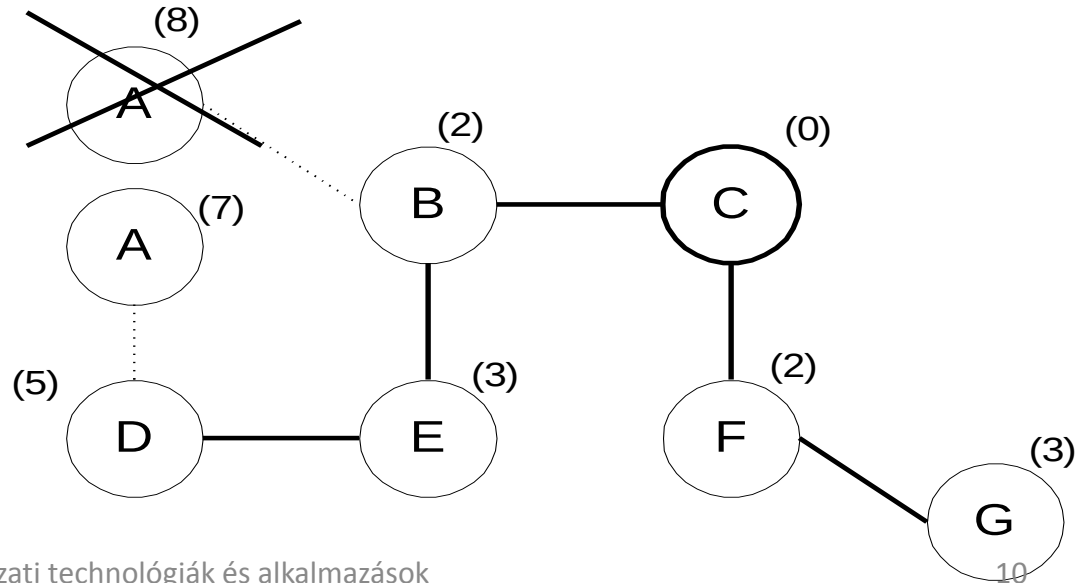
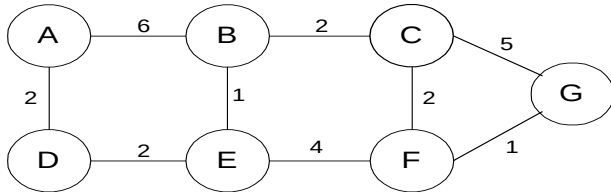
Dijkstra algoritmus 5

- Vegyük be G-t, és számoljuk ki G szomszédainak költségét
- Nincs változás



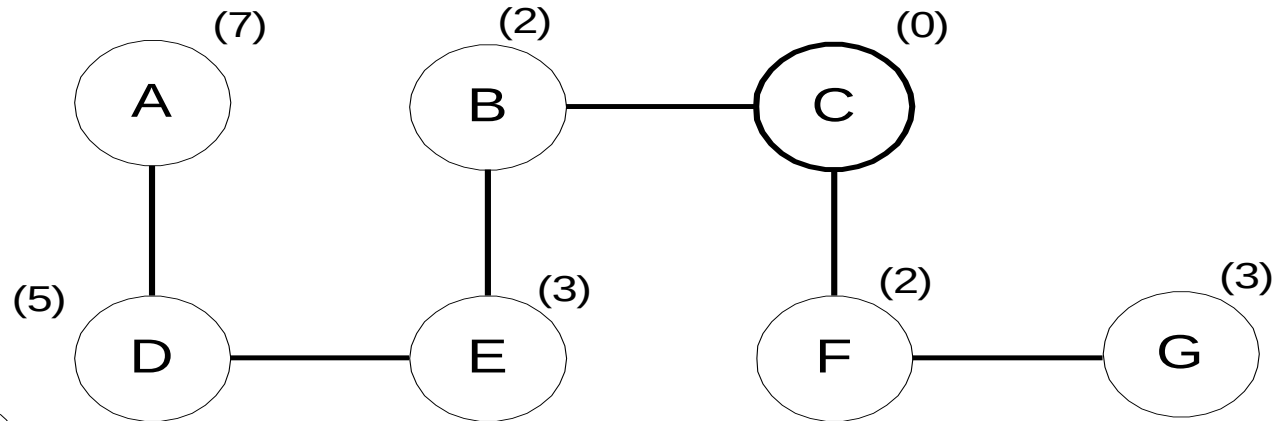
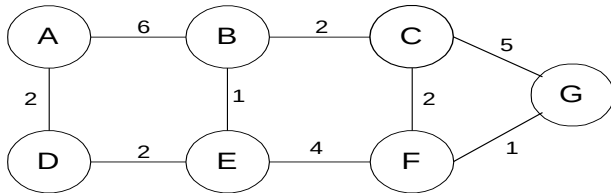
Dijkstra algoritmus 6

- Vegyük be D-t, és számoljuk ki D szomszédainak költségét
- Rövidebb út A-hoz!



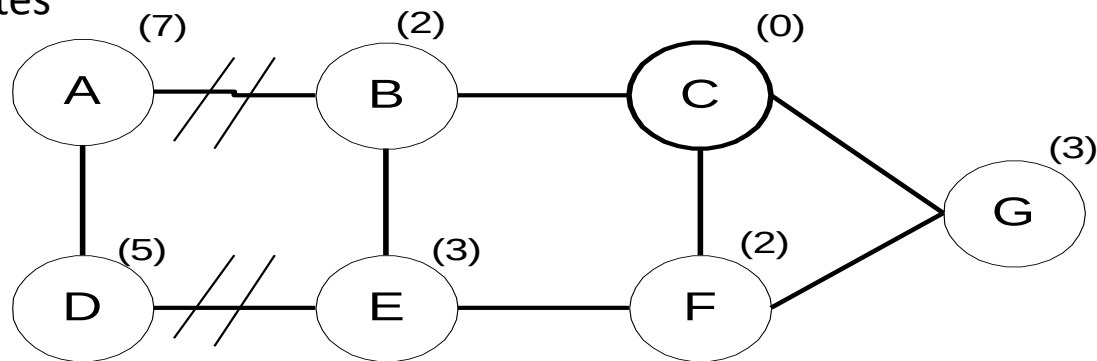
Dijkstra algoritmus 7

- Vegyük be A-t és számold ki A szomszédjainak költségét
- Nincs több szomszéd
- Befejezés



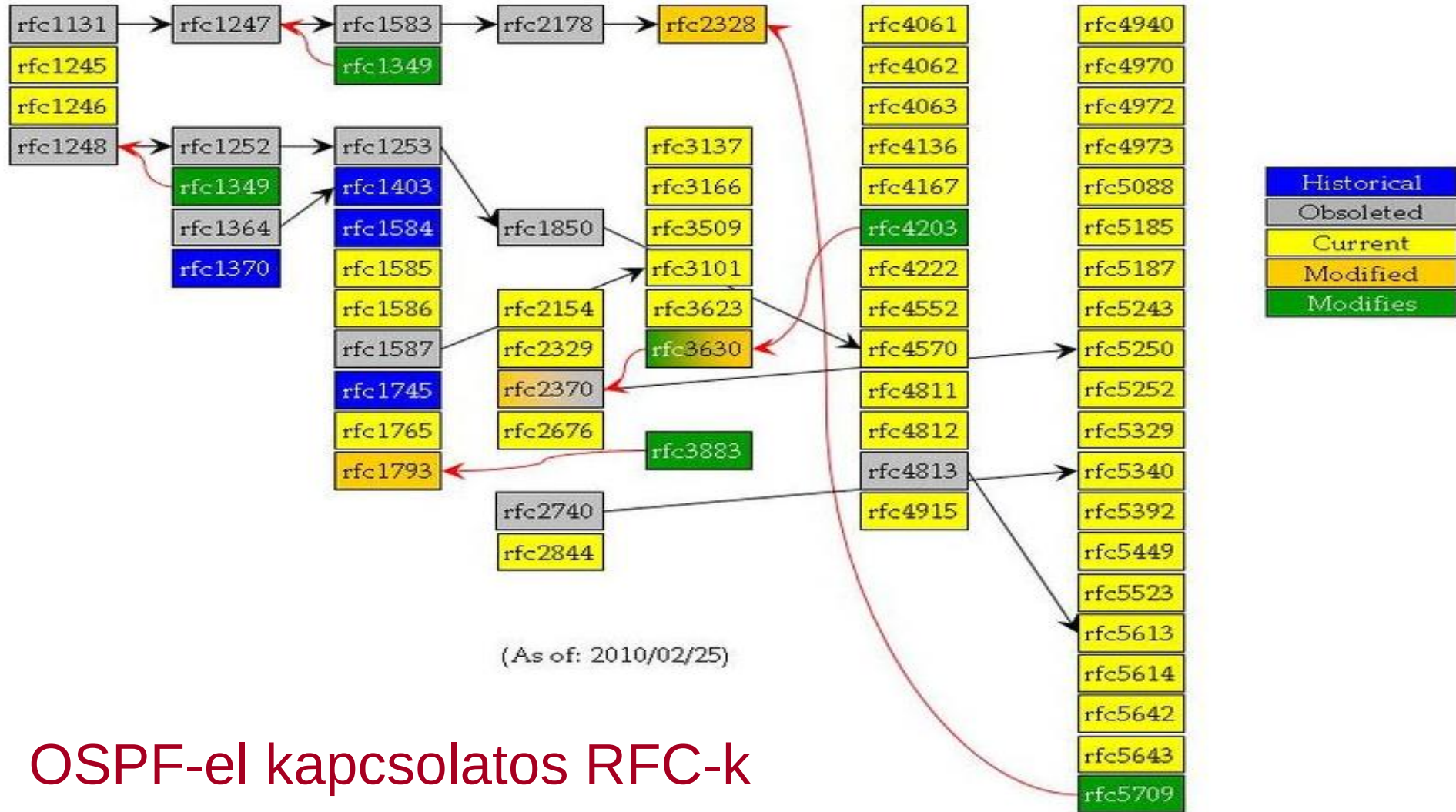
Hibás link következménye

- A-B és D-E linkek megsérülnek
 - hálózat kettészakad
 - két rész képtelen értesíteni egymást a változásokról
- A és D elérhetetlennek nyilvánítja a hálózat többi részét
- Hibás link megjavulása után a routerek szinkronizálják adatbázisukat
 - Topológia frissítés



Link-state protokollok

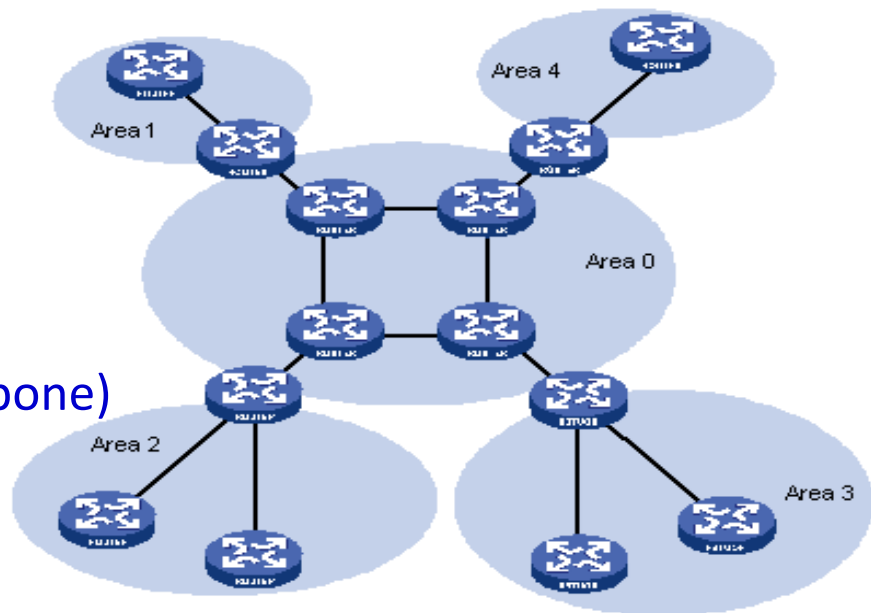
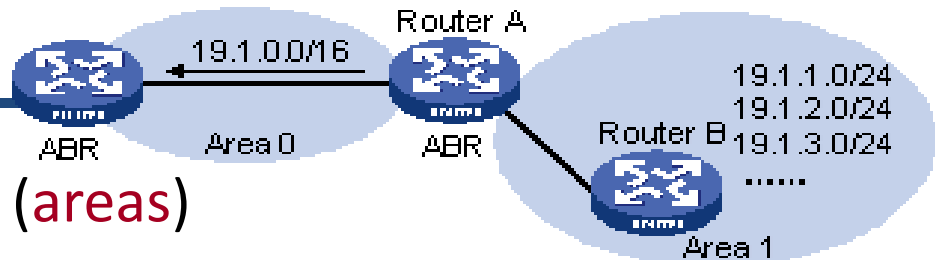
- OSPF – Open Shortest Path First
 - Első szabvány – RFC 1131 ('89)
 - OSPFv2 – RFC 2178 ('97)
 - OSPFv3 – RFC 2740 ('99)
 - IPv6-os verzió



OSPF-el kapcsolatos RFC-k

2 szintű hierarchia

- Egy OSPF tartomány területekre (**areas**) oszlik
 - Skálázhatósági szempontok miatt
- **LSA (Link State Advertisement)** terjesztés a területeken belül
- Területek között aggregáció
 - A változások egy területen belül nem látszanak kívülre
 - Speciális terület - **Gerinchálózat (backbone)** területe (AreaID=0)



OSPF protokoll összetevők

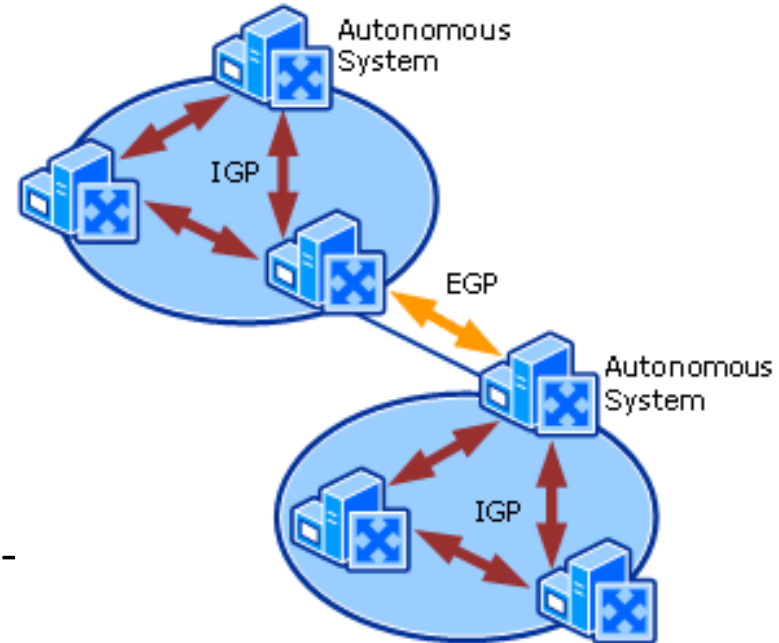
- Szomszéd felismerés
 - Hello protokollal
- Designated Router (DR), Backup Designated Router (BDR) kiválasztása
 - Prioritás alapú
 - 0-től 254-ig
 - Ha 0 prioritás, akkor soha nem lehet DR vagy BDR
 - Egyenlőség esetén a nagyobb Router ID nyer
 - RID = a legnagyobb konfigurált loopback cím a routeren (127.x.x.x)
 - Ha nincs loopback cím konfigurálva, RID = a legnagyobb aktív interfész cím
 - Ha a DR választás után egy nagyobb prioritású router megjelenik (bekapcsolják), nem veszi át a DR szerepét amíg a DR és a BDR jól működik
 - Ha a DR „meghal”, a BDR átveszi a szerepet
 - Új BDR választás

OSPF protokoll összetevők

- Szomszédosság meghatározása (Forming adjacencies)
 - Adatbázis szinkronizálás és LSA terjesztés csak a szomszédok között
 - A DR csökkenti a hálózati forgalmat egy üzenetszórásos hálózaton
 - A DR karbantart egy táblázatot a teljes hálózati topológiáról
 - Minden router egy területen belül master-slave kapcsolatban a DR-rel
 - A routerek a 224.0.0.6 multicast címre küldik a változásokat
 - All OSPF DR and BDR routers
 - A DR a 224.0.0.5 címre küldi az új táblázatot
 - All OSPF routers

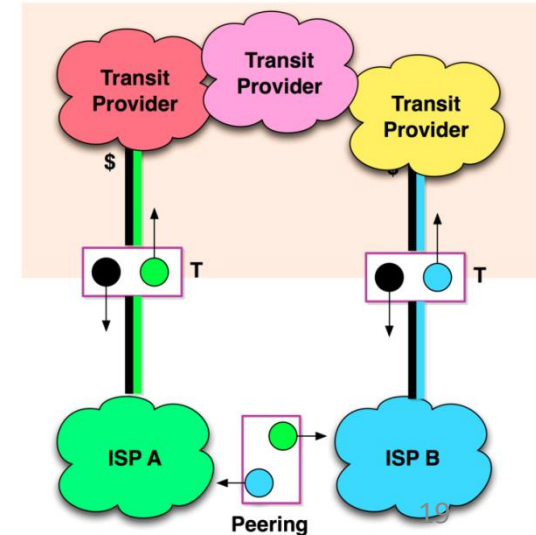
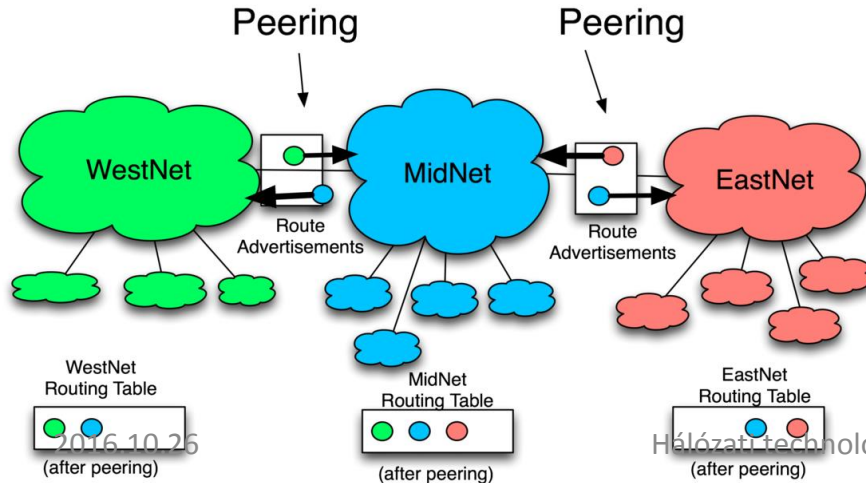
Autonóm rendszer

- Útválasztási tartomány = autonóm rendszer (**AS – autonomous system**)
 - Útválasztók összessége egy technikailag összetartozó területen
 - Egy szolgáltató, egy adminisztráció
 - Valamilyen IGP (Interior Gateway Protocol) protokollt használ (belül)
 - pl. RIP, OSPF
 - Exterior Gateway Protokollok (EGP) az AS-ek közötti útválasztáshoz
 - pl. BGP-4



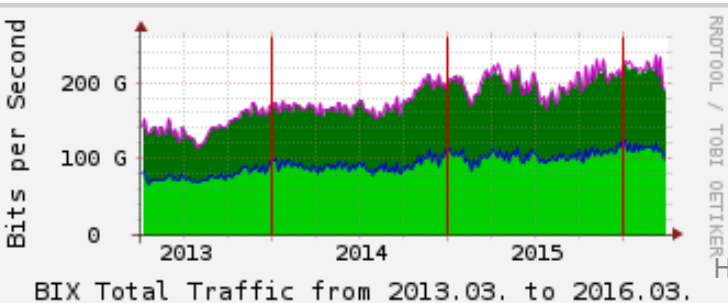
Internet topológia

- Autonóm rendszerek hálózata
 - Vevő – szolgáltató (customer-provider) kapcsolat
 - **Tranzit kapcsolat** – kapcsolódás a globális hálózathoz
 - **Peering kapcsolat** - két egyenrangú AS, két egyenrangú szolgáltató között
 - Nem tranzitív



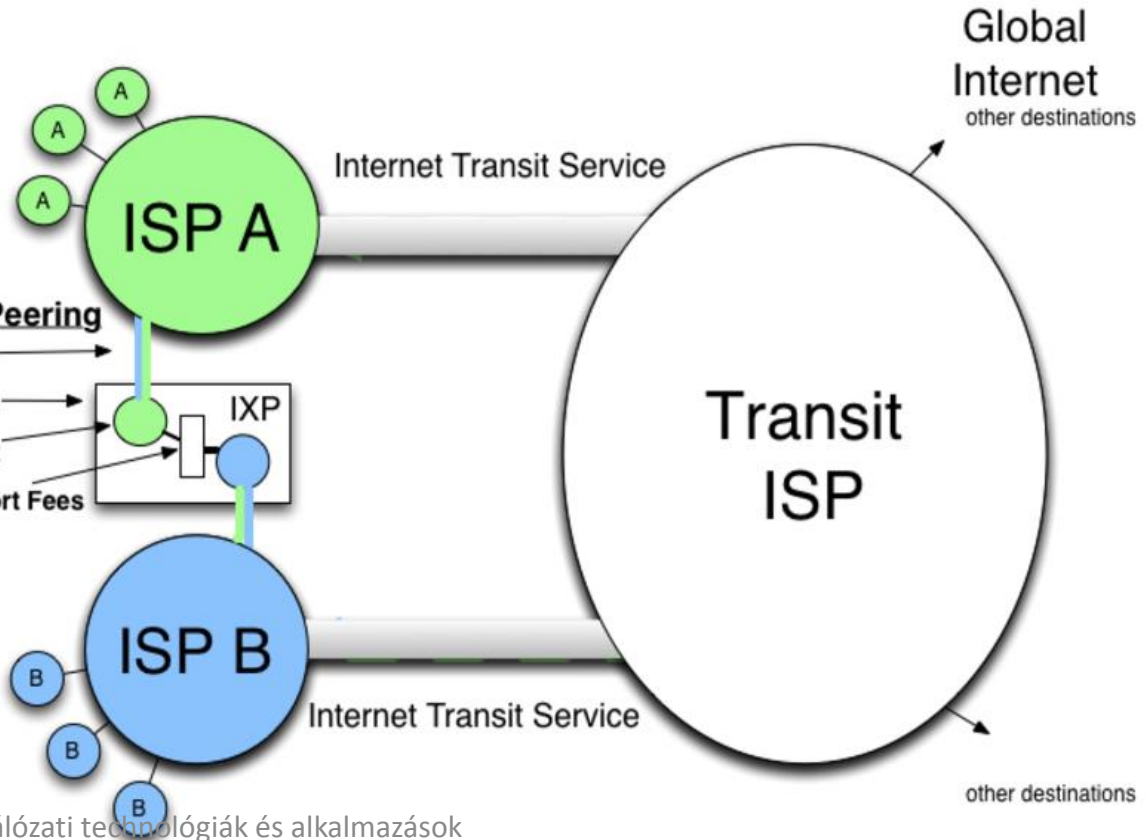
Tranzit vs. Peering

BIX2
(Budapest Internet Exchange)
210 Gbit/s (2014)



Costs of Peering

- 1) Transport
- 2) Colocation
- 3) Equipment
- 4) Peering Port Fees



Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. október 28.



Internet topológia

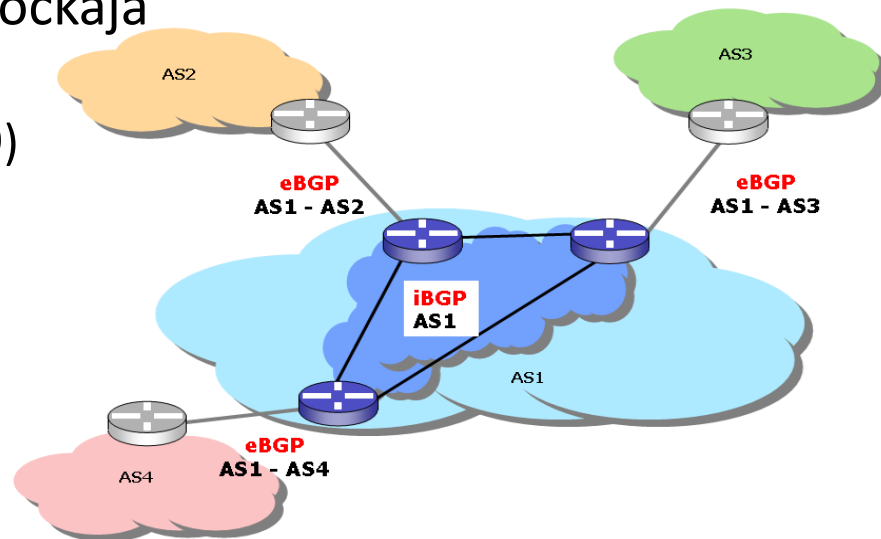
- IGP-EGP hierarchia előnyei
 - Skálázhatóság nagy hálózatokra
 - Kevesebb prefix terjesztése
 - Gyorsabb konvergencia
 - Hibák terjedését korlátozza
 - Adminisztratív autonómia
 - Minden AS-en belül igény szerinti IGP protokoll

IGP vs. EGP

- IGP-ben automatikus szomszéd felderítés
- EGP-ben specifikusan konfigurált peerek
- IGP esetén bizalom az útválasztókban
- EGP esetén (viszonylag) megbízhatatlan kapcsolatok más hálózatokkal
- IGP-ben a prefixek terjesztése a teljes hálózatban
- EGP-ben a prefixek terjesztése adminisztratíván korlátozott
- Egy IGP protokoll egy AS routereit köti össze
- Egy EGP protokoll AS-eket köt össze

Border Gateway Protocol

- Az Internet egyik legfontosabb építőköve
- BGP kronológia
 - Eredeti szabvány – BGP – RFC 1105 ('89)
 - BGP-3 – RFC 1267 ('91)
 - BGP-4 – RFC 1771 ('95)
 - Legújabb verzió – RFC 4271 ('06)
- **External BGP (eBGP)**
 - BGP viszony egy szomszédos útvásztóval más AS-ből
- **Internal BGP (iBGP)**
 - BGP viszony egy szomszédos útvásztóval ugyanabból az AS-ből



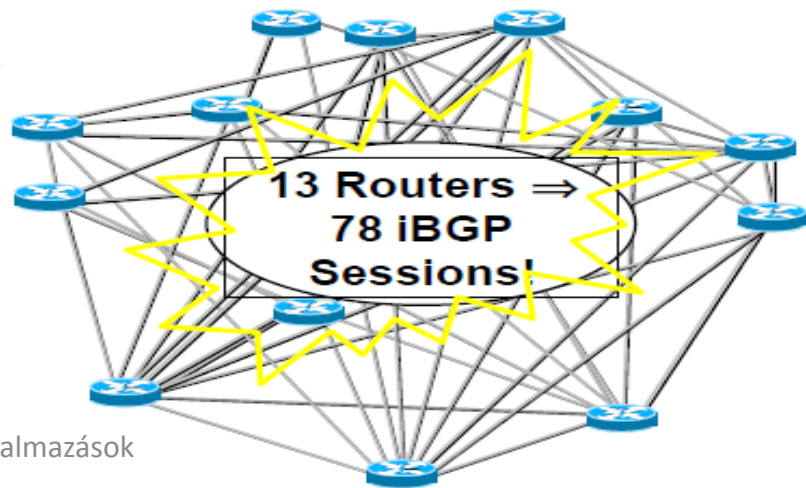
BGP jellemzők

- **CIDR (Classless Inter-Domain Routing)** támogatás
 - Változó hosszúságú hálózati prefixek
 - Hatékony címtartomány aggregáció
- Manuális szomszéd beállítás
 - Nincs automatikus felfedezés
- Nincsenek periodikus frissítések
 - UPDATE üzenetek – NLRI bejegyzések
 - Network Layer Reachability Information
 - (Destination prefix, AS útvonal, next hop)
 - Hurkok elkerülése az AS-ek felsorolásával
 - A nem elérhető útvonalakat is explicit módon jelzik

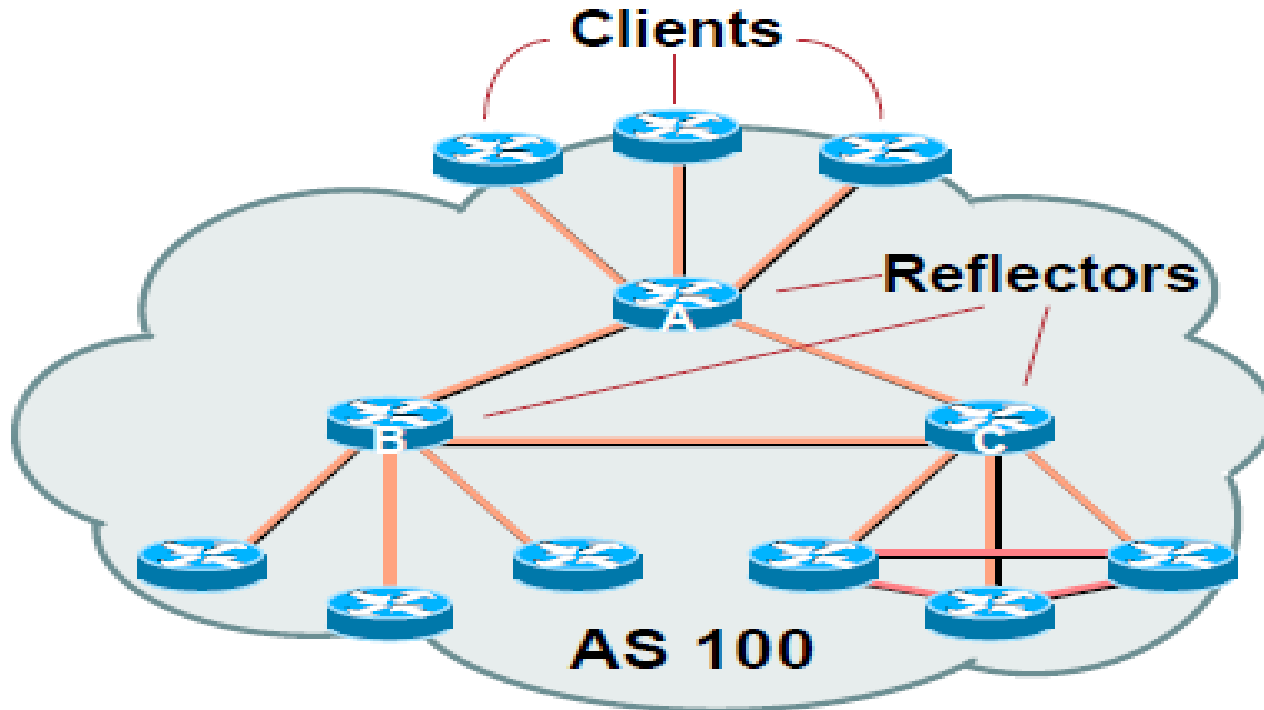
iBGP

- eBGP szomszédoktól való címeket terjeszti
- iBGP csomópontok - teljes összeköttetés
 - iBGP nem szükséges útválasztás
- Hátrány – egy teljes mesh nem skálázódik
 - pl. ha $n=1000$,
$$n(n-1)/2 = 499.500$$

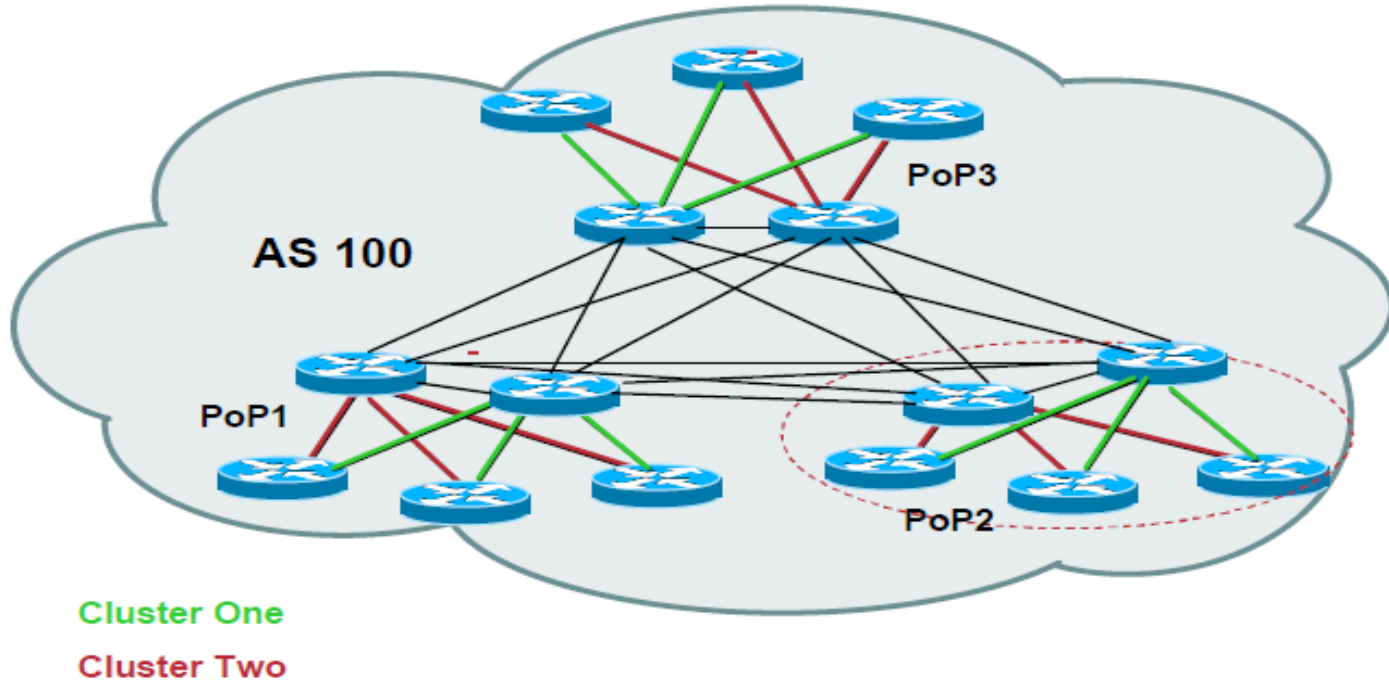
iBGP kapcsolat



Route reflector



Route reflector redundancy



Csoportos kommunikáció

- Cél: egy egyedi célállomás helyett egy célállomás halmazzal (csoporttal) kommunikálni
 - „természetes” általánosítása a pont-pont kommunikációnak (unicast)
- **Multicast** = többesadás



Mi is a multicast?

- **Unicast**

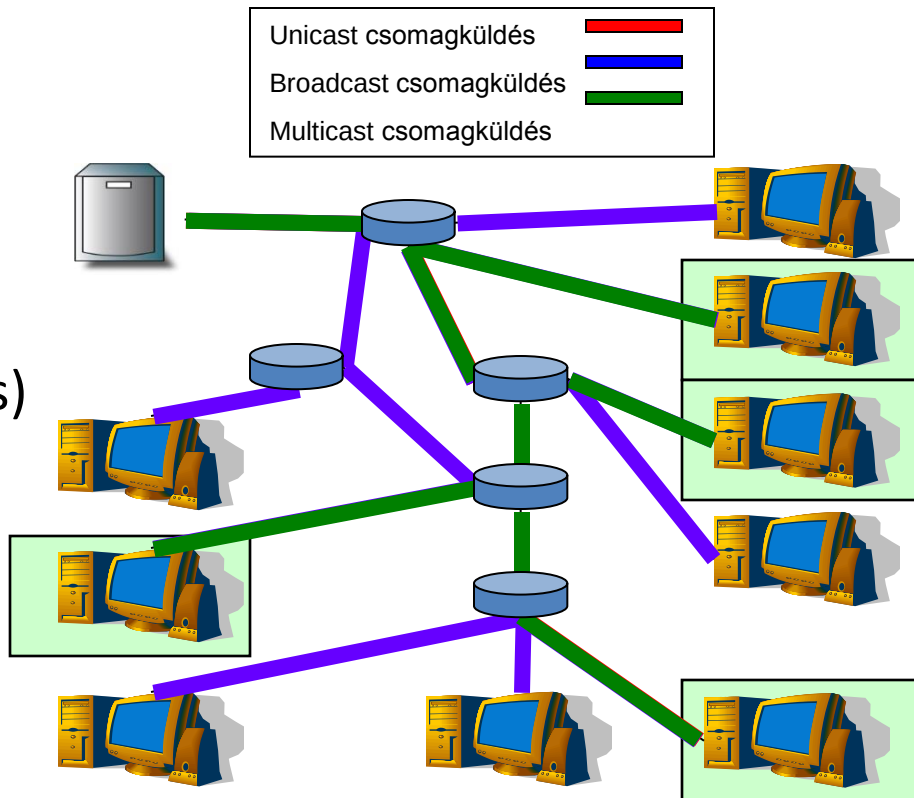
- Pont - pont
- Célcím: egyedi vevő címe

- **Broadcast**

- Pont - mindenki (csomagszórás)
- Célcím: a hálózat címe

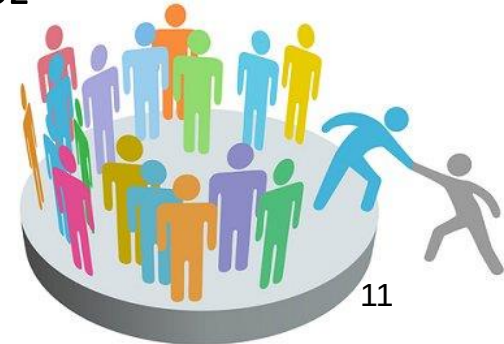
- **Multicast**

- (Több)pont - többpont
- Célcím: a csoport címe



Csoportos kommunikáció

- A csomagokat egy csoport minden tagjához el kell juttatni, nem csak egy célállomáshoz
 - A csoport felépítése (tagsága) dinamikus lehet
- Működési alapelv: miután egy csoport létrejön
 - Érdeklődő vevők (receiver) csatlakoznak a csoporthoz
 - A hálózat foglalkozik a csoport karbantartásával és a csomagok továbbításával

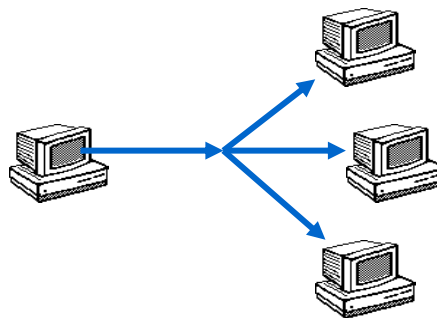


Multicast alkalmazások

- Számos alkalmazás nem pont-pont alapú

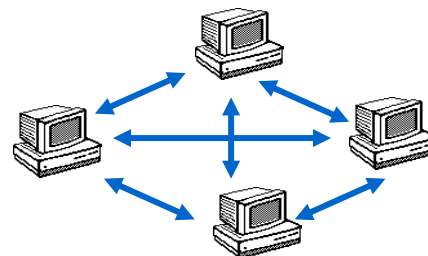
- Pont-többpont

- Távköztetés
 - Cache update
 - Video on demand



- Többpont-többpont

- Videokonferencia, Audiokonferencia, Chat,
 - Elosztott hálózati játékok
 - Kooperatív alkalmazások



Követelmények

- Nincs olyan megoldás, mely minden környezetben ideális lenne
- A követelmények nagyban változnak
 - az alkalmazás igényeitől függően
 - a csoport méretétől függően
 - a hálózati szolgáltatásoktól függően
 - a résztvevők heterogeneitásától függően



Részvételi feltételek

- Csoporttagság ellenőrzés
 - Nyitott csoport: bárki lehet tag
 - Zárt csoport: tagsági korlátozások
- Forrásellenőrzés
 - Bárki küldhet csomagot a csoportnak
 - Csak egy csoporttag küldhet csomagokat
 - Csak egy kijelölt forrás küldhet csomagokat



Adatátviteli feltételek

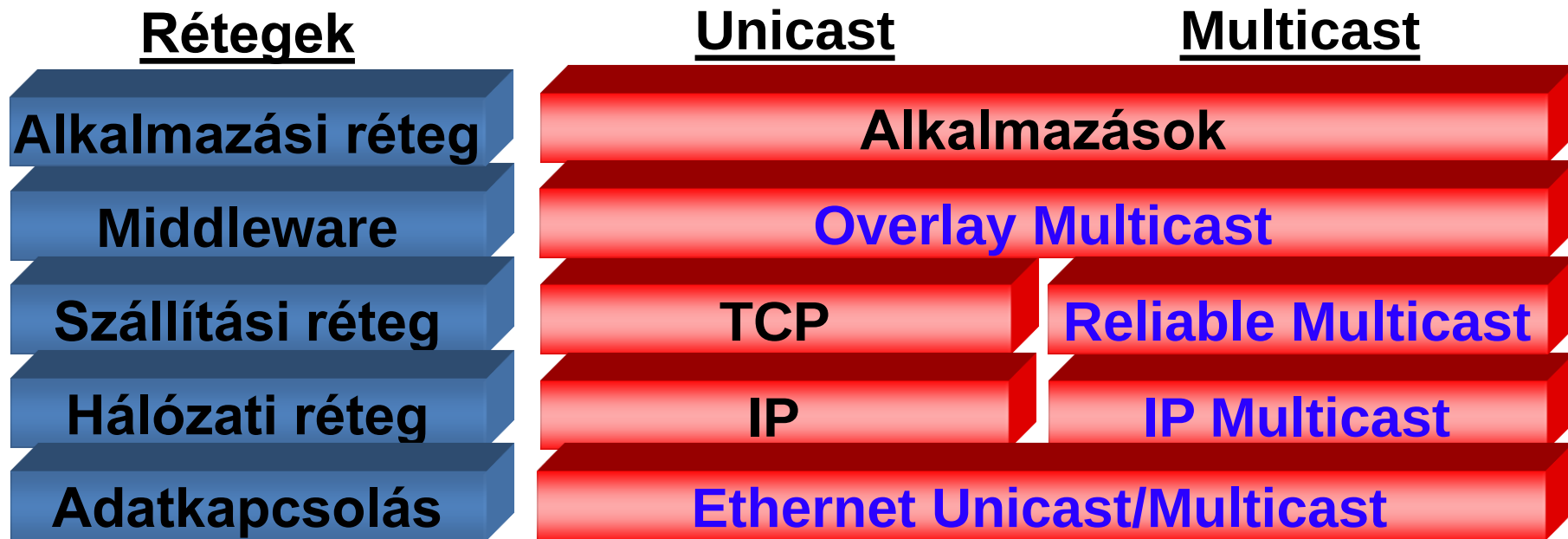
- Pont-pont kommunikáció
 - megbízható vagy best-effort (garanciák nélkül)
 - A cél ellenőrzi a kapott csomagot: vagy OK, vagy nem
- Pont-multipont kommunikáció
 - minden vevő másként értékelheti a szolgáltatást
- Különböző megbízhatósági szintek
 - 0-megbízhatóság: egyetlen vevőnek sem garantált a megbízható átvitel
 - 1-megbízhatóság: legalább egy vevőnek garantált
 - k-megbízhatóság: legalább k vevőnek garantált
 - totális megbízhatóság: minden vevőnek garantált a megbízható szolgáltatás



Multicast különböző rétegekben

- A multicast (többesadás) szolgáltatást különböző rétegekben lehet implementálni
 - Adatkapcsolási réteg (data link layer)
 - pl. Ethernet multicast
 - Hálózati réteg (network layer)
 - pl. IP multicast, Xcast
 - Alkalmazási réteg (application layer)
 - pl. Narada, TBCP
- Melyik megoldás a jobb?
 - Attól függ, nincs általános megoldás

Multicast különböző rétegekben



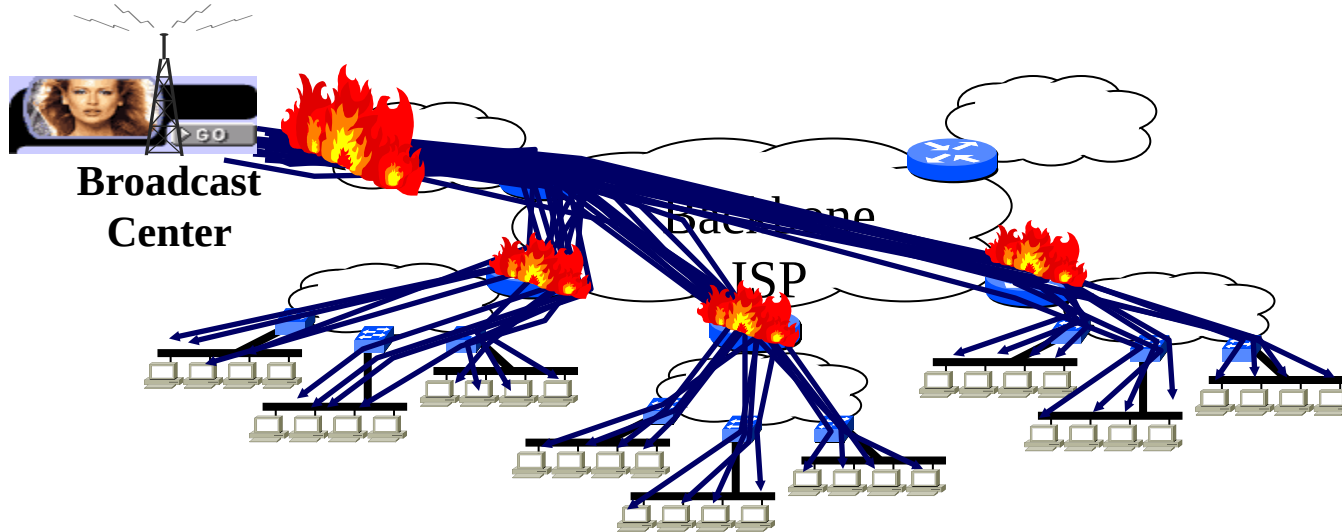
Ethernet Multicast

- Néhány Ethernet MAC cím multicast számára elkülönítve
- Ha egy G csoporthoz akarunk csatlakozni
 - A hálózati kártya (network interface card, NIC) elvileg csak a unicast és broadcast címre küldött csomagokat hallgatja
 - A csatlakozáshoz a G multicast címet is hallgatnia kell
 - Hardware megoldás, hatékony
- Csomagküldés a G csoportban
 - A csomag eláraszt minden LAN szegmenst
 - mint broadcast esetén
 - A kártyák melyek nem hallgatják a G multicast címet, eldobják a csomagot

Hálózati rétegű multicast

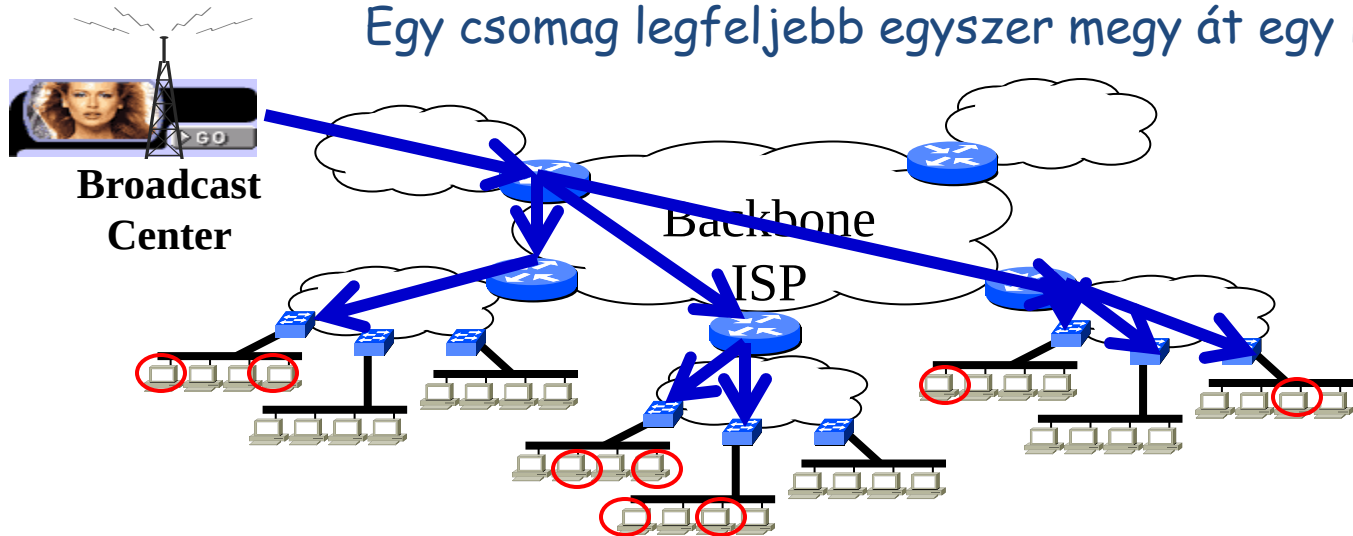
- A cél a hálózati erőforrások optimális kihasználása
 - Egy csomag egy link-en csak egyszer megy át
- A router-ek egy multicast fát tartanak fenn
 - A multicast fa mentén történik az adatátvitel
 - A router-ek duplikálják a csomagokat ha szükséges
 - Elágazási pontok a fán

A csoportos unicast nem skálázható

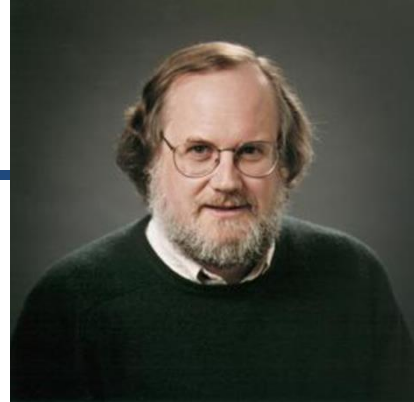


Helyette építsünk fákat

A router-ek duplikálják a csomagokat
Egy csomag legfeljebb egyszer megy át egy linken



IP Multicast



- Steve Deering PhD disszertációja (1990)
 - Any Source Multicast (ASM)
- Nyitott csoportkommunikációs modell
 - Bárki csatlakozhat egy csoporthoz, bármilyen engedélyezés nélkül
 - Egy felhasználó több csoportnak is tagja lehet egyszerre
 - Bárki küldhet adatokat a csoportnak, ha nem is tagja annak
 - A csoport tagsága dinamikus
 - Senki nem ismeri a csoport méretét, vagy a tagok kilétét

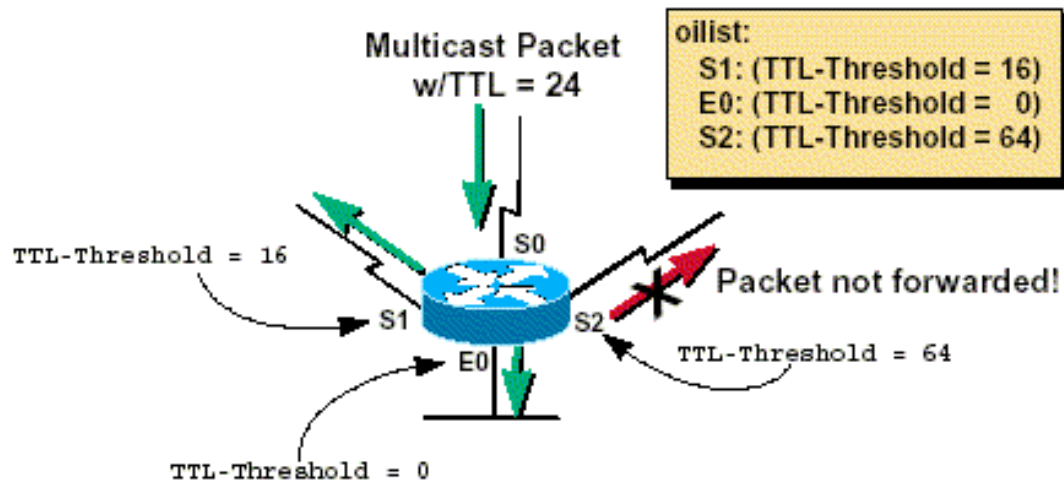
IP Multicast

- S. Deering, "Host Extensions for IP Multicasting", RFC 1112, 1989.
- A forrás csomagjait egy virtuális csoportcímre küldi
- Bárki aki csatlakozik a csoporthoz „elérhető” ezen a címen
 - Megkapja az ezen címre küldött csomagokat
- Egy multicast csoportot egy **class D** IP cím azonosít
 - **224.0.0.0 – 239.255.255.255**
 - 1110 + 28 bites csoport azonosító

Bits:	1	8	9	16	17	24	25	32
Class A	0NNNNNNN	Host	Host	Host				
	Range (1-126)							
Bits:	1	8	9	16	17	24	25	32
Class B	10NNNNNNN	Network	Host	Host				
	Range (128-191)							
Bits:	1	8	9	16	17	24	25	32
Class C	110NNNNNN	Network	Network	Host				
	Range (192-223)							
Bits:	1	8	9	16	17	24	25	32
Class D	1110MMMM	Multicast Group	Multicast Group	Multicast Group				
	Range (224-239)							

Multicast Scoping

- Egy IP multicast csoport hatóköre szabályozva van:
 - TTL alapú szabályozás
 - Adminisztrációs szabályozás
- TTL alapú szabályozás
 - Node-local 0
 - Link-local 1
 - Site-local < 32
 - Region-local < 64
 - Continent-local < 128
 - Global Scope < 255



Multicast Scoping

- Adminisztrációs szabályozás
 - link-local scope 224.0.0.0 - 224.0.0.255
 - Egy router sohasem küldi tovább
 - global scope 224.0.1.0 - 238.255.255.255
 - A teljes Interneten érvényes
 - administrative scope 239.0.0.0 - 239.255.255.255
 - Nem küldik egy szervezet Intranet-jén kívülre

Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. november 3.



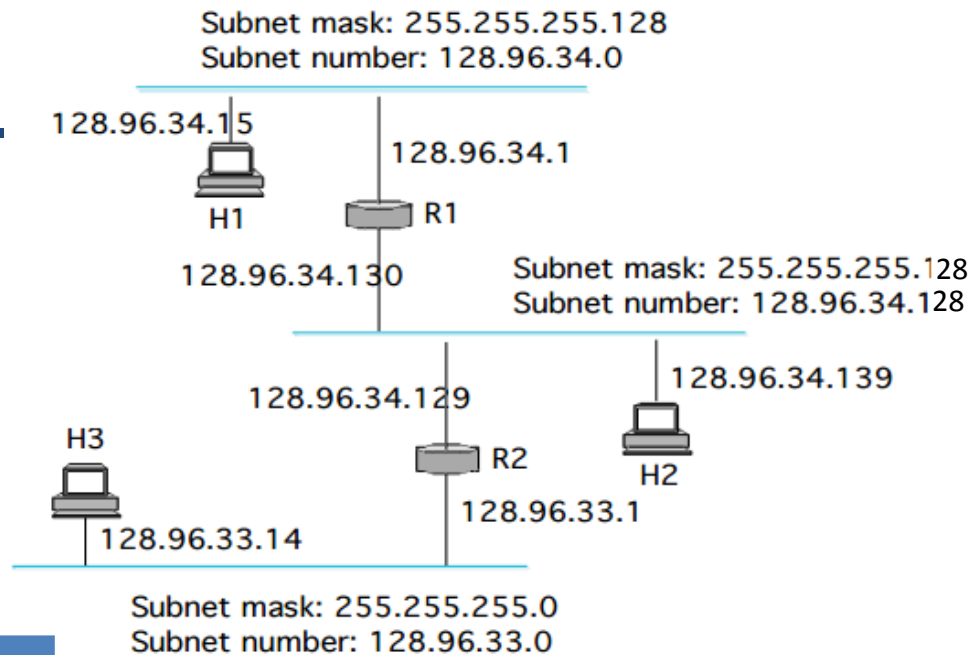
1. feladat

Írjuk fel az R2 routing tábláját,
a következő formátumban

Dest | NextHop | Subnet Mask

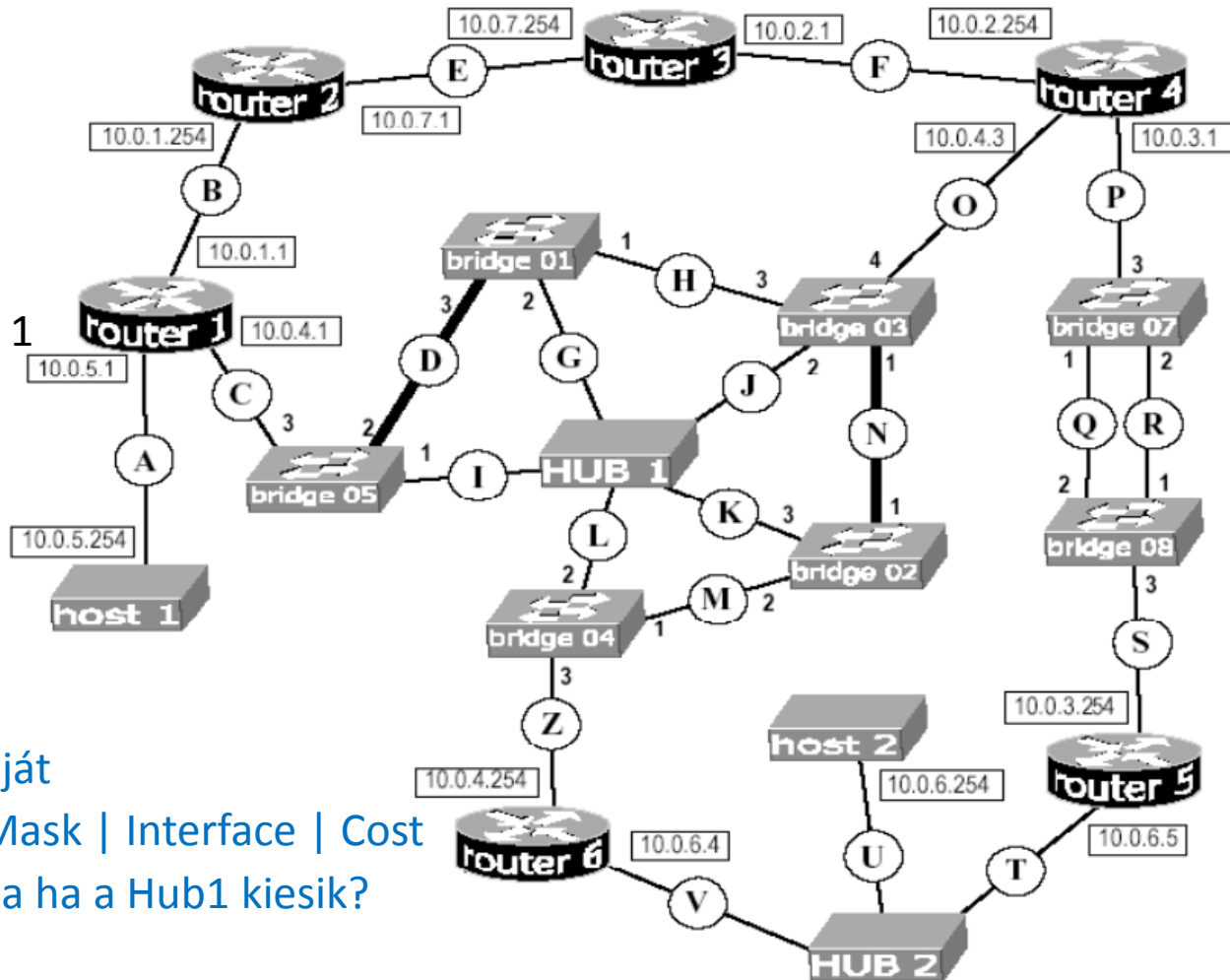
Megoldás

DestNet	Next Hop	Subnet Mask
128.96.34.128	Direct	255.255.255.128
128.96.33.0	Direct	255.255.255.0
128.96.34.0	128.96.34.130	255.255.255.128



2. feladat

- D és N kapacitása 2
- Minden más link kapacitása 1
- RIP routing



Írjuk fel az R1 útválasztó tábláját

DestNet | NextHop | NetMask | Interface | Cost

Hogyan változik a routing tábla ha a Hub1 kiesik?

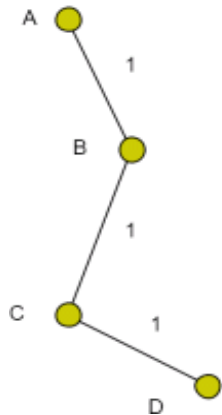
2. feladat - megoldás

DestNet	NextHop	Netmask	Interface	HopCount
10.0.5.0	Direct	255.255.255.0	A	0
10.0.1.0	Direct	255.255.255.0	B	0
10.0.4.0	Direct	255.255.255.0	C	0
10.0.7.0	10.0.1.254	255.255.255.0	B	1
10.0.2.0	10.0.4.3	255.255.255.0	C	1
10.0.6.0	10.0.4.254	255.255.255.0	C	1
10.0.3.0	10.0.4.3	255.255.255.0	C	1

- Ha a Hub1 kiesik, nem történik semmi!
 - Léteznek alternatív útvonalak a 10.0.4.0 hálózaton

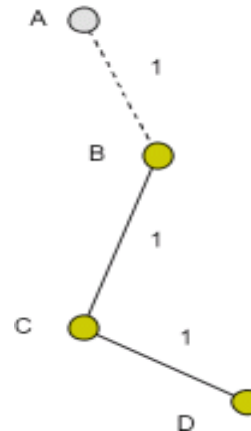
3. feladat

- Végtelenig számolás
 - A routerek a célcím költségek hirdetésekor végtelenig inkrementálhatnak



B	C	D
1	2	3

Távolság A felé



B	C	D
3	2	3

B	C	D
3	4	3

B	C	D
5	4	5

4. feladat

Az alábbi routing táblát hogyan lehet tömöríteni CIDR-t használva?

DestNet	NetMask	Interfész
194.100.0.0	255.255.255.0	I1
194.100.1.0	255.255.255.0	I1
194.100.2.0	255.255.254.0	I1
194.100.4.0	255.255.252.0	I1
194.100.8.0	255.255.248.0	I1
194.100.48.0	255.255.240.0	I1
194.100.64.0	255.255.240.0	I1

4. feladat - megoldás

- 194.100.0.0/24 -> 194.100.0.0-tól 194.100.0.255-ig (256 cím)
- 194.100.1.0/24 -> 194.100.1.0-tól 194.100.1.255-ig (256 cím)
- 194.100.2.0/23 -> 194.100.2.0-tól 194.100.3.255-ig (512 cím)
- 194.100.4.0/22 -> 194.100.4.0-tól 194.100.7.255-ig (1024 cím)
- 194.100.8.0/21 -> 194.100.8.0-tól 194.100.15.255-ig (2048 cím)
- Folytonos címtartományok, megegyezik az első 20 bit, 4096 cím
 - CIDR-rel 194.100.0.0/20
- 194.100.48.0/20 -> 194.100.48.0-tól 194.100.63.255-ig (4096 cím)
- 194.100.64.0/20 -> 194.100.64.0-tól 194.100.79.255-ig (4096 cím)
- Folytonos címtartomány, de 48 -> 00110000, 64 -> 01000000, nem egyezik meg az első 19 bit, nem lehet aggregálni egy /19-es címbe

DestNet	NetMask	IF
194.100.0.0	255.255.255.0	l1
194.100.1.0	255.255.255.0	l1
194.100.2.0	255.255.254.0	l1
194.100.4.0	255.255.252.0	l1
194.100.8.0	255.255.248.0	l1
194.100.48.0	255.255.240.0	l1
194.100.64.0	255.255.240.0	l1

4. feladat - megoldás

- Tömörített CIDR routing tábla

DestNet	IF
194.100.0.0/20	I1
194.100.48.0/20	I1
194.100.64.0/20	I1

DestNet	NetMask	IF
194.100.0.0	255.255.255.0	I1
194.100.1.0	255.255.255.0	I1
194.100.2.0	255.255.254.0	I1
194.100.4.0	255.255.252.0	I1
194.100.8.0	255.255.248.0	I1
194.100.48.0	255.255.240.0	I1
194.100.64.0	255.255.240.0	I1

5. feladat

Az alábbi routing táblát hogyan lehet tömöríteni CIDR-t használva?

DestNet	NetMask	IF
200.0.0.0	255.255.192.0	A
200.0.64.0	255.255.192.0	A
200.0.128.0	255.255.128.0	A
200.1.0.0	255.255.0.0	A
193.0.2.0	255.255.255.0	B
193.0.3.0	255.255.255.0	B
193.0.4.0	255.255.255.0	B
193.0.5.0	255.255.255.0	B

5. feladat - megoldás

- 200.0.0.0/18 -> 200.0.0.0-tól 200.0.63.255-ig (16384 cím)
- 200.0.64.0/18 -> 200.0.64.0-tól 200.0.127.255-ig (16384 cím)
- 200.0.128.0/17 -> 200.0.128.0-tól 200.0.255.255-ig (32768 cím)
- 200.1.0.0/16 -> 200.1.0.0-tól 200.1.255.255-ig (65536 cím)
- Folytonos címtartomány, megegyező első 15 bit, 131072 cím
 - CIDR-rel 200.0.0.0/15
- 193.0.2.0/24 -> 193.0.2.0-tól 193.0.2.255-ig (256 cím)
- 193.0.3.0/24 -> 193.0.3.0-tól 193.0.3.255-ig (256 cím)
- 193.0.4.0/24 -> 193.0.4.0-tól 193.0.4.255-ig (256 cím)
- 193.0.5.0/24 -> 193.0.5.0-tól 193.0.5.255-ig (256 cím)
- 193.0.2.0 és 193.0.3.0 folytonos címtartomány, megegyező első 23 bit -> 193.0.2.0/23
- 193.0.4.0 és 193.0.5.0 folytonos címtartomány, megegyező első 23 bit -> 193.0.4.0/23

DestNet	NetMask	IF
200.0.0.0	255.255.192.0	A
200.0.64.0	255.255.192.0	A
200.0.128.0	255.255.128.0	A
200.1.0.0	255.255.0.0	A
193.0.2.0	255.255.255.0	B
193.0.3.0	255.255.255.0	B
193.0.4.0	255.255.255.0	B
193.0.5.0	255.255.255.0	B

5. feladat - megoldás

- Tömörített CIDR routing tábla

DestNet	IF
200.0.0.0/15	A
193.0.2.0/23	B
193.0.4.0/23	B

DestNet	NetMask	IF
200.0.0.0	255.255.192.0	A
200.0.64.0	255.255.192.0	A
200.0.128.0	255.255.128.0	A
200.1.0.0	255.255.0.0	A
193.0.2.0	255.255.255.0	B
193.0.3.0	255.255.255.0	B
193.0.4.0	255.255.255.0	B
193.0.5.0	255.255.255.0	B

6. feladat

- Egy ISP rendelkezik a 194.48.0.0/16 címtartomány felett
- 3 szervezetnek (A, B és C) 2048, 8192 illetve 4096 címre van szüksége
- Az A és C szervezetek az 1. interfészen keresztül érhetőek el, a B szervezet a 2. interfészen
- Optimizáljuk CIDR-t használva a címkiosztást a szervezeteknek. Hogyan fog kinézni az útválasztó tábla?

6. feladat - megoldás

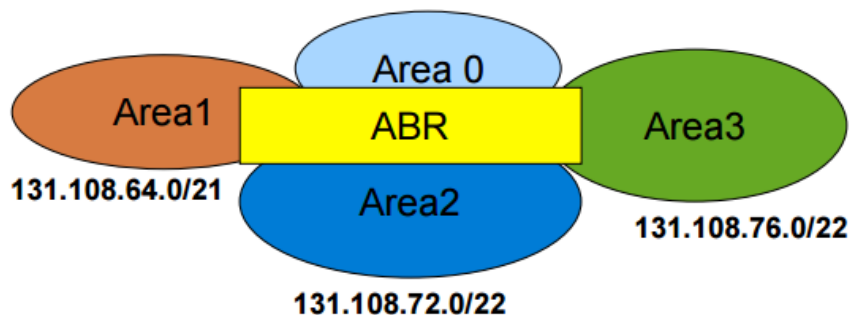
- A és C egy interfészen, B egy másikon -> legalább két bejegyzés a routing táblában
- A -> 2048 cím -> /21
- C -> 4096 cím -> /20
- Nem lehet aggregálni, mert nem töltene ki egy teljes /19-es tartományt
- A -> 194.48.0.0, subnet mask 255.255.248.0 , 194.48.0.0-tól 194.48.7.255-ig, IF 1
- C -> 194.48.16.0, subnet mask 255.255. 240.0 , 194.48.16.0-tól 194.48.31.255-ig, IF 1
- 2048 címek blokk az A és C között fenntartva egy későbbi lehetséges aggregálásra
- B -> 8192 cím -> /19
- B -> 194.48.32.0, subnet mask 255.255.224.0 , 194.48.32.0-tól 194.48.63.255-ig, IF 2

7. feladat

- Egy X cégnek legyen a címtartománya 131.108.64/20
- A tartományt a cég 3 egysége között kell felosztani, a következőképpen:
 - Az A egységnek max. 2000 címre van szüksége
 - A B és C egységeknek max 1000-1000 címre van szüksége
- Javasoljunk egy megfelelő hálózati architektúrát, feltételezve hogy OSPF-et használunk IGP protokollként
- Hogyan osztjuk fel a címtartományt?

7. feladat - megoldás

- Címtartomány mérete: 4096 cím -> OK
- 3 OSPF tartomány (+ area 0), egységenként egy
- Cím kiosztás:
 - A: 2048 cím -> /21 -> 131.108.64.0/21 (131.108.64.0-tól 131.108.71.255-ig), area1
 - B: 1024 cím -> /22 -> 131.108.72.0/22, (131.108.72.0-tól 131.108.75.255-ig), area2
 - C: 1024 cím -> /22 -> 131.108.76.0/22, (131.108.76.0-tól 131.108.79.255-ig), area3

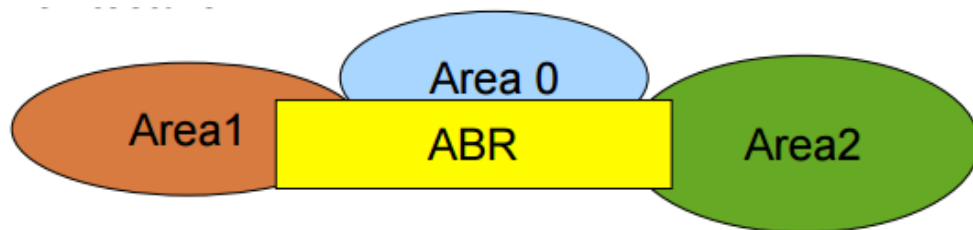


8. feladat

- Egy hálózati adminisztrátor egy B címosztályt kezel
- Úgy dönt, hogy OSPF-et használ, és 3 tartományt definiál (az area 0-át beleértve)
- Egyenlő módon osztja fel a címtartományt
 - Az összes páros cím az 1. tartományba kerül
 - Az összes páratlan cím a 2. tartományba
- Nagyságrendileg hány bejegyzés lesz az ABR router táblájában?
- Hozzávetőleg mennyi memóriára lesz szükség, ha egy bejegyzést 80 byte-on tárolunk?

8. feladat - megoldás

- Architektúra



- Páros IP címek: minden cím ami 0-ra végződik
- Páratlan IP címek: minden cím ami 1-re végződik
- Nem lehet aggregálni
- $2 \times 2^{15} = 2^{16}$ bejegyzés $\rightarrow 2^{16} \times 80 \sim 5\text{MB}$

Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. november 9.



IP Multicast adatátvitel

- A csatlakozás egy multicast csoporthoz két lépésben történik
 - A helyi hálózaton (LAN)
 - Egy felhasználó értesíti a helyi multicast router-ét hogy szeretne csatlakozni egy csoporthoz
 - IGMP (IPv4), MLD (IPv6)
 - A nagy kiterjedésű hálózaton (WAN)
 - A helyi router közreműködik a hálózat többi multicast router-ével a multicast fa kiépítésében és a csomagok továbbításában
 - DVMRP, MOSPF, CBT, PIM-DM, PIM-SM, PIM-SSM

IGMP

- **Internet Group Management Protocol**
- IPv4 protokoll, a végső felhasználók és a helyi multicast router-ek között a helyi hálózaton
 - A multicast csoportokban való tagságot kezeli
 - Aszimmetrikus protokoll
 - Felhasználói rész
 - Router rész
- A router megtanulja hogy milyen csoportokat hallgatnak a saját helyi hálózatán
 - Nem érdekli hányan hallgatják, a fontos hogy legyen legalább egy valaki
 - Nem érdekli ki hallgatja

IGMPv1

- S. Deering, "Host Extensions for IP Multicasting", RFC 1112, 1989.
- A multicast router rendszeres **Query** üzeneteket küld az összes felhasználó közös multicast címére (224.0.0.1)
- A felhasználók **Report** üzenettel válaszolnak, melyben beszámolnak az általuk hallgatott csoportokról
 - A Report-ot a hallgatott csoportok multicast címére küldik
- A Report csomagok számának csökkentése érdekében:
 - Időzítők (timer) használata
 - Egy felhasználó nem válaszol azonnal
 - Host Suppression
 - Ha valaki más már válaszolt, törli a saját Report üzenetét
- **Unsolicited Report**
 - Ha egy felhasználó egy új csoportot akar hallgatni

IGMPv1 Router

- Egy IGMPv1 router fenntart egy multicast tagsági táblát
 - Milyen multicast csoportokat hallgatnak a hálózaton
 - Mikor volt az utolsó Report egy csoporttal kapcsolatban
- **Soft-state** protokoll
 - Ha egy adott időn belül nem erősíti meg senki egy csoport iránti érdeklődését, a csoportot törli a táblájából
- Csak azokat a multicast csomagokat küldi tovább a helyi hálózatra, melyeket egy a táblájában szereplő multicast címre küldtek

IGMPv2

- W. Fenner, "Internet Group Management Protocol, Version 2", RFC 2236, November 1997.
<http://www.ietf.org/rfc/rfc2236.txt>
- IPv6-os változata: MLD (Multicast Listener Discovery)
 - S. Deering, W. Fenner, B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, November 1999.
<http://www.ietf.org/rfc/rfc2710.txt>
- Bevezet egy gyors kilépési mechanizmust (**Fast Leave**)
 - Nem kell várni az időzítők lejáráásáig ahhoz, hogy a router „levágjon” egy csoportot

IGMPv2 üzenetek

- Membership Query
 - General Query
 - Group Specific Query
- Membership Report
- **Leave Group Message**
- Ha egy felhasználó ki akar lépni egy csoportból, küld egy Leave üzenetet az összes multicast router közös multicast címére (224.0.0.2)
- Mielőtt a router levágja a csoportot, megkérdezi, van-e valaki más aki hallgatja a csoportot
 - Group Specific Query
 - Ha egy adott időn belül nem érkezik válasz, a router törli a csoportot
- **IGMPv3** – később...

Multicast Routing

- A forrás egy multicast csoportcímre küldi csomagjait
- A hálózat multicast routerei kialakítanak és karbantartanak egy multicast fát
 - Az adatátvitel ezen fa mentén történik
- A helyi multicast router, az IGMP tagsági táblája alapján csatlakozik a fához, vagy elhagyja azt
- A hálózat multicast router-ei között egy útválasztó protokoll működik
 - MOSPF, DVMRP, CBT, PIM

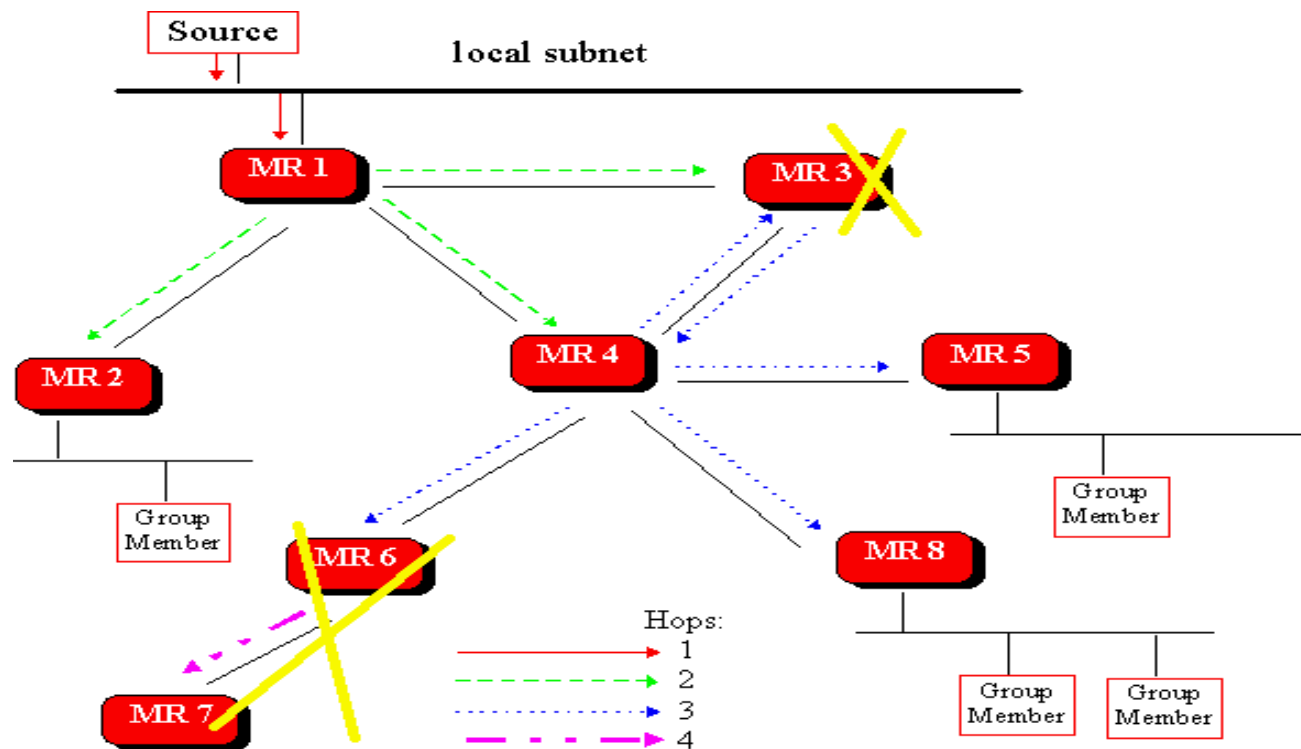
MOSPF

- **Multicast Open Shortest Path First**
 - J. Moy, „Multicast Extensions to OSPF”, RFC 1584, March 1994
<http://www.ietf.org/rfc/rfc1584.txt>
- Kapcsolatállapot (Link State) protokoll
- Az OSPF unicast útválasztó protokollt bővíti ki
 - Multicast csoportinformációt is küldenek a routerek egymásnak
 - Minden MOSPF router megtudja hogy melyik helyi hálózaton melyik csoportot hallgatják
 - Az információ alapján forrásonként és csoportonként egy legrövidebb útvonalú fát (shortest path tree) építenek fel
- Nagy a jelzés többletterhelés
- Nehezen alkalmazkodik a topológia változásokhoz
 - Újra kell számolni a fákat

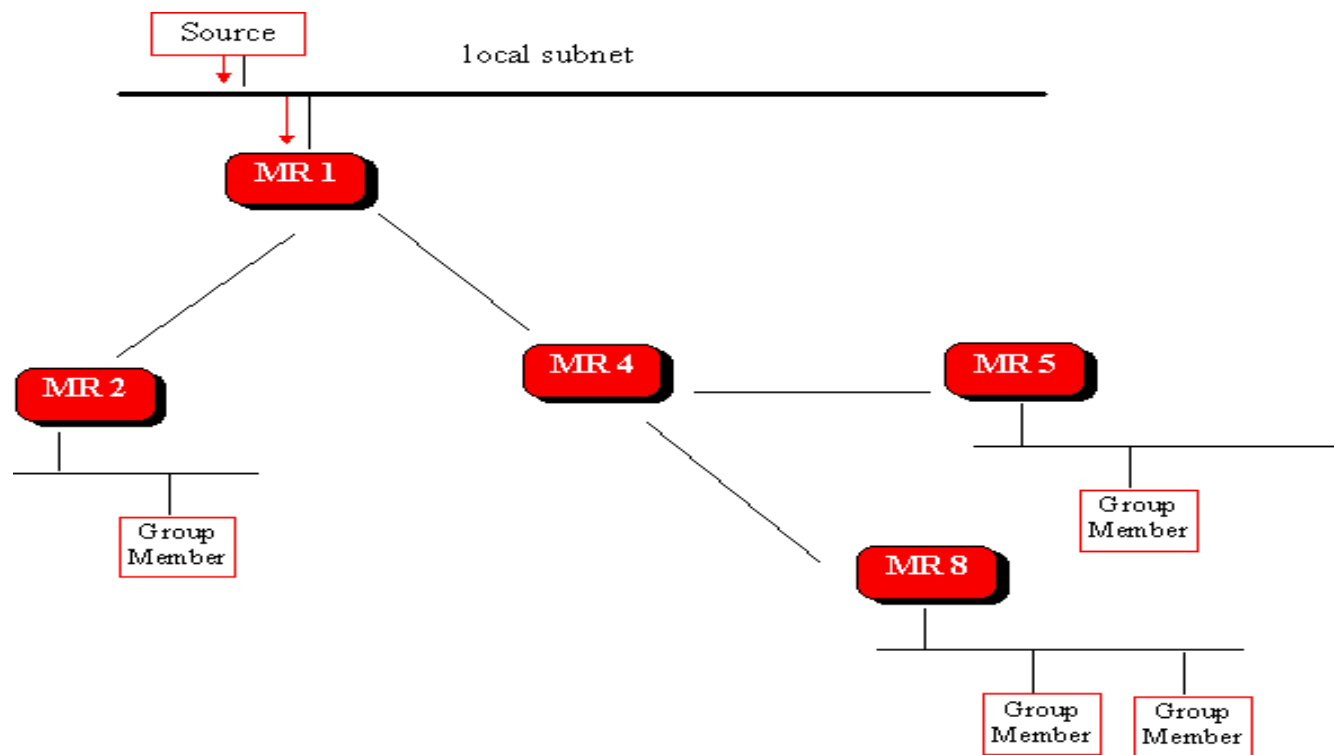
- **Distance Vector Multicast Routing Protocol**
 - D. Waitzman, C. Partridge, S. Deering, "Distance Vector Multicast Routing Protocol", RFC 1075, November 1988
<http://www.ietf.org/rfc/rfc1075.txt>
- Távolság-vektor (distance vector) alapú protokoll
 - A RIP unicast útválasztó protokollt használja

- Elárasztás és metszés (flood and prune)
 - Elárasztás
 - Ellenőrzi a csomag bejövő interfészét
 - Ha nem a legrövidebb út a forrás felé, eldobja a csomagot
 - Ha igen, továbbküldi a csomagot az összes többi interfészen
 - Metszés
 - Ha nincs érdekelt felhasználó egy helyi hálózaton
 - Ha nem a legrövidebb úton jött a csomag
 - Egy közbeeső router megjegyzi azokat az interfészeit, ahol Prune érkezett
 - Azokra az interfészekre nem küldi ki a további csomagokat
 - A Prune bejegyzések percenként elavulnak

DVMRP elárasztás



DVMRP metszés



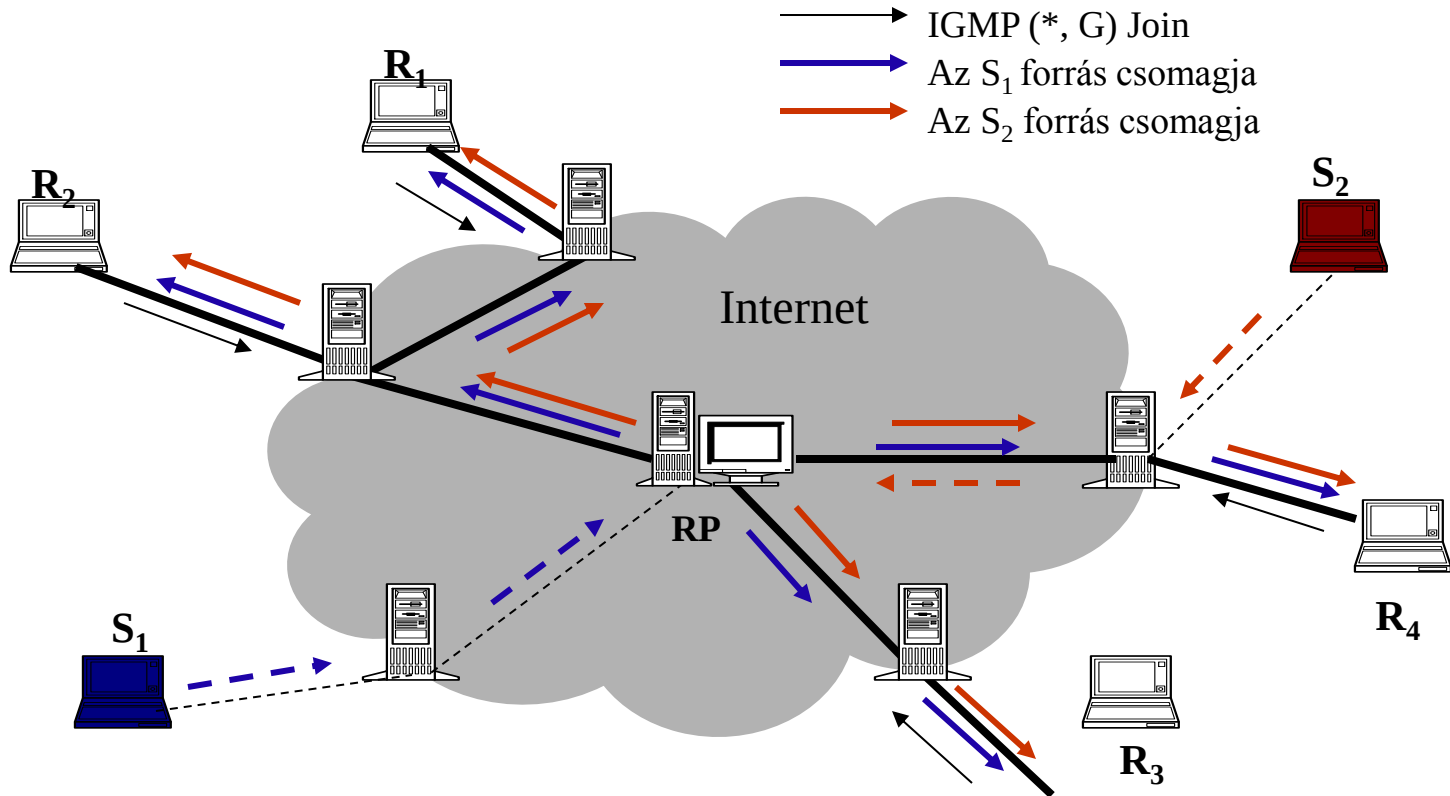
PIM

- Protocol Independent Multicast
 - PIM Dense Mode (PIM-DM)
 - PIM Sparse Mode (PIM-SM)
- PIM-SM
 - W. Fenner et al., „Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)” , RFC 4601, August 2006
 - A napjainkban legelterjedtebb multicast útválasztó protokoll

PIM-SM

- Egy közös multicast fát használ (**shared tree**)
- Kiválaszt egy „randevú” pontot (RP)
 - Az RP a közös fa gyökere
 - „Explicit join” – nem mindenki akarja hallgatni
 - Minden forrás az RP-hez küldi csomagjait
 - Az RP továbbküldi azokat a közös fán
 - Ha nagy forgalom, Source Register üzenettel bejelentkezik a forrás az RP-nél
 - az RP feliratkozik a forrás saját fájára
 - Bizonyos számú csomag után áttérés a közös fáról a forrás fájára
 - Nagy forgalmú forrásoknál az RP-nek csak közvetítő szerepe van, megismerteti a forrást a vevőkkel
 - Ha minden érdeklődő vevő már a forrás fáján van, az RP lecsatlakozik a forrás fáról

PIM-SM csomagküldés



Az ASM modell hátrányai

- Az ASM modell elterjedését több gazdasági és technikai tényező gátolta
 - Bonyolult címkiosztás
 - Dinamikus címválasztás a forrás által
 - Komplex címallokációs megoldások az ütközések elkerülésére
 - GLOP (RFC 3180) – AS-ekhez statikusan rendelt multicast címek
 - » Autonomous System – pl. ISP hálózata
 - MALLOC - Multicast Address Allocation Architecture (RFC 2908)
 - » MADCAP – Multicast Address Dynamic Client Allocation Protocol
 - » AAP – Multicast Address Allocation Protocol
 - » MASC – Multicast Address Set Claim

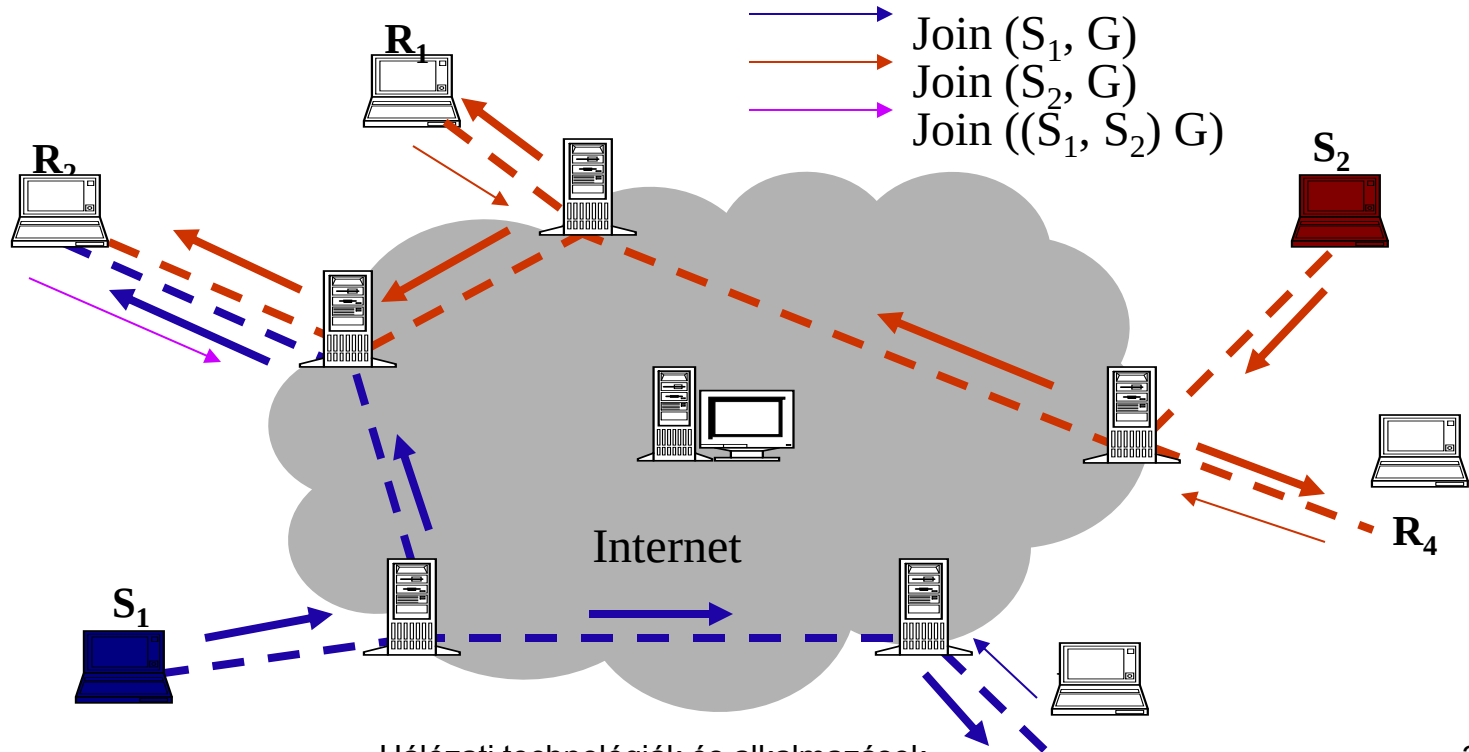
Az ASM modell hátrányai

- Túl nyílt modell a szolgáltatók számára
 - A források és vevők ellenőrizhetetlensége
 - Nehezen megoldható számlázás
- Nem skálázható a tartományok közötti útválasztás
 - PIM-SM csak egy tartományon belül
 - Egy ISP nem szereti ha forgalmát egy másik ISP-n belüli RP ellenőrzi
 - A tartományok között más protokollok
 - MSDP – Multicast Source Discovery Protocol
 - MBGP – Multicast Border Gateway Protocol

Az SSM modell

- Egy egyszerűbb modellre volt szükség
- **SSM - Source Specific Multicast**
 - Az Express modellre alapul
 - H. Holbrook, D. Cheriton, "IP Multicast Channels: Express Support for Large-Scale Single-Source Application", in *Proceedings of ACM SIGCOMM'99*, Cambridge, MA, USA, Sept. 1999.
- A $(*,G)$ multicast csoport helyett az (S,G) multicast csatornát használja
 - S a forrás unicast címe
 - G a csoport multicast címe
 - Csak az S forrás küldhet csomagokat az (S,G) csatorna vevőihez
 - Az adatátvitel egy forrás-specifikus fa mentén történik

SSM csomagküldés



Forrás szűrés

- Az SSM-hez szükség van forrás szűrésre
 - A felhasználó nem csak azt mondja meg a helyi router-nek, hogy melyik csoportot hallgatja, hanem hogy azon belül melyik forrást is
- IPv4 – IGMPv3
 - B. Cain, et. Al, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.
<http://www.ietf.org/rfc/rfc3376.txt>
- IPv6 – MLDv2
 - R. Vida, L. Costa, „Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004.
<http://www.ietf.org/rfc/rfc3810.txt>

Üzenet típusok

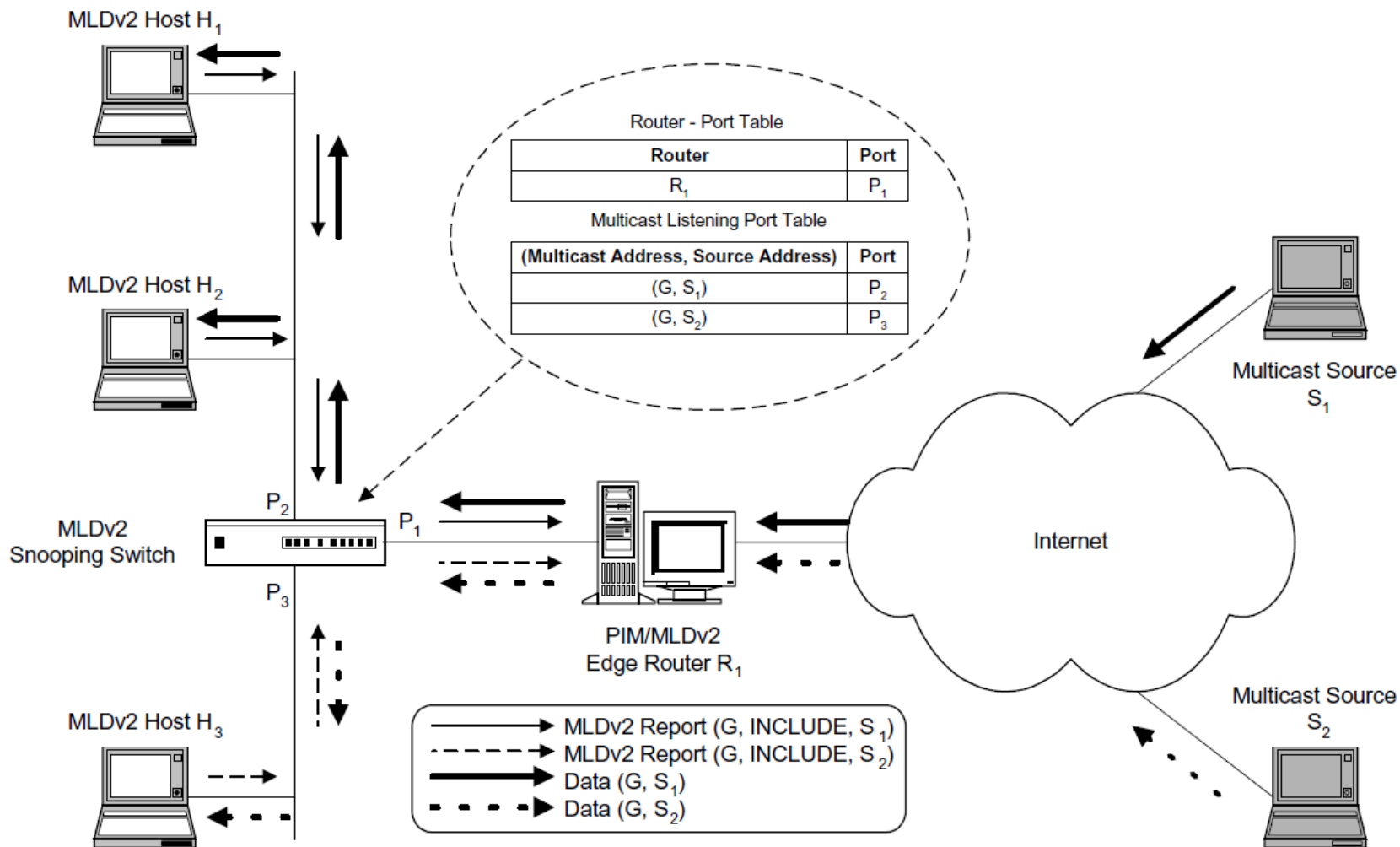
- IGMP/MLD **Query**
 - General Query
 - Ki mit hallgat?
 - Group Specific Query
 - Hallgatja-e valaki ezt a csoportot?
 - **Group and Source Specific Query**
 - Hallgatja-e valaki ezt a forrást ebben a csoportban?
- IGMP/MLD **Report**
 - Current State Record
 - Mit hallgatok – pl. Include (A) vagy Exclude (B)
 - A és B forráscím halmazok
 - Filter Mode Change Record
 - Szűrési mód váltása (Include vagy Exclude)
 - Source List Change Record
 - Allow (A) vagy Block (B)

Nincs host suppression!

- Túl komplex lenne

IGMP/MLD Snooping Switch

- A helyi hálózaton nem szeretnénk minden részt elárasztani a multicast forgalommal
 - Egy snooping switch kezeli ezt
- **Control message snooping**
 - Az IGMPv3 / MLDv2 Report üzeneteket csak a routerek felé küldjük
 - Mivel nincs host suppression, nem kell a többi vevő felé is küldeni
- **Multicast data snooping**
 - Csak oda küldjük az adatot, ahol vannak érdeklődők



IP Multicast

- Több éven át folyamatosan a jövő „forradalmi technológiájának” tartották
- Előnyök
 - Hatékony adatátvitel
 - A legrövidebb úton (DVMRP, MOSPF, PIM-SSM)
 - Figyelembe véve a fizikai topológiát
 - Hatékony erőforráskihasználás
 - Egy csomagot egy link-en csak egyszer küld át
 - Skálázható megoldás nagyméretű csoportok kommunikációjára
 - A csoportot egy virtuális cím azonosítja
 - Senki nem tartja számon a csoporttagok számát és kilétét

IP Multicast

- Mégsem terjedt el a várt mértékben
 - Technikai és gazdasági tényezők miatt
- Technikai hátrányok
 - Bonyolult címezés
 - Skálázható, tartományok közötti útválasztó megoldás hiánya
 - Rossz skálázhatóság a csoportok számát illetően
 - Egy router csoportonként egy bejegyzést tárol az útválasztó táblájában
 - A multicast címek nehezen aggregálhatók
 - Magasabb szintű szolgáltatások nehézkes támogatása
 - IP multicast egy *best-effort (több)pont-többpont* adatátviteli szolgáltatás
 - A végfelhasználók felelősek a felsőbb szintű szolgáltatások kezeléséért
 - Bonyolult torlódás vezérlés és megbízható adatátvitel

IP Multicast

- Gazdasági tényezők
 - Lassú és nehézkes telepítés a hálózatban
 - Noha a router-ek ma már képesek a multicast kezelésére, az ISP-k nem mindig aktiválják a hálózatukon
 - Csak akkor működik hatékonyan, ha minden router alkalmazza
 - Különben alagutazásra van szükség
 - „Tyúk-tojás” probléma
 - Az ISP-k nem támogatják, mert nincs elegendő multicast alkalmazás, nincs kellő kereslet
 - A szoftware cégek nem fejlesztenek multicast alkalmazásokat, mert nincs hálózati támogatás, nem lehet majd őket eladni
 - Nincs megfelelő gazdasági modell mögötte
 - Az ISP számára nehezen ellenőrizhető az erőforrásfelhasználás
 - A tartalom-szolgáltató számára nehezen ellenőrizhető ki használja a szolgáltatást
 - Nincs megfelelő számlázási megoldás

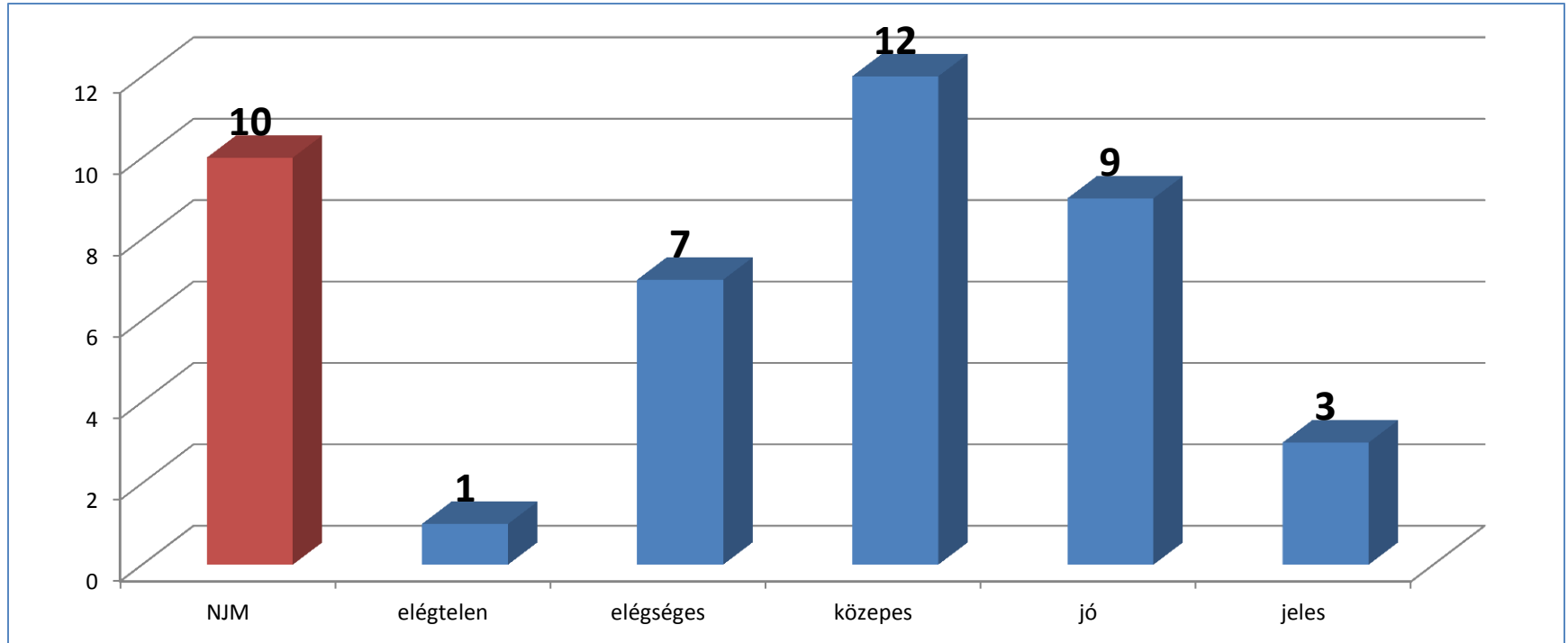
Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

2016. november 16.

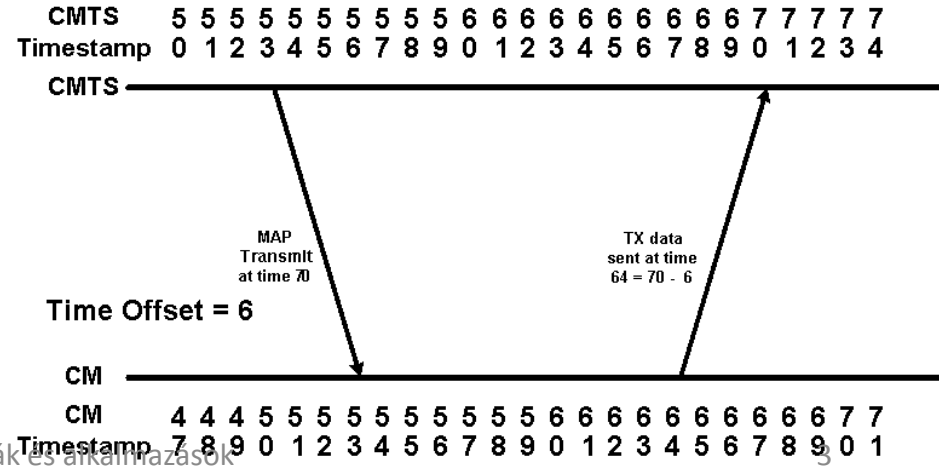
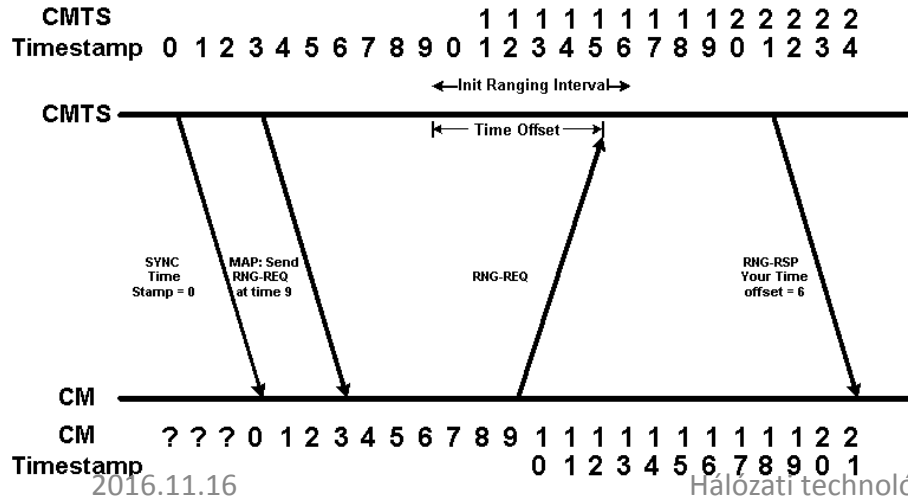


ZH eredmények



Ranging

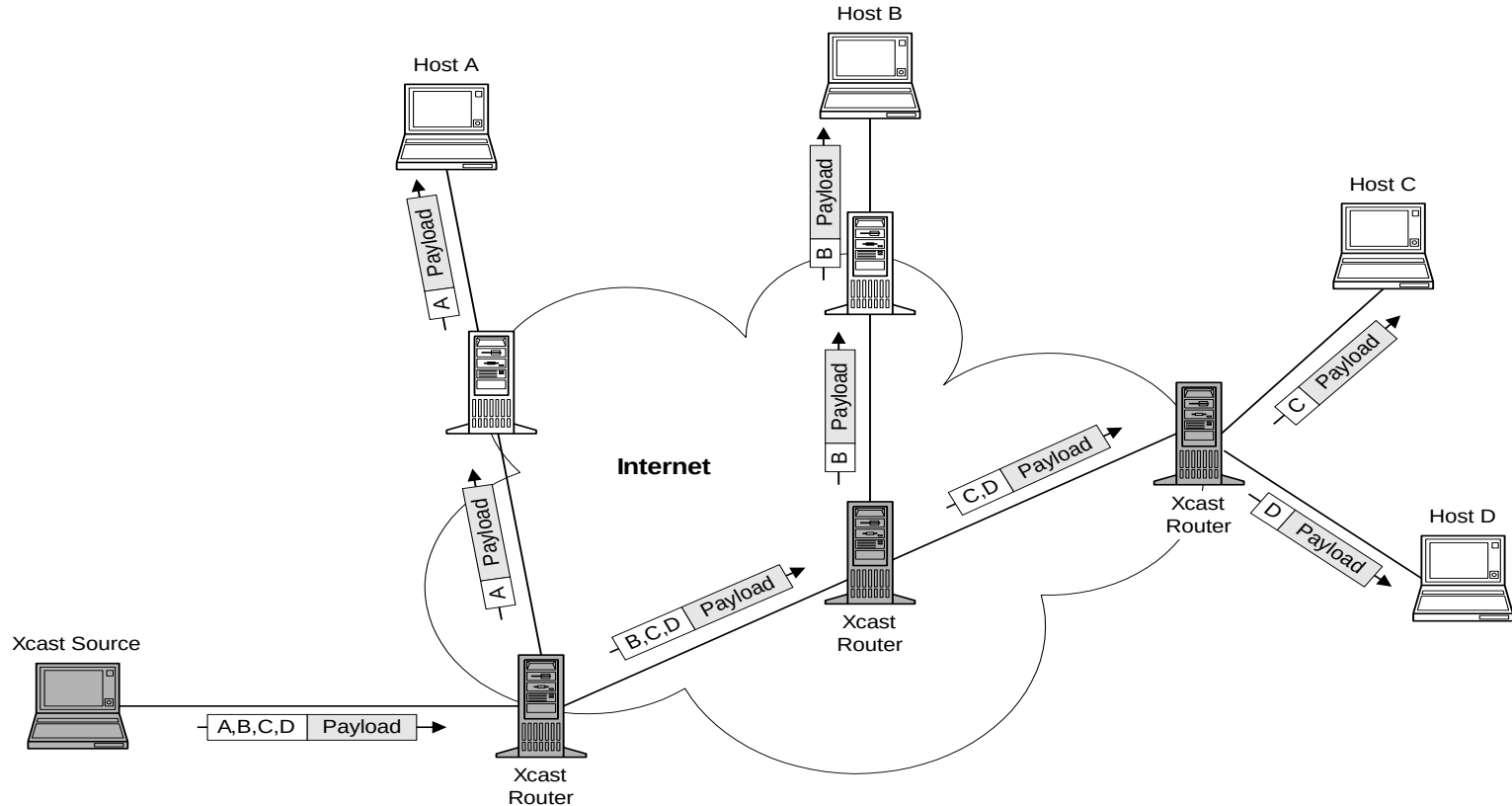
- A modem megméri milyen távol van a fejállomás
 - Távolságbecslés (ranging) – mint a ping
 - Szükség van rá az időzítések miatt



Explicit Multicast (Xcast)

- Hálózati rétegbeli multicast megoldás
- Nem használ multicast címezést
 - A forrás a csomag fejlécében tárolja az összes célállomás unicast IP címét
- A közbeeső Xcast router-ek duplikálással új csomagokat hoznak létre, a saját unicast útválasztó tábláik bejegyzései alapján
 - A router ellenőrzi hogy a kapott csomag fejlécében lévő célállomások felé melyik interfészen kell továbbítani a csomagot
 - Ennek megfelelően készíti el a duplikált csomagok fejléceit

Explicit Multicast (Xcast)



Explicit Multicast (Xcast)

- Nem skálázható megoldás nagy csoportokra
 - A fejlécbeli címek több helyet foglalhatnak el mint a tényleges adat
- Jól skálázható viszont nagyszámú kis csoport támogatására
 - A router-eknek nincs szükségük multicast útválasztó táblákra
- R. Boivie, N. Feldman, C. Metz, "Small Group Multicast: A New Solution for Multicasting on the Internet", *Internet Computing*, vol. 4, no. 3, May/June 2000, pp. 75-79.

Alternatív multicast megoldások

C. Diot et al, "Deployment Issues for the IP Multicast Service and Architecture", *IEEE Network Magazine, Special Issue on Multicasting*, vol. 14, no. 1, January/February 2000, pp. 78-88.

Van-e olyan csoportkommunikációs megoldás, mely ne szoruljon az ISP-k hálózati rétegbeli támogatására?

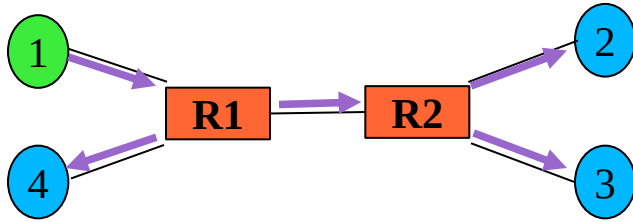
ALM – Application Layer Multicast
vagy...

ESM – End System Multicast
vagy..

HBM – Host-based Multicast

IP multicast - ALM

- IP multicast



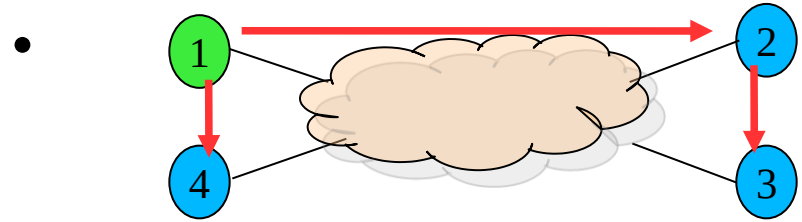
- Duplikálás a router-ekben

- Hálózati támogatás

- A topológia függ...

- az útválasztó tábláktól
- a fizikai topológiától

- ALM



- Duplikálás a végfelhasználóknál

- Nem igényel hálózati támogatást

- Virtuális topológia

- A fizikai topológia egy „fekete doboz”

ALM: motiváció

- Adatátvitel

- Nem szükséges IP multicast támogatás
 - Kizárólag unicast kommunikációra épít
- Kis csoportok
 - Az IP multicast nem mindig a legjobb megoldás
- Az adatok aktív felhasználása
 - Az adatokat lehet módosítani/értékelni az átvitel folyamán
 - Az átviteli struktúra módosítható az adatok függvényében

- Kontroll

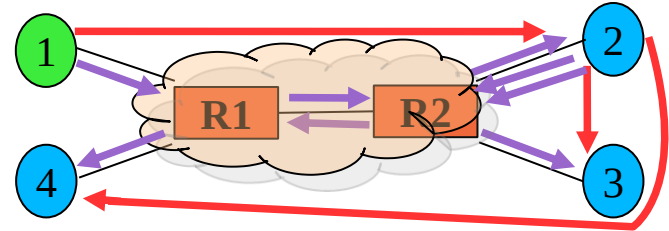
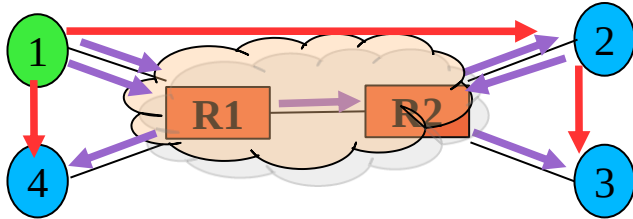
- A kontroll adatok aggregálása (Megbízható multicast)

ALM: előnyök

- **Általános megoldás**
 - Nincs szükség hálózati támogatásra, bármilyen hálózat felett működhet
 - Felhasználhatóak a létező kommunikációs mechanizmusok
 - Pl. a TCP torlódás vezérlése
- **Skálázhatóság**
 - A router-ek nem tárolnak csoportonkénti bejegyzéseket
 - A peer-ek igen, de ők kevés csoportban vesznek részt
- **Egyszerűen telepíthető**
 - Nem szükséges a hálózat belső elemeit (router-ek) módosítani
 - Csak a végső felhasználónál kell telepíteni
- **Különböző metrikák a szomszédok kiválasztására**
 - Különböző topológiák
 - Nagyon precíz topológia kontroll lehetősége

ALM: hátrányok

- Hatékonyság
 - End-to-end “ágak”
 - A késleltetés nagyon nagy lehet
 - Nagy erőforrás (sávszélesség) pazarlás



- Skálázhatóság
 - Peer-ek közötti kapcsolatok folyamatos értékelése
 - Teljes gráf: $n \cdot (n-1)$ virtuális kapcsolat egy n tagú csoportban

ALM: hátrányok (2)

- Stabilitás
 - A résztvevők stabilitása
 - Az overlay hálózatban a résztvevők („router-ek”) a végfelhasználók
 - Kevésbé megbízhatóak mint egy valódi router
 - Jönnek-mennek a hálózatban
 - A mérési adatok stabilitása
 - Az overlay hatékonysága a választott metrika stabilitásától is függ
 - RTT, sávszélesség, stb.
 - Mérlegelni kell a hatékony adatátvitelt a többletterhelés függvényében

Általános megjegyzések

- A legtöbb ALM megoldás megpróbálja kiküszöbölni az előbbi hátrányokat
 - Egy dolog biztos: **ALM soha nem lesz olyan hatékony mint az IP Multicast**
 - Különböző megoldások különböző kompromisszumokra alapulnak
- Mivel „multicast” kommunikációról van szó, a cél egy adatátviteli fa építése
 - A fa lehet szabályozott (maximalizálva a lehetséges gyerekek száma) vagy nem
 - A fa „minősége” attól függ, hogy mire akarjuk használni

ALM - Általános koncepció

- Az ALM megoldások két topológiába szervezik a résztvevőket
 - Kontroll topológia („mesh” - szövevény)
 - A kontroll topológia tagjai periódikus frissítő üzenetekkel ellenőrzik a „szomszédai” jelenlétét
 - hibák felderítése, kezelése
 - Adatátviteli topológia („tree” – fa)
 - Az adatátviteli topológia a kontroll topológia egy része, melyet adatátvitelre használunk
- Ezen topológiák kiépítésének sorrendje alapján, az ALM megoldások lehetnek:
 - Mesh-first: Narada
 - Tree-first: Yoid, HMTP, TBCP, Overcast, ALMI
 - Implicit: CAN-Multicast, Scribe, Bayeux, NICE

Narada

Hindu bölcs

<http://en.wikipedia.org/wiki/Narada>

Y. Chu, S. Rao, H. Zhang, "A case for End System Multicast", Proceedings of ACM Sigmetrics, June 2000

<http://www.cs.cmu.edu/~srini/Papers/2002.Chu.jsac.pdf>



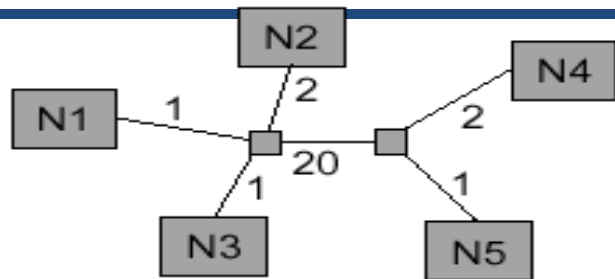
Narada

- Elosztott algoritmust használó, önszervező és önoptimalizáló overlay megoldás
- Mesh-first algoritmus
 - Először egy kétirányú mesh-t épít ki a résztvevők között
 - Egy Reverse Path Forwarding (RPF) megoldást használva kiépít egy legrövidebb útvonalú fát (Shortest Path Tree, SPT) a mesh felett
- Következmények:
 - A multicast fa minősége a mesh minőségétől függ
 - A fa kiépítése nem központosított
 - Forrásonkénti egyirányú fákat épít

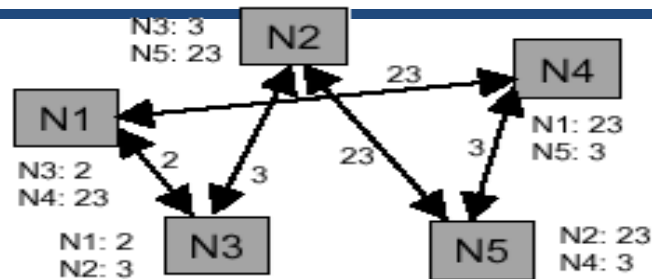
Narada

- Egy új résztvevő megszerzi valahogy néhány aktív résztvevőnek a címét
 - www, e-mail, randevú pont (RP)
- Kiépít egy parciális gráfot (csatlakozik a mesh-hez)
 - Az elején véletlenszerűen választ néhány szomszédot
 - Minden résztvevőnek szabályozva van hány szomszédja lehet
 - Ha valaki elfogadja szomszédjának, a csatlakozás megtörtént

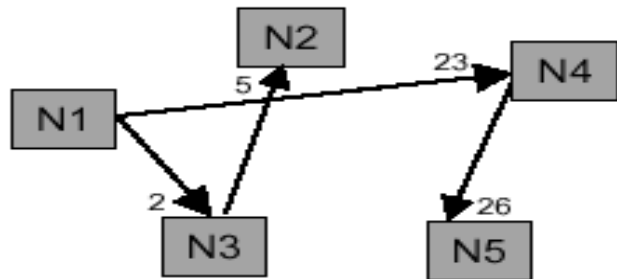
Narada



1. Fizikai topológia



2. Overlay mesh (max 2 szomszéd)



3. Adatátviteli fa

- ☐ A mesh kétirányú
- ☐ Forrásonként egy külön, egyirányú fa
- ☐ Ha N1 a forrás, akkor N2 nem küld csomagokat N5 felé, mert a legrövidebb út N1-től az N4-en át vezet

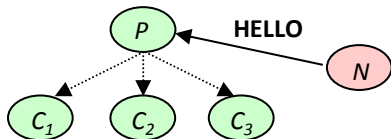
TBCP

- Tree Building Control Protocol
- L. Mathy, R. Canonico, D. Hutchison. *An overlay tree building control protocol*. In Proceedings of International Workshop on Networked Group Communication (NGC), London., 2001.
- <http://citeseer.ist.psu.edu/mathy01overlay.html>
- Tree-first protokoll
- Peer párok közötti mérésekre alapul
- Az adatátviteli fát lokális „teljes mesh”-re alapuló döntések sorozata alapján állítja össze

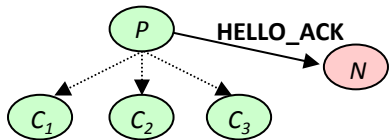


TBCP algoritmus

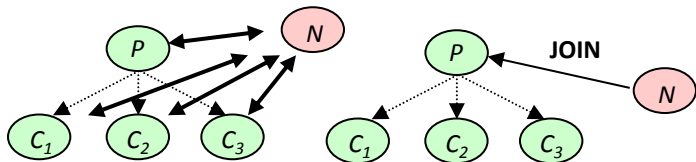
- Alap ötlet:
 - Minden peer a forráshoz küldi először csatlakozó üzenetét
 - A peer-ek „dominóként hullanak” a fa mentén



1) N **HELLO** üzenetet küld a P gyökérnek

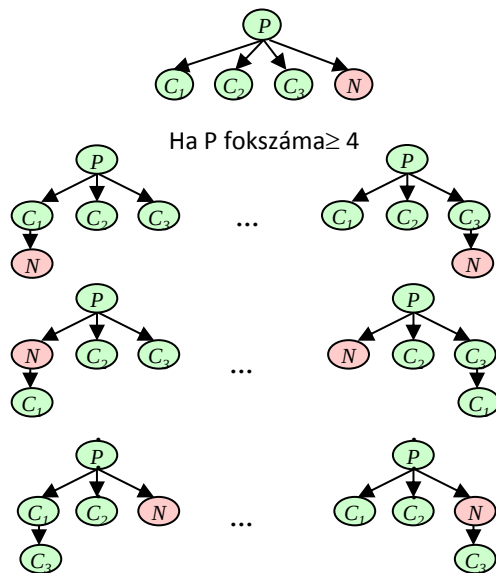


2) P válaszol egy **HELLO_ACK** üzenettel melyben elküldi a gyermekei listáját (C_i)



3) N leméri a távolságát a P és a C_i gyerekek felé, és elküldi azt egy **JOIN** üzenetben

TBCP algoritmus (2)

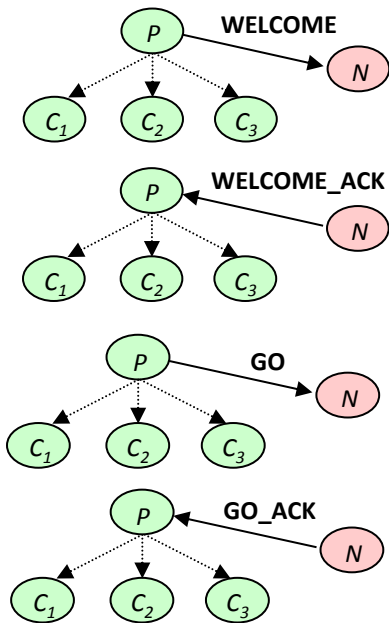


- Meghatározunk egy érték függvényt, mely a konfigurációk minőségét jellemzi
- Ezt a függvényt használva P értékeli az összes lehetséges konfigurációt, és kiválasztja a „legjobbat” (helyi döntés)
- Különböző metrikákat használhatunk
- Különböző értékek $\rightarrow$ különböző fák
- Melyik a jobb fa? Attól függ mire akarjuk használni...

Előny/hátrány:

a fát helyi döntések sorozata alapján építjük ki

TBCP algoritmus (3)



- Ha P elfogadja N -t mint saját gyereket, küld egy **WELCOME** üzenetet
- Ha P „meg akar szabadulni” N -től (vagy valamelyik gyerekétől), küld egy **GO(C_k)** üzenetet
- P , vagy a C_i peer amelyik megkapja a **GO(C_k)** üzenetet újratekdi az algoritmust, és elküld egy **HELLO** üzenetet C_k -nak
 - C_i nem csatlakozik le azonnal a fáról, kap egy kis időt az új kapcsolat felépítésére

Értékelés

- Viszonylag jól skálázható megoldás
 - Semmilyen információ nem szükséges a fizikai topológiáról
 - Nem szükséges ismerni a csoport összes tagját
 - Elosztott megoldás
 - Több peer akár egyszerre is csatlakozhat
 - Egyszerre csak egy peer csatlakozhat egy gyökérhez
- Könnyen telepíthető
 - Viszonylag jó fát épít viszonylag gyorsan
- Implementációs hack a hatékonyság javítására
 - Ha túl sokáig tart egy cél felé a mérés, akkor állítsuk az értéket végtelenre

Hálózati Technológiák és Alkalmazások

Vida Rolland
BME TMIT

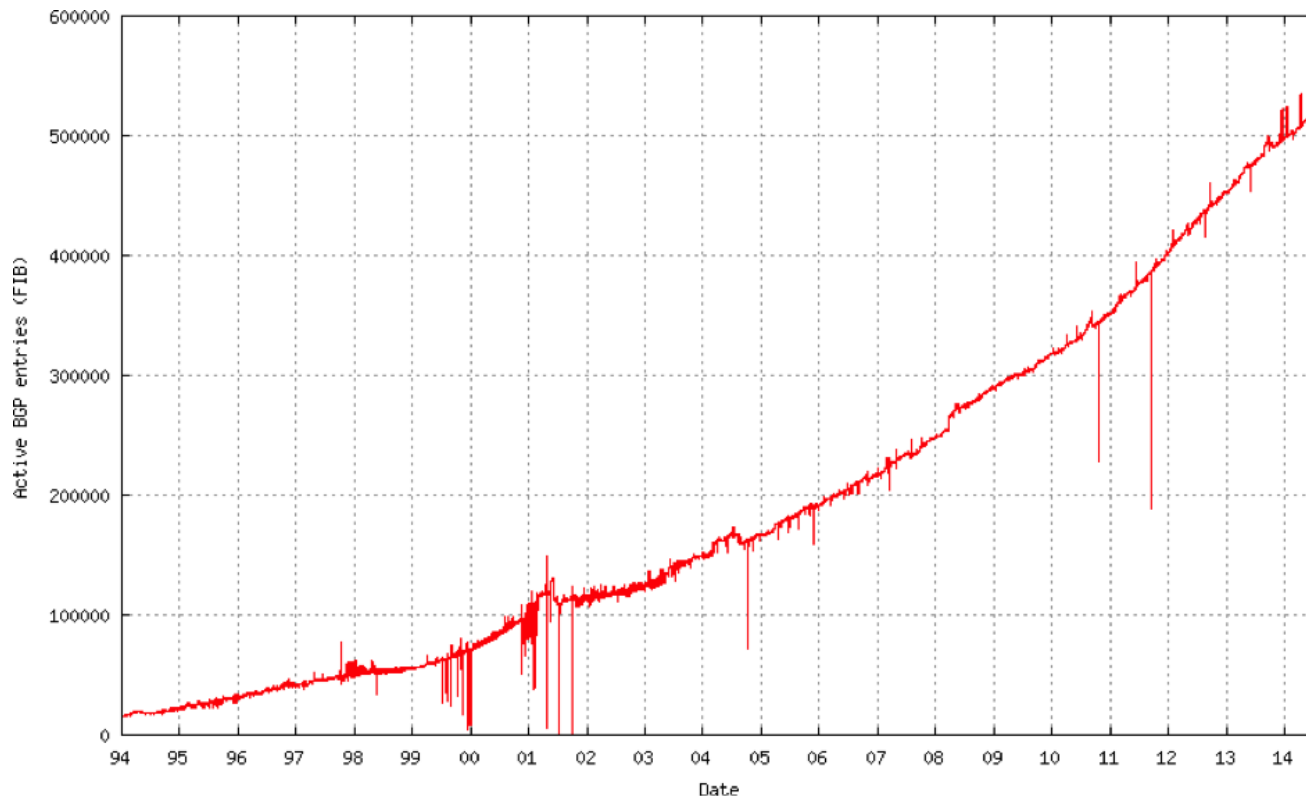
2016. november 18.



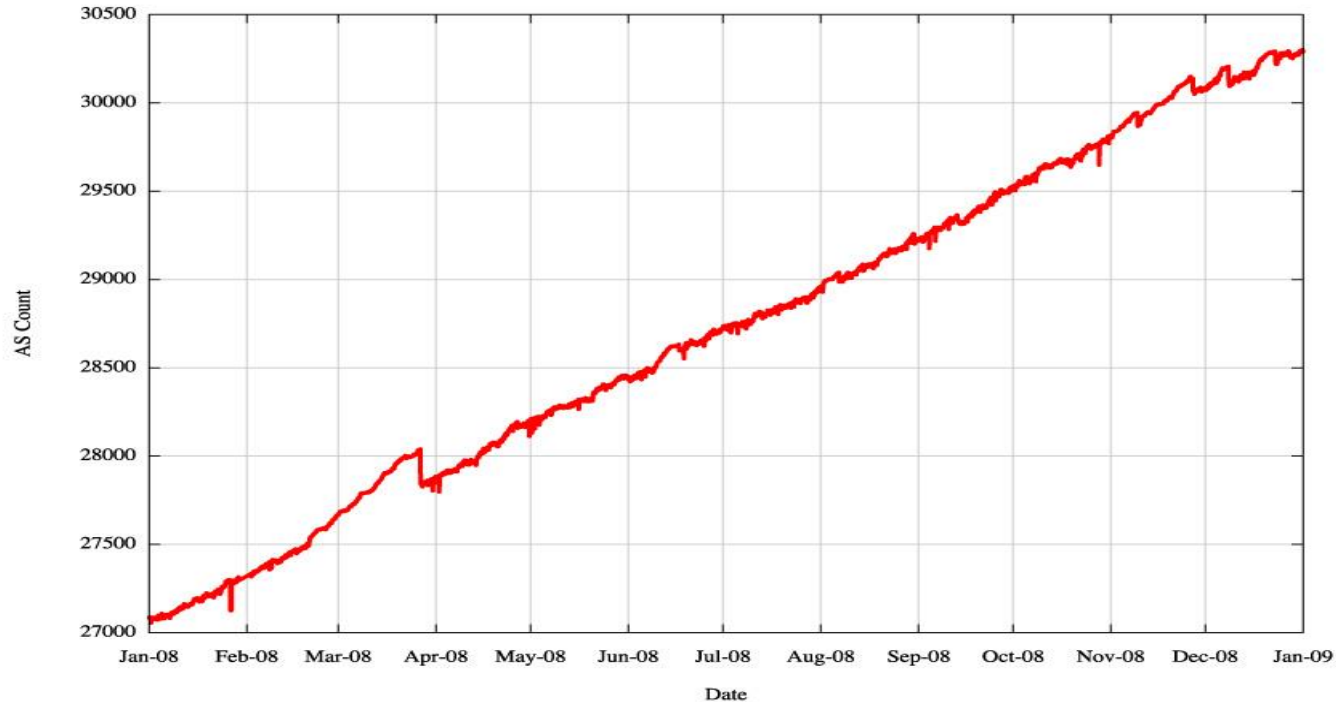
Miért IPv6?

- Gondok az IPv4-el
 - Az IPv4-et a 70-es évek elején dolgozták ki
 - Folyamatos foltozgatás
 - MobileIP
 - QoS
 - Security (IPsec)
- A 80-as években optimalizálás és aggregálás nélküli „esztelen” címkiosztás
 - Következmény
 - nagy útválasztó táblák a gerinchálózatban
 - Lehetséges megoldás
 - IP újracímzés és a felhasználatlan címtartomány újraosztása
 - Nem egyszerű kivitelezni

IPv4 – BGP táblák növekedése



IPv4 – egyre több AS



Elvárások az IPv6-al szemben

- Nagyobb címtartomány
- Hierarchikus címkiosztás (útválasztás támogatása)
- QoS architektúrák támogatása
- Mobilitás támogatása
- Végpontok közötti biztonságos adatátvitel támogatása
- Egyszerű hálózatmenedzsment
- Automatikus konfiguráció
- Multicast támogatás

IPv6 kronológia

- TUBA (1992)
 - TCP and UDP over Bigger Addresses
 - OSI CLNP (Connection-Less Network Protocol) protokollra épülve
 - elvetették
- SIPP (1993)
 - Simple IP Plus
 - 64 bites címek
- IPng egy kiterjesztett SIPP verzióra épülve (1994)
 - 128 bites címek
 - 1995 decemberétől hivatalosan IPv6

Az IPv6 címezési rendszere

- Az IPv6 címtér rendkívül nagy
 - $2^{128} = 340.282.366.920.938.463.463.374.607.431.768.211.456$
 - 67 milliárd milliárd cím a Föld területének minden cm^2 -ére
 - 10^{30} cím a Föld minden lakosának
- Az IPv6 cím típusát a cím kezdő bitjei szabják meg
 - hosszuk változó - Format Prefix (FP)

IPv6 címek írásmódja

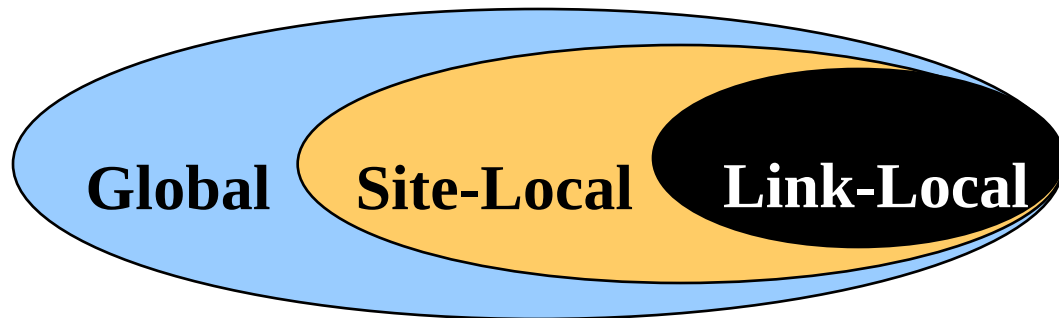
- 128 bit = 16 byte = 32 x 4bit hexadecimális írásmóddal
FECD:BA98:0000:0000:00CD:BA98:0000:3200
- A kezdő nullák elhagyhatóak
FECD:BA98:0000:0000:00CD:BA98:0000:3200 helyett
FECD:BA98:0:0:CD:BA98:0:3200
- A sorozatos nullák kihagyhatóak
FECD:BA98::CD:BA98:0:3200
- Hálózati prefix jelölés a CIDR-ben használttal megegyező módon
teljes IPv6 cím/prefix hossz bitekben
12AB:0000:0000:CD30:FFFF:DEC8:0000:0000/60
12AB:0:0:CD30:0:0:0:0/60
12AB:0:0:CD30::/60

Az IPv6 címzési rendszere

- Három típus:
 - Unicast címek
 - egyedi interfészt azonosítanak
 - Multicast címek
 - interfészek egy csoportját azonosítják, a csomagot ezek mindegyikéhez eljuttatják
 - Helyettesítik a broadcast címeket is
 - Anycast címek
 - interfészek egy csoportját azonosítják, a csomagot ezek egyikéhez juttatják el.

Az IPv6 címzési rendszere

- Egy interfésznek több címe is lehet, különböző hatáskörrel:
 - Link Local
 - Site Local
 - Global



Unicast címek

- **Unspecified Address**

- Helyettesítőként használt cím, ha nincs más
 - Pl. kezdeti DHCP kérés
- Mint a 0.0.0.0 IPv4-ben
- 0:0:0:0:0:0:0:0 vagy ::

- **Loopback cím**

- Saját magad azonosítására
- Mint a 127.0.0.1 IPv4-ben
- 0:0:0:0:0:0:0:1 vagy ::1
- Pl. hogy ellenőrizzük, hogy az IPv6 stack működik-e
 - Ping6 ::1

Unicast címek

- **Scope = lokális link**
 - Ugyanazon a linken levő csomópontok közötti kommunikációra
 - Csak linken egyedi, nem lehet vele a linken kívül kommunikálni
 - Automatikusan konfigurálva minden interfészen
 - Minden IPv6-os eszköznek egy kezdeti címe, amivel elkezdhet kommunikálni
 - Szomszéd felderítés, router felderítés
 - Formátum:
 - FE80:0:0:0:<interface identifier>
 - Az interfész ID – EUI (64) cím
 - A korábbi 48 bites MAC cím kiterjesztése

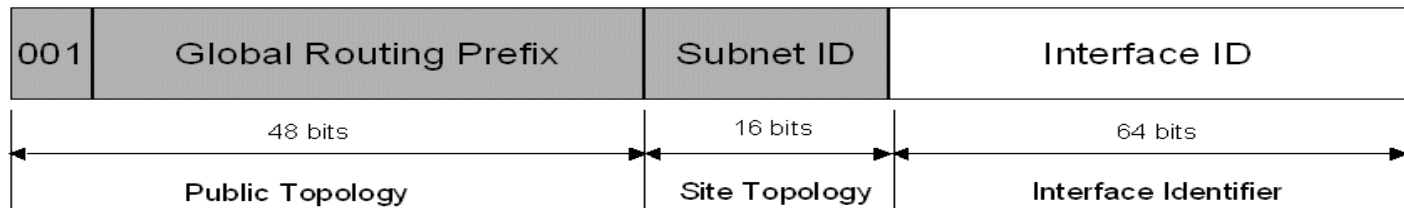
Unicast címek

- **Scope = site local**
 - Ugyanazon a site-on belül lehet csak kommunikálni vele
 - A router-ek nem küldik ki a site-on kívülre (az Internetre)
 - Nagyon hasonló az IPv4-es privát címekre
 - Nem automatikusan konfigurált cím
 - Formátum:
 - FEC0:0:0:<subnet id>:<interface id>
 - Subnet id = 16 bit = 64K subnet
 - Egy teljes szervezeti hálózat (cég, egyetem) címzését lehetővé teszi
 - Pl. megcímezzük a hálózatot site-local címekkel
 - Újracímezzük (renumbering) amikor az IPv6 hálózatra csatoljuk
 - Lecseréljük az első 48 bitet (site ID)
 - Újracímézhethetjük ha más szolgáltatóhoz csatoljuk

Unicast címek

- **Global Unicast Address**

- Globális kommunikációra használják
- Hierarchikus globális prefix
 - Az RIR-ek és az ISP-k strukturálják
- Alhálózat azonosító
 - Hierarchikusan struktúrált, a hálózati adminisztrátor által
- Interfész azonosító



Unicast címek

- IPv4 összerendelt címek (mapped addresses)
 - Az IPv4/IPv6 áttérés megkönnyítésére
 - Egy IPv4 címből egy IPv6 címet csinálunk
 - Az első 80 bit 0
 - A következő 16 bit 1
 - A maradék 32 bit az IPv4-es címnek
- Kevert írásmód

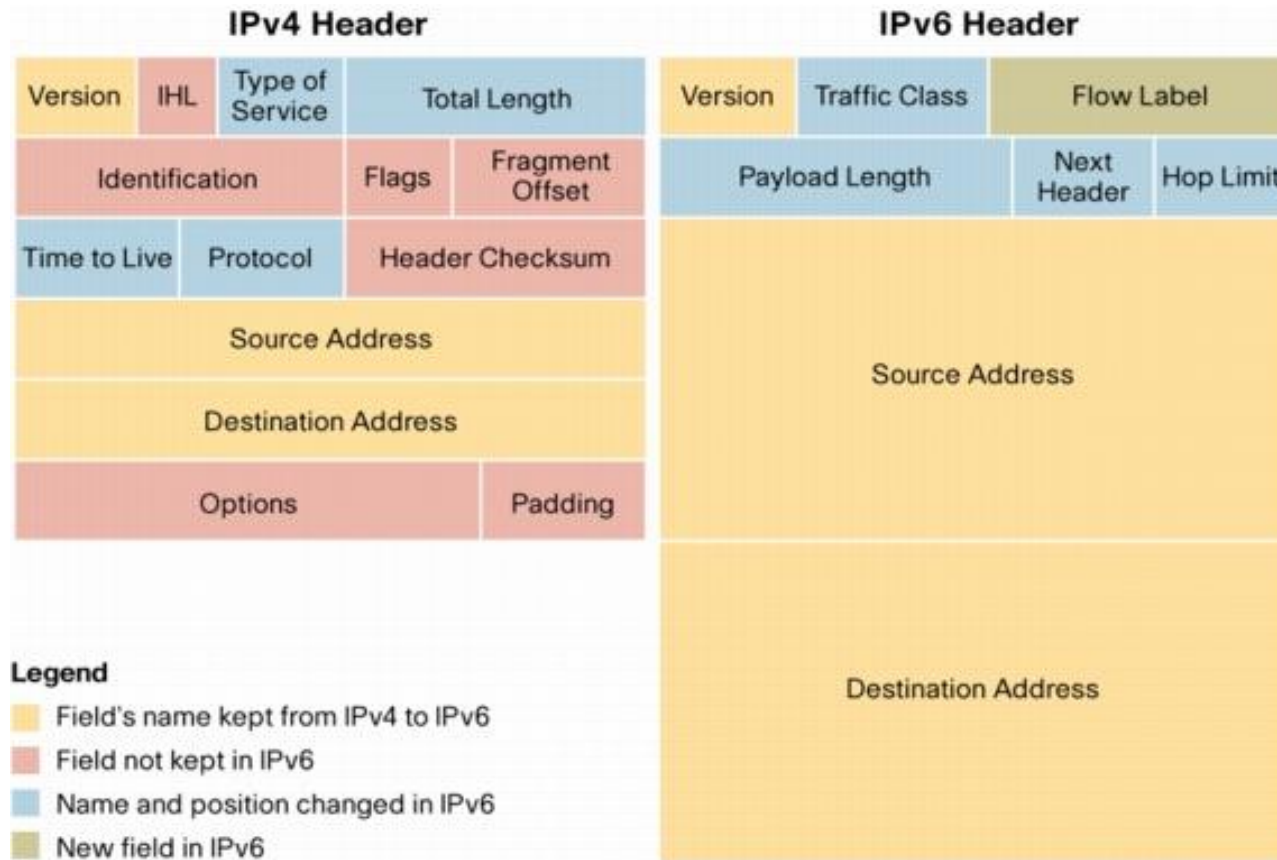
0:0:0:0:0:FFFF:192.0.3.128

::FFFF:192.0.3.128

Multicast címek

- A broadcast címzés helyett is
- Korlátozott hatókörű címek (scoped addresses)
 - Node, link, site, organisation, global
- Formátum:
 - `FF<flags><scope>::<multicast group>`
- Flag:
 - 0 – permanens
 - 1 – dinamikus
- Scope:
 - 1 – node
 - 2 – link
 - 5 – site
 - 8 – organisation
 - E – global
- Pl.
 - `FF02::1` – all nodes on the local network
 - `FF02::2` – all routers on the local network

Az IPv6 alap fejléc formátum



IPv6 alap fejléc

- **Version** – verzió (4 bit)
 - IP verziószáma
- **Class** - prioritás osztály (8 bit)
 - csomag prioritását definiálja
 - ToS (Type of Service) mező az IPv4-ben
- **Payload Length** (16 bit)
 - hasznos információ hossza, byte-ban

IPv6 alap fejléc

- Folyam címke (Flow Label)
 - A címke speciális QoS követelményű adatfolyamhoz rendelhető
 - 20 bit hosszú
 - A folyam címke kulcsként használható az útvonalválasztók tárolójában (cache) a feldolgozási idő csökkentésére
 - Egy datagram először érkezik az útvonalválasztóhoz
 - Elmenti a folyam címkéjét a tárolójában
 - Ha ezután ugyanilyen folyam címkéjű datagram érkezik...
 - nincs szükség az útvonalválasztó táblában való keresésre
 - azonnal továbbítható a csomag a folyam címke alapján.

IPv6 alap fejléc

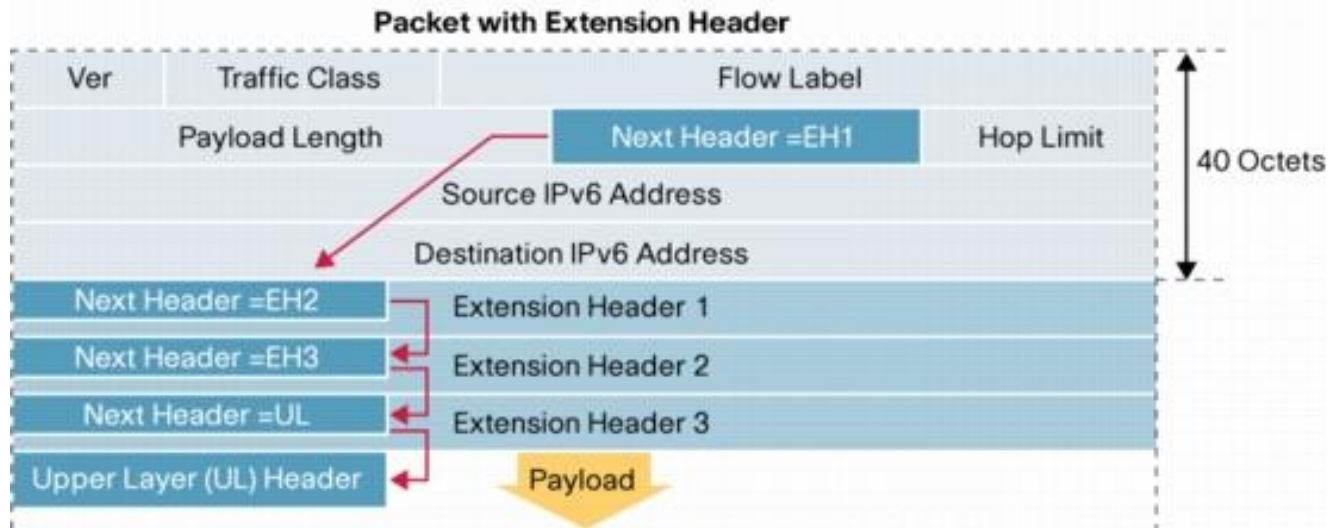
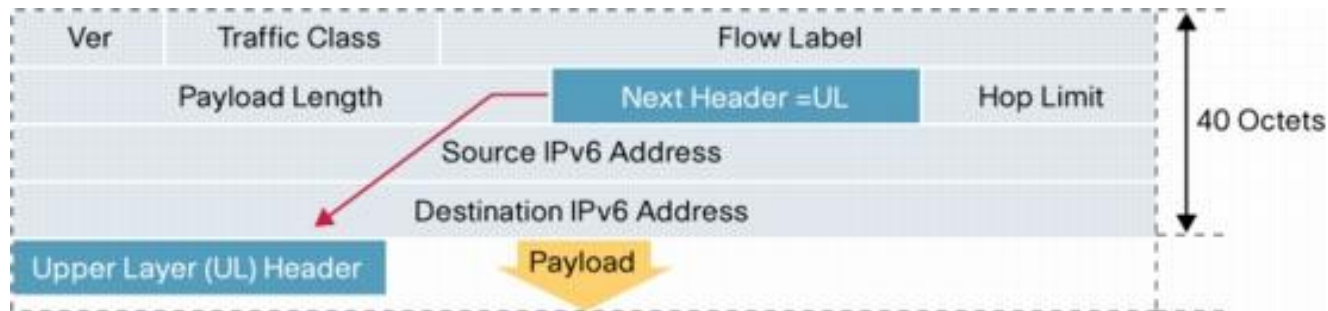
- **Next-Header** - következő fejléc (8 bit)
 - azonosítja az alap IP fejlécet közvetlenül követő fejlécet
 - Ez lehet kiegészítő fejléc vagy egy felső rétegbeli protokoll
- **Hop Limit** - ugrás szám (8 bit)
 - Megadja, hogy a csomag milyen messzire utazik
 - megegyezik az IPv4 Time To Live (TTL) mezőjével
- **Source Address** - forrás cím (128 bit)
 - a csomag eredeti küldőjének címe
- **Destination Address** - cél cím (128 bit)
 - a csomag szándékolt vevőjének a címe
 - nem biztos, hogy az utolsó vevő címe, ha opcionális Routing Header-t is tartalmaz a csomag

IPv6 kiegészítő fejlécek

- IP csomag - 40 byte hosszú alap fejléccel kezdődik
- A közbenső hálózatra vonatkozó extra információhoz kiegészítő fejlécek - **Extension Headers**
- Legtöbb kiegészítő fejléct az útvonalon található útvonalválasztók nem vizsgálják és nem dolgozzák fel, csak a célállomás.
- A kiegészítő fejlécek mindegyike saját egyedi értékkel rendelkezik a **next header** mező számára
 - így több kiegészítő fejléct is használhat egyszerre
 - az utolsó kiegészítő fejléc next header mezője azonosítja a felsőbb réteg protokollt
 - A fejléc tetszőleges hosszúságú lehet

IPv6 kiegészítő fejlécek

Összeláncolható
fejlécek



IPv6 kiegészítő fejlécek

- Fejléc sorrend
 - Fontos: kiegészítő fejlécek sorrendje az ajánlott sorrendben
 - egyszerűsödik az útvonalválasztók számára a csomag feldolgozása
 - legtöbb esetben az útvonalválasztók csak a hop-by-hop opciót és az útvonalválasztó fejlécet dolgozzák fel
 - Kivétel: **cél (destination) opció**
 - közvetlenül a felső réteg opció előtt áll
 - Ha a cél kiegészítő fejlécek az útvonalon található néhány útvonalválasztó számára is érdekesek, akkor közvetlenül az útvonalválasztó fejléc elé kell őket helyezni, mert azzal együtt dolgozandók fel
 - Egy csomag mindkét helyen tartalmazhat cél opciót.

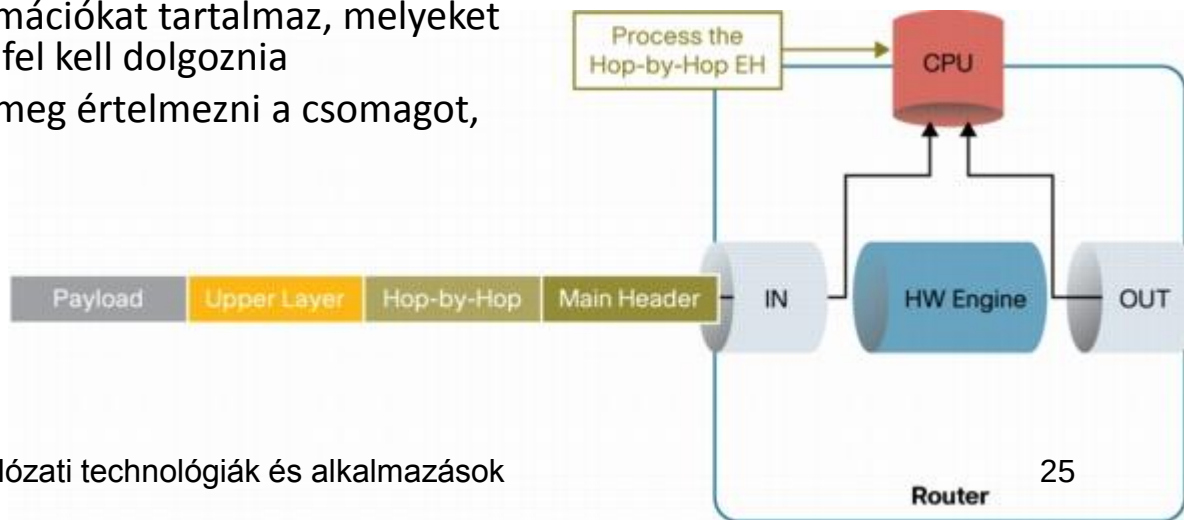
IPv6 kiegészítő fejlécek

- Az ajánlott fejlécsorrend:
 - IPv6 Header
 - Hop-by-hop Options Header (type = 0)
 - Destination Options Header (1)
 - Routing Header (type = 43)
 - Fragment Header (type = 44)
 - Authentication Header (type = 51)
 - Encapsulating Security Payload (ESP) (type = 50)
 - Destination Options Header (2) (type = 60)
 - Upper Layer Header (pl. TCP vagy UDP)

IPv6 kiegészítő fejlécek

- **Hop-by-hop Options Header**

- A csomag útvonalán található gépek számára tartalmaz IP opciókat
- Az útvonal minden útvonalválasztójának meg kell vizsgálnia és fel kell dolgoznia a Hop-by-hop Header-t
- **Router Alert opció** riasztja a tranzit útvonalválasztókat
 - Ha a csomag olyan információkat tartalmaz, melyeket egy közbeeső routernek fel kell dolgoznia
 - Különben nem próbálja meg értelmezni a csomagot, csak továbbküldi



IPv6 kiegészítő fejlécek

- **Routing Header**
 - Normál esetben az IP csomag forrása a hálózatra bízta a csomag eljuttatását a célhoz
 - Lehetőség van forrás oldali útvonal megadására az útválasztók címeivel
 - A teljes lista a Routing Header-ben (pl. A, B, C, D)
 - A célcím mindig a következő útválasztó címe (A), kivéve az utolsó útválasztót
 - A célcímet minden útválasztó átírja továbbítás előtt

IPv6 kiegészítő fejlécek

- **Fragment Header**
 - IPv4 – tördelés és visszaállítás automatikusan, ha explicit módon nem tiltják
 - IPv6 - alapértelmezésben a csomagokat nem tördelik
 - ha a csomag túl nagy a közeg számára, eldobják és ICMP üzenet
 - a host felderíti az átviteli közegre jellemző MTU-t
 - Maximum Transmission Unit
 - alapesetben megpróbál kisebb csomagokat küldeni, mint az MTU
 - ha tördelésre van szükség, az a Fragmentation Extension Header-el oldható meg
- **Authentication Header**
 - Garantálja, hogy a ...
 - vett csomag hiteles
 - nem változtatták meg az út során
 - megadott küldőtől érkezett
- **Destination Option Header**
 - A cél opció a cél számára tartalmaz IP opciókat

Áttérés az IPv6-ra

- IP-re épülő hálózati szolgáltatások
 - RIPv6(ng), OSPFv6 (v3), BGPv6
- IP-re épülő hálózati és szállítási protokollok
 - TCPv6, UDPv6, RSVPv6
- Alkalmazások
 - Minden alkalmazás, mely közvetlenül használta az IPv4-es címeket, nem független az alatta lévő rétegektől, így az IPv6 támogatást implementálni kell
- Fokozatos áttérés
 - Nincs „D-day”
- Elvárások az áttérést illetően
 - Ne legyenek áttérési függőségek
 - egy-egy csomópont áttérése függetlenül történhet
 - legfontosabb szempont a visszafelé kompatibilitás
 - A végfelhasználó számára minél egyszerűbb legyen
 - Az áttérési technikák egymástól függetlenül legyenek alkalmazhatóak
 - legalább a tartományok szintjén

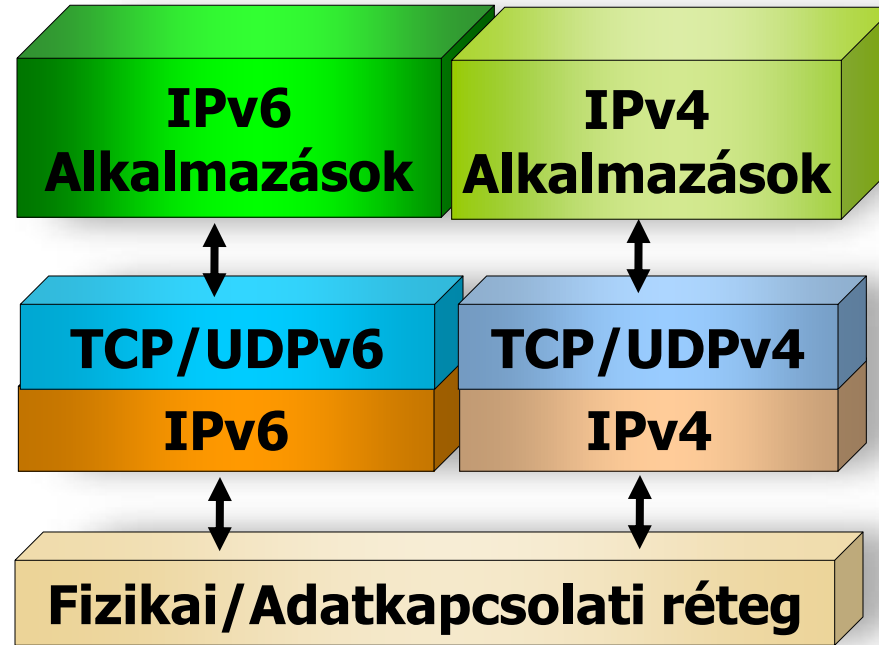
Áttérési megoldások

- **Dual Stack** (dupla protokoll verem)
 - IPv4/IPv6 egyszerre ugyanazon az eszközön
- **Alagutak**
 - Kezdetben, IPv6 csomagok alagutazása IPv4 felhőkben
 - Később IPv4 csomagok alagutazása IPv6 felhőkben
- **Protokoll fordítás**
 - Protokoll információkat hordozó fejlécből másik protokoll fejléc létrehozása fordítási szabályok alkalmazásával
 - IPv6 <-> IPv4

Dupla protokoll verem

- Az IPv6 felé tett első lépés olyan rendszerek telepítése, melyek támogatják az IPv6-ot.
 - ezek a rendszerek a kettős stack stratégián alapulnak, amely az IPv4 és IPv6 használatát is támogatja.
- Ezek a rendszerek
 - IPv6-ot használnak más IPv6 rendszerekkel való kommunikációra
 - képesek visszalépni IPv4 módba régi rendszerekkel való párbeszédhez

Dual Stack



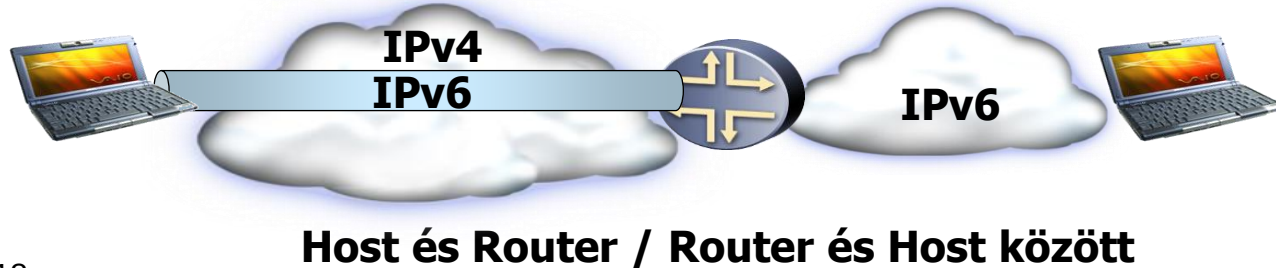
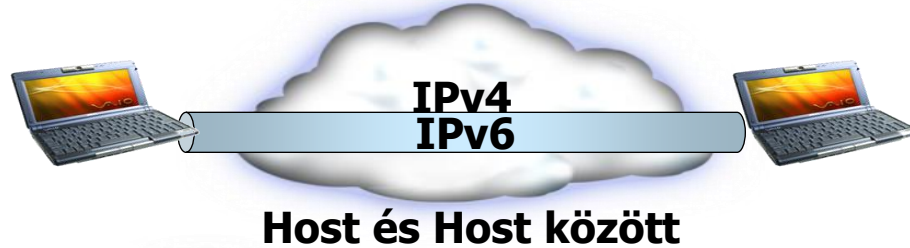
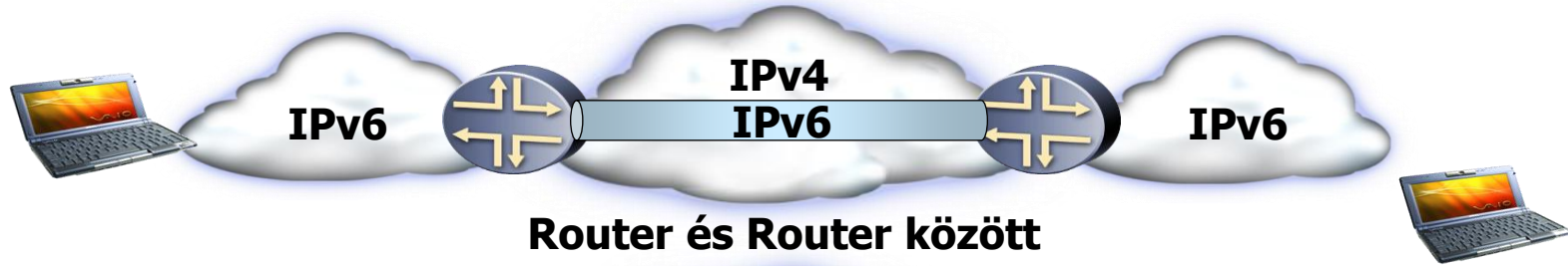
Dupla protokoll verem

- Alkalmazásának előnyei
 - egyszerű installálni, konfigurálni, karbantartani
 - az IPv6 teljes funkcionalitása kihasználható
 - bármely két csomópont tud egymással kommunikálni csak IPv4, vagy csak IPv6-os csomagokkal
 - átlátszó, az áttérés a felhasználók számára észrevétlenül történhet
- Alkalmazásának hátrányai
 - nem skálázható: minden csomópontnak kell rendelkeznie IPv4-es és IPv6-os címmel is, az IPv4-es címtartomány korlátozza a megoldás elterjedését
 - a hálózati útválasztókban megnövekszik az útválasztási tábla mérete
 - nem flexibilis: nincs kommunikációs lehetőség a csak IPv4-es és a csak IPv6-os csomópontok között

Alagutazás

- IPv6 csomag egy IPv4 csomagba foglalva
- Az alagút végpontjai végzik a becsomagolást
- A folyamat „transzparens” a közbeeső csomópontoknak
- **Konfigurált alagutak**
 - Az alagút végpontjait explicit módon konfigurálják
 - A végpontok dual stack csomópontok
- **Automatikus alagutak**
 - Az alagút végpontjait a hálózat automatikusan felfedezi
 - **Alagút ügynökök** (Tunnel Brokers) (RFC3053)
 - **6to4** (RFC3056)
 - **ISATAP** (Intra-Site Automatic Tunnel Addressing Protocol)
 - **6over4** (RFC2529)
 - **Teredo**: IPv4 NAT-on keresztüli alagutak

Alagutazás

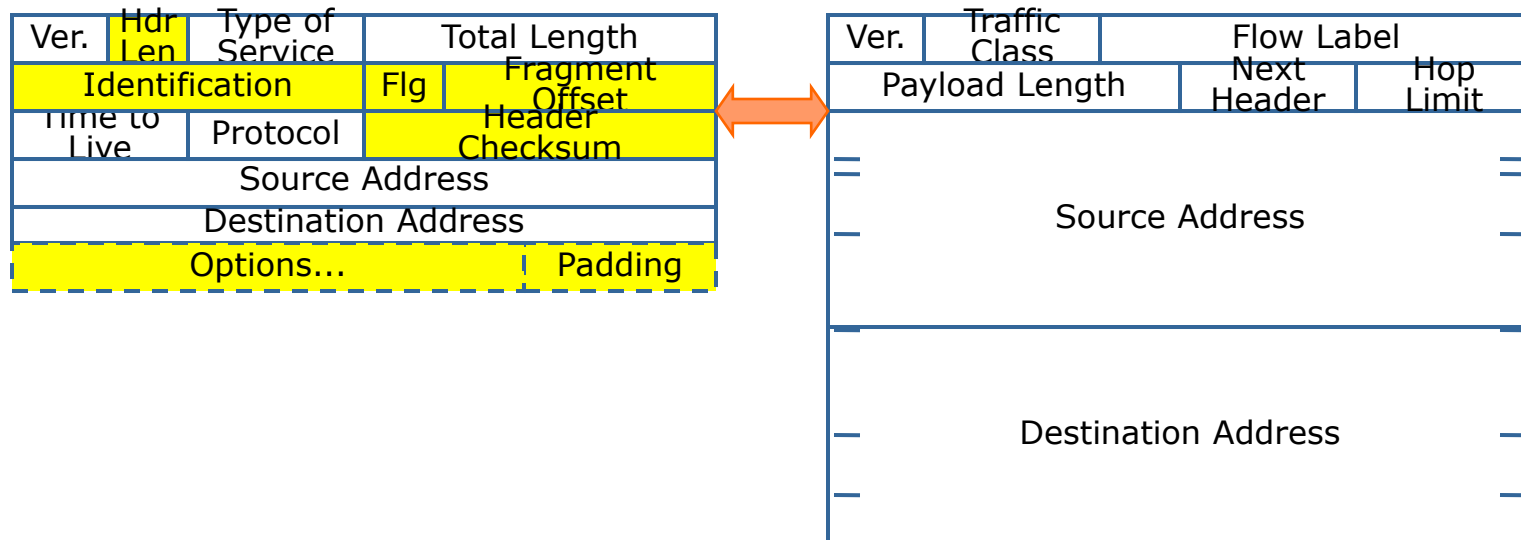


Fordítók

- Hálózati szintű fordítók
 - SITT (Stateless IP/ICMP Translator Algorithms) (RFC2765)
 - NAT-PT (Network Address Translator-Protocol Translator) (RFC2766)
 - BIS (Bump int the Stack) (RFC2767)
- Átviteli szintű fordítók
 - TRT (Transport Relay Translator) (RFC3142)
- Alkalmazási szintű fordítók
 - BIA (Bump in the API) (RFC3338)
 - SOCKS64 (RFC3089)
 - ALG (Application Level Gateway)

Hálózati szintű fordítók

- Az IPv4 és IPv6 csomagok protokoll üzeneteit fordítják egymásba
 - Elsősorban a fejléceket



Archipelago August 2010



MPLS

Moldován István

moldovan@tmit.bme.hu



BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
TÁVKÖZLÉSI ÉS MÉDIAINFORMATIKAI TANSZÉK

- MPLS Bevezető
- Label Distribution – címke kiosztás
- QoS támogatás
- Traffic Engineering
- Védelem és helyreállítás
- MPLS VPN szolgáltatások
- GMPLS

- MPLS: MultiProtocol Label Switching
- Alapvető cél
 - A vezérlés és továbbítás szétválasztása
- „Label-switching” paradigma
 - Az L2 címkekapcsolás és az L3 routing összekapcsolása

Az MPLS kiterjeszti a hagyományos IP-t a következő területeken::

- Egyszerűsített továbbítás
- Hatékony Explicit útvonalak
- Traffic Engineering
- QoS Routing
- A csomagok nem triviális módon történő útvonalakba rendezése

Összehasonlítva a létező IP gerinchálózati és aggregációs technológiákkal, az MPLS a következő előnyökkel rendelkezik:

- Az L2 réteg teljesítményét nyújtja
- Ugyanakkor L3 rétegbeli kapcsolatot és kapcsolódó szolgáltatásokat
- Növeli az teljesítmény/ár arányát a hálózati szintű útvonalválasztásnak
- Skálázhatóság

MPLS előnyök folyt.



BME-TMIT

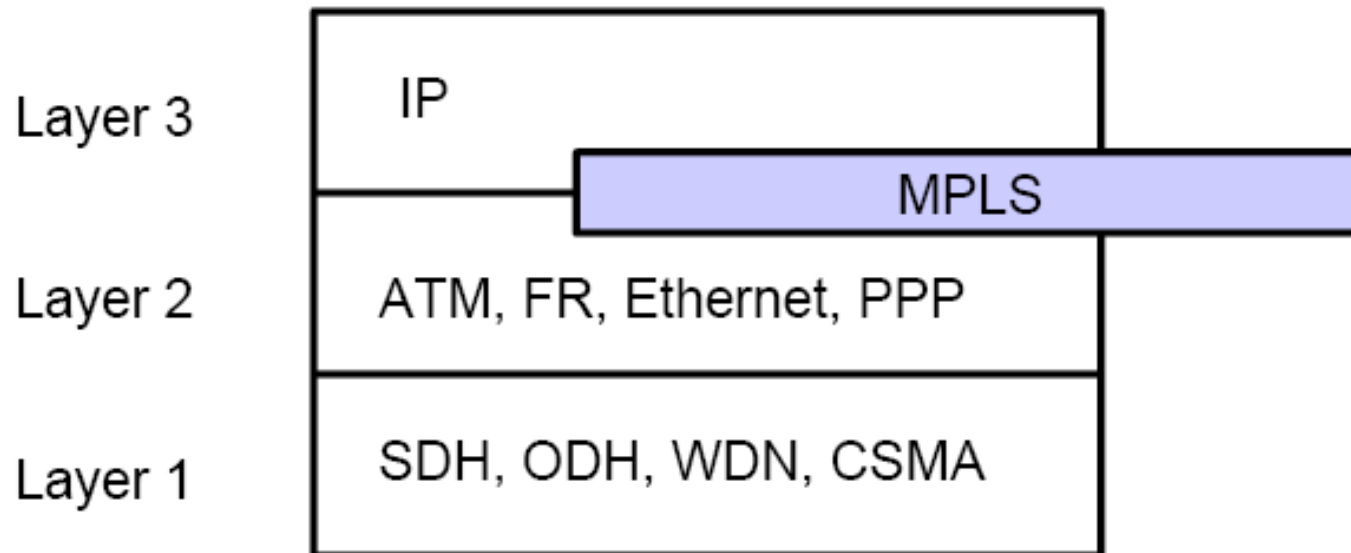
- Traffic engineering lehetőségek
- Szolgáltatások QoS garanciáinak biztosítása
- Kiküszöböli az L2 és L3 címek és routing információk koordinálását

Helye a Protocol Stack-ban



BME-TMIT

- MPLS az IP és Layer 2 között van
 - Az L2 protokollok széles körét támogatja
 - Támogatja a felsőbb szintű mechanizmusokat is



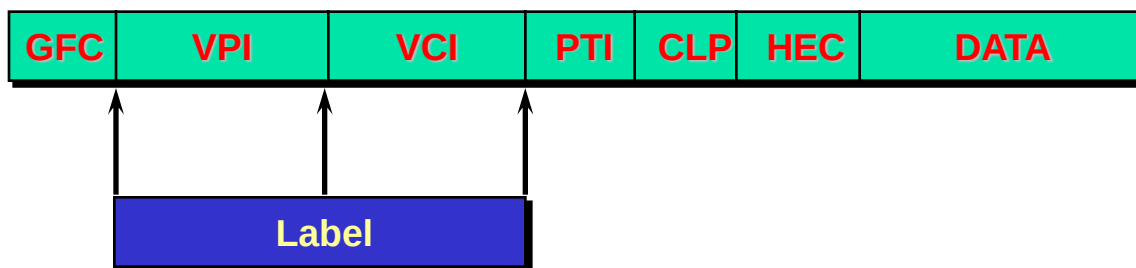
- Decomposition of network layer routing into control and forwarding components **applicable**
- Label switching forwarding component algorithm uses
 - forwarding table
 - label carried in the packet
- What is a Label ?
 - Short fixed length entity

Label – címke hordozása

- Bizonyos L2 technológiák képesek a címkét a saját fejlécük részeként kezelni
 - pl. VPI/VCI az ATM esetén, DLCI a FR esetén vagy MPLS címke PPP/Ethernet esetén
- Azon L2 rétegek amelyek nem támogatják a címkéket egy külön „shim” fejlécben hordozzák

Link layer fejléc	“Shim” label fejléc	Network layer fejléc	Network layer fejléc
----------------------	------------------------	-------------------------	-------------------------

ATM Cella Fejléc



PPP Fejléc (Packet over SONET/SDH)



LAN MAC Label Fejléc



Az MPLS Shim Header



BME-TMIT

- A címke - Label (Shim Header) valójában egy is Label Stack Entry szekvencia, azaz egy vagy több bejegyzés
- Minden Label Stack Entry 4 byte (32 bit) hosszú
- 20 Bit fenntartva a címke azonosítónak (Label Identifier) – ez a „Label”, címke

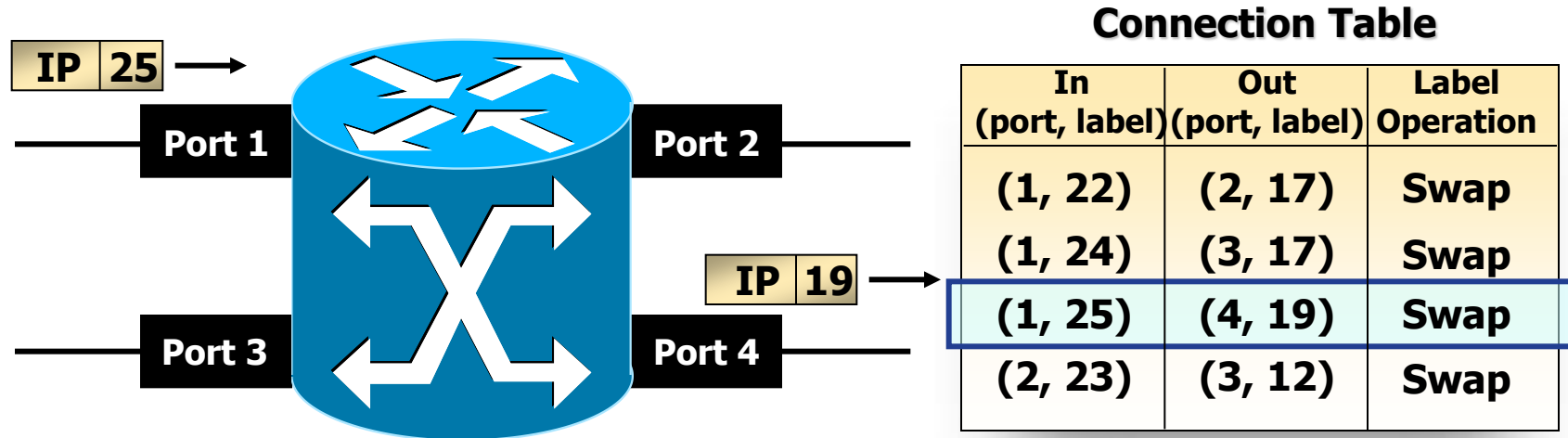


Label : Label value (0 to 15 are reserved for special use)

Exp : Experimental Use

S : Bottom of Stack (set to 1 for the last entry in the label)

TTL : Time To Live



- Label Swapping

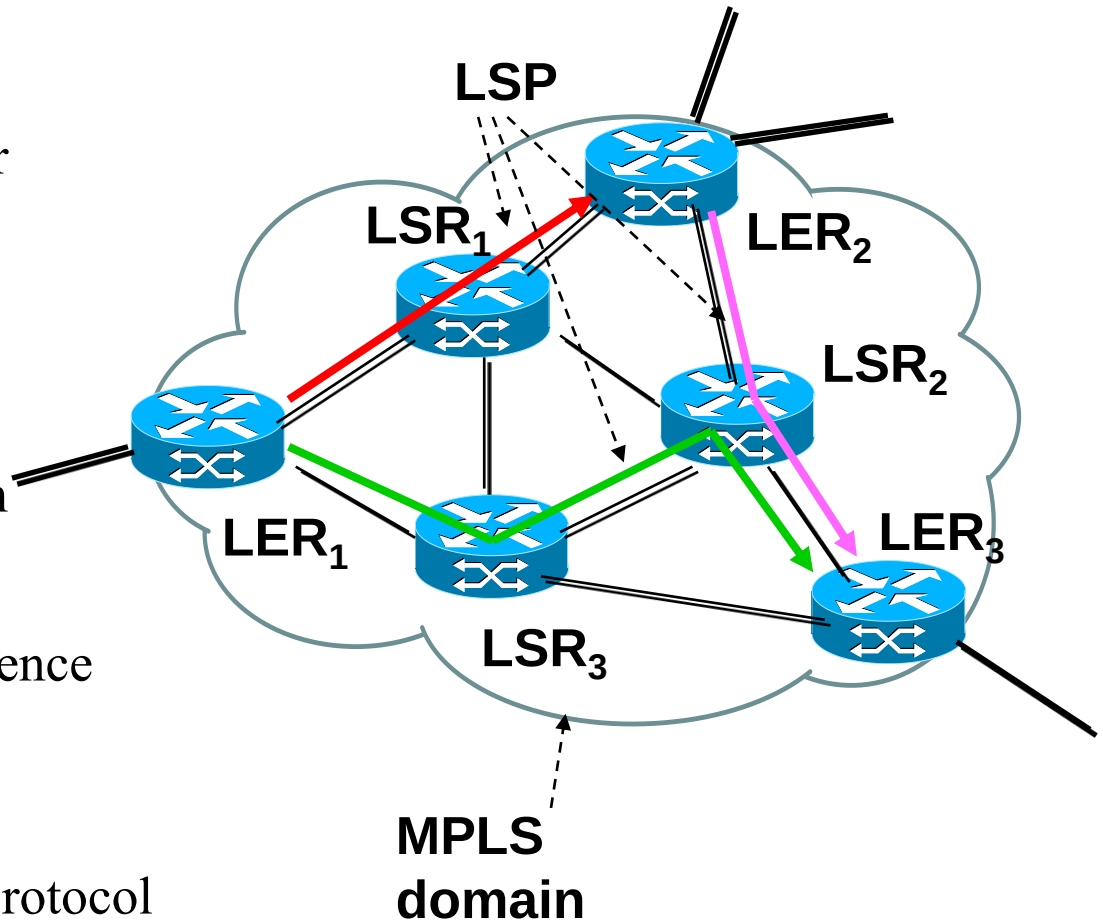
- A kapcsolat tábla tartalmazza a bejegyzéseket („Connection table”)
- Exakt találat keresés (nem mint IP routing esetén)
- Bemenet: Input (port, label) meghatározza:
 - Címke művelet
 - Kimenet - Output(port, label)
- Hasonló továbbítást használ az ATM és Frame Relay is

MPLS Komponensek



BME-TMIT

- LSR
 - » Label Switch Router
- LER
 - » Label Edge Router
- LSP
 - » Label Switched Path
- FEC
 - » Forwarding Equivalence Class
- LDP
 - » Label Distribution Protocol



Label Edge Router - LER



BME-TMIT

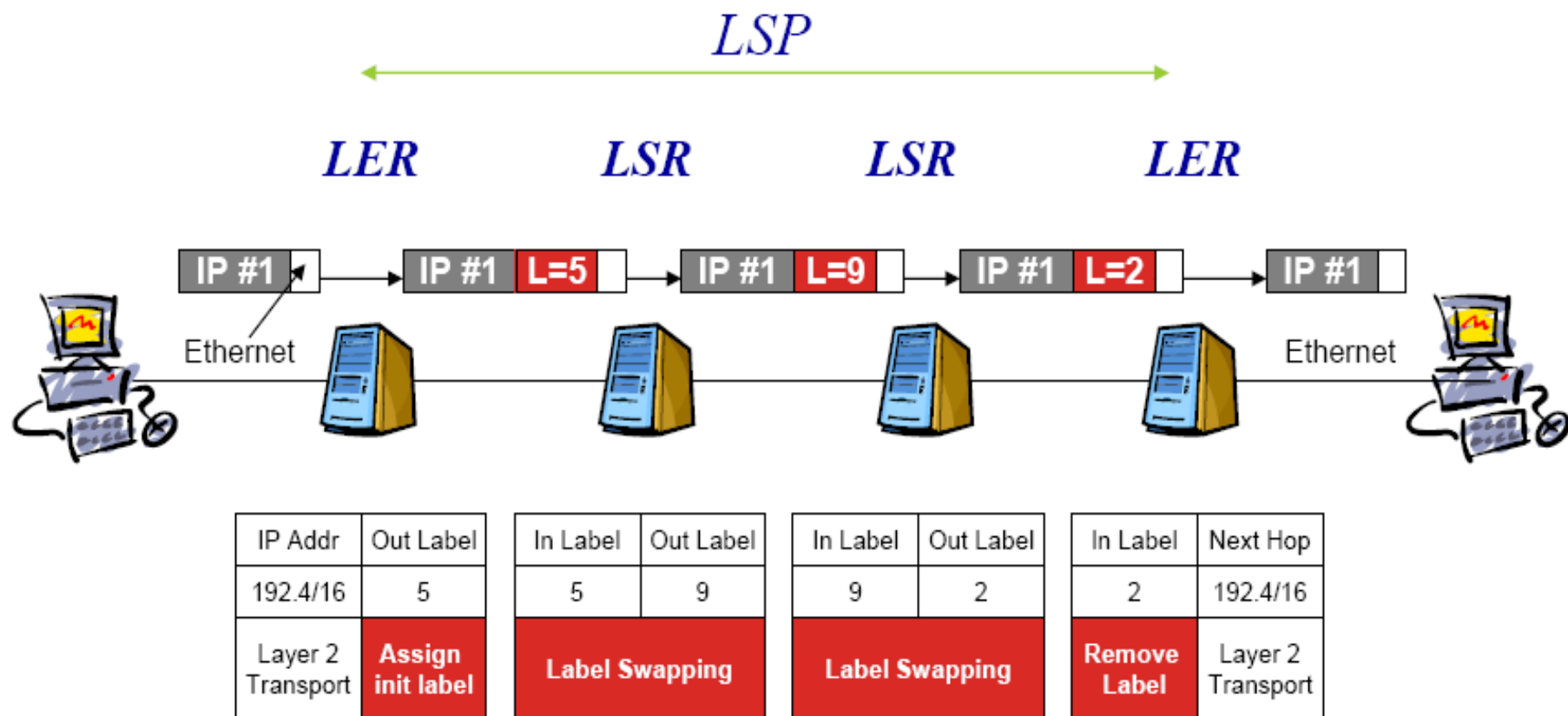
- Az MPLS hálózat szélén található és ő rendeli hozzá/veszi le a címkéket a csomagokról
- Többféle port típus használatát teszi lehetővé (pl. frame relay, ATM, és Ethernet).

- Nagysebességű kapcsolók amelyek a hálózat magját alkotják
- Címkekapcsolást végeznek nagy sebességgel
- ATM kapcsolók használhatók LSR-ként hardver változtatás nélkül. Az ATM/MPLC címkekapcsolás megegyezik az ATM VP/VC kapcsolással.

A LER & LSR-ek helyei



BME-TMIT



“ROUTE AT EDGE, SWITCH IN CORE”

- Olyan csomagcsoportok összessége, amelyeket a hálózatban ugyanúgy kezelünk (útvonal, QoS garanciák)
 - Például ugyanazt a szolgáltatást igénybe vevő felhasználók forgalma
- Egy csomag FEC-hez rendelését egyszer tesszük meg (amikor beérkezik a hálózatba)

- Egy csomagot a következő kritériumok szerint tudunk FEC –hez rendelni:
 - destination IP address,
 - source IP address,
 - TCP/UDP port,
 - Inter-AS-MPLS esetén: Source-AS and Dest-AS,
 - class of service,
 - applikáció,
 - ...
 - Az előzőek kombinációi

Ingress Label	FEC	Egress Label
6	138.120.6/24 - xxxx	9

- FEC – manuálisan beállítva az operátor által
- Egy FEC –hez legalább egy címke lesz rendelve

Ingress Label	FEC	Attribute	Egress Label
6	138.120.6/24 - xxxx	A	9
6	138.120.6/24 - xxxx	B	12

- Egy címkekapcsolt útvonal, melyet a kommunikáció megkezdése előtt létre kell hozni.
- Egy LSP lehet egy FEC reprezentációja

- Az MPLS két lehetőséget nyújt az LSP-k kihúzására
 - hop-by-hop routing – IP routing
 - Minden LSR individuálisan választja ki a következő célt. Az LSR bármilyen útvonalválasztó algoritmust használhat.
 - explicit routing
 - A source-routinghoz hasonló. Az Ingress LSR (vagy egy központi entitás) meghatározza az útvonalat adó csomópontok listáját
- Egy LSP amely egy FEC-hez van rendelve mindig egyirányú. A visszafele irányú forgalom számára külön LSP szükséges.

Hogy „húzzunk ki” egy LSP-t?



BME-TMIT

- Label Distribution – címke kiosztás
- Label Distribution – protokoll segítségével
- Label Distribution ami az IGP útvonalát követi
 - Label Distribution Protocol (LDP)
- Label Distribution explicit útvonalak mentén:
 - Explicit útvonalon való kihúzás
 - Sáv szélesség foglалás (opcionális)
 - Class of Service (DiffServ stílusban)
- Label Distribution BGP használatával
 - Az AS-ek közötti BGP/MPLS VPN-ek megvalósításához

- Az IETF MPLS architektúra nem feltételez egyetlen címkekiosztó protokollt
- LDP
 - Ugyanazt az útvonalat követi mint az IGP, vagy explicit útvonalat
- RSVP
 - Explicit útvonalválasztás – megszorításokkal
 - Constraint based routing – egyéb (pl. QoS) paraméterek figyelembe vétele
 - Traffic Engineering és Fast Reroute
- BGP
 - Címkekiosztás IPv4 útvonalak számára

- Topológia vezérelt
 - Címkekiosztás a routing protokollok által jelzett topológia változások követésére
- Kontroll vezérelt
 - Címkekiosztás az RSCP, CR-LDP protokollok kéréseire
- Forgalom vezérelt
 - Címkekiosztás új folyamatok észlelése esetén

- MPLS Bevezető
- **Label Distribution – címke kiosztás**
- QoS támogatás
- Traffic Engineering
- Védelem és helyreállítás
- MPLS VPN szolgáltatások
- GMPLS

- Az LSP-ket elosztottan hozzuk létre és tartjuk fenn
- Minden LSR megegyezik egy címkében minden FEC (Forwarding Equivalence Class) számára a felfele és lefele irányú szomszédjaival, valamilyen kiosztási módszert használva
- Az eredmény: Label Information Base (LIB)

- Az RFC 3035 és 3036 írja le
- Címkek kiosztására használják MPLS hálózatban
- Forwarding Equivalence Class
 - Meghatározza hogy a csomagok hogyan lesznek LSP-hez rendelve
- FEC szerint hirdeti a címkéket
 - Pl. az a.b.c.d cím x label –el érhető el
- Neighbor discovery
 - Szomszédos LSR-ek felderítése
 - Basic és Extended Discovery

- Label Distribution Protocol (LDP)
 - Műveletek gyűjteménye melyekkel egy LSR LSP-
ket hoz létre
 - Hálózati szintű útvonalak adatkapcsolati szintű
kapcsolt útvonalakra
- LDP peer-ek:
 - Szomszédos LSR-ek melyek címke/stream
kiosztást végeznek
 - Ez az információ csere "LDP Session" néven
ismert

- Discovery üzenetek – egy LSR jelenlétének jelzésére és fenntartására
- Session üzenetek – kapcsolatok létrehozására, fenntartására és bontására LDP peer-ek között
- Advertisement üzenetek – címke kiosztás létrehozására, változtatására és törlésére
- Notification üzenetek – további információk, hibaüzenetek átvitelére

- Címkek kiosztásánál használt tartományok
- Kétféle tér lehetséges
 - **Per interface label space:** Interfész specifikus, interfészenként történik a kiosztás
 - **Per platform label space:** Platform-specifikus, a bejövő címkeken osztoznak az egyforma interfészek

- Mechanizmus az LDP peer-ek felderítésére
- Elkerüli hogy explicit módon kelljen az LSR peer-eket megadni
- Két lehetőség a felderítésre:
 - basic discovery mechanism: link szinten csatlakoztatott LSR szomszédok felderítésére
 - extended discovery mechanism: azon LSR szomszédok felderítésére, amelyek nem link-lokálisan észlelhetők

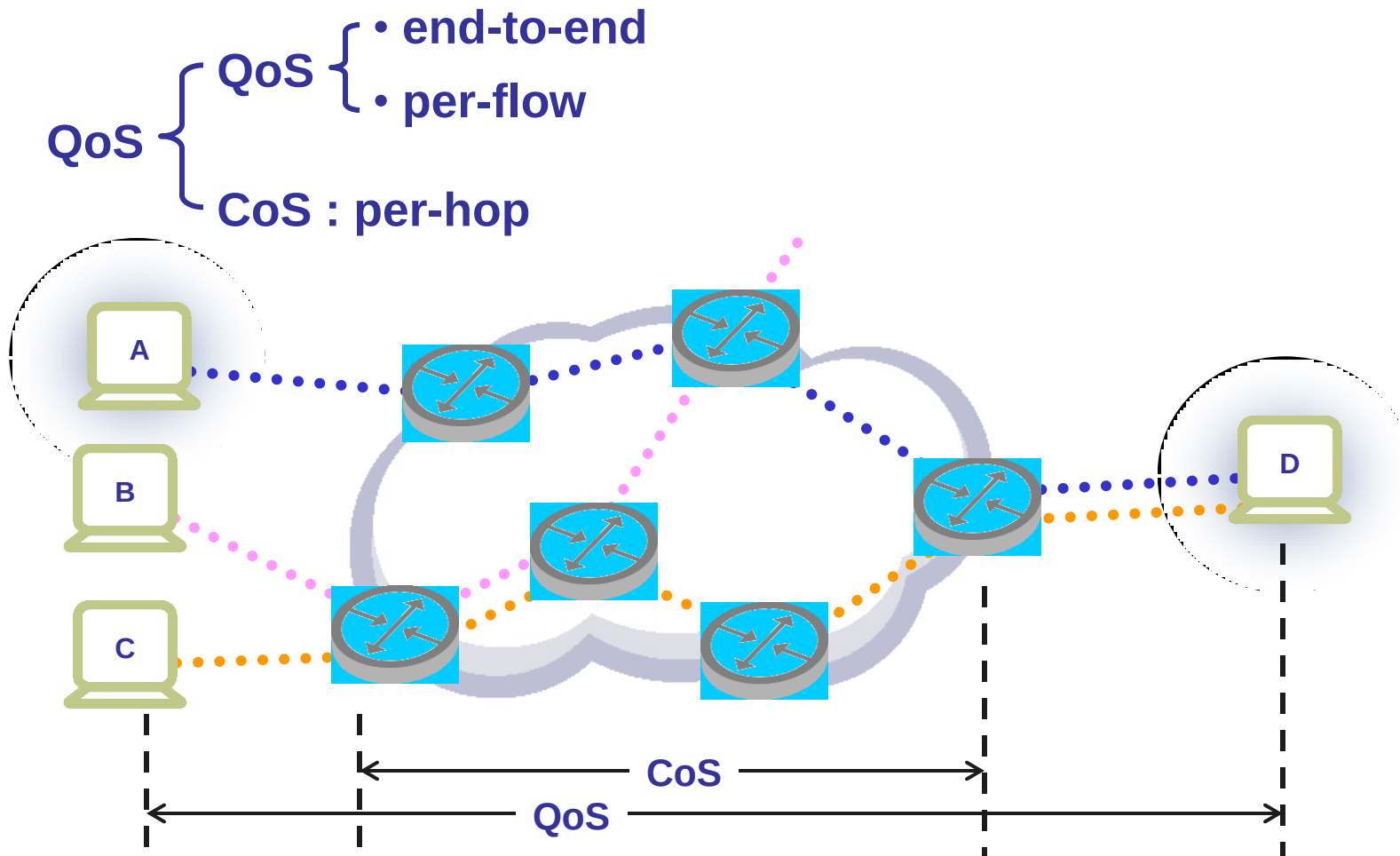
- Az LSR a megtanult címkéket egy Label Information Base (LIB) felépítésére használja
- Minden bejegyzés a LIB-ben egy FEC-et rendel egy (LDP azonosító, címke) pároshoz

- MPLS Bevezető
- Label Distribution – címke kiosztás
- **QoS támogatás**
- Traffic Engineering
- Védelem és helyreállítás
- MPLS VPN szolgáltatások
- GMPLS

QoS vs CoS



BME-TMIT

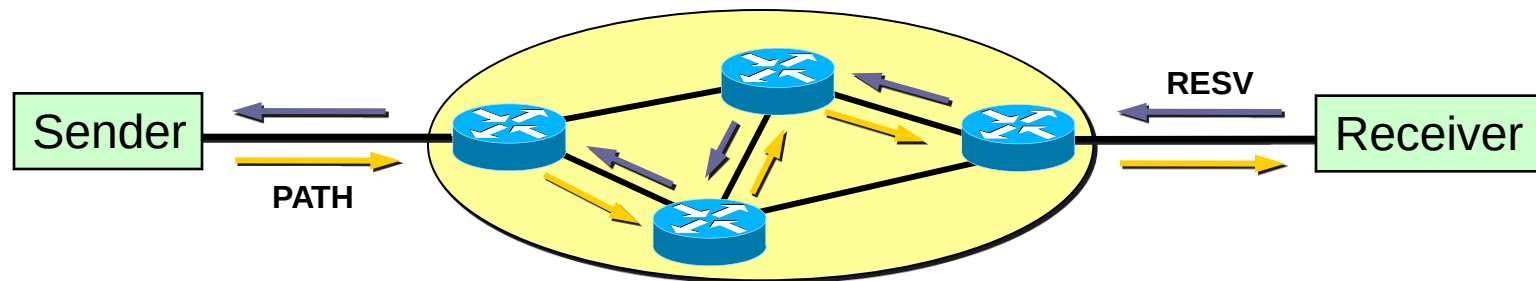


Integrated Services (IntServ)



BME-TMIT

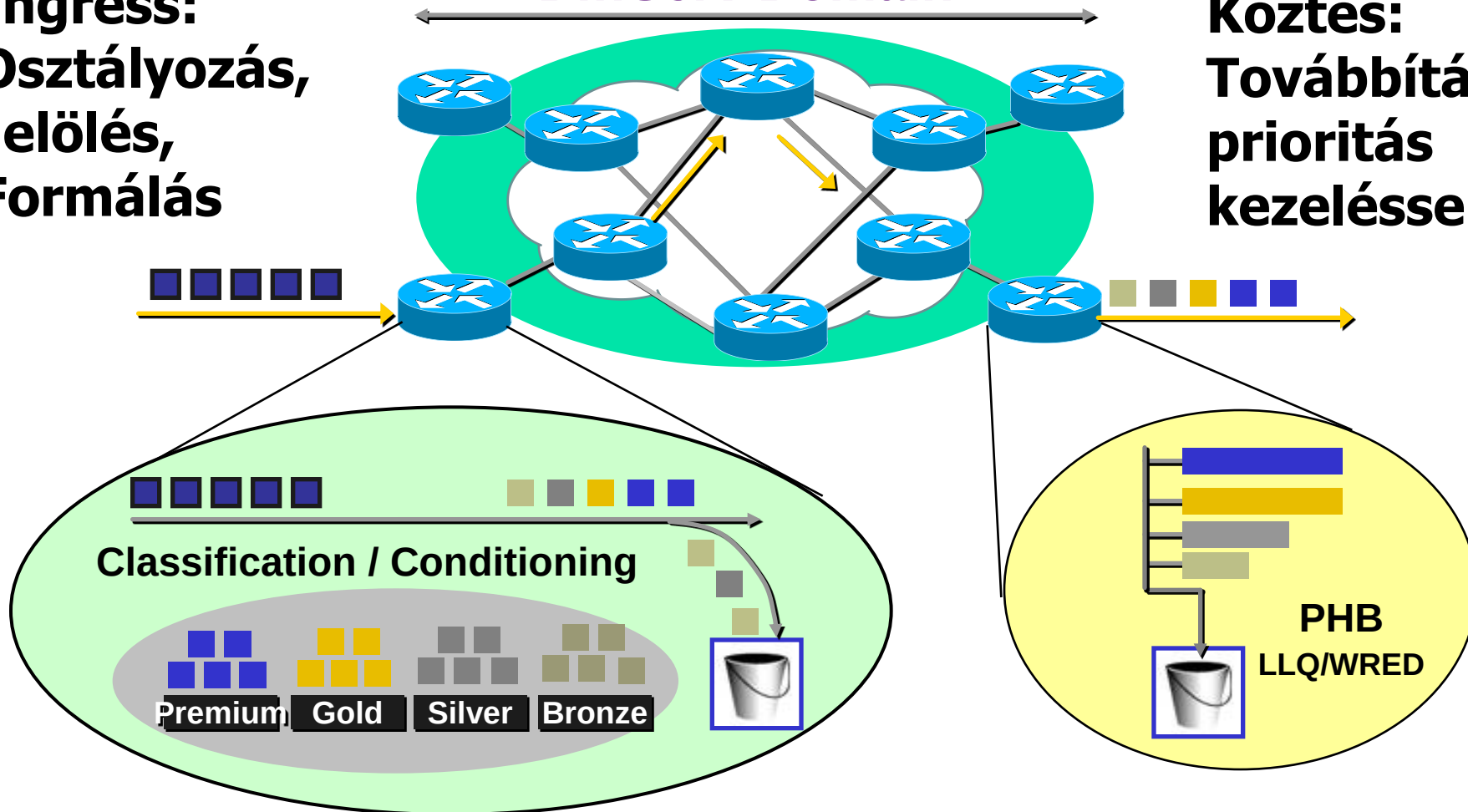
- Az Integrated Services (IntServ) az Resource Reservation Protocol (RSVP) protokollon alapul
- **per simplex flow** erőforrás foglalás
- Az alkalmazások kérnek erőforrásokat a hálózattól
- A küldő leírja az erőforrás követelményeket a PATH üzenetben amelyet a vevőnek küld
- A vevők lefoglalják az erőforrásokat egy RESV üzenettel amely a visszafele irányban halad



Ingress:
Osztályozás,
Jelölés,
Formálás

DiffServ Domain

Köztes:
Továbbítás
prioritás
kezeléssel



- Az MPLS önmagában nem ad QoS-t
- MPLS könnyebbé teszi az IntServ megvalósítását
 - Még mindig skálázhatósági problémát jelent
- MPLS támogatja a DiffServ-et, Traffic Engineering lehetőségekkel
 - Diffserv-aware TE

- RSVP-TE segítségével
- RSVP – erőforrások lefoglalása
 - PATH üzenet – egyben címke kérés is
 - RESV üzenet – erőforrás foglalás, és egyben címke válasz is
- RSVP ugyanakkor beengedés vezérlést is végez
 - A PATH üzenet ellenőrzi az erőforrásokat
 - az RESV üzenet végzi a tulajdonképpeni foglalást



● Class of Service (CoS)

- A hálózat különböző minőségi osztályokat implementál
- A forgalmat forgalmi osztályokba soroljuk
 - Layer 3 szinten: applikáció, célállomás,...
- Egyszerűbb és hatékonyabb mint egy csomó virtuális kapcsolat kezelése (pl. ATM-nél), és megvalósul az L2-L3 megfeleltetés is
- Két módszer a szolgáltatási osztály jelölésére:
 - IP precedence -> MPLS header (CoS field)
 - Összesen 8 osztály kezelhető (3 bit)
 - Külön címke használata minden osztályra
 - Nincs határ a címkék számára vonatkozóan (QoS specifikusan)

- MPLS és DiffServ kombináció
 - Diffserv az erőforrások kezelésére
 - MPLS a gyors továbbításra
- Ingress:
 - Diffserv – FEC összerendelés
- Köztes csomópontok
 - Explicit LSP-k
 - Könnyebben nyomon követhetők az erőforrások

- Használja a meglévő IP QoS mechanizmusokat
 - Queuing – LLQ, CBWFQ
 - Policing
 - WRED
- Osztályozás és jelzés az EXP bitek segítségével
- A címke fejléc QoS jelölése eltérhet az IP DSCP jelöléstől

- MPLS Bevezető
- Label Distribution – címke kiosztás
- QoS támogatás
- **Traffic Engineering**
- Védelem és helyreállítás
- MPLS VPN szolgáltatások
- GMPLS

Minek Traffic Engineering?



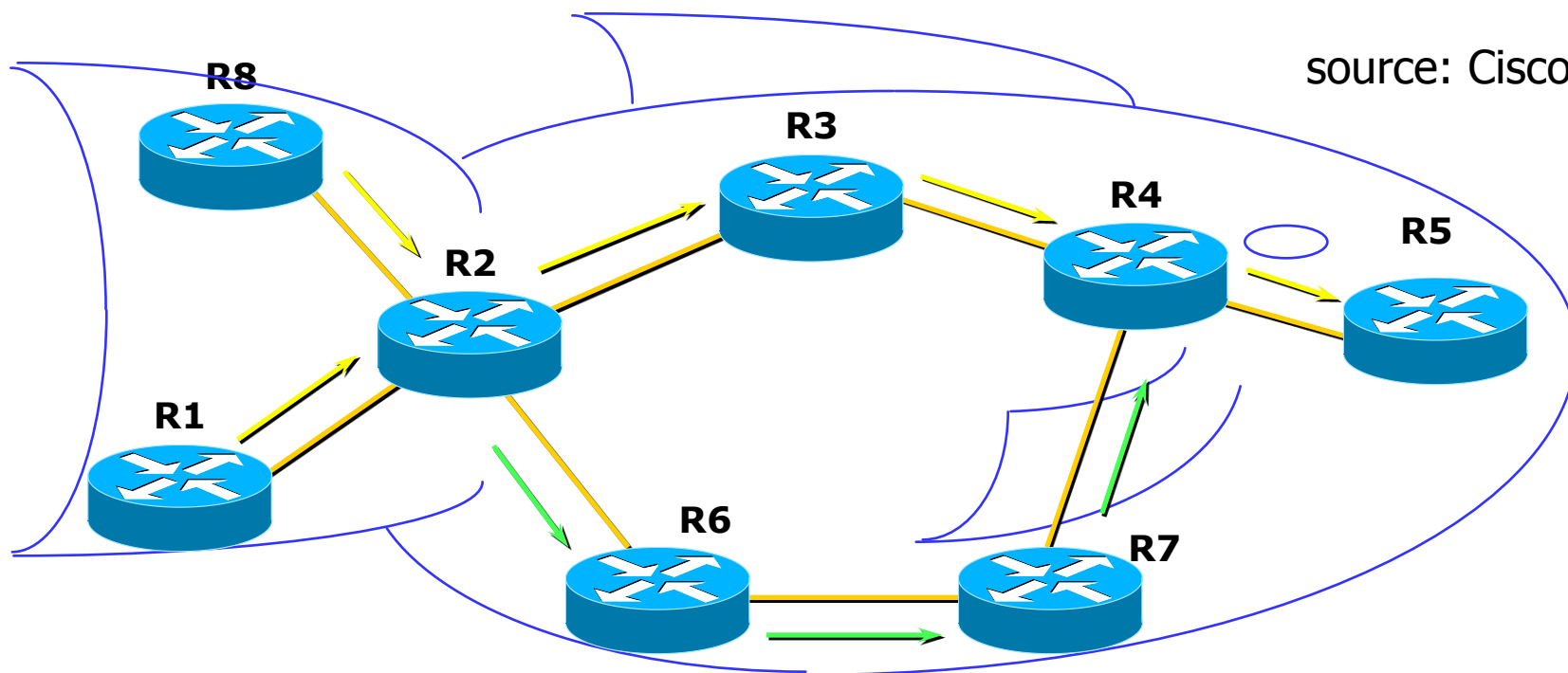
BME-TMIT

- Szűk keresztmetszetek a hálózatban a változó forgalom hatására
- Jobb sáv szélesség kihasználás
 - Útvonal nem a legrövidebb útvonalon
- Hibás linkek/csomópontok kikerülése
 - Fast rerouting – a felhasználó számára transzparens
 - Védelem
- Új szolgáltatások megvalósítása – Virtuális bérelt vonal
 - point-to-point sáv szélesség garanciával
- Kapacitás tervezés
 - Lehetővé teszi az aggregátumok kezelését, számolni lehet az útvonalakon az igényelt sáv szélességet

IP Routing és a Hal



BME-TMIT



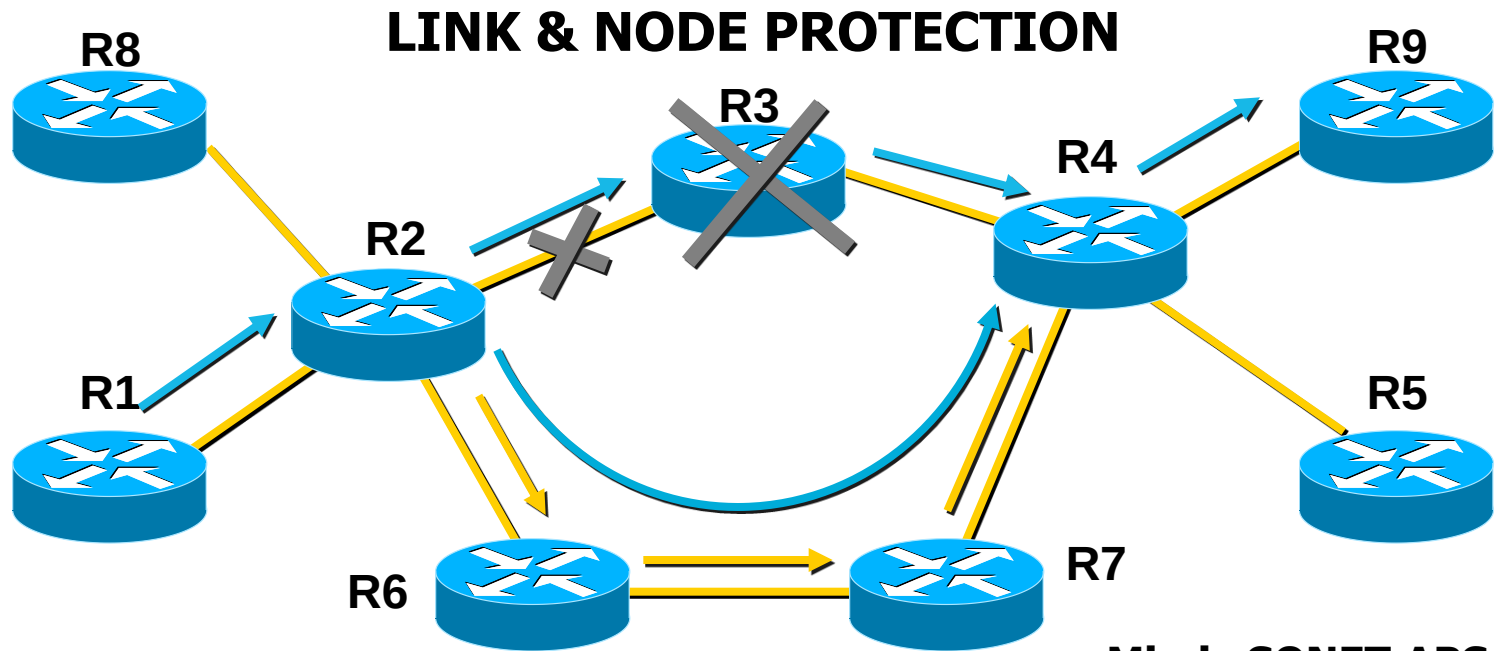
IP (legtöbbször) a legrövidebb útvonalat választja
Az R8 és R1-ből induló folyamatok R2-nél keverednek, és a későbbiekben nem megkülönböztethetők

Az alternatív útvonal nem kihasznált

Az MPLS TE előnye



BME-TMIT



source: Cisco

Mimic SONET APS
Re-route in 50ms or less

- Több csomópont kikerülhető. R2 kicserélheti az R3 irányába mutató címkét az R6 irányúra

TE performancia célok

- Traffic Engineering (TE) célja az optimális hálózati kihasználás
- A fő célok
 - Forgalom szempontjából: javítja a QoS-t, pl. a csomagvesztés minimizálásával
 - Erőforrás szempontjából: javítja az erőforrások felhasználását, jobb hálózati kihasználtságot tesz lehetővé

- Főbb komponensek
 - Traffic Trunk – azonos osztályú forgalmi folyamatok aggregálása melyek egy LSP-t használnak
- Indukált MPLS Gráf
 - Virtuális topológia – akár egy overlay model
 - Logikai hozzárendelés egy fizikai útvonalhoz LSP traffic trunk segítségével
 - Az LSP-k pont-pont kapcsolatot nyújtanak két LER között, több LSR-en keresztül

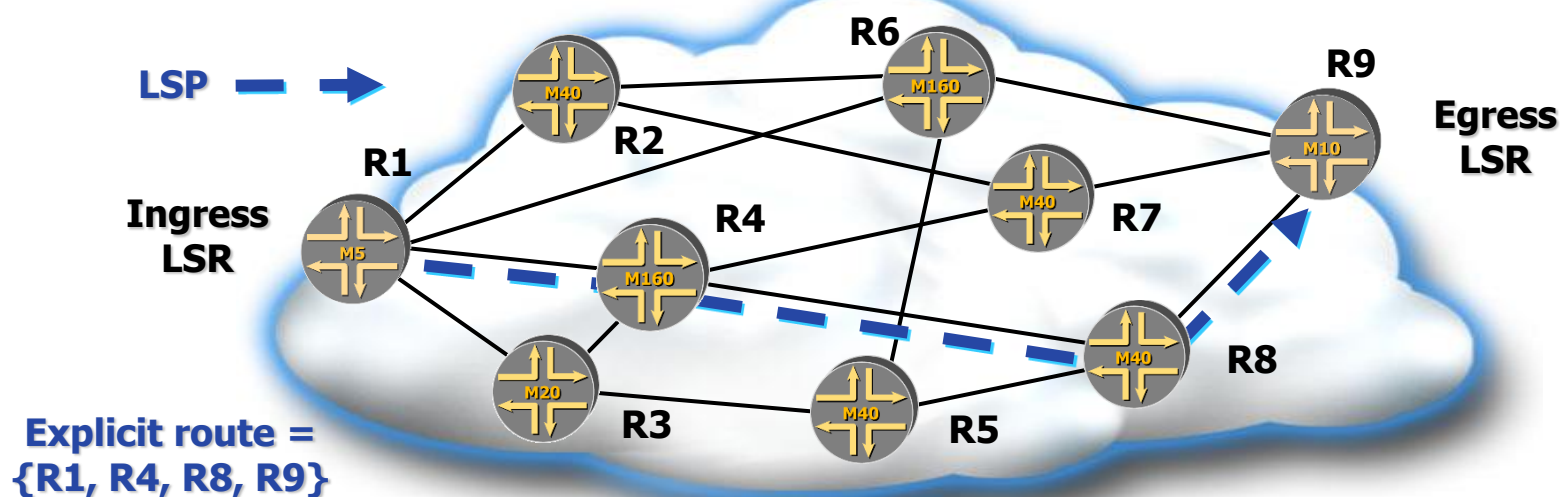
- Két opció az útvonalválasztásra:
 - Hop by hop routing
 - Explicit routing
- Explicit Routing (azaz Source Routing) – nagyon hasznos
 - IP explicit routing: minden csomagban benne az útvonal - túl nagy overhead
 - MPLS: csak az LSP kihúzásánál kell az explicit útvonalat megadni
 - Az MPLS praktikussá teszi az explicit útvonalak kezelését

- Offline Útvonal számítás
 - PCE, Path Computation Element
 - Központi eszköz
- Online Útvonal számítás
 - OSPF/IS-IS kiterjesztések
 - Constraint based routing - megszorításokkal
- LSP jelzés – explicit útvonal
 - RSVP-TE
 - CR-LDP

Offline útvonal számítás



BME-TMIT



source: Juniper

- Bemenet az offline útvonal számítónak:
 - Ingress és egress csomópont
 - Fizikai topológia
 - Forgalmi mátrix (statisztikai)
- Kimenet:
 - Fizikai útvonal szett, explicit útvonalként

- CR kiterjesztést igényel a routing protokollokhoz
 - Constraint-based SPF algoritmus: figyelembe veszi a megszorításokat
- Link State Információk - kiterjesztve
 - Hálózati topológia, folyam követelmények, a linkeken az elérhető erőforrások
 - Az LSA – Link State Advertisment üzenetek kiterjesztése OSPF és IS-IS protokollokban
 - Kompromisszum az információ részletessége és pontossága illetve a gyakori frissítés között

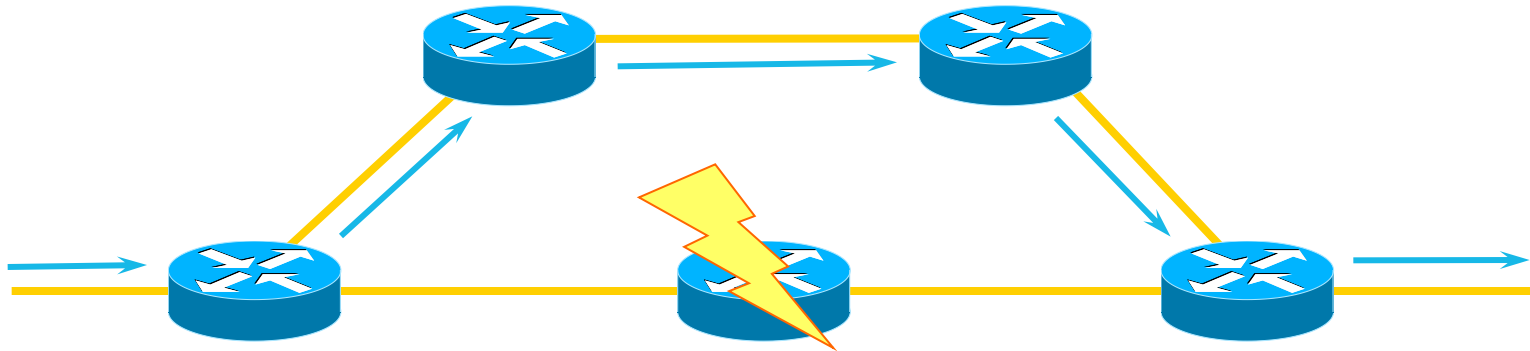
- MPLS Bevezető
- Label Distribution – címke kiosztás
- QoS támogatás
- Traffic Engineering
- **Védelem és helyreállítás**
- MPLS VPN szolgáltatások
- GMPLS

- IGP alapú: helyreállítás (restoration)
- RSVP-TE lehetővé teszi a védelmet
 - 1+1
 - a forgalom egyszerre mindkét útvonalon megy
 - Nincs csomagvesztés
 - 1:1
 - Előre kihúzott védelmi LSP, amelyre átkapcsol hiba esetén
 - Nagyon gyors $\sim 50\text{ms}$, de lehet csomagvesztés
- Szegmens védelem lehetséges
 - Fast reroute

MPLS Link és Node védelem

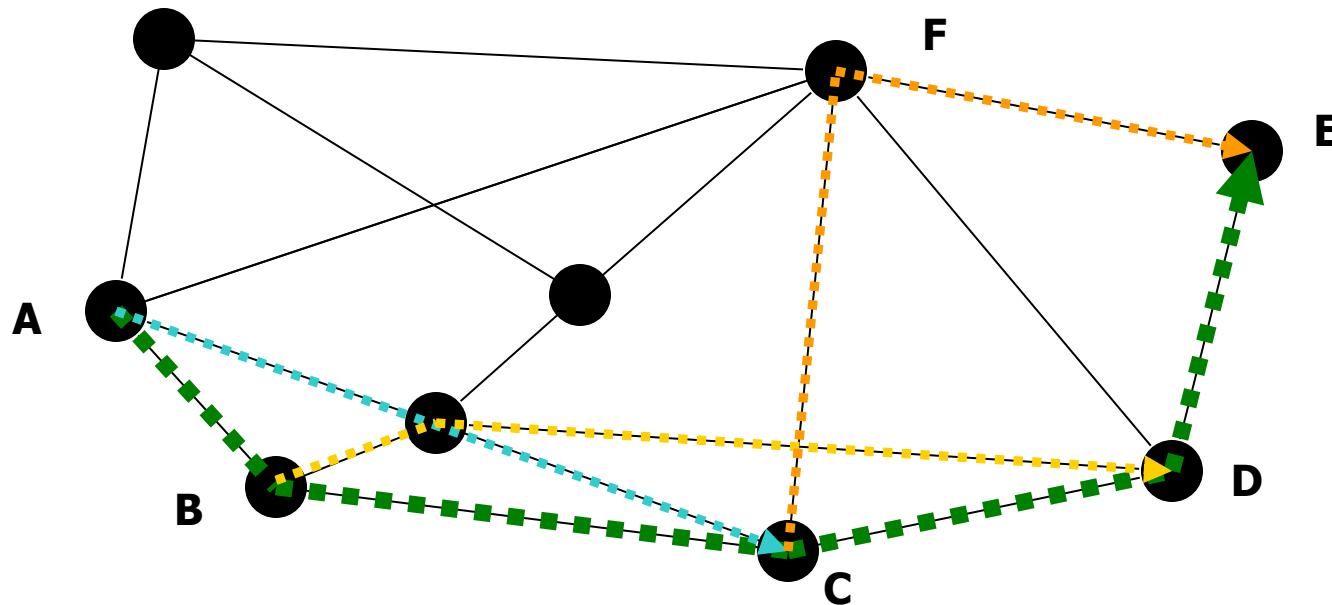


BME-TMIT

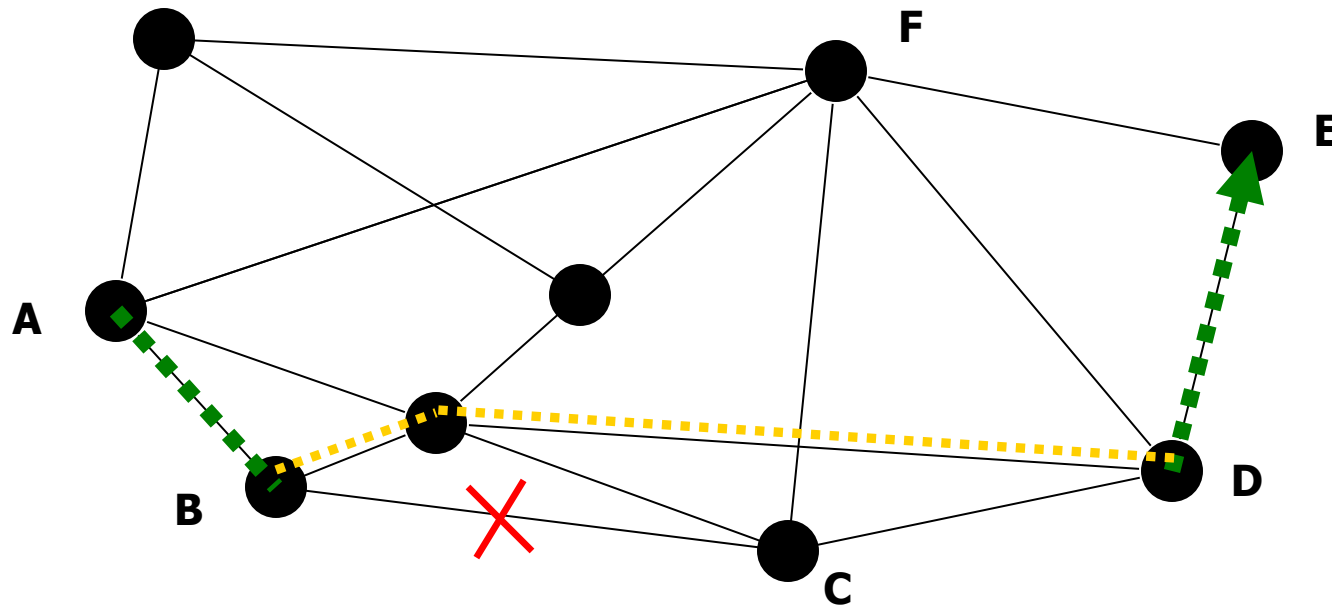


- Időszakos kikerülés egy hibás csomópontnál
 - 50 msec alatti
 - Skálázható több ezer LSP-re

- Gyors útvonalkeresés az ingresztől
 - A kikerüli B-t
 - B kikerüli C-t
 - C kikerüli D-t



- B - C link megszakad
 - B azonnal kikerüli C-t
 - B jelzi A-nak hogy hiba történt

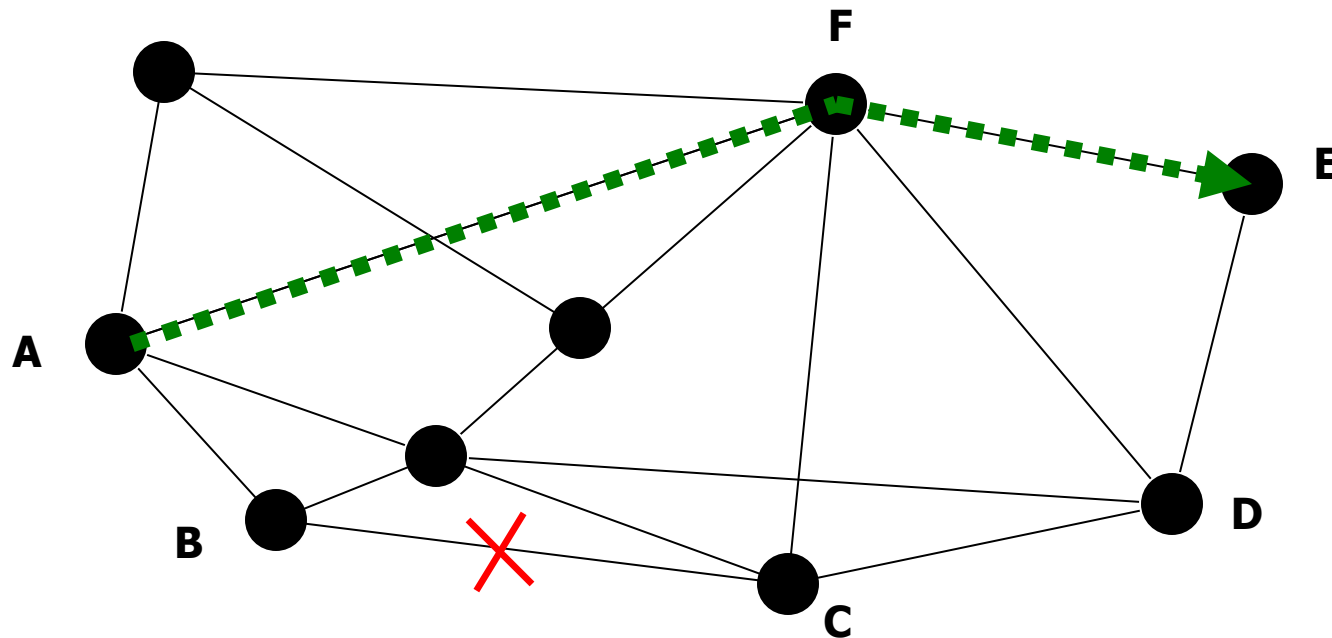


Hosszú távú megoldás



BME-TMIT

- A kiszámol és kihúz egy optimális elsődleges útvonalat



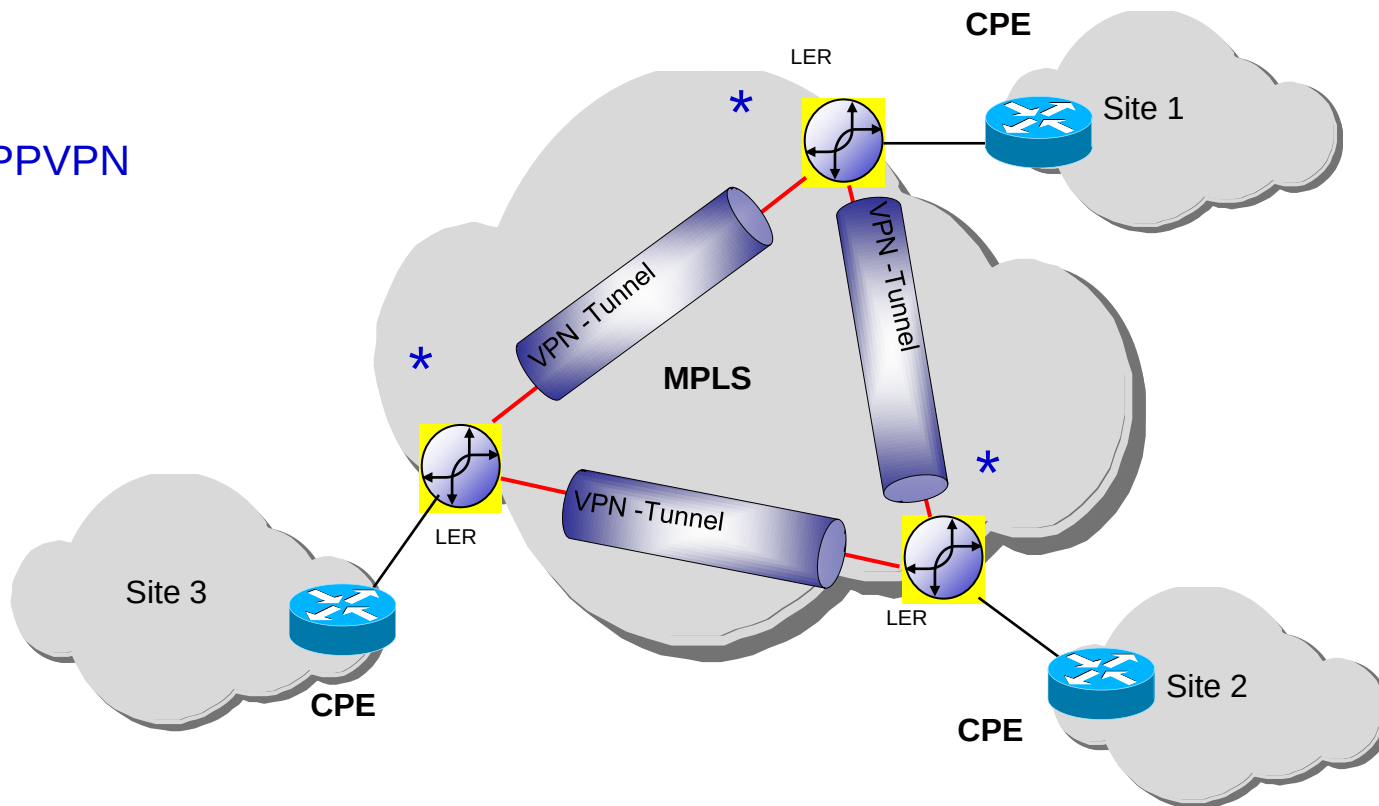
- MPLS Bevezető
- Label Distribution – címke kiosztás
- QoS támogatás
- Traffic Engineering
- Védelem és helyreállítás
- **MPLS VPN szolgáltatások**
- GMPLS

- Traffic Engineering
 - Optimális hálózati kihasználás
 - Explicit és policy routing
- Védelem és gyors helyreállítás
 - Fast reroute
 - 1:1 védelem
- Szolgáltatások
 - IP VPN-ek (RFC 2547bis: BGP/MPLS VPN)
 - Layer 2 VPN-ek
 - Layer 2 Transport: '*' MPLS felett, * = ATM, FR, Ethernet, stb.

- Virtuális router modell
 - L2 & L3 Provider Provisioned – szolgáltató által beállított VPN megoldás (PP)

Provider Provisioned VPN

IETF - PPVPN

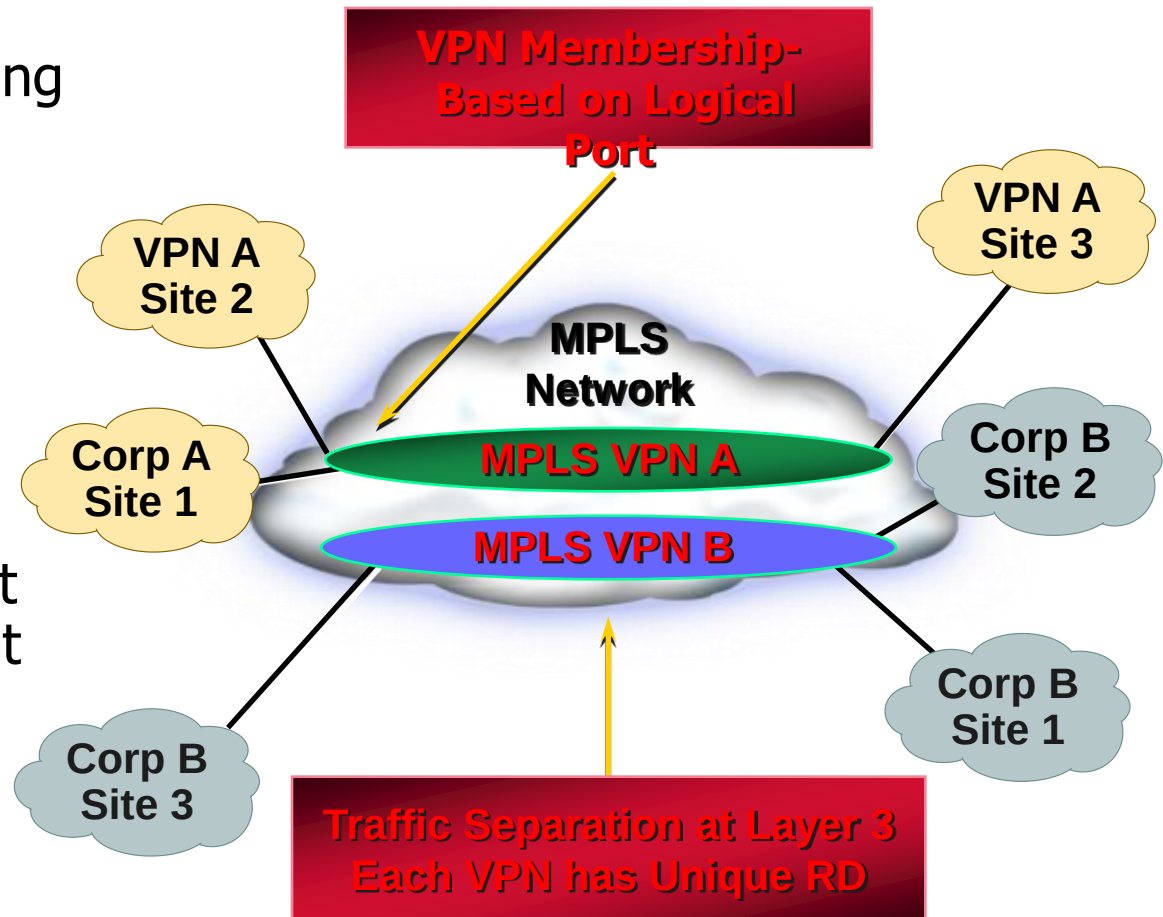


MPLS alapú IP-VPN Architektúra



BME-TMIT

- Skálázható VPN-ek
- IP QoS és traffic engineering
- Könnyen menedzselhető
- Magas megbízhatóság – hasonló az ATM és Frame Relayhoz
- Lehetővé teszi értéknövelt alkalmazások szolgáltatását
- Felhasználói IP címek szabad megválasztása

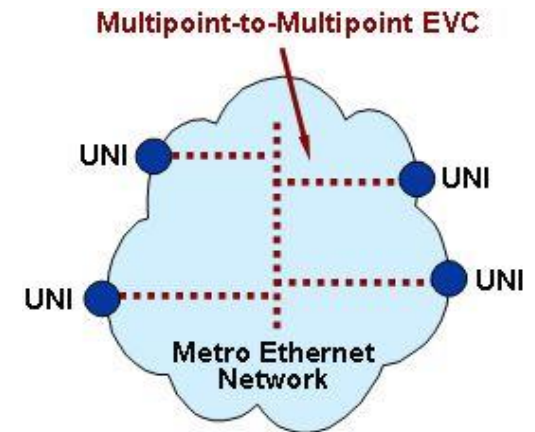
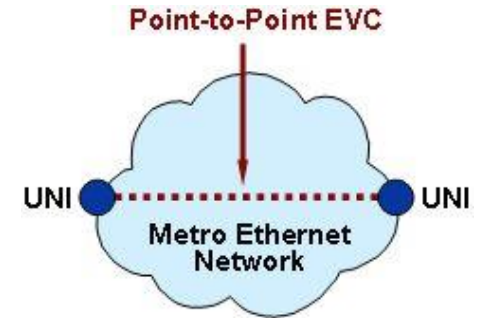


Ethernet szolgáltatások



BME-TMIT

- Virtual Private Wire Service, **VPWS**
 - Bérelt vonali szolgáltatás
 - Pont-pont kapcsolat
- Virtual Private LAN Service, **VPLS**
 - Virtuális LAN szolgáltatás
 - Multipont-multipont
- **UNI**-kapcsolódási pont
 - Virtuális kapcsolat - EVC



Forrás: MEF

- MPLS Bevezető
- Label Distribution – címke kiosztás
- QoS támogatás
- Traffic Engineering
- Védelem és helyreállítás
- MPLS VPN szolgáltatások
- **GMPLS**

Az MPLS evolúciója

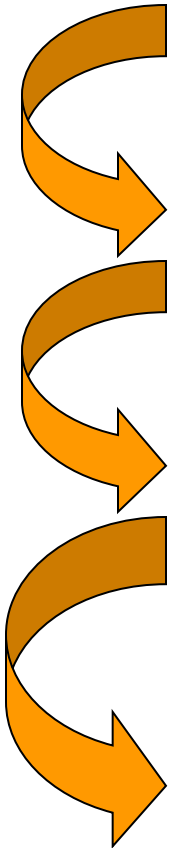


BME-TMIT

IETF
46-48

IETF
48-49

IETF
50-51



- **MPLS**: MultiProtocol Label Switching
 - IP csomag alapú
 - Packet Traffic Engineering (**MPLS-TE**)
- **MP λ S**: MultiProtocol Lambda Switching
 - MPLS vezérlés kiterjesztve a WDM technológiára (wavelength/lambda) és IGP TE kiterjesztések
- **GMPLS**: Generalized MPLS
 - MPLS vezérlés kiterjesztve a vonalkapcsolt technológiákra (SDH/Sonet) és optikai rétegre
 - Új protokoll: LMP
- **GMPLS**: Technológia függetlenség bevezetése
 - LMP kibővítve "passzív eszközökre" : LMP-WDM

- GMPLS: 5 interfész típus
 - PSC - Packet Switching Capable: IP/MPLS
 - L2SC - Layer-2 Switching Capable: ATM, FR, Ethernet
 - TDM - Time-Division Multiplexing: Sonet, SDH, G.709 ODU
 - LSC - Wavelength Switching: Lambda, G.709 OCh
 - FSC - Fiber Switching
- GMPLS kiterjeszti az MPLS/MPLS-TE vezérlést
 - Az LSP kihúzása lehetővé válik különböző interfészeken át

- MPLS is a strategy for streamlining the backbone transport of IP packets across a layer 3/layer 2 network. Although it does involve QoS issues, that is not its main purpose.
- MPLS is focused mainly on improving internet scalability through better traffic engineering.
- MPLS will help to build backbone networks that better support QoS traffic, but it entails significant changes in existing network architecture.

Köszönöm a figyelmet!

- End -

Szállítási réteg (L4)



Budapest University of Technology and Economics

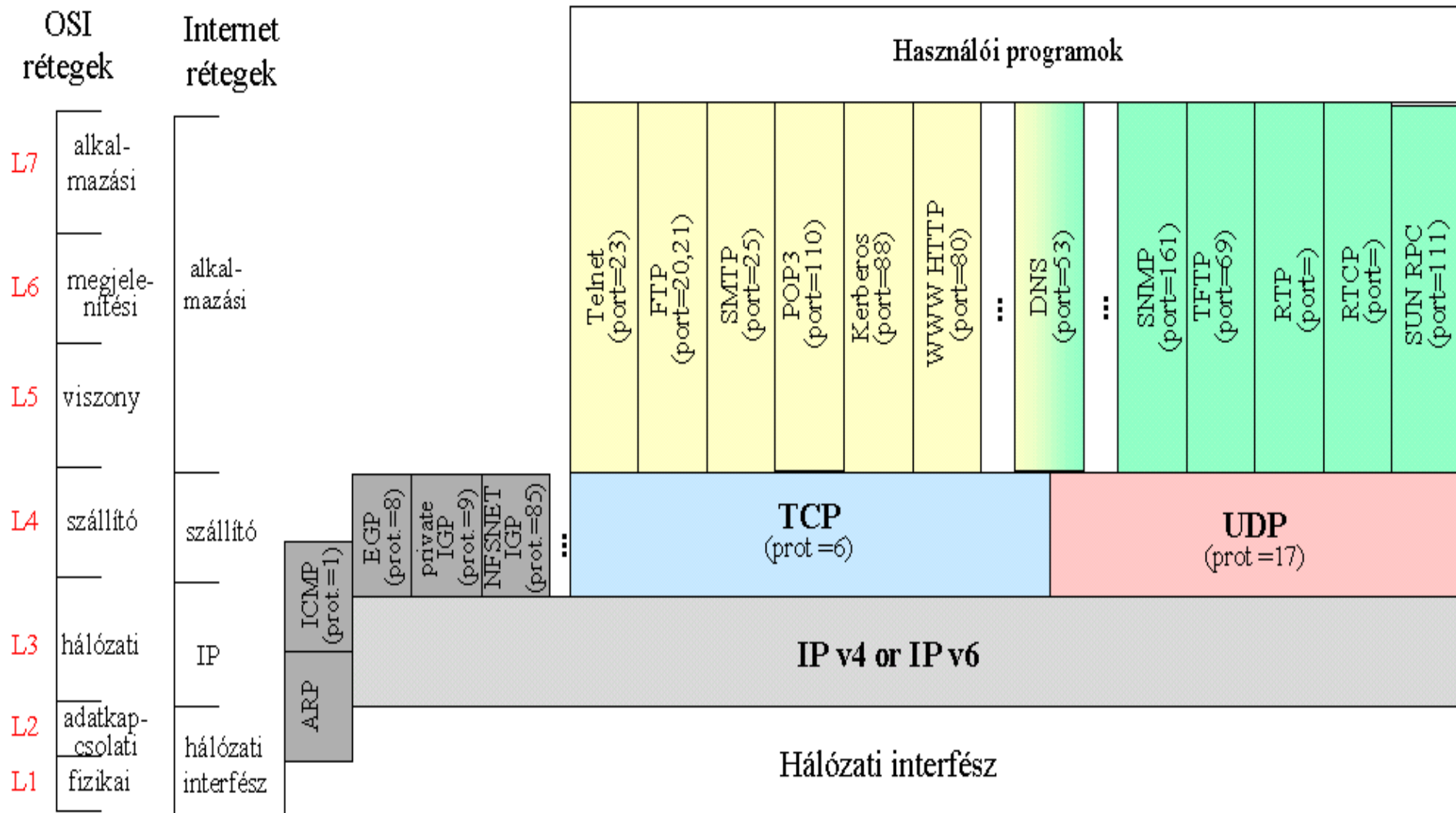
**Department of
Telecommunications and Media Informatics**



Protokoll stack



BME-TMIT



- Iteratív szerver
 - Vár, hogy érkezzen egy kliens igény
 - Feldolgozza a kliens igényét
 - Elküldi a választ az igényt küldő kliensnek
 - Ugrás az első lépéshez!
- Konkurrens szerver
 - Vár, hogy érkezzen egy kliens igény
 - Elindít egy új kiszolgálót, hogy kezelje a kliens igényét (új processz, vagy új szál)
 - Ha kész, e szerver működése megáll
 - Ugrás az első lépéshez!

- TCP szerverek általában **konkurrenszek**
- UDP szerverek általában **iteratívak**

- Az UDP Szerver iteratívan dolgozik
 - Általános eset:
 - Egy db well-known port az UDP forgalomnak
 - A sort a küldő végtelennek tekinti
- Ha mégis megtelik a sor:
 - Nincs figyelmeztetés az alkalmazástól a küldőnek
 - Nincs figyelmeztetés a csomageldobásra
 - UDP input sora FIFO

- TCP
 - Kapcsolat-orientált
 - Megbízható kapcsolat
 - Automatikus torlódás vezérlés
 - A küldési sebesség automatikus
 - Az alkalmazás nem tudja vezérelni
- UDP
 - Kapcsolat nélküli
 - A sebességet az alkalmazás szabja meg

- Két host között egyszerre több TCP/UDP folyamat lehet
- Azonosításra a portokat használják:
 - 16 bites számok a transzport fejlécben
- A forrásnál általában a számítógép választja ki
- A cél általában egy ismert szám
 - HTTP: 80, FTP: 21, SIP: 5060

- Egy kapcsolat azonosításához szükséges:
 - Forrás IP, forrás port
 - Cél IP, cél port
- Így egyszerre több kapcsolat lehet ugyanarra a portra – akár ugyanarról a forrás gépről
- Encapsulation – beágyazás

Szállítási réteg - UDP

UDP- User Datagram Protocol



User Datagram Protocol



BME-TMIT

- Egyszerű
- Datagram orientált
- Szállítási réteg beli protokoll
- RFC 768



- Kapcsolat nélküli protokoll
 - Nincs állapot információ
- Nem garantálja a csomagok megérkezését
 - Nem küld újra
 - Nem állít a küldési sebességen torlódás esetén
- Kis fejléc (8 byte)
- Főleg multimédia alkalmazások használják
 - Valós idejű átviteleknel jó választás

UDP kapcsolat kiépülés



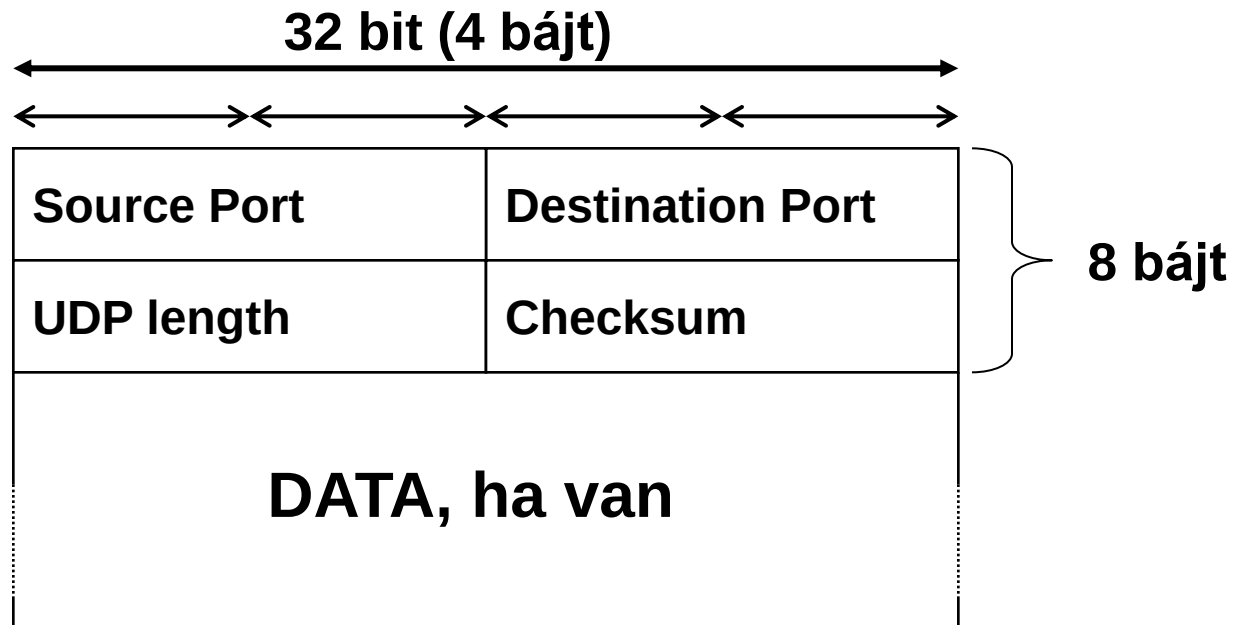
- A kapcsolat kiépülése a portok hozzárendeléséből áll
- A másik számítógép nem kell válaszoljon
 - Ha hiba van ICMP üzenettel jelzi
- A protokoll szegmensekre bontja az adatot
 - Az UDP szegmensek sorszámot kapnak
 - a szegmenseket azonnal küldi

- Nem megbízható
 - Nincs garancia, hogy a csomag elér a célba
- Nincs folyamvezérlés
- Egy UDP csomag – egy IP datagramm
 - Az alkalmazásnak kell odafigyelnie a helyes csomagméret használatra
 - Fregmentáció!
 - DF bit nem használható
 - A fregmentáció mindig tiltott

UDP csomagformátum



BME-TMIT



UDP packet structure



BME-TMIT

Source Port	Destination Port
-------------	------------------

- Source port (16 bit):
azonosítja a küldő alkalmazást
- Destination port (16 bit):
azonosítja a fogadó alkalmazást
- TCP, UDP port számok függetlenek
 - A TCP/UDP demultiplexáció az IP protokoll mező alapján!

UDP packet structure



BME-TMIT

UDP length	Checksum
------------	----------

- UDP length (16 bit): Az UDP fejléc és az UDP adat hossza
 - bájtban
 - Minimum érték: 8 bájt – nincs UDP adat
 - Elméleti max: 65507
- Checksum (16 bit): fejléc és az adatra számítva
 - A 16 bites szavak egyes komplementű összege
 - „pseudo fejléc” alapján számolt, bele tartozik:
 - IP csomagból: IP címek, protokoll azonosító mezők
 - Teljes UDP fejléc

- Pszeudo fejléc:
 - Kettős ellenőrzése a helyes átvitelnek
 - Szükséges, mert nincs folyamvezérlés
- Nem kötelező (teljes 0 – nincs checksum)
 - Ha a checksum csupa 0
 - Helyettesíti 65535 – egyes komplement aritmetika
 - Ha a checksum hibás:
 - A csomag csendben (*silently*) eldobásra kerül
 - Nincs hibaüzenet!

Szállítási réteg - TCP

TCP - Transmission Control
Protocol



- Két alkalmazás között nyújt
 - Megbízható végpont-végpont adattovábbítást
 - Kapcsolat-orientált adatfolyam szolgáltatás
 - Folyamvezérlő algoritmus
- Két végponti alkalmazás
 - Az adatközvetítés előtt fel kell építenie a TCP kapcsolatot
- Broadcastingra és multicastingra nem alkalmazható a TCP

Transmission Control Protocol



BME-TMIT

- A TCP adat IP csomagba enkapszulálva
- A TCP által az IP-hez továbbított adategység neve **szegmens**
- TCP logikai kapcsolatokat használ processz párok között:
 - TCP szegmens tartalmazza a forrás és a cél port számait
- Az IP cím és a megfelelő TCP port számok kombinációját hívjuk a kapcsolat **socketjének**, transzport címének
 - Socket párok

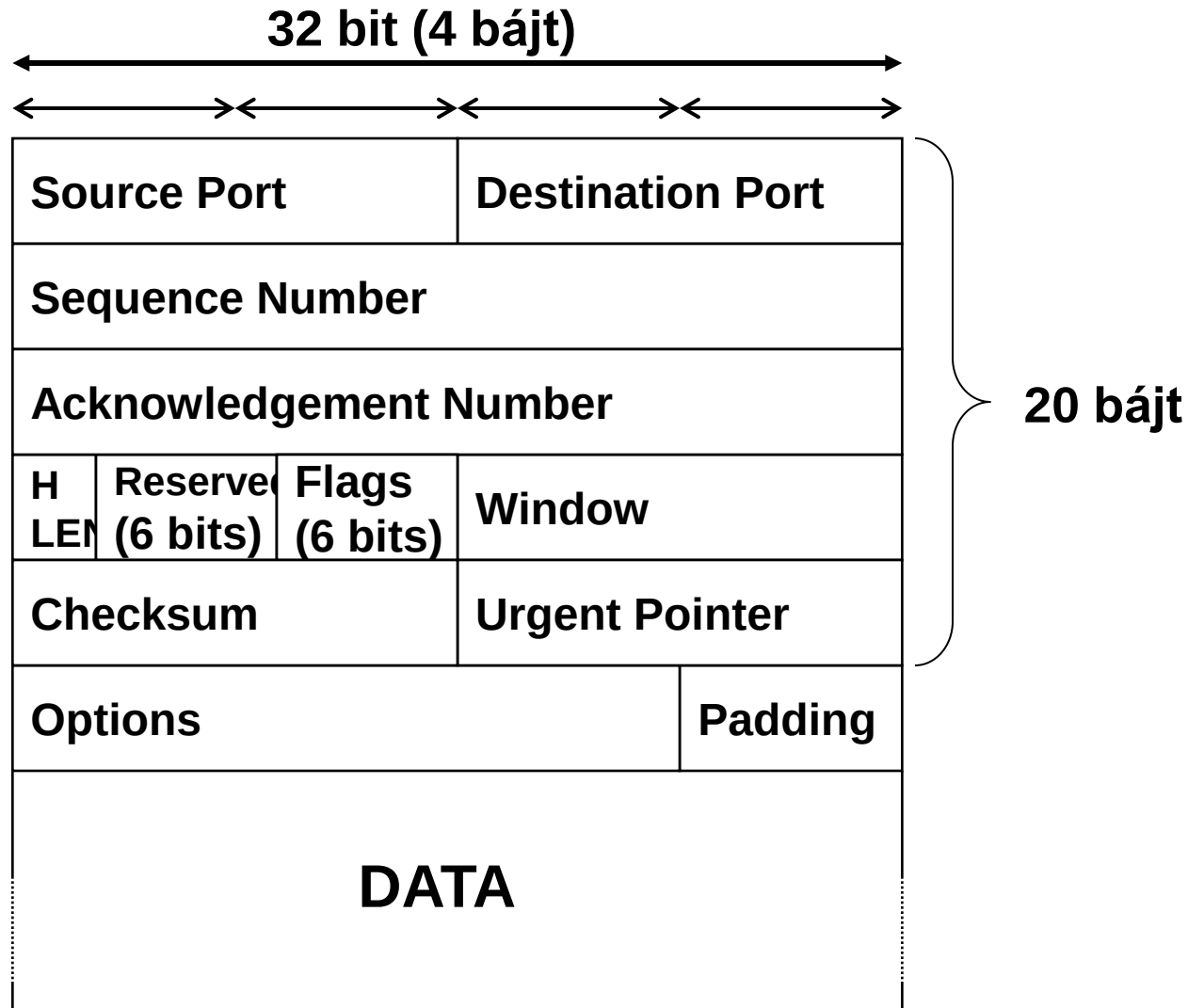
- ***full duplex*** szolgáltatást nyújt az alkalmazási rétegnek
 - Kétirányú adatátvitel
 - Mindkét végpont sorszámozást végez az adataikon
- ***nincs selective*** ACK
 - ack jelentése: eddig a bájtig (de a küldöttet nem beleértve) sikeres a vétel
- ***nincs negative*** ACK
- ***nem interpretálja*** a bájtfolyamot
 - Továbbadja az alkalmazásnak

- Csomag formátum
- Kapcsolat kiépítés
- Csúszó ablakos átvitel
- Torlódás vezérlés
 - Slow Start
 - Fast Retransmit
 - Congestion avoidance

TCP csomagformátum



BME-TMIT



TCP csomagformátum



BME-TMIT

Source Port	Destination Port
Sequence Number	
Acknowledgement Number	

- **Source port** (16 bit):
 - A TCP port száma a küldőnél
- **Destination port** (16 bit):
 - A TCP port száma a fogadónál
- **Sequence number** (32 bit):
 - A bájtfolyam adott szegmensének sorszáma
- **Acknowledgement number** (32 bit):
 - A fogadó által következőként várt szegmens sorszáma

- Azonosítják az alkalmazásokat (16 bit)
- well-known port számok (1-1023)
 - szerverek, pl. Telnet 23, FTP 21
- ephemeral port számok (1024-5000)
- Internet Assigned Numbers Authority, IANA

TCP csomagformátum



BME-TMIT

H LEN	Reserved (6 bits)	Flags (6 bits)	Window
----------	----------------------	-------------------	--------

- **Header Length** (4 bit):
 - A TCP fejléc 32 bites szavainak száma
 - Az options mező vááltozó hossza miatt szükséges
- **Reserved** (6 bits):
 - MBZ
 - Jövőbeni használatra foglalt
- **Flags** (6 bits):
 - 6 flag, melyek szabályozzák a TCP csomag viselkedését
 1. Urgent (URG)
 2. Acknowledgement (ACK)
 3. Push (PSH)
 4. Reset connection (RST)
 5. Synchronous (SYN)
 6. Finish (FIN)

TCP flagek



BME-TMIT

- Urgent flag (URG)
 - A végpontok üzenhetnek, hogy sürgős adat van az adatfolyamban
- Acknowledgement flag (ACK)
 - Megadja, hogy a nyugtaszám a szegmensben érvényes
- Push flag (PSH)
 - A szegmens adatokat tartalmaz, melyeket az alkalmazásnak kell továbbítani

Source Port		Destination Port	
Sequence Number			
Acknowledgement Number			
H LEN	Reserved (6 bits)	Flags (6 bits)	Window
Checksum		Urgent Pointer	
Options			Padding

TCP flagek



BME-TMIT

- Reset flag (RST)
 - Reset szegmenst küld a TCP
 - ha nem megfelelő portra érkezik kapcsolatkerés
 - Ha az egyik fél meg akarja szakítani a kapcsolatot
- Synchronous flag (SYN)
 - A SYN flag bekapcsolt azokban a szegmensekben, melyek a kapcsolatfelépítéshez szükségesek
- Finish flag (FIN)
 - A végpontok kapcsolat lezárásra használják ezt a flaget

Source Port		Destination Port	
Sequence Number			
Acknowledgement Number			
H LEN	Reserved (6 bits)	Flags (6 bits)	Window
Checksum		Urgent Pointer	
Options			Padding

TCP csomagformátum



BME-TMIT

H LEN	Reserved (6 bits)	Flags (6 bits)	Window
Checksum			Urgent Pointer

- **Window** (16 bit):
 - Az adatfolyam vezérléshez szükséges
 - Megadja, hogy a fogadónak mennyi bájt adat fogadására képes a buffere.
- **Checksum** (16 bit):
 - A TCP fejléc integritásának megőrzésére
 - A checksum pszeudo fejléc alapján számítható, információkat véve az IP fejlécből is

TCP csomagformátum



BME-TMIT

Checksum	Urgent Pointer
Options	Padding

- **Urgent Pointer** (16 bit):
 - Ha sürgős adat van a szegmensben ez a pointer mondja meg, hogy hol kezdődik az az adatrészben
- **Options**:
 - A leggyakoribb opció mező az MSS - maximum segment size
 - Megadja a legnagyobb szegmensméretet, melyet a fogadó fogadni szeretne

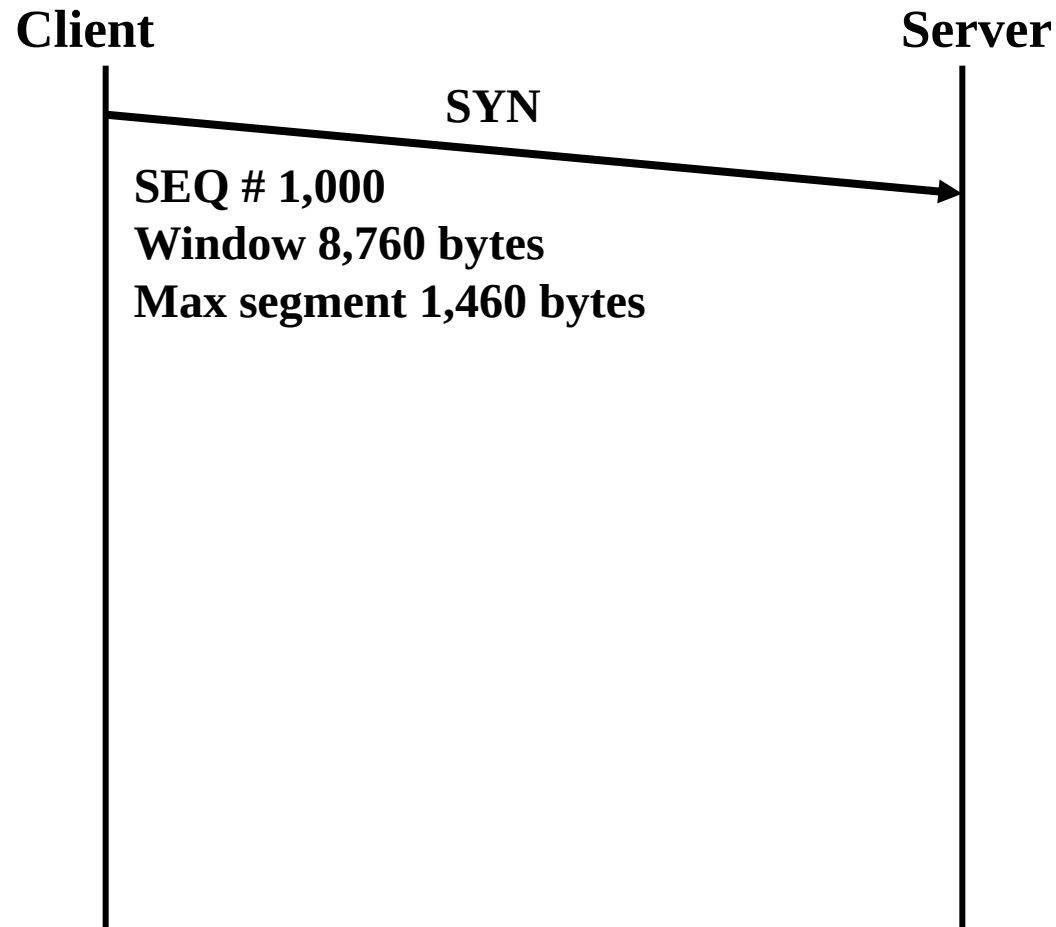
TCP kapcsolat felépítés és bontás



- A TCP az adatszegmensek továbbításakor a következőket végzi:
 - Kapcsolat felépítés
 - Ablakméret (Advertised window size), Maximum szegmens méret meghirdetése
 - Adatok továbbítása
 - Nyugták küldése a fogadott szegmensekre
 - Kapcsolat lezárása

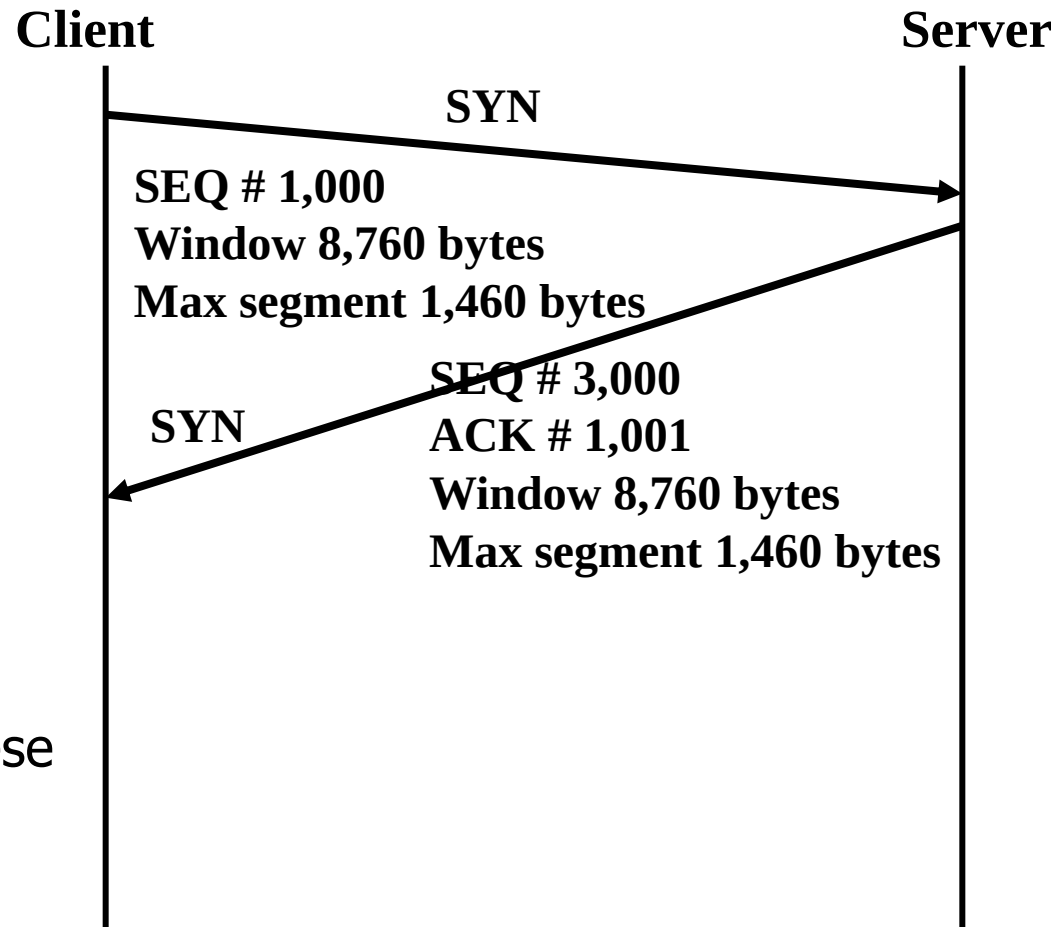
1. Kezdeményező végpont - kliens

- SYN szegmens küldése
 - Szerve port számának megadása – ahová kapcsolódni szeretne
- Kezdeti saját, sorszám
 - ISN – initial seq. num.
- Saját ablakméret hirdetése
- MSS hirdetése



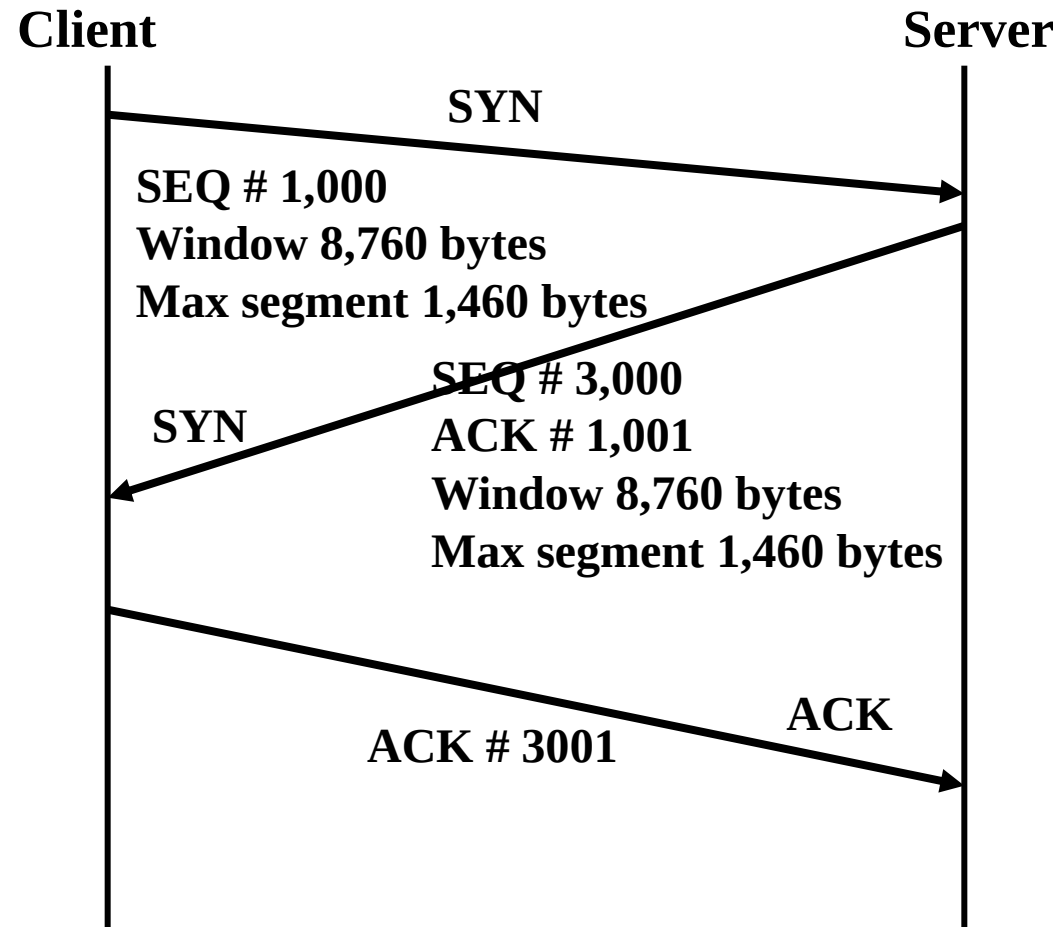
2. The server válaszol

- SYN szegmens tartalmazza
 - Szerver ISN
 - Nyugta a kliens szegmensére
 - Várja az 1001-et
- Saját ablakméret hirdetése
- MSS hirdetése



3. A kliens nyugtáz

- Kötelező a kliensnek nyugtát küldeni a szerver SYN szegmensére

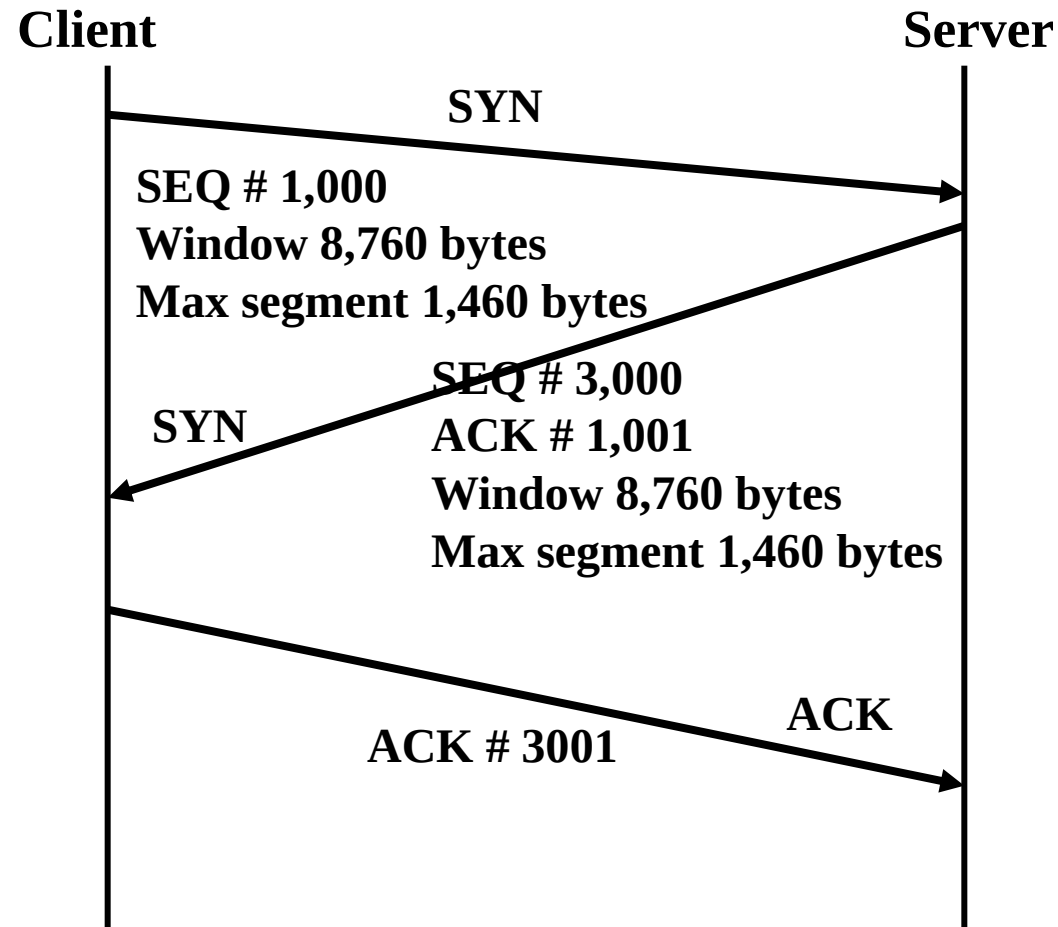


Three-way handshake



BME-TMIT

- TCP nagyon érzékeny a SYN szegmens elvesztésekre
 - Hosszú időzítések kapcsolat felépítéskor
- SYN szegmens 1 sorszámot foglal
 - Adatoknál bájtokat számlál a seqnum

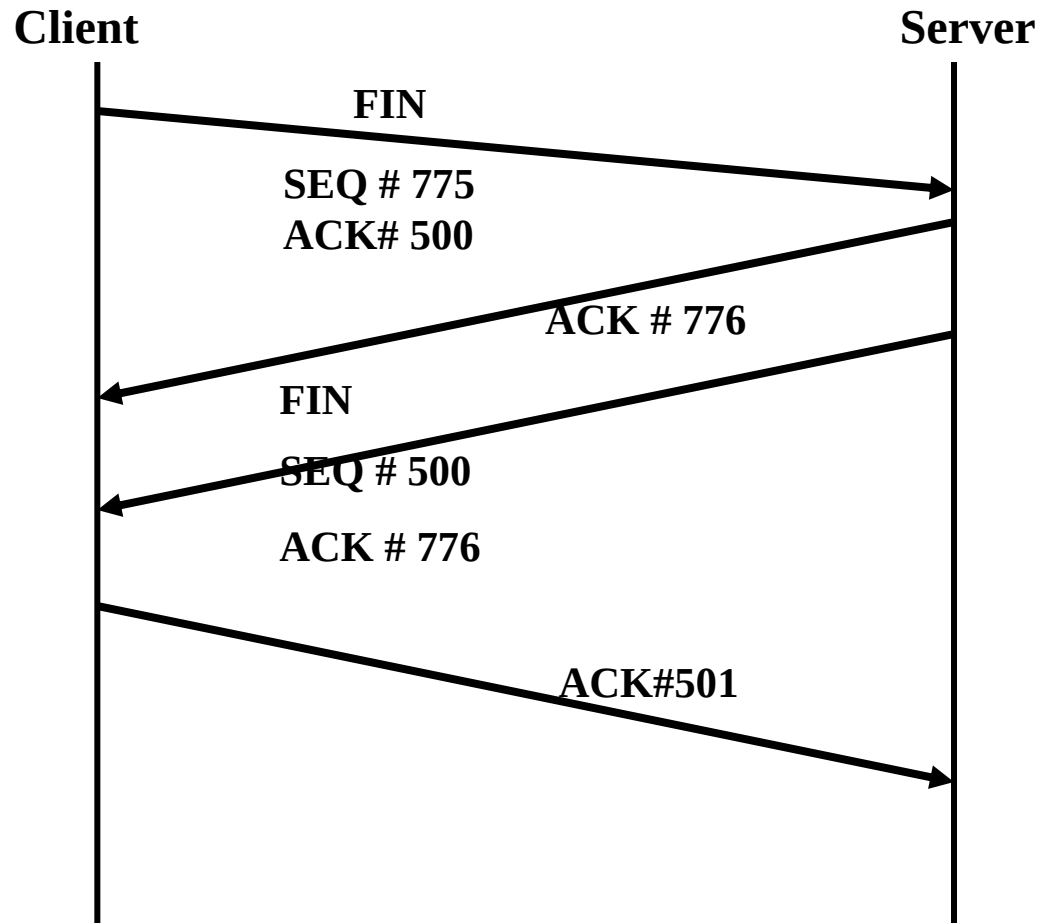


TCP kapcsolat lezárása



BME-TMIT

- 4 szegmens a kapcsolat szabályos lezárásakor
- Mindkét végpont egymástól függetlenül lezár
- FIN fogadásakor
 - A TCP-nek értesíteni kell az alkalmazást, hogy a túloldal lezárta a kapcsolatot
 - TCP FIN fogadás után továbbra is tud adatokat küldeni



Maximum Segment Size, MSS



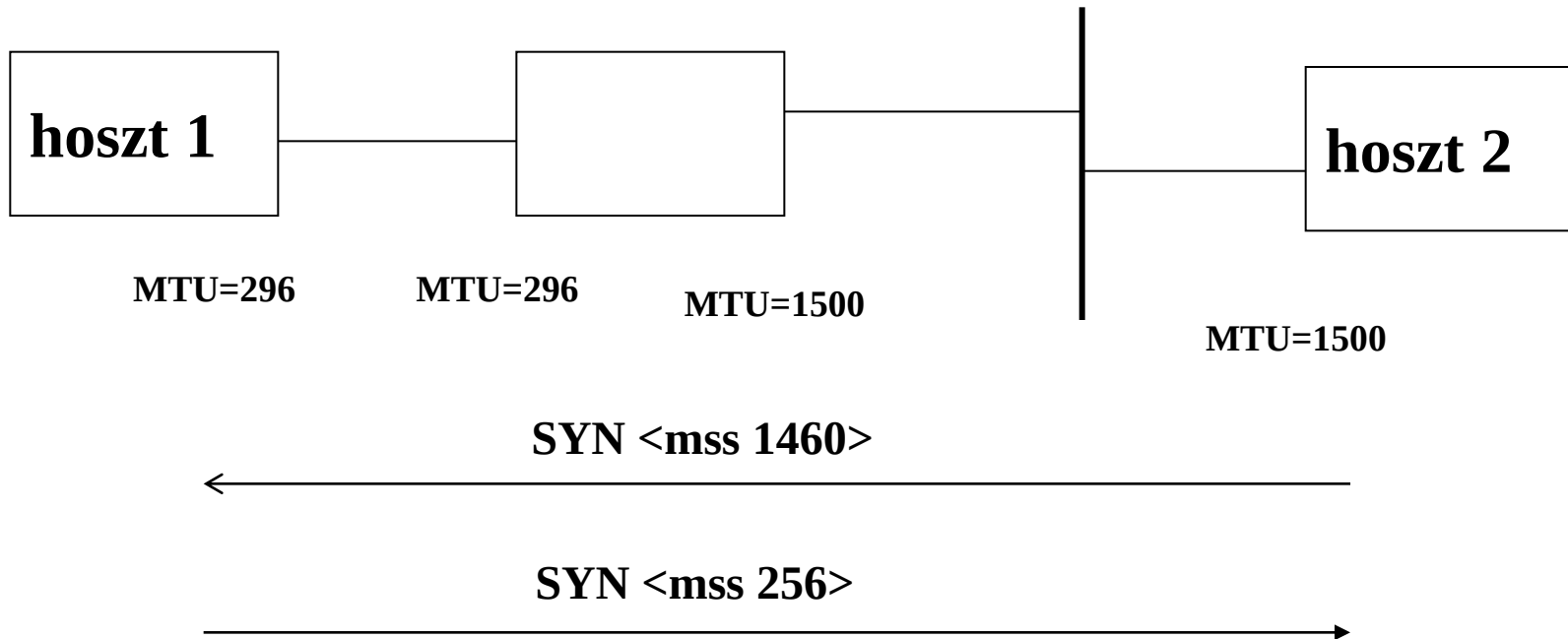
BME-TMIT

- Legnagyobb szegmens méret, melyet a másik oldal küldhet
- Mindkét végpont meghirdeti, hogy milyen méretet vár el
- MSS opció a SYN szegmensben
 - Ha nincs ilyen, akkor alapértelmezett (536 bájt)
- MSS max értéke:
 - A kimenő interfész MTU csökkentve az IP és TCP fejléc méretével
 - Ethernet: $1500 - 20(\text{IP}) - 20(\text{TCP}) = 1460$ bájt max

MSS értékek



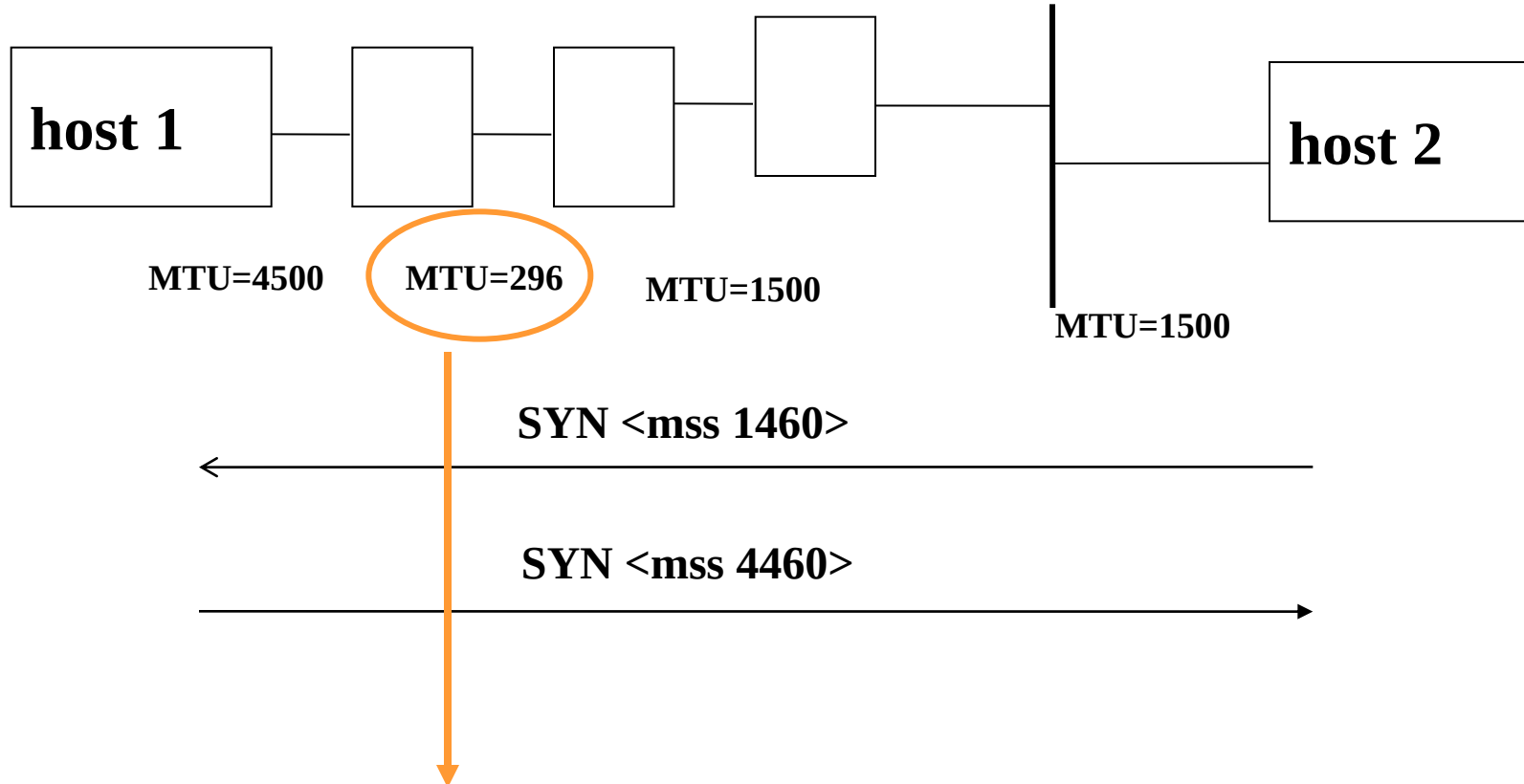
BME-TMIT



MSS értékek 2



BME-TMIT



MTU Path discovery!

- Kapcsolat felépítés után a TCP
 - Saját MSS
 - Túloldal által hirdetett MSS
 - Minimumát használja a szegmensek méretéhez
- Előfordulhat az útvonalon kisebb MTU-jú hálózat
 - Kommunikáció közben derülhet ki
 - Felderítéséhez a TCP beállított DF bites IP csomagokat küldhet

- Ha fregmentáció történik ICMP üzenet érkezik vissza:
 - "DF set, can't fragment"
- TCP csökkenti a szegmens méretet és újraküld
 - Az újabb ICMP tartalmazza a következő hop MTU-ját
 - Régebbi ICMP nem tartalmazza, következő kisebb MTU-val próbálkozik a TCP
- Nagyobb MTU érték próbája MSS növeléséhez (~10 percenként)
 - Routing változás esetén lehet
 - Kis szegmens – nagyobb overhead

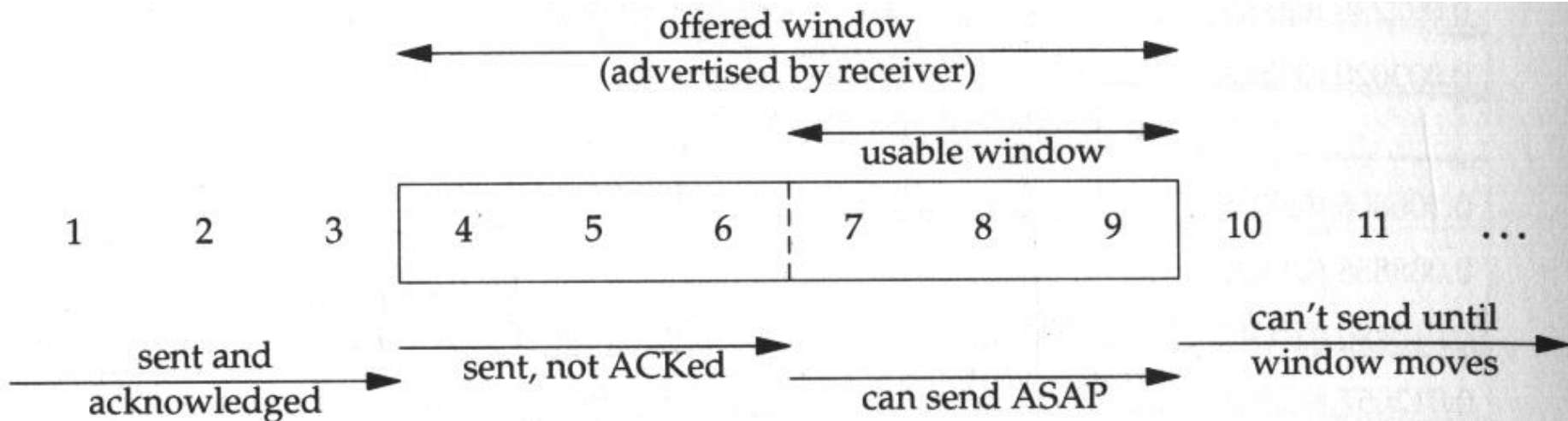
Csúszó ablak

„TCP Sliding window”



Csúszó ablak – Sliding Window

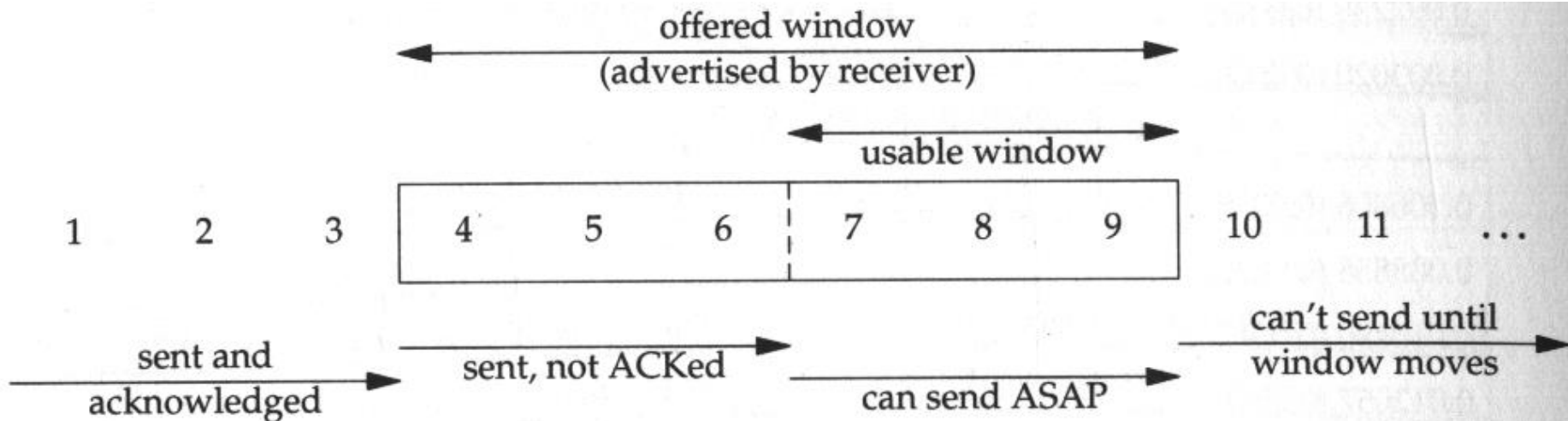
BME-TMIT



- Csúszóablak jellemzői:
 - Mérete **bájtban** adott
 - Ábrán szegmens számok!!!
- Előző példa:
 - Window méret 4096
 - Ebbe 4 db szegmens fér bele, ha egy szegmens 1024
 - Mert pl. az MSS = 1460

Csúszó ablak – Sliding Window

BME-TMIT

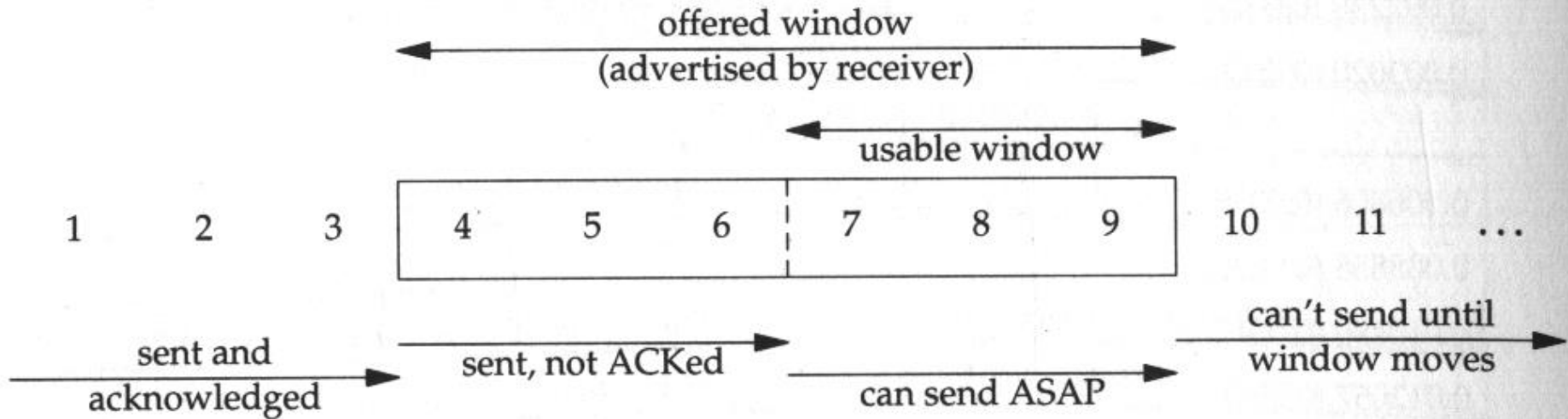


- 1-3 szegmensek már nyugtázottak
- **offered window:**
 - A fogadó által meghirdetett ablakméret
 - 4-9 szegmensek:
 - 4-6 elküldött, de még nem jött nyugta
 - 7-9 azonnal küldhető szegmensek
- **usable window:**
 - az azonnal küldhető szegmensek

Csúszó ablak – Sliding Window



BME-TMIT

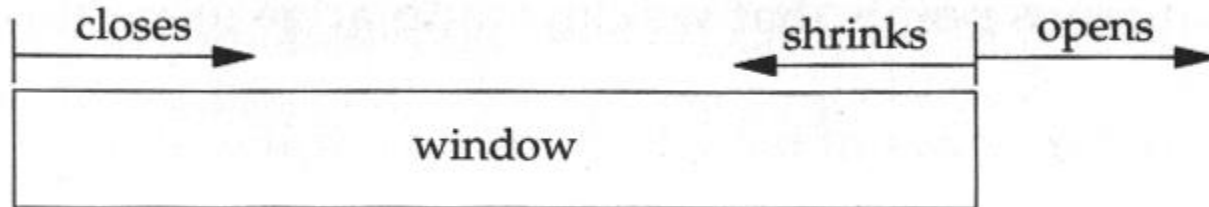


- 10 ... szegmensek
 - Elküldésre váró szegmensek
 - Csak ha érkezik nyugta és az ablak tovább csúszik

A csúszóablak végeinek mozgása



BME-TMIT



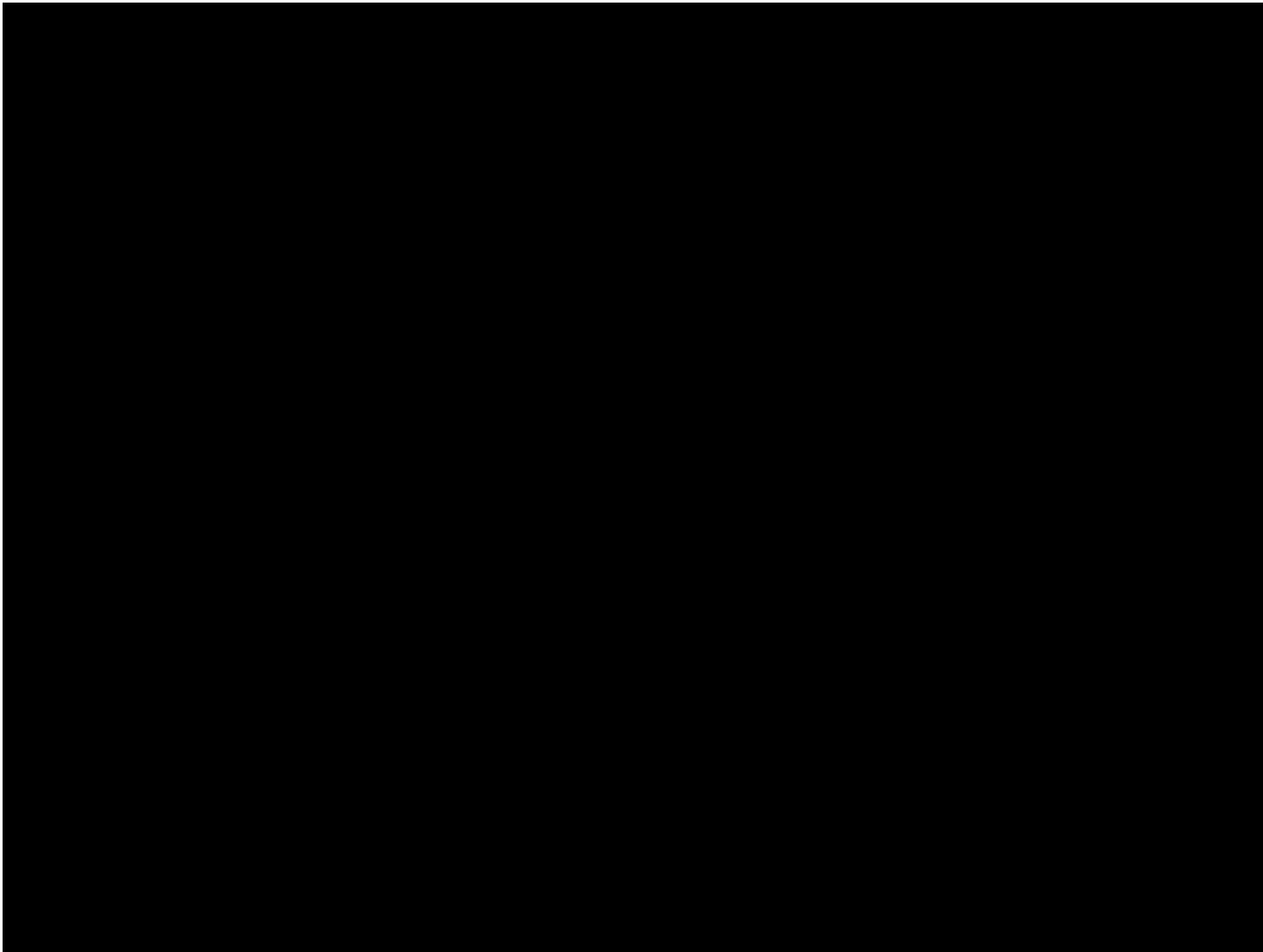
- **Zár - Closes:** ha nyugta érkezik
- **Nyit - Opens:** ha a fogadó oldalon az alkalmazás fogadja az adatokat – ürül a buffer
- **Összehúzódik - Shrinks:**
 - Normál esetben nincs
 - De a TCP-nek kell tudni kezelni!

Csúszóablak működése



BME-TMIT

http://www2.rad.com/networks/2004/sliding_window/



- Az ablak méretét a fogadó oldal kezeli
 - TCP teljesítményére hatással van
- Általában (window size)
 - Az általános **alapértelmezett érték 8 kb-át**
 - **Nem mindig optimális**
 - **Korlátozó tényező lehet!**
- Az **optimális ablak mérete** függ
 - A kommunikációs média **sávszélességétől**
 - A két hoszt közötti **körülfordulási időtől**
 - Round trip time

Slow Start



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**

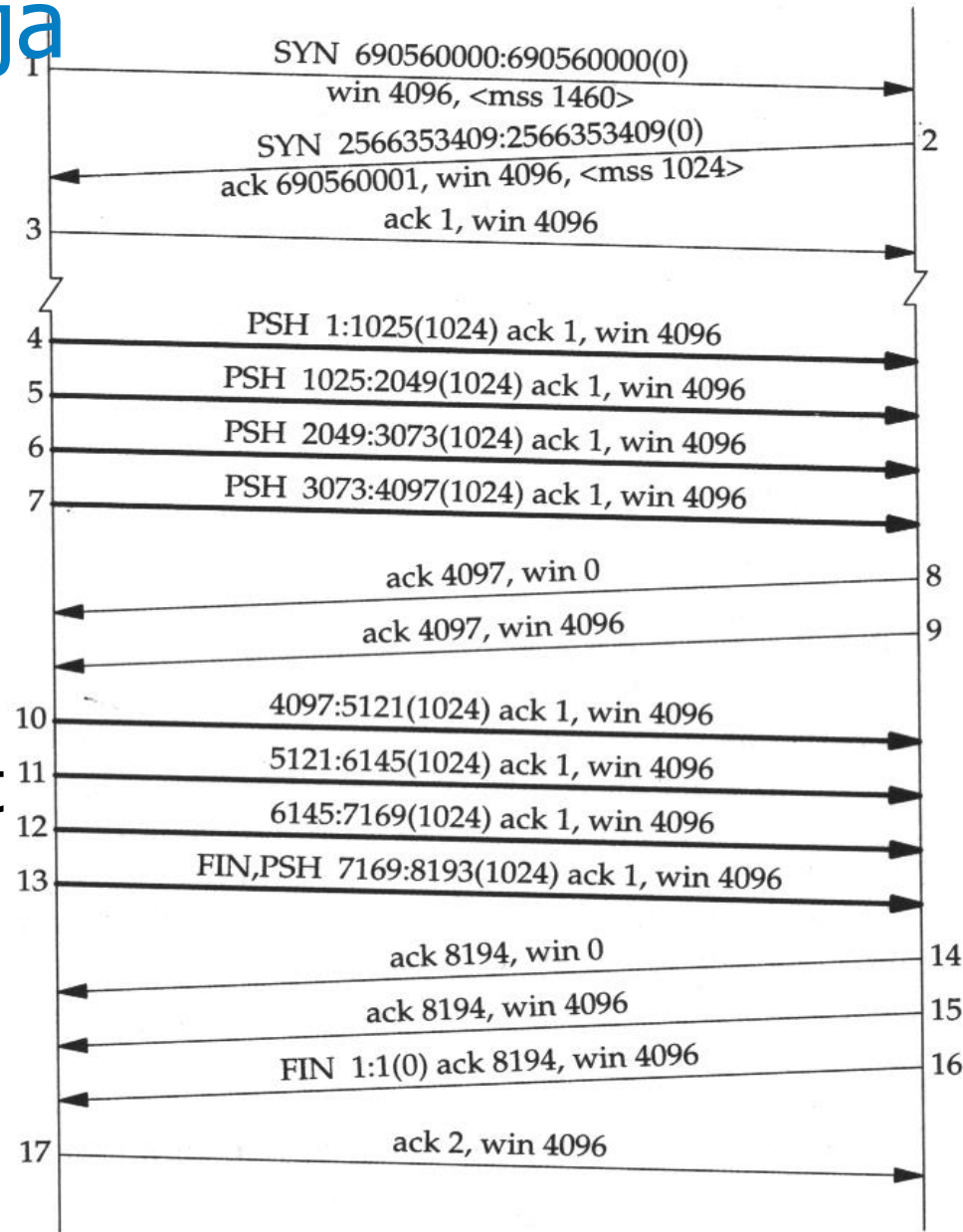




Többszörös csomagok küldésének problémája

- Korábbi példa:

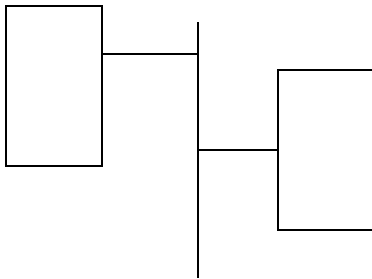
- A küldő a kapcsolat elején rögtön elküldi a meghirdetett ablakméret szerinti szegmensmennyiséget
- Milyen problémákat okozhat ez?



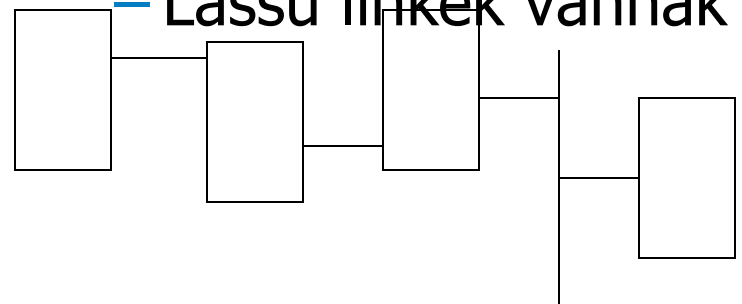
Többszörös csomagok küldésének problémája

Többszörös csomag küldés

- Alkalmazható
- ha a két fél egy LAN-on van



- Nem alkalmazható
- ha a két fél nem egy alhálózaton van,
- köztük
 - Routerek
 - Lassú linkek vannak

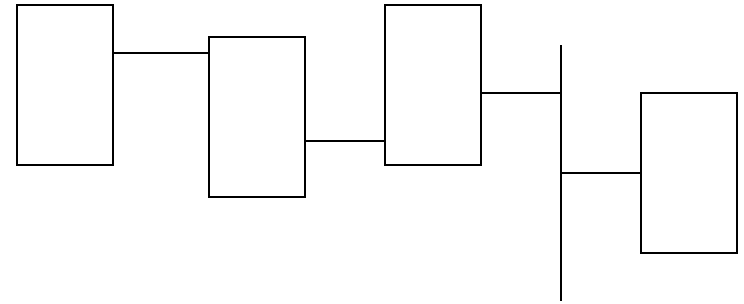


Többszörös csomag küldés nem alkalmazható



BME-TMIT

- A közbenső routerekben felsorakozhatnak a csomagok
 - Csomageldobás következhet be



- Jelentős teljesítménycsökkenést okoz

- **Megoldás: TCP Slow start algoritmus**



Slow Start – lassú indítás



BME-TMIT

- Mire való?
 - Elkerülhető vele a meghirdetett ablakméret szerinti többszörös csomagküldésből származó problémák
- Slow Start algoritmus alapja
 - Meghirdetett ablakméreten felül
 - Definiál egy ***congestion window (cwnd)*** változót a kapcsolathoz
 - Megadja az adott pillanatban elküldhető maximális szegmensszámot
 - A kapcsolat állapotától függ
 - Számlálása bájtban (példákban szegmensszámmal!)
- A cwnd - küldő oldal forgalomszabályzója

- Új kapcsolat létrehozatalakor
 - Congestion window; **cwnd = 1 szegmens**
 - Jelentés: a küldő 1 szegmenst küldhet
- Minden alkalomkor, ha ACK érkezik
 - cwnd növelhető egy szegmensnyivel
 - (cwnd bájtban számol, de a növekedések mindig szegmens egységnyik).
- A küldő egyszerre mindig ***min(cwnd, advertised window)*** mennyiségű adatot küldhet
- A cwnd maximuma az advertised window
- **Exponenciális növekedés** az átviteli sebességben
- Egy bizonyos határ felett a közbenső hálózat elérhető sávszélességét eléri a kapcsolat
 - A közbenső routerek elkezdik eldobni a csomagokat
 - Vissza szabályzás!

Az exponenciális növekedés

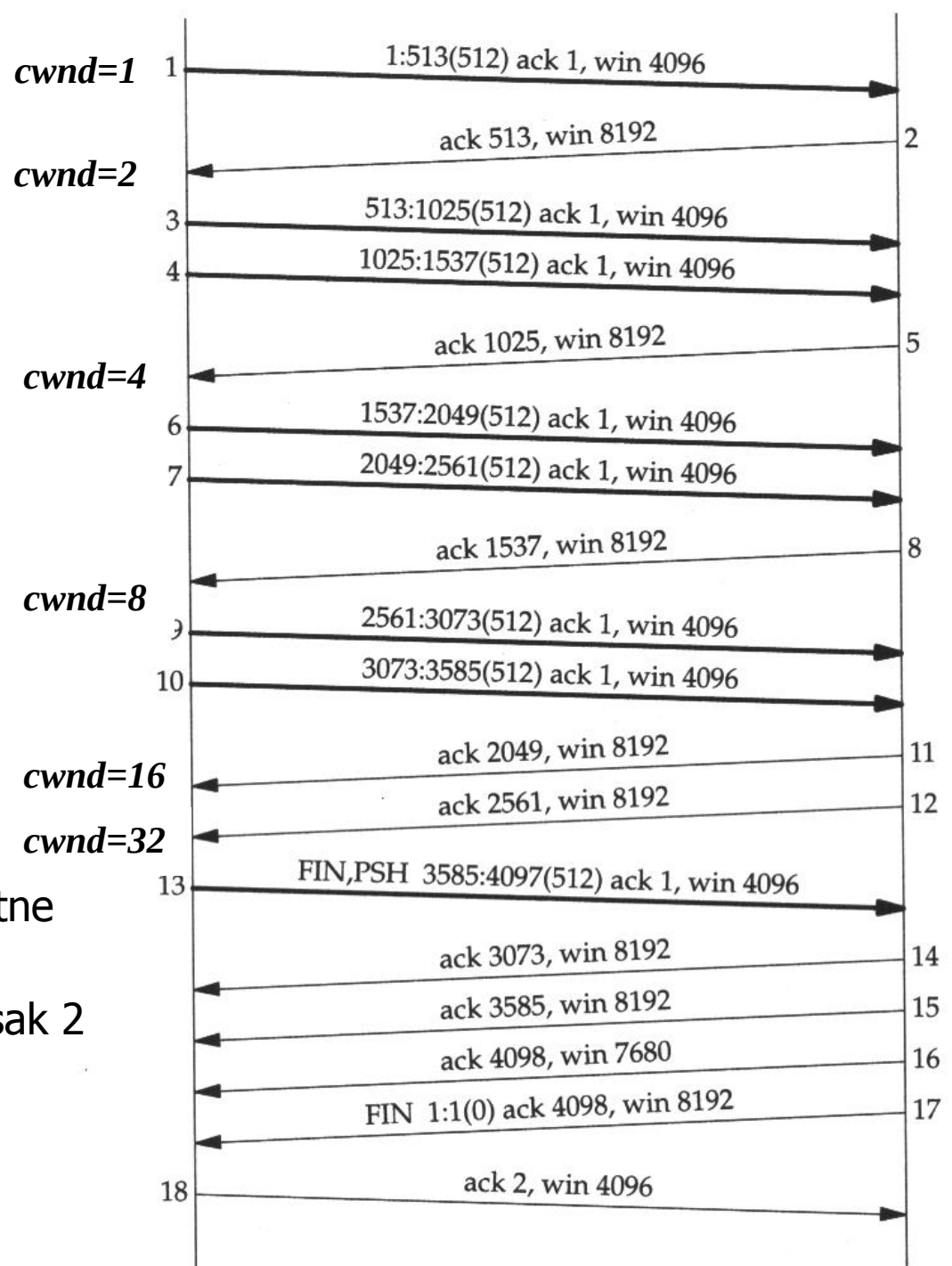


BME-TMIT

- Küldő kiküld 1 szegmenst ($cwnd=1$)
 - Vár az ACK-ra
- ACK megérkezik
 - **cwnd** megnövelhető 2-re
 - 2 szegmens küldhető
- Megérkezik a nyugta
 - **cwnd** 4-re növekszik
- ...
- $\approx$ minden körülfordulási időnként (RoundTripTime)
 - A **cwnd** megduplázódik

Példa a Slow Startra

- Cwnd=1
 - 1 szegmens küldése
 - Bár a window=8192 a fogadónál
- Nyugta
- Cwnd=2
 - 2 szegmens küldése
- Nyugta
- Cwnd=4
 - 4 szegmens küldése lehetne
 - DE a nyugta csak 1-et nyugtázott, így most is csak 2 szegmens
- ...



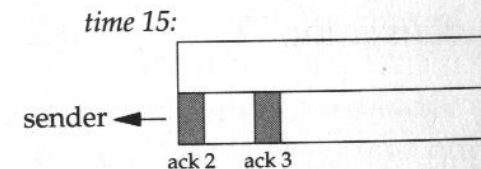
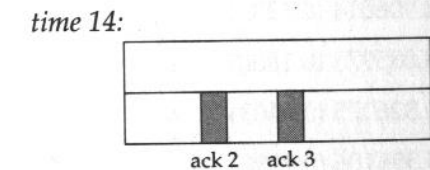
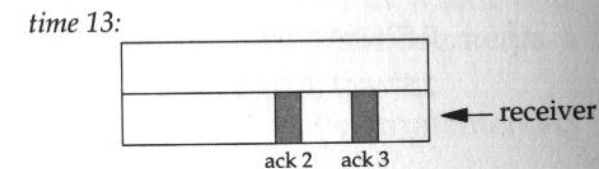
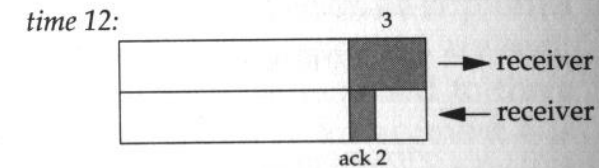
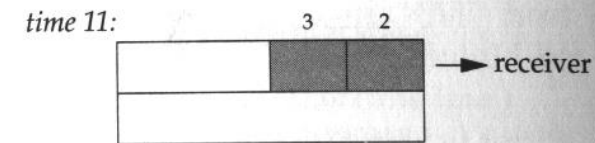
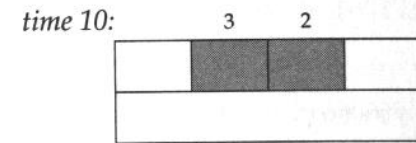
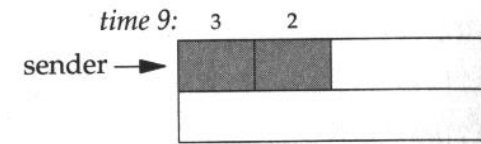
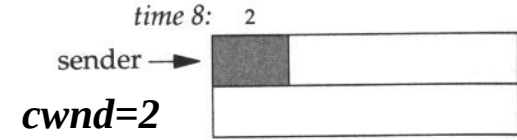
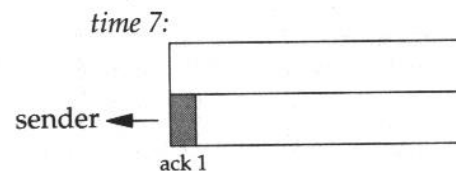
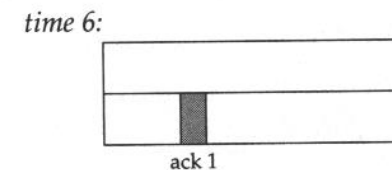
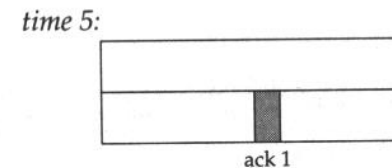
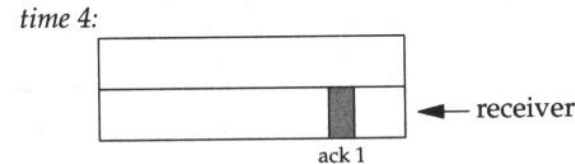
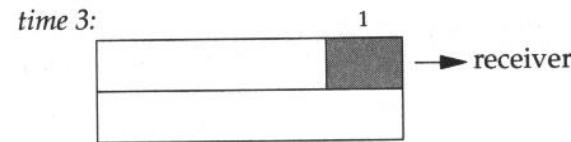
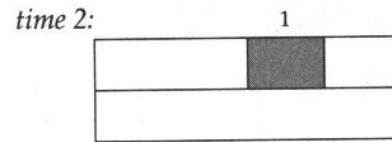
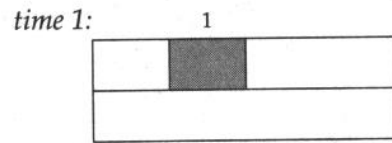
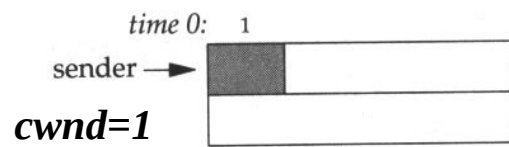
Ideális álapot elérése

- **time 0:**

- küldő egy szegmenst továbbít
- A slow start miatt ACK-ra vár
- RTT: 8 időegység

- **ACK megérkezése**

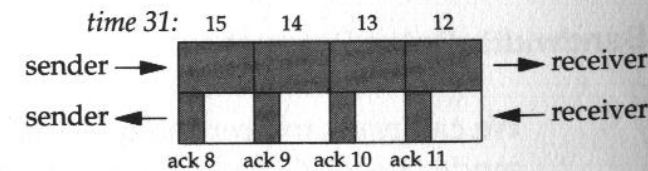
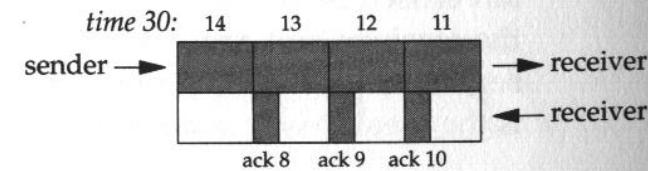
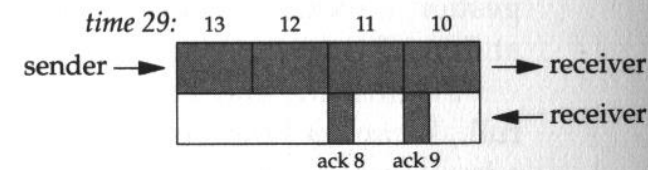
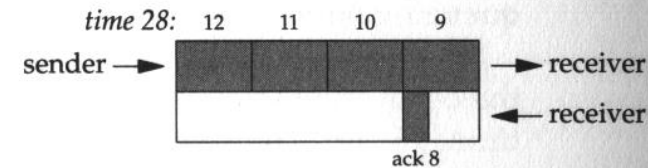
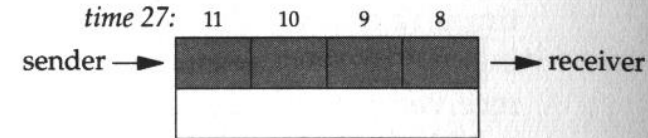
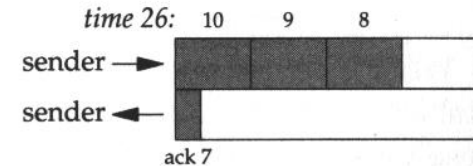
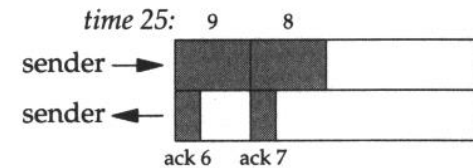
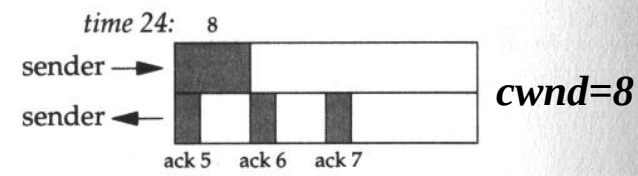
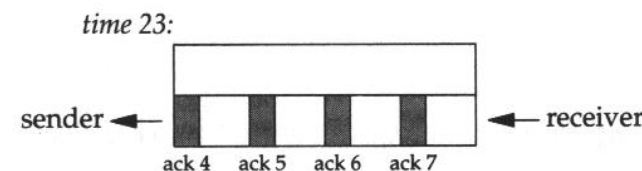
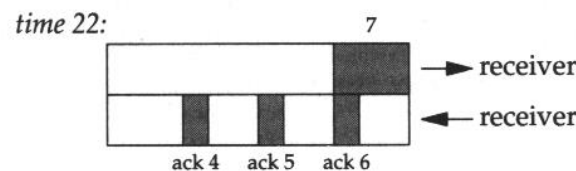
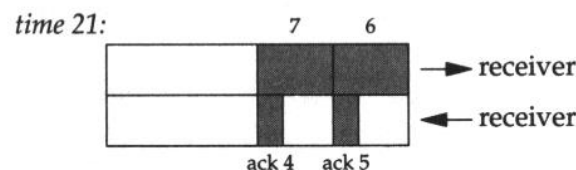
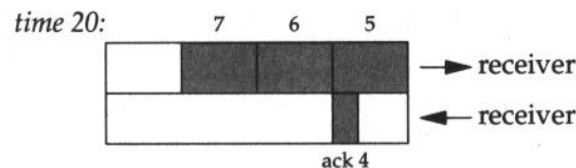
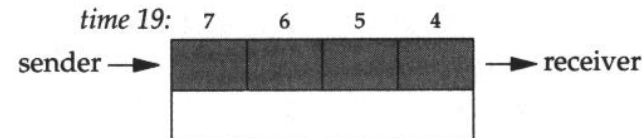
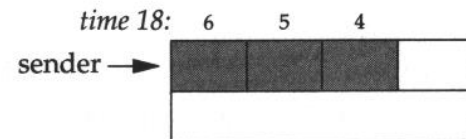
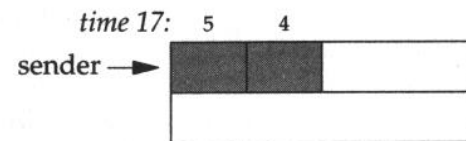
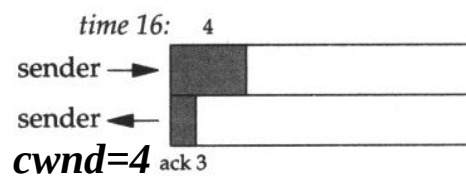
- két szegmens küldhető (cwnd=2)



Ideális állapot elérése

- 31-től a csatorna a küldő és a fogadó között telített

- Ideális állapota a kapcsolatr



TCP interaktív adatfolyam

TCP interactive data flow



- Bulk – tömeges
 - ftp, e-mail, ... (általában maximális méretű szegmensek)
- Interaktív
 - telnet (minimális, kb. 10 bájtot szállítanak)
- Csomagok száma alapján
 - Összes TCP szegmens 50%-a bulk adat
- Szállított bájtok alapján
 - 90% bulk adat
 - 10% interaktív adat

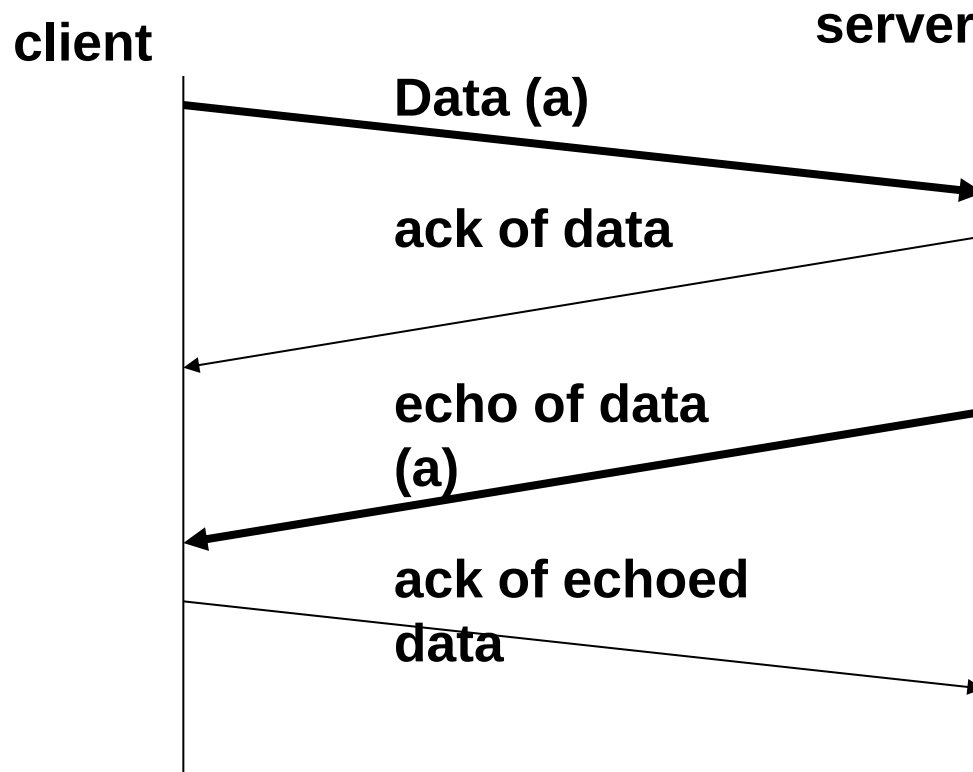
- Adatfolyamvezérléshez különböző típusú algoritmusok
 - Kis csomagok (kevés hasznos adat)
 - Küldés: ritkán

Interaktív adatátvitel (telnet)



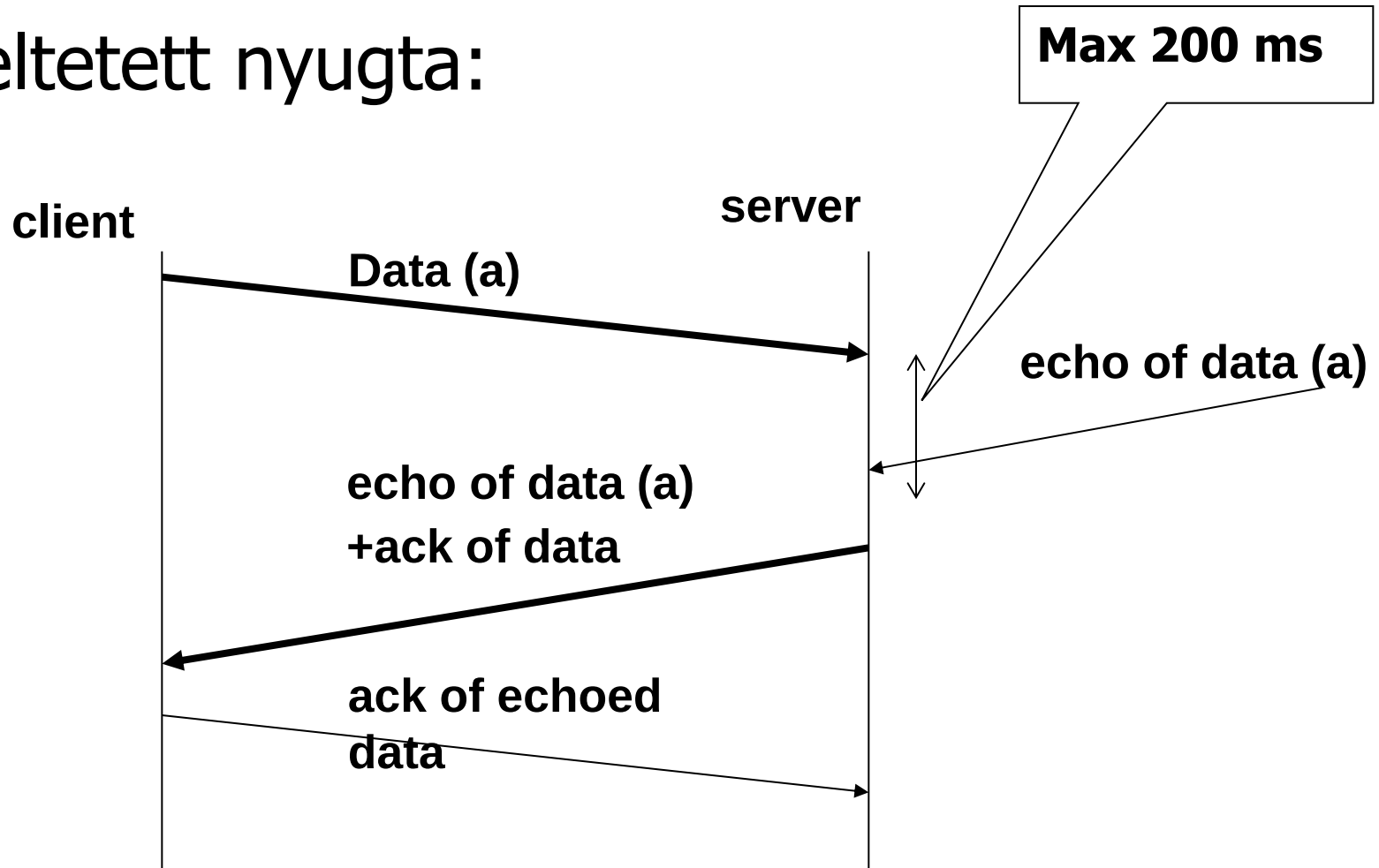
BME-TMIT

- „a” karakter átvitele és megjelenése a képernyőn
 - általában a 2. és a 3. szegmensek összevonhatók



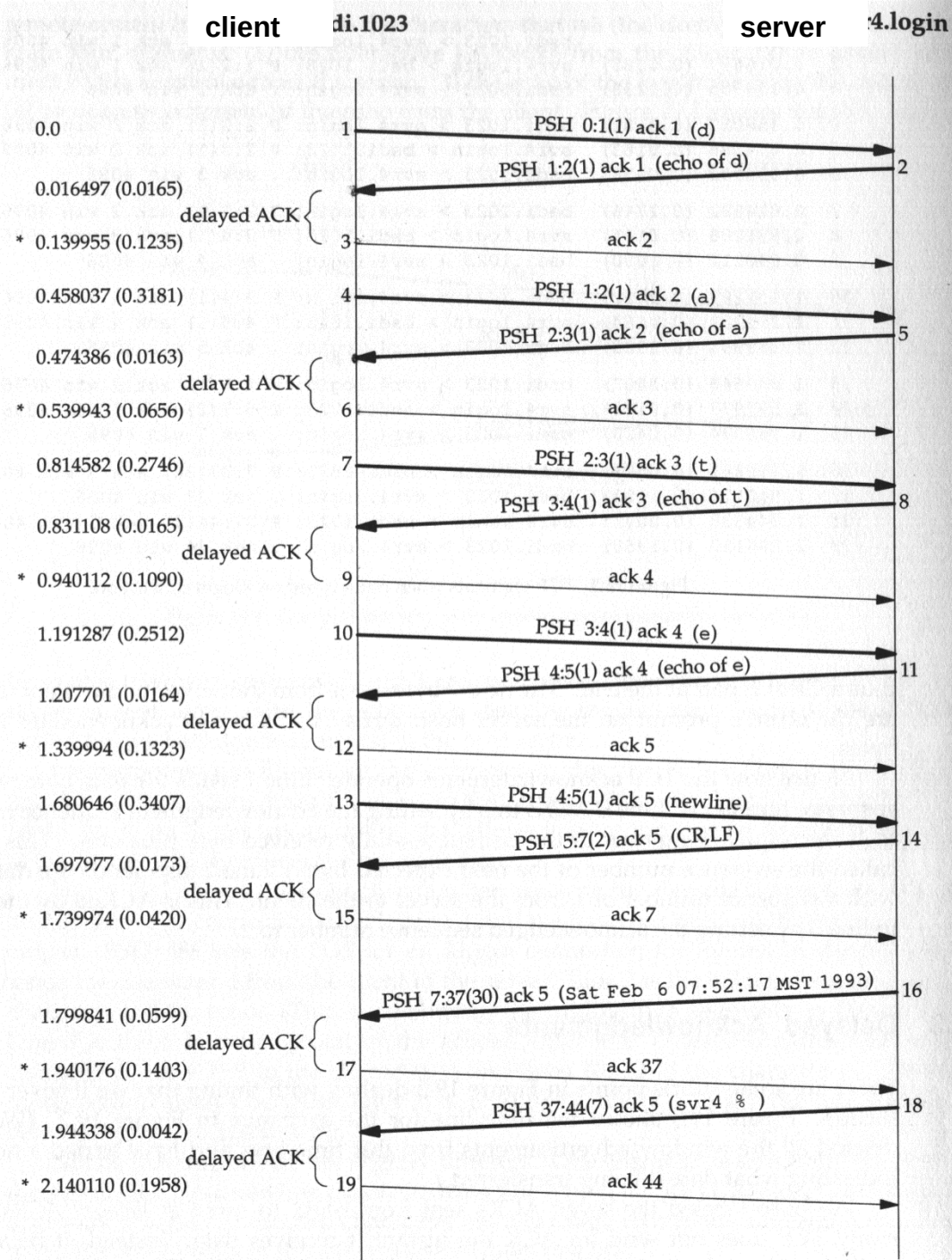
- TCP az adat fogadásakor nem küld rögtön nyugtát
 - Késlelteti az ACK kiküldését:
 - Várja, hogy jön-e hasznos adat – majd annak a fejlécében nyugtáz
 - (ACK ***piggyback*** with the data)
- Pl. 200 ms-os időzítő az ACK-ra:
 - TCP az adat fogadása után maximum 200 ms-ig késlelteti az ACK küldést
 - Ha nem jön adat ACK magában megy

- Késleltetett nyugta:



Delayed ACK példa

- Date
parancs...
- ...és a válasz



- Gyakran mennek 1 bájtos csomagok interaktív forgalom esetén
 - 1 bájt adat - 41 bájtos IP datagrammba kerül
 - tinygram
- Nagy overhead: hasznos adat - összes adat arány alacsony
 - Torlódás okozó
 - LAN-on nem probléma
 - WAN-on sok ilyen forgalom probléma lehet
 - Rossz sáv szélesség kihasználás

- Nagle algorithm
 - Ha a TCP kapcsolatnak van kintlévő adata, amelyet még nem nyugtáztak, kis szegmensek nem küldhetők, amíg a nyugta meg nem érkezik
 - Helyette, a kismennyiségű adatokat összegyűjti és nyugta vételekor egy szegmensben küldi el őket
- Minél gyorsabban jön a nyugta, annál gyorsabban küldi az adatokat

- Előfordul, hogy szükséges kis szegmensek átvitele:
 - X-window egér pozíciók
 - Speciális terminálfunkciók
 - Több bájtos adatok (escape karakterrel kezdve)
 - Előfordulhat, hogy az első bájtot elküldi a kliens, majd vár a nyugtára
 - De a szerver is vár a nyugtával
 - Észrevehető késleltetés!

TCP tömeges adatfolyam

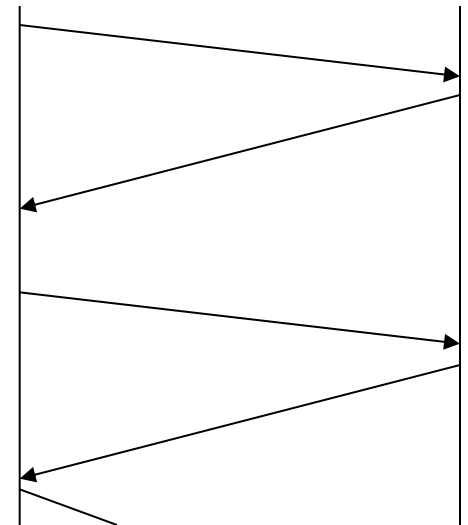
TCP bulk data flow



- Stop-and-wait protocol
 - Adatszeletet elküldése
 - Nyugtára vár
 - Nyugta megérkezik – első lépés (következő adatszelettel)
 - Nyugta nem érkezik (időzítés) – előző adatszelet újraküldése, majd nyugtára vár

- Jellemzők

- Minden adatszeletre nyugta
- Nagy távolság esetén nagyon lassú



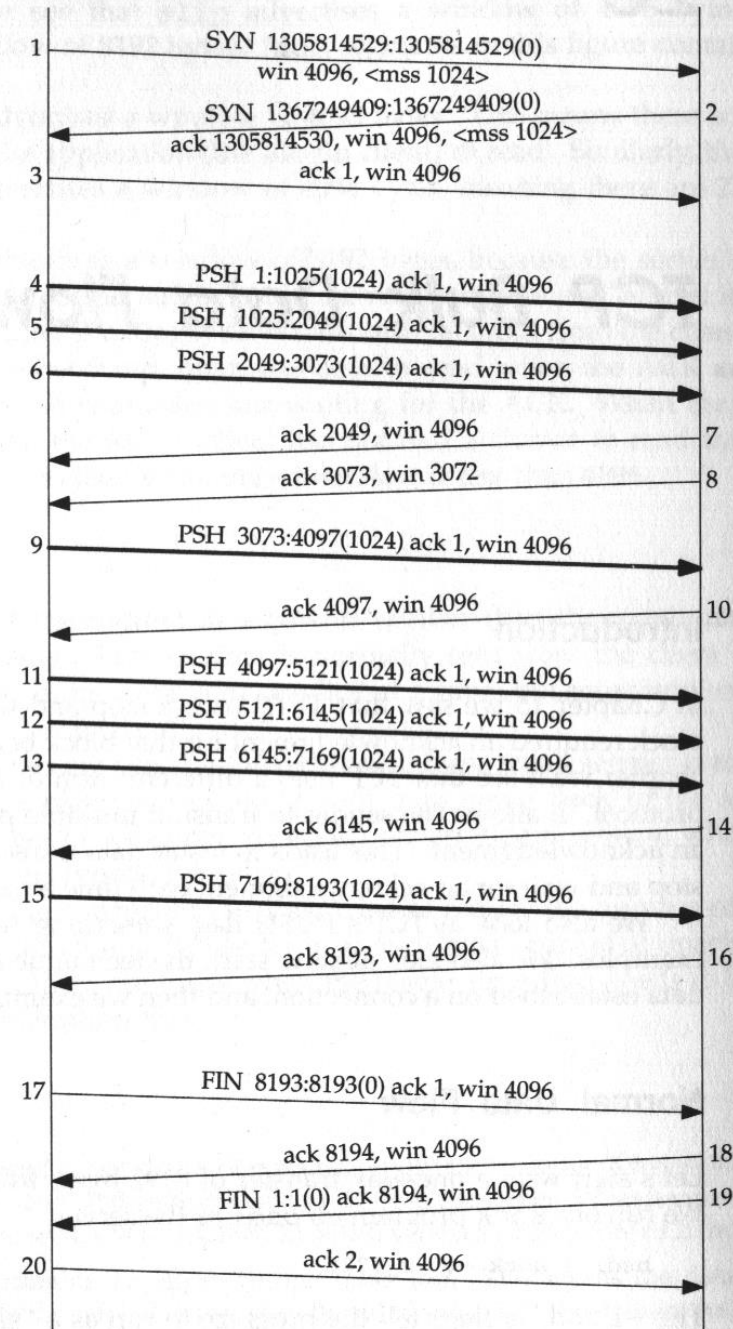
- Sliding Window – csúszóablak
- Nem vár minden szegmensre külön nyugtát
 - Több nyugtázatlan csomag is lehet a hálózaton
 - Többszörös szegmensküldés
 - Nyugták csoportosan is érkezhetnek
- Gyorsabb átvitel
 - Ha megfelelően szabályozott a kint lévő szegmensek száma

Példa - Normál adatfolyam

svr4.1056

bsdi.7777

- A küldő 3 szegmenst küld (4-6)
- A 7. szegmens nyugtázza az első kettőt – magyarázat:
 - **4,5,6 megérkezik a szerverhez** és az IP bemeneti sorába kerülnek
 - IP input queue
 - TCP feldolgozza a **4.** szegmenst
 - A kapcsolat **késleltetett nyugta küldésre megjelölést** kap
 - Nyugtaküldő időzítő elindul
 - TCP feldolgozza az **5.** szegmenst
 - 2049-es bájtra **nyugtát generál, mert már összesen két kimenő szegmens várakozik**

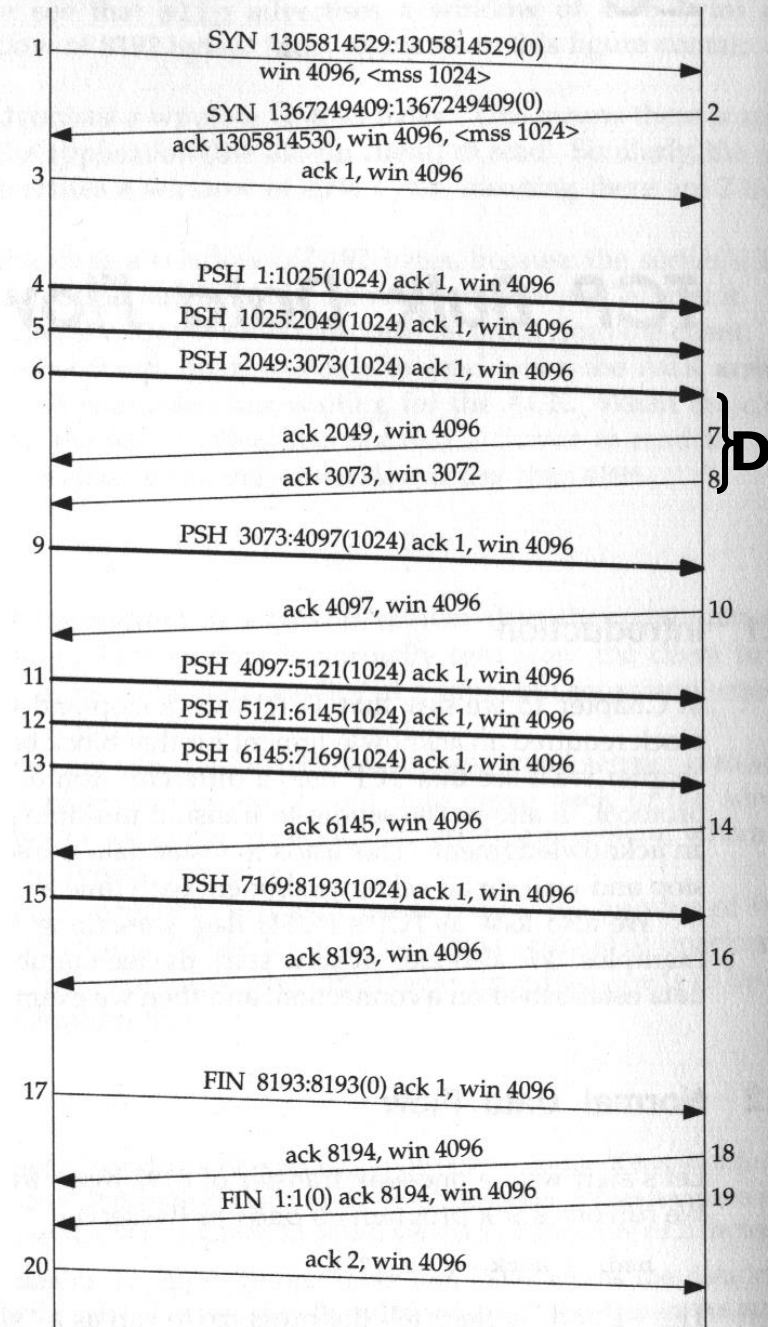


Példa - Normál adatfolyam

svr4.1056

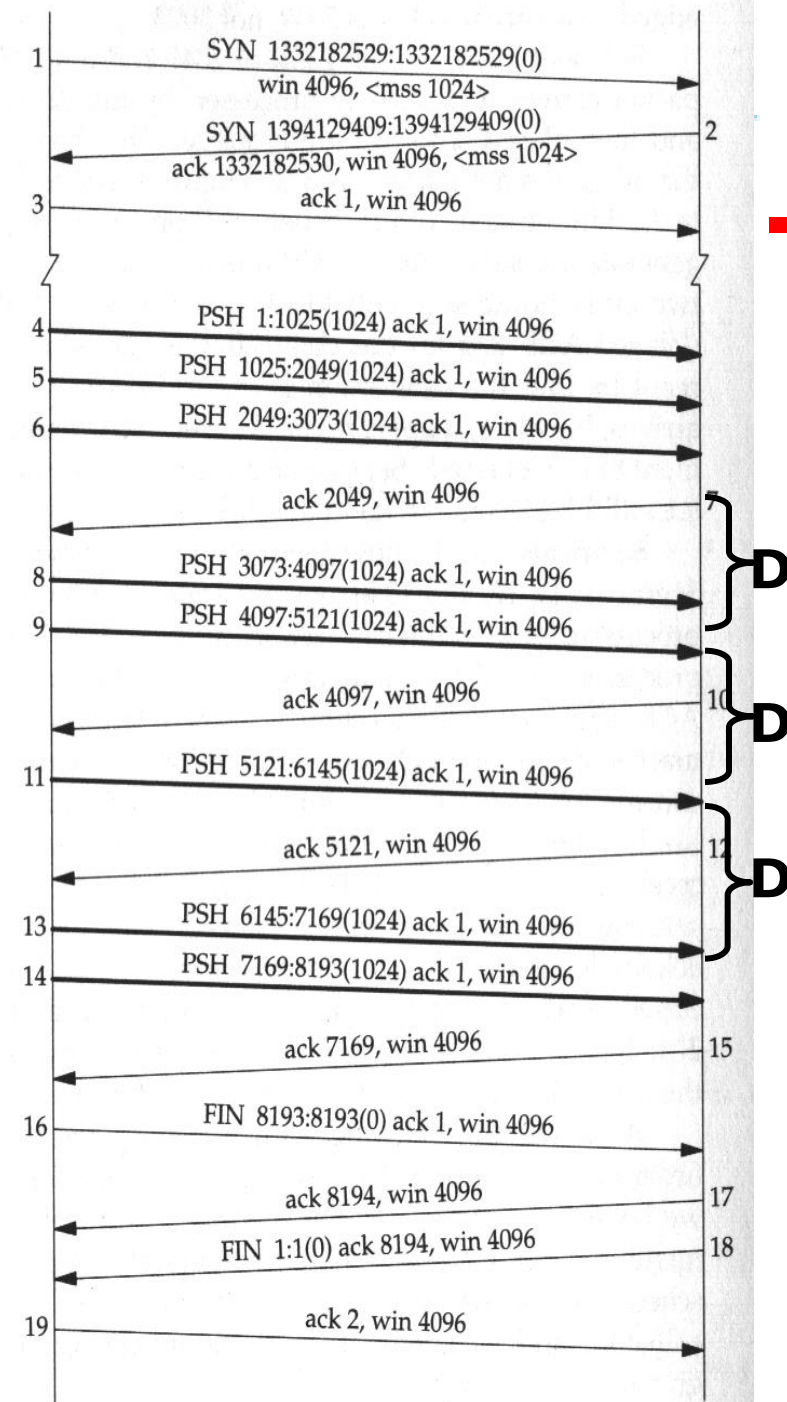
bsdi.7777

- TCP feldolgozza a 6. szegmenst
- A kapcsolat **késleltetett nyugta küldésre** ismét **megjelölést** kap
 - Nyugtaküldő időzítő elindul
- Nyugtaküldő **időzítő lejár (D)**
 - 9. szegmens még nem jött meg
 - **Nyugtát** (8) a 3073 bájtra **kiküldi**
- Szegmens 8, **win 3072**
 - A Windows méret kevesebb:
 - A nyugta küldésekor még 1024 adat (6. szegmensé) még benne van a TCP fogadó bufferében, melyet az alkalmazás még nem vett át



Példa2 - Normál adatfolyam

- Itt a 8. szegmens a nyugta késleltetés időzítése előtt megérkezik
 - Nyugta 4097-ig kiküldésre kerül
 - Bár még a kiküldés előtt megérkezik a 9. szegmens
- 12. nyugta
 - Csak a 9. Szegmens – időzítő!

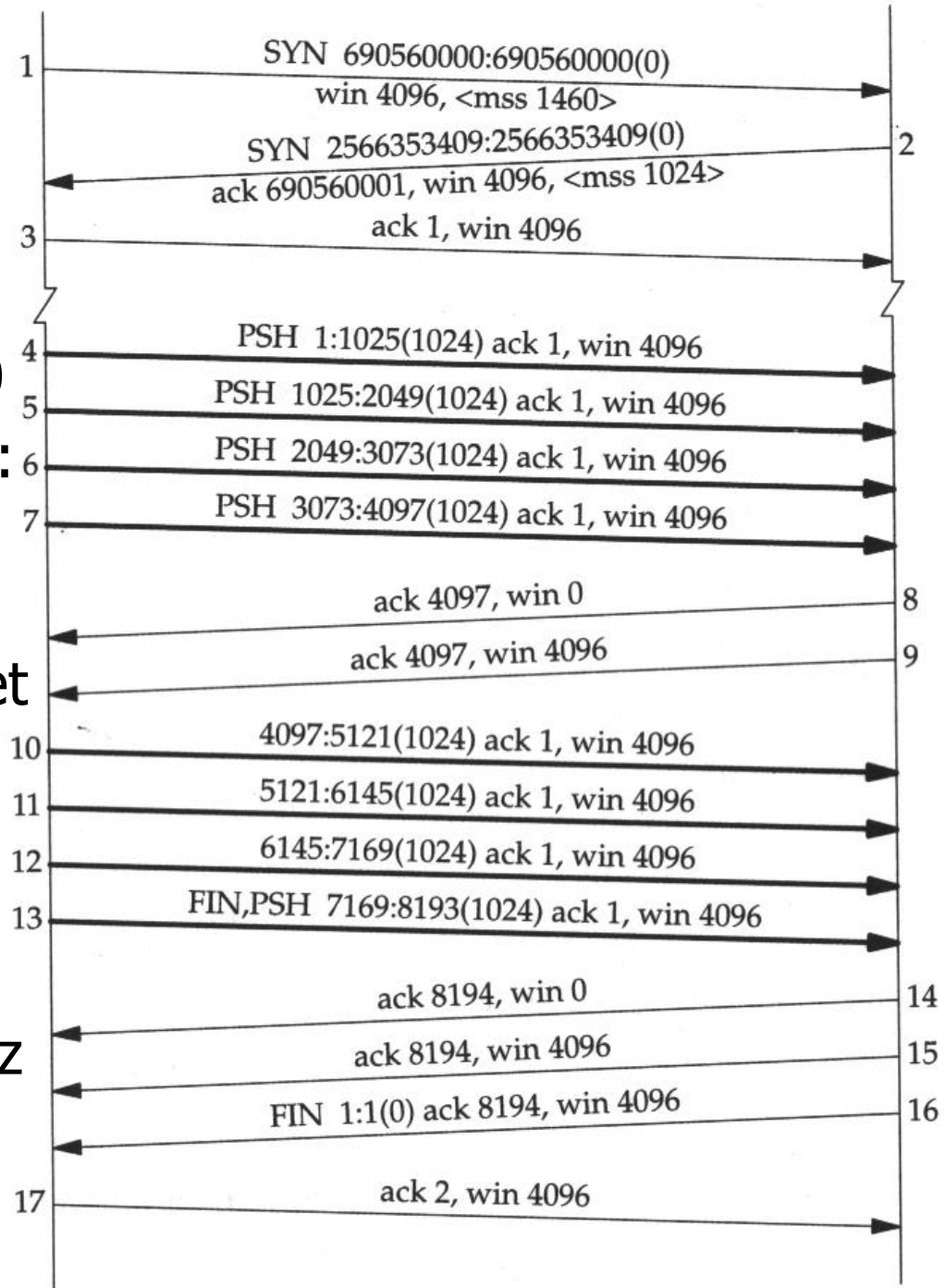


- Küldő a meghirdetett ablakméret szerint küld:
 - Megtölti a fogadó bufferét
 - Küldő vár a nyugtára
 - A fogadó oldal lassú, nem tudja az alkalmazásnak tovább adni az adatokat – buffere még mindig tele
- A fogadó nyugtájában az „advertised window size” = 0
 - A küldő nem küld több szegmenst
- Ebből az állapotból ki kell billenteni a küldőt:
 - Ha már ürül a buffere
 - Window Update üzenet:
 - Újabb nyugta ugyanarra a szegmensre
 - De új window mérettel

Példa2

window update

- Egymás utáni 4 szegmens küldése (**4-7**)
 - Fogadó buffere tele lesz:
 - Ack: 4097 (**8**)
 - Window: **0** !!!
 - **Window update** üzenet
 - még egy ACK: 4097 (**9**),
 - Window: 4096 !!!
- 
- A TCP fogadó bufferét az alkalmazás kiürítette



- A TCP szegmensek és nyugták elveszhetnek a hálózatban
 - Torlódások miatt a közbeeső routerekben csomageldobás lehet
- **retransmission timer** – újraküldési időzítés
 - Szegmens megküldésekor időzítő indul
 - Lejáratáig az ACK-nak meg kell érkezni
 - Ha nem érkezik meg az időzítés lejáratáig (timeout)
 - A szegmenst újraküldi
 - Cwnd=1 lesz ismét

- Milyen nagy legyen az ablakméret, hogy ideális állapot legyen?
- Előző példa:
 - A küldőnek 8 szegmenst kell a csatornára engedni nyugták nélkül, a maximális teljesítményhez
- **Csatorna kapacitása = sávszélesség $\times$ körülfordulási idő**
- (bandwidth-delay product)
- A fogadó meghirdetett ablakméretének egyenlőnek kell lenni a csatorna kapacitásával

- RTT – becsléssel számítja a TCP
 - Az előző érték (RTT)
 - És a mért (M) értékből származik

$$RTT \leftarrow \alpha RTT + (1 - \alpha) M$$

- α - csillapító faktor (≈ 0.9)
- Minden méréskor frissítés

Retransmission TimeOut

$$RTO = RTT \times \beta$$

- β : késleltetési variancia faktor
 - Ajánlott érték 2
- **probléma:** a késleltetés nagy ingadozásait nem tudja követni
- Jacobson féle számítás
 - Késleltetés ingadozás jobb követésére más RTT számítás

Jacobson féle számítás

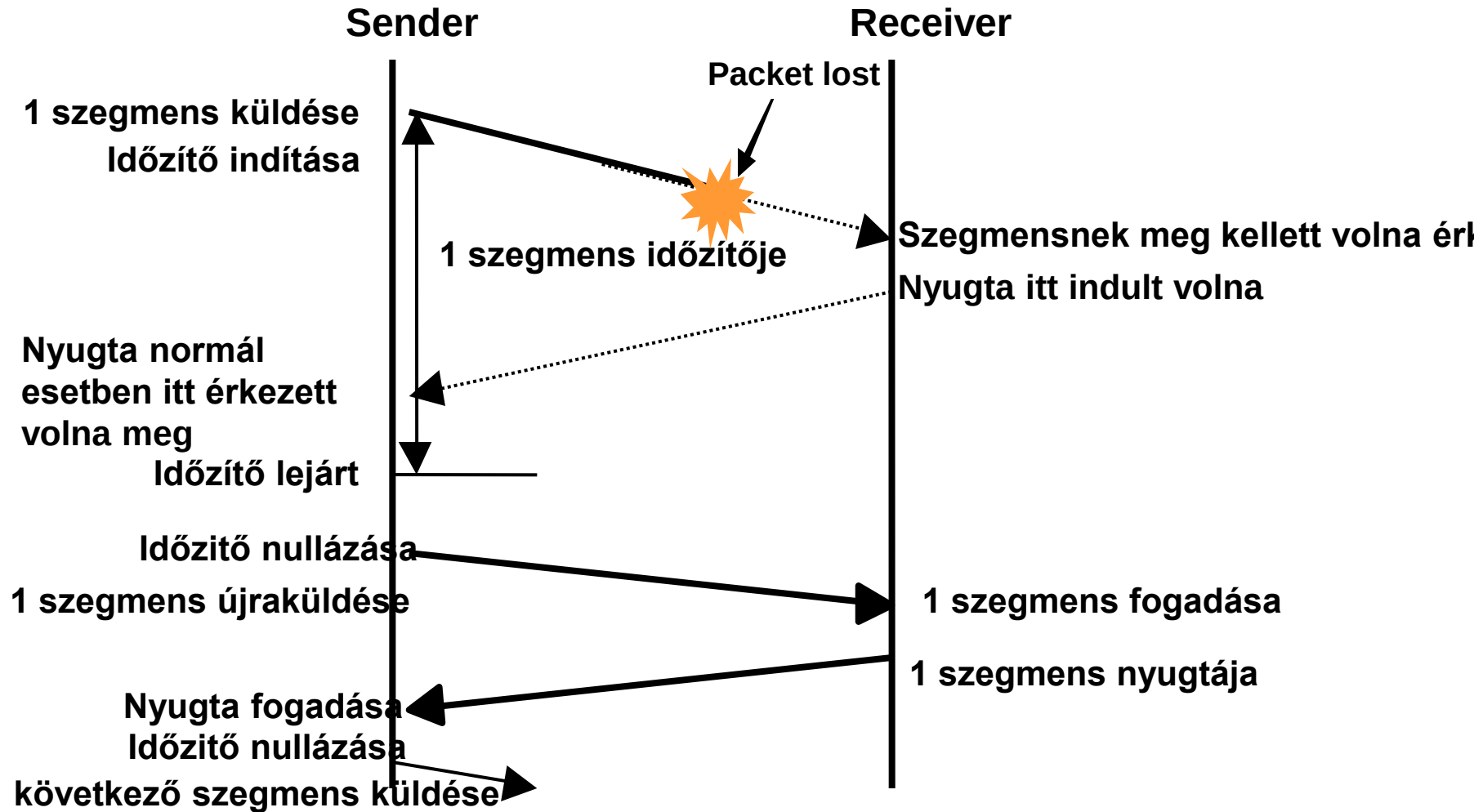


BME-TMIT

- $Err = M - A$
 - M Mért érték
 - A csillapított RTT (átlag becslése)
- $A \leftarrow A + g Err$
 - $g = 0.125$
- $D \leftarrow D + h (Err - D)$
 - D csillapított átlagos eltérés
 - $h = 0.25$
- **$RTO = A + 4D$**

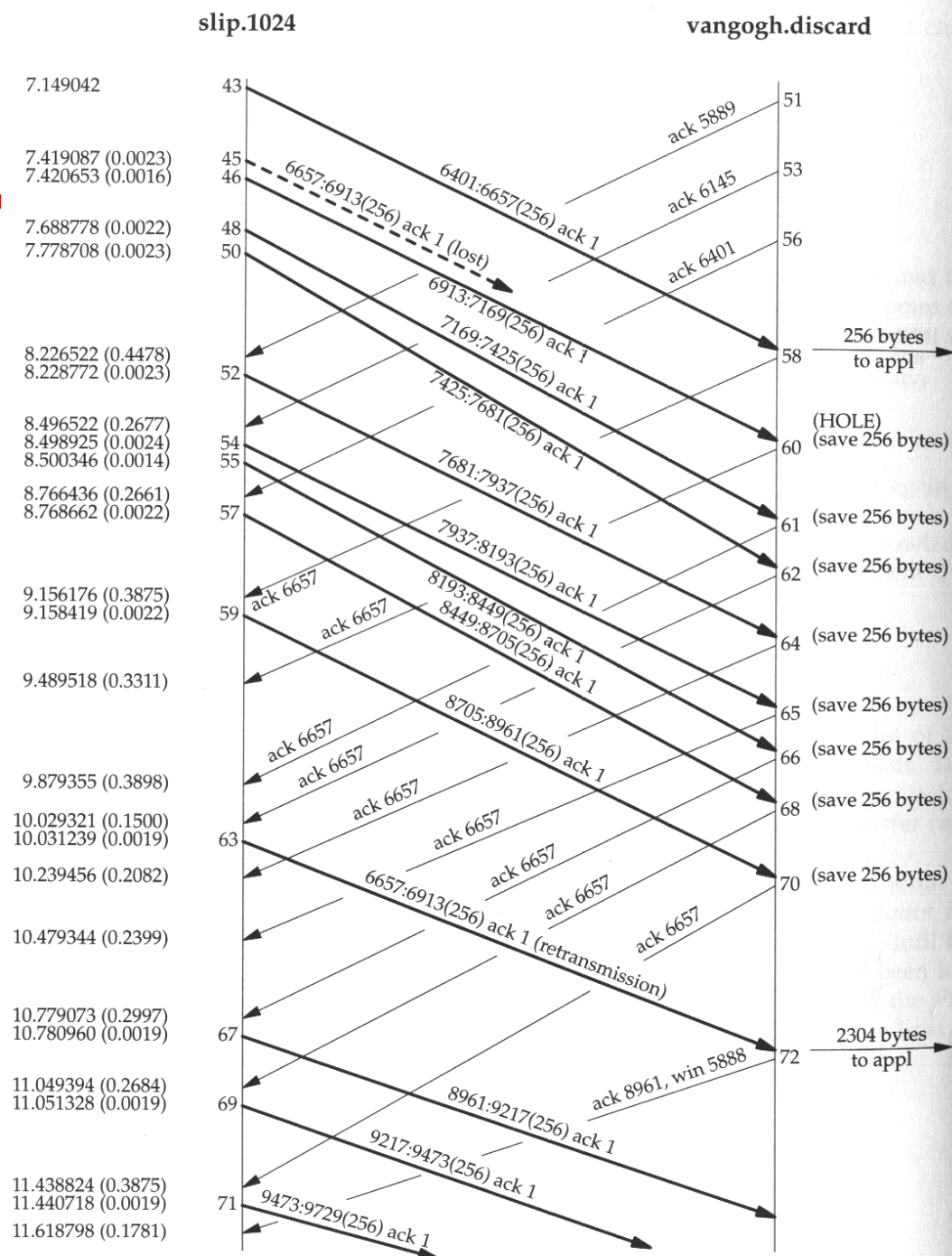
Példa: újraküldés időzítés lejártakor

BME-TMIT



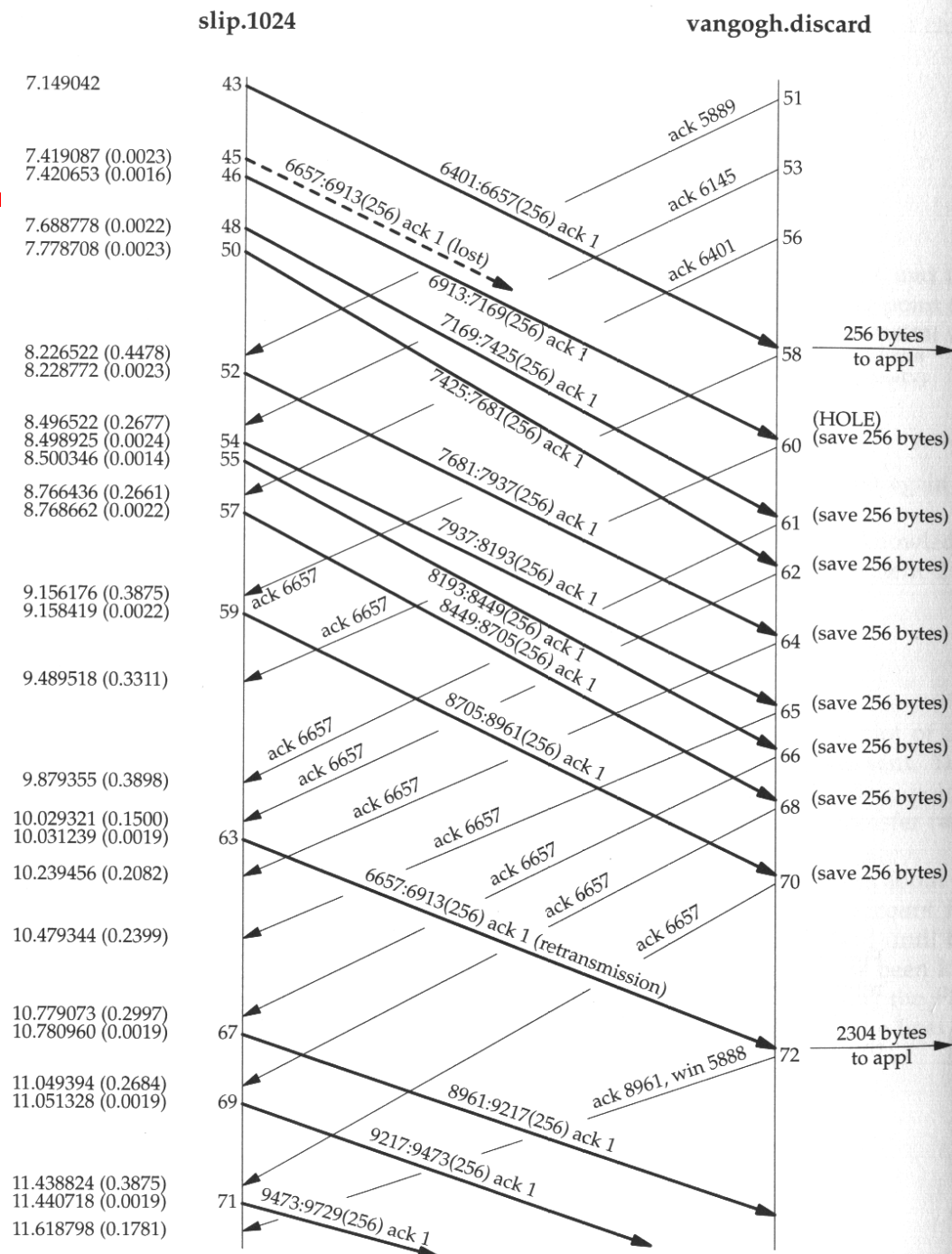
Újraküldés ACK-k száma miatt

- 45. szegmens elveszett
- A 62. szegmens (nyugta) hatására újraküldés
 - Ez volt a 3. egyforma ACK 6657-ből
- Fogadó folyamatosan menti a további szegmenseket
 - A hiányzó utániak így nem vesznek el



miatt

- A TCP implementáció számolja az egyforma ACK-k számát
- Ha a harmadik egyforma is megérkezik
 - Következtet: a hivatkozott szegmens elveszett
 - Reméli, hogy csak az az egy
 - Csak azt az egy szegmenst küldi újra
- Jacobson féle ***fast retransmit algorithm***
 - ***Gyors újraküldés algoritmus***



Újraküldés ACK-k száma miatt 2

- gyorsítás:
 - A fogadó egy szegmens hiányakor rögtön duplikált ACK küld vissza
 - A küldő gyorsabban újraküldi a hiányzó szegmenst
 - Gyorsabban visszaáll a rendes adatátvitel
- Jelenleg a TCP nem tud
 - Egy szegmens hiányát jelezni
 - Sorrend eltéréseket jelezni

Gyors újraküldés, gyors visszaállítás algoritmus



BME-TMIT

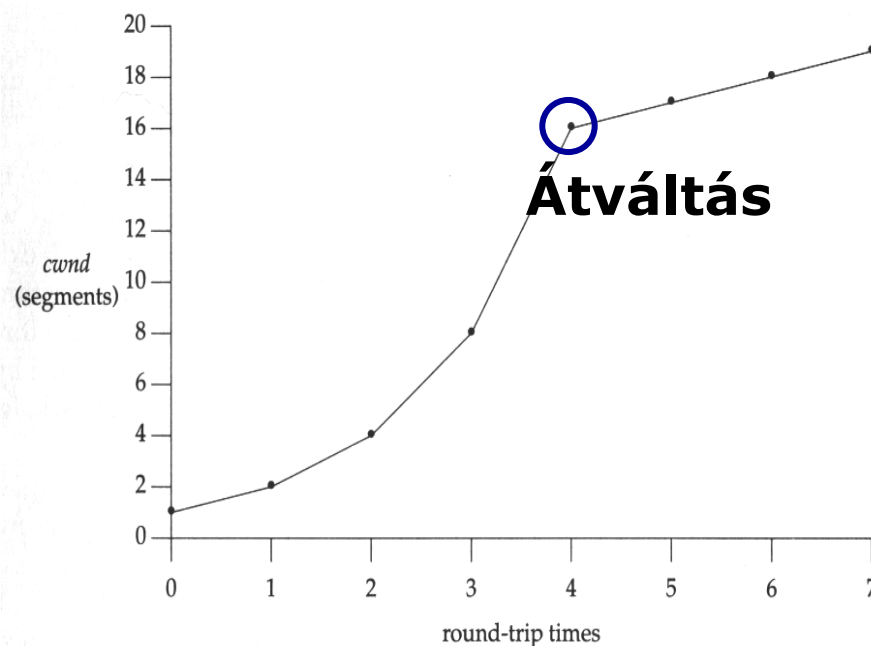
- Fast retransmit – fast recovery
- 3 vagy több duplikált ACK
 - Gyors jelzése egy szegmens elvesztésének
- Újraküldés (már az RTO lejáratára előtt)
 - ***fast retransmit – gyors újraküldés***
- De utána:
 - **congestion avoidance** – torlódás elkerülés
 - NEM slow start – lassú indítás
- Ez a ***fast recovery*** algoritmus

Congestion avoidance alg.



BME-TMIT

- A Slow Starttal a **cwnd** exponenciálisan nő
- Az exponenciális növekedés vége: ha csomagvesztés fordul elő
 - Ezután újraküldés
 - **cwnd** lecsökken 1-re, és indul újra a slow-start
- **Congestion avoidance** algoritmus lehetővé teszi, hogy az exponenciális növekedés additív növekedéssé váljon
- Ezzel a TCP a cwnd növekedését 1-re tudja maximalizálni egy RTT alatt
- A csomagvesztések kevésbé gyakoriak lesznek



CWND növekedése a congestion avoidance alatt



BME-TMIT

- Minden ACK fogadásakor

$$cwnd = cwnd + 1/cwnd$$

- Additív növekedés
 - A slow start exponenciális növekedésével szemben
- cwnd gyakorlatilag 1 szegmens értékkel nő RTT-ként

$$cwnd = cwnd + \text{szegmensméret} \times \text{szegmensméret} / cwnd$$

- cwnd valódi értéke szintén bájtban!

Congestion avoidance alg.



BME-TMIT

- *cwnd*: Congestion window
 - *ssthresh*: slow start threshold size
1. Kapcsolat kezdetén a kiindulási értékek
cwnd = 1 szegmens,
ssthresh = 65536 bájt
 2. TCP küldő maximum
min(*cwnd*; advertised window)
szegmenst küldhet

Congestion avoidance alg.



BME-TMIT

3. cwnd értéke slow start szerint (exponenciálisan) nő, míg csomagvesztés nem történik

ssthresh új értéke: **$\min(\text{cwnd}, \text{advertised window})/2$**
de legalább 2

Ha a torlódás időzítő lejáratára miatt következett be, a cwnd értéke 1 lesz és újra slow start

- 4.

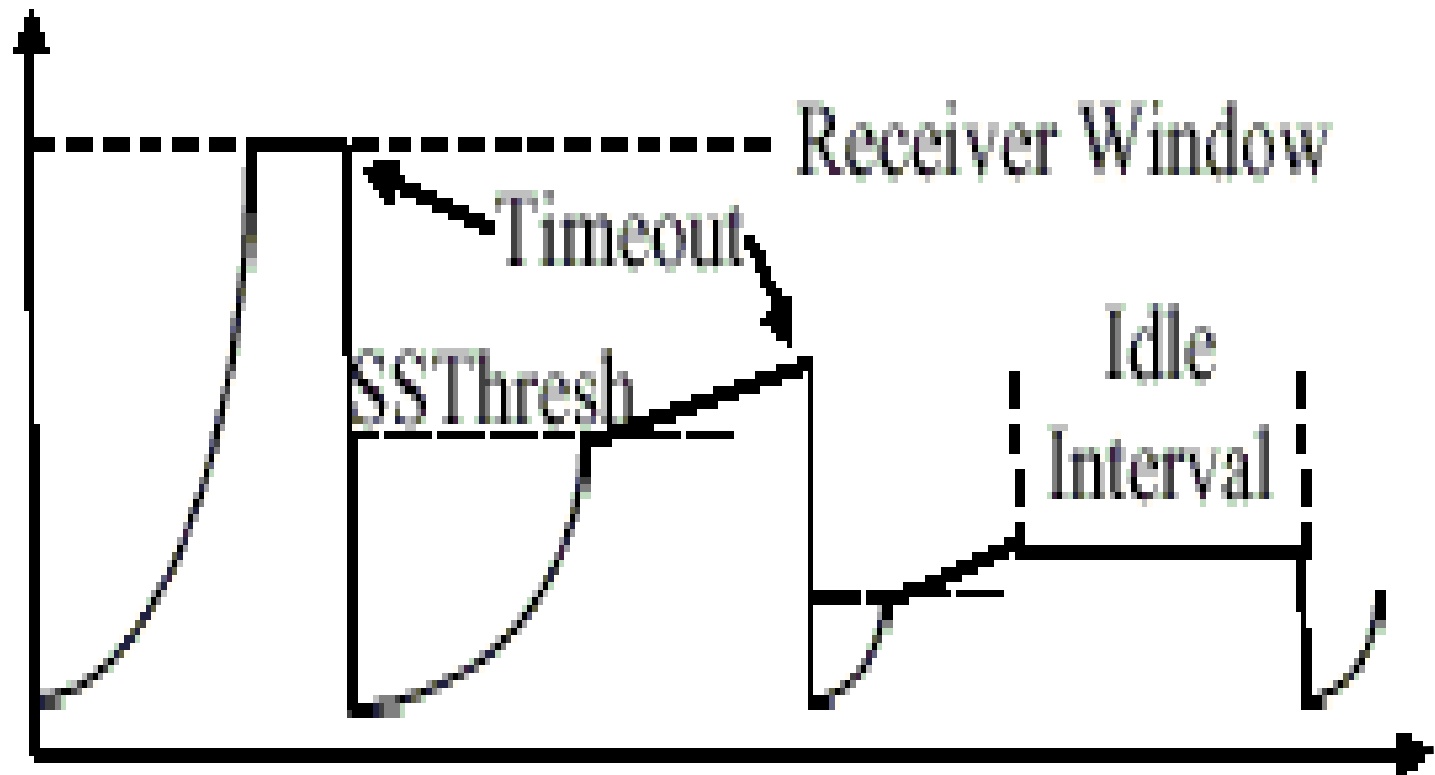
Ha a cwnd kevesebb vagy egyenlő ssthresh-sel, slow start szerinti növekedés

Ha a cwnd nagyobb, mint az ssthresh, congestion avoidance lép működésbe, cwnd legfeljebb 1 szegmenssel növekszik RTT-ként

Congestion avoidance



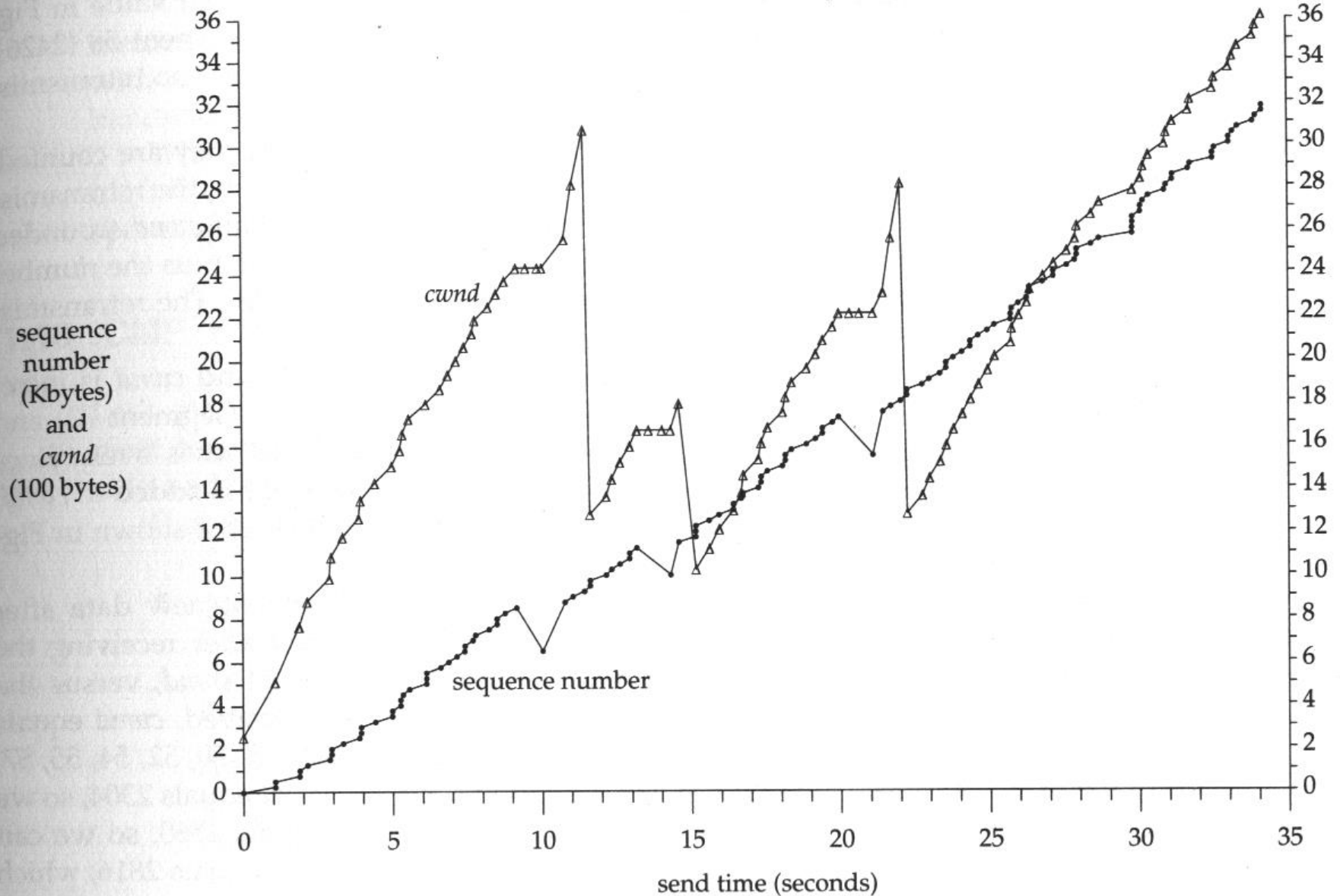
BME-TMIT



Valós példa



BME-TMIT



Köszönöm a figyelmet

- Vége -



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**



Szállítási réteg (L4)

Gyakorlat



- A TCP-t nagyon sok környezetben használják
- A főbb mechanizmusok ismerete fontos
 - A programozónak
 - A hálózati szakembernek
 - Wired
 - Wireless
- Lassan az ipari környezetbe is beszivárog

- Kapcsolat fogalma
 - 5-tuple
- Kapcsolat kiépítése
 - UDP - nincs
 - TCP
 - 3 way handshake
 - Miért van rá szükség?

- Netstat parancs

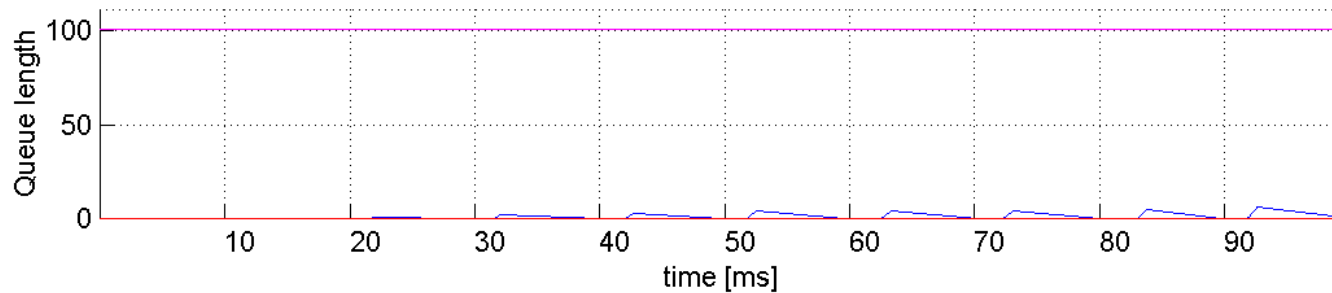
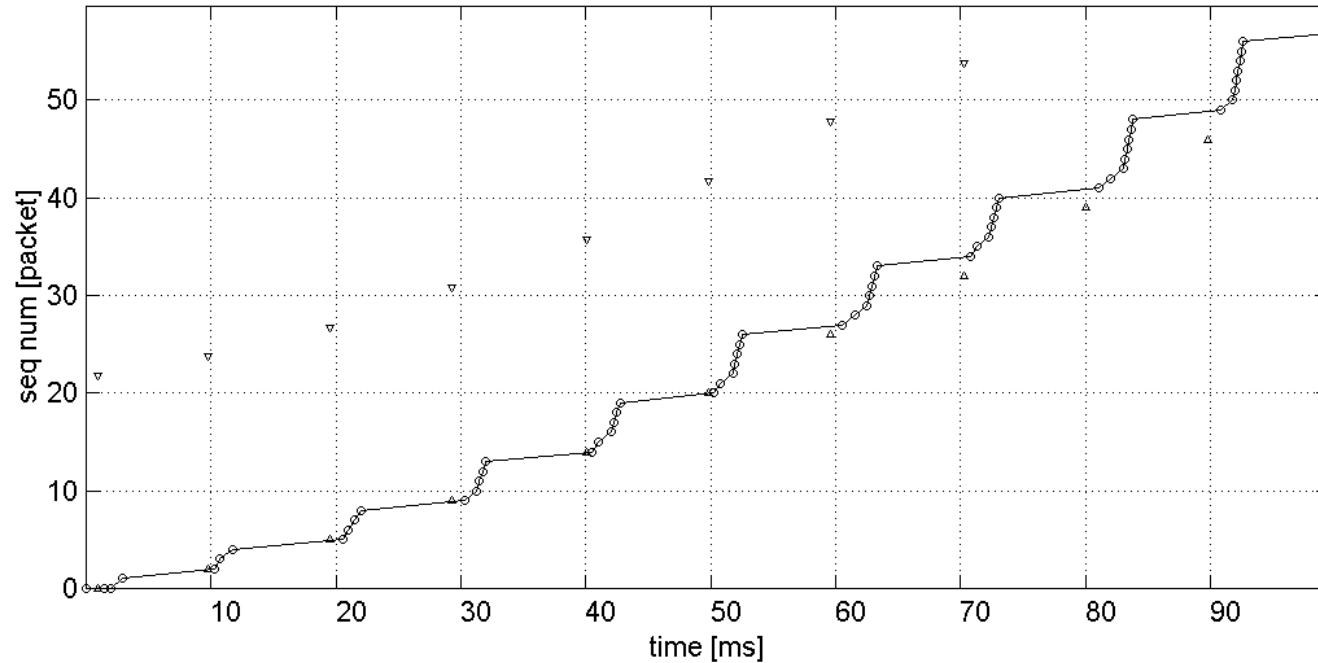
TCP	127.0.0.1:1906	localhost:1907	ESTABLISHED
TCP	192.168.1.147:53699	13.77.87.52:https	ESTABLISHED
TCP	192.168.1.147:53703	91.190.216.57:12350	ESTABLISHED
TCP	192.168.1.147:53737	64.4.23.152:40008	ESTABLISHED
TCP	192.168.1.147:53759	108.177.96.188:5228	ESTABLISHED
TCP	192.168.1.147:53772	40.77.226.192:https	ESTABLISHED
TCP	192.168.1.147:54512	a104-96-129-73:https	CLOSE_WAIT
TCP	192.168.1.147:54513	a104-96-129-73:https	CLOSE_WAIT
TCP	192.168.1.147:54514	a104-96-129-73:https	CLOSE_WAIT

- Adatküldés: szegmensek
- Hibakezelés
 - ICMP: port nem elérhető
 - Loss: nincs visszajelzés
- Sáv szélesség, késleltetés

- Bandwidth delay product
- 8k – korlátozott sávszélesség
 - Telítődés
- 64K
 - jobb
 - Window scale option

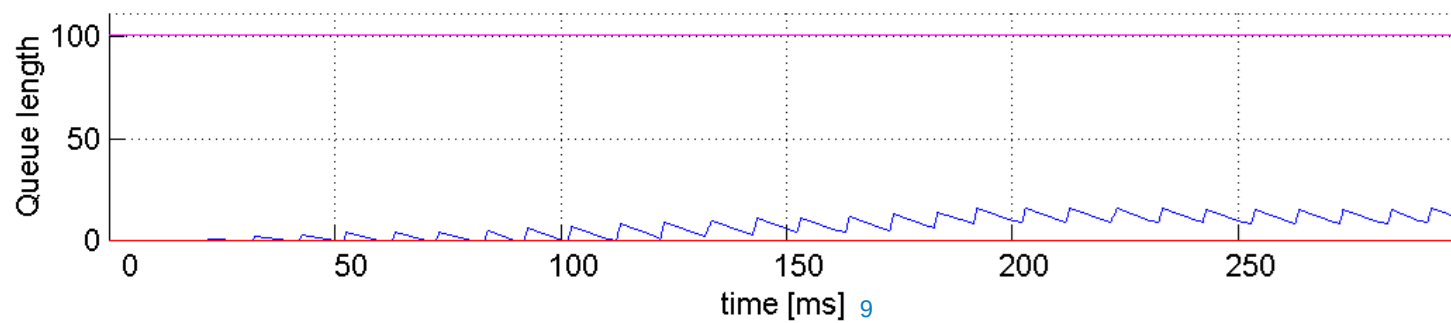
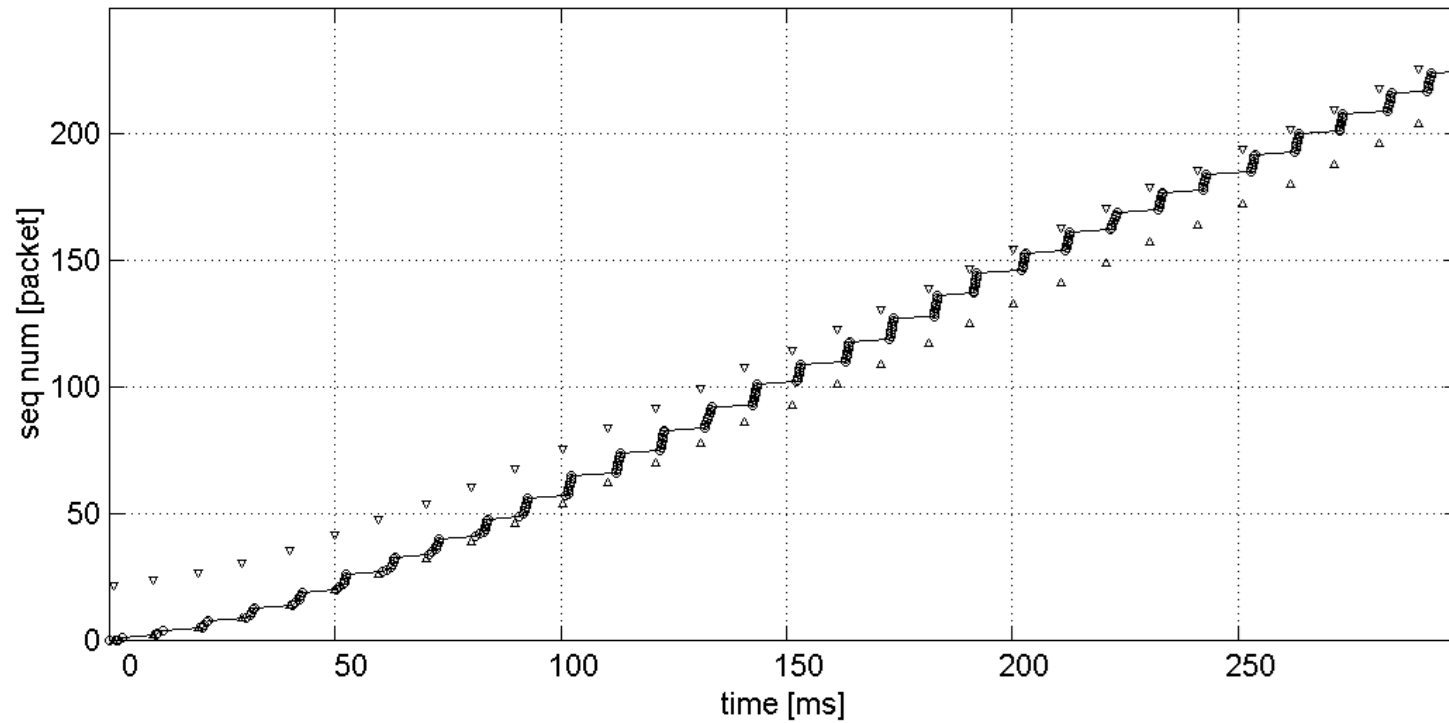
No.	Algo.	Sender	Traffic	# conn.	Receiver	Kbytes/sec	# ovfl.
413	None	Linux 2.1	one.1024	1	Linux 2.0	1097.7 (gw)	0

Trace: 413



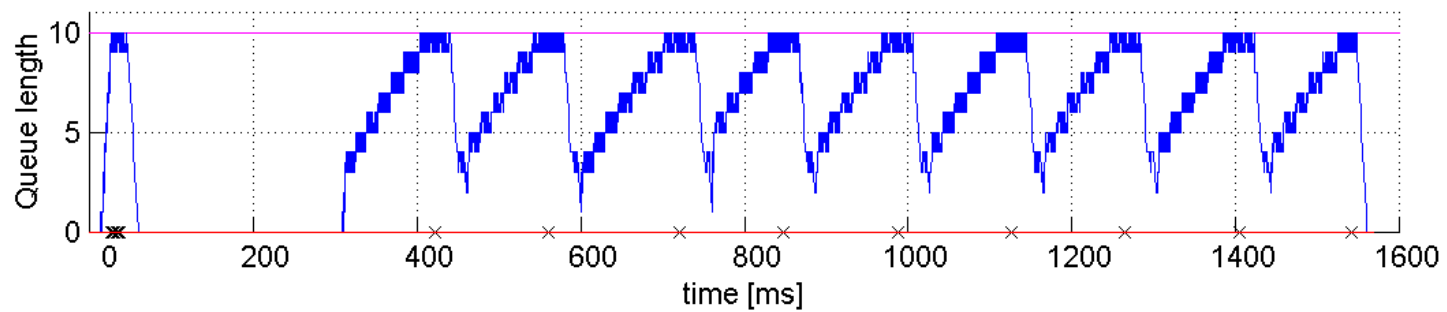
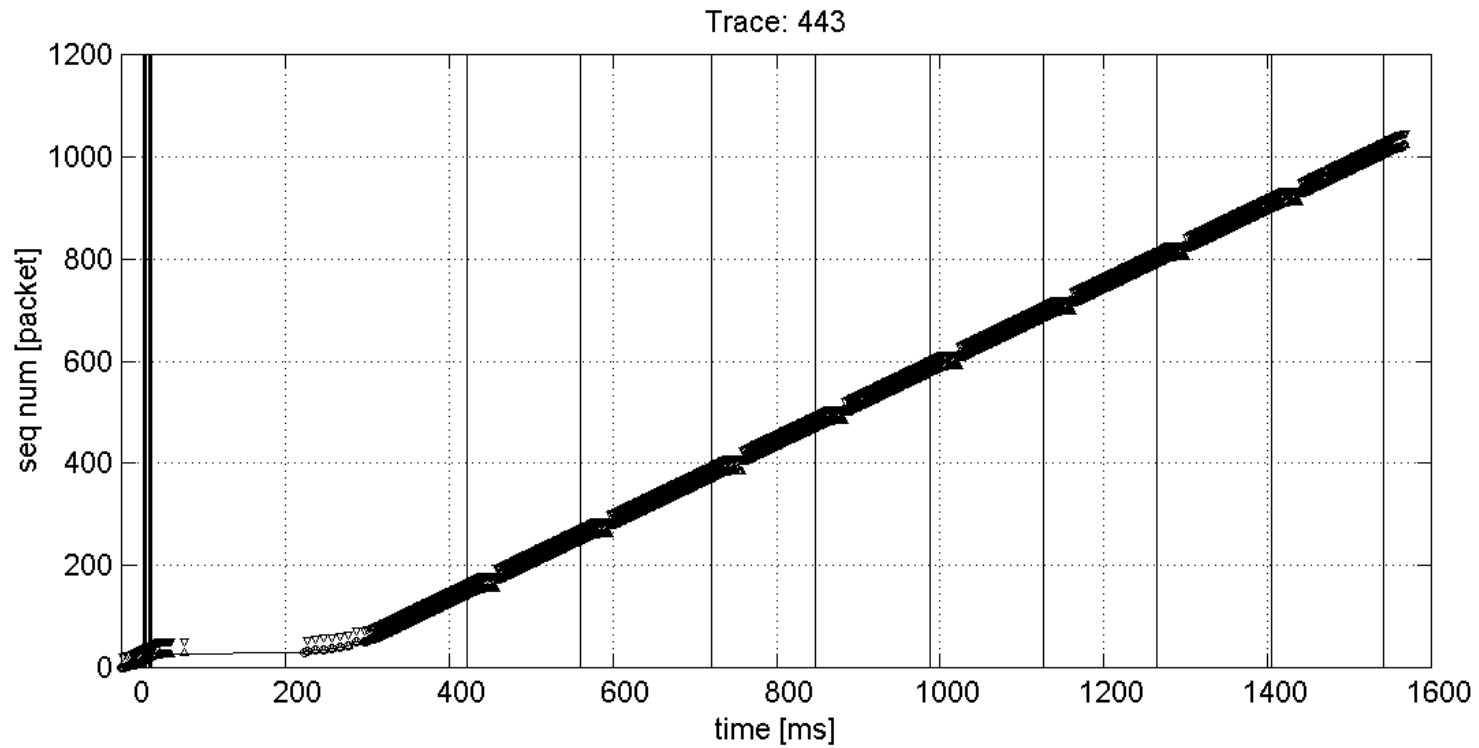
- A fogadó csak egy ACK-t küld egy csomagcsoportra
- Minden ACK érkezése a küldőnél egy csomagbőrszt küldését eredményezi
- A küldő a *cwnd*-t minden egyes ACK érkezésekor eggyel növeli
- A sorok hossza növekszik a bőrsztök érkezésekor, a bőrsztök méretei pedig növekednek az idővel
- ACK érkezések 10 ms-ként

Trace: 413



- Kb 200ms után, a küldő eléri a fogadó meghirdetett ablakméretének felső határát
- A sorhosszúság kb. 15 körül stabilizálódik
- A sorok hossza a börsztök érkezésekor növekszik – ezeket a fogadó egy-egy csomagcsoportot nyugtázó ACK-ra küldött
- A sorok hossza fokozatosan csökken függően a 10 Mbps-os Ethernet kapacitásától

- Ablak – gyors felfutás
 - Hatása: többszörös loss
 - Timeout
- Sok timeout – lassú kapcsolat
 - Főleg a kapcsolat elején nagy gond
 - Inkább fast retransmit kellene

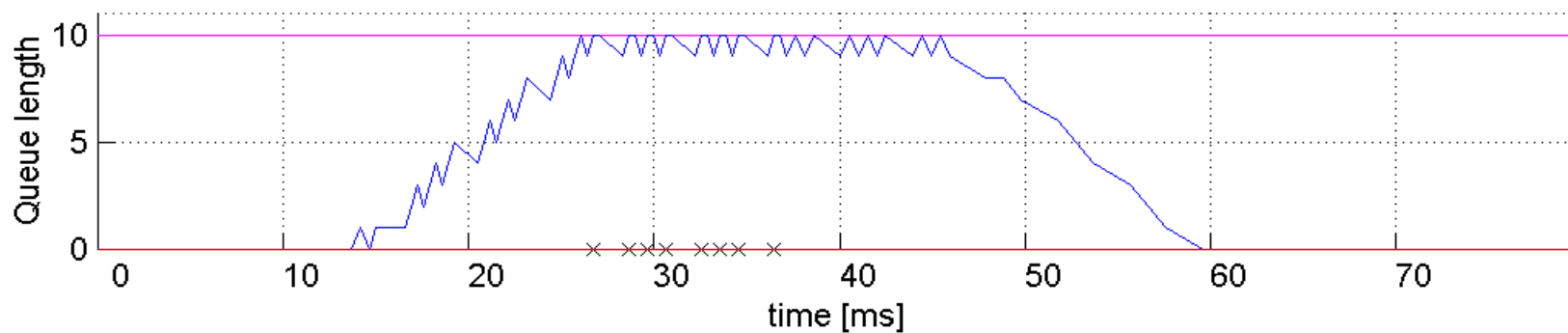
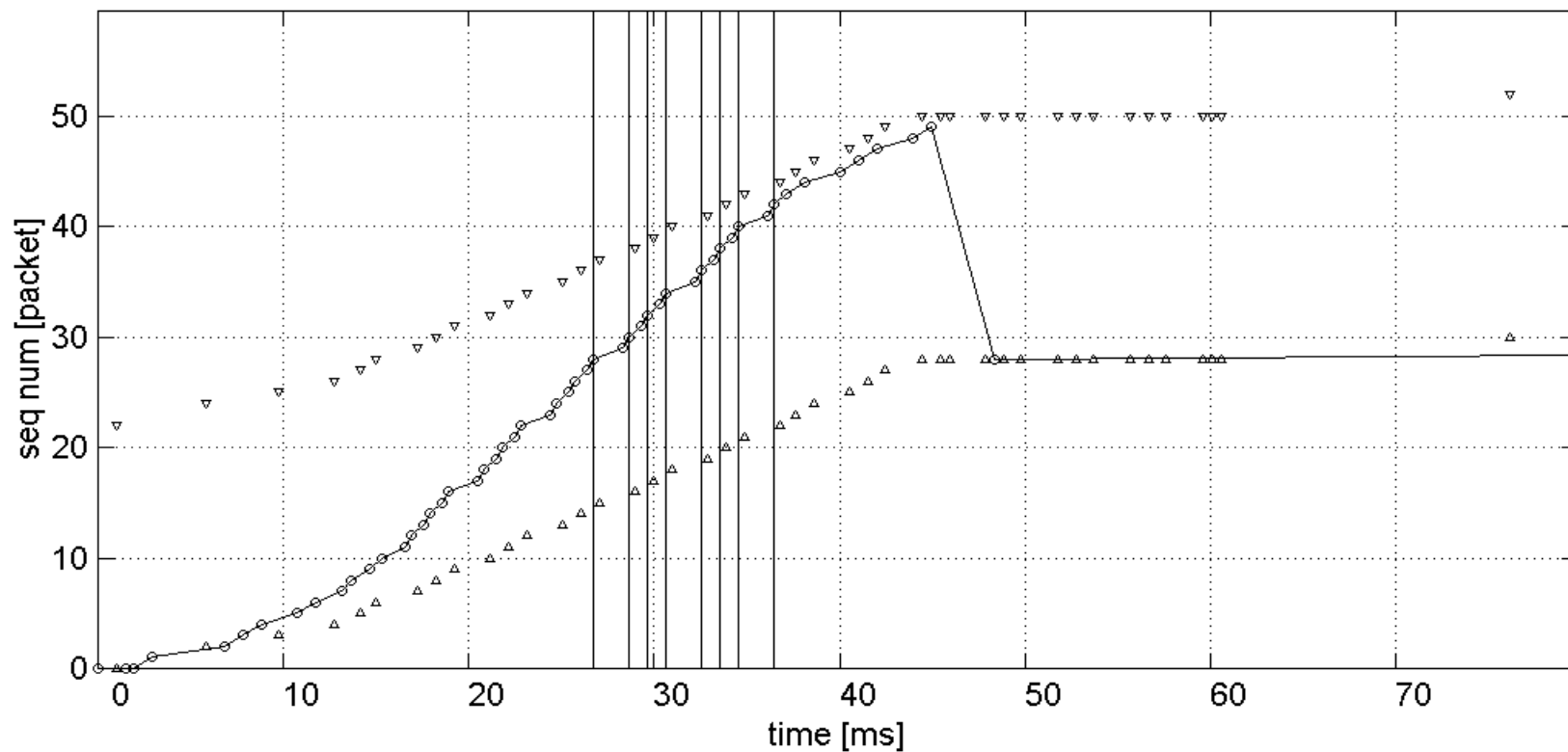


- A slow start börsztös csomagvesztéssel és az adás szüneteltetésével ér véget
- Periódikus veszteségek láthatók, melyeket a küldő gyorsan javít – a veszteségek nem okoznak jelentős teljesítménycsökkenést
- A sorhosszban periódikus minta van 300 ms után: lineáris növekedés, egy veszteség, egy esés. Ez mutatja a congestion avoidance mechanizmus működését a küldő oldalon: a veszteség észrevételekor csökkenti a congestion window értékét, majd ismét lineárisan (additívan) növeli

Az első 80 ms



Trace: 443

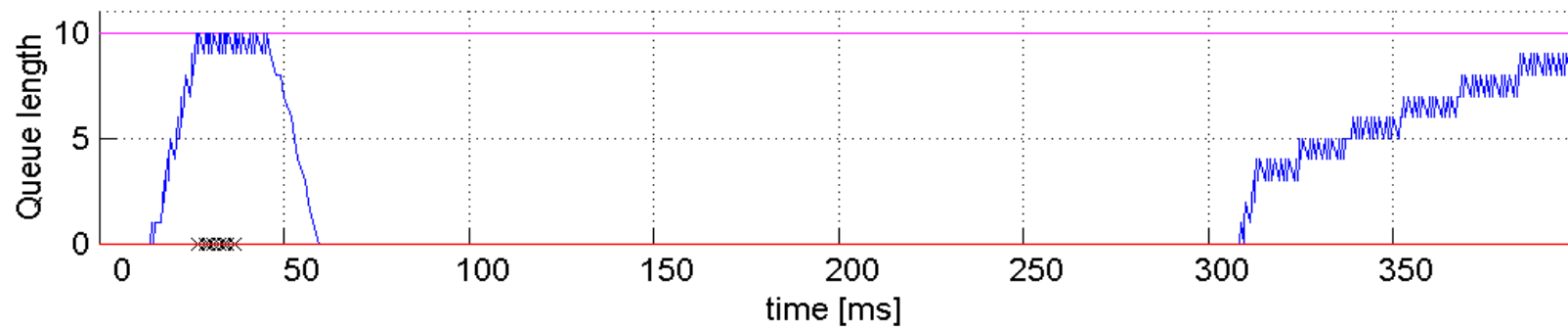
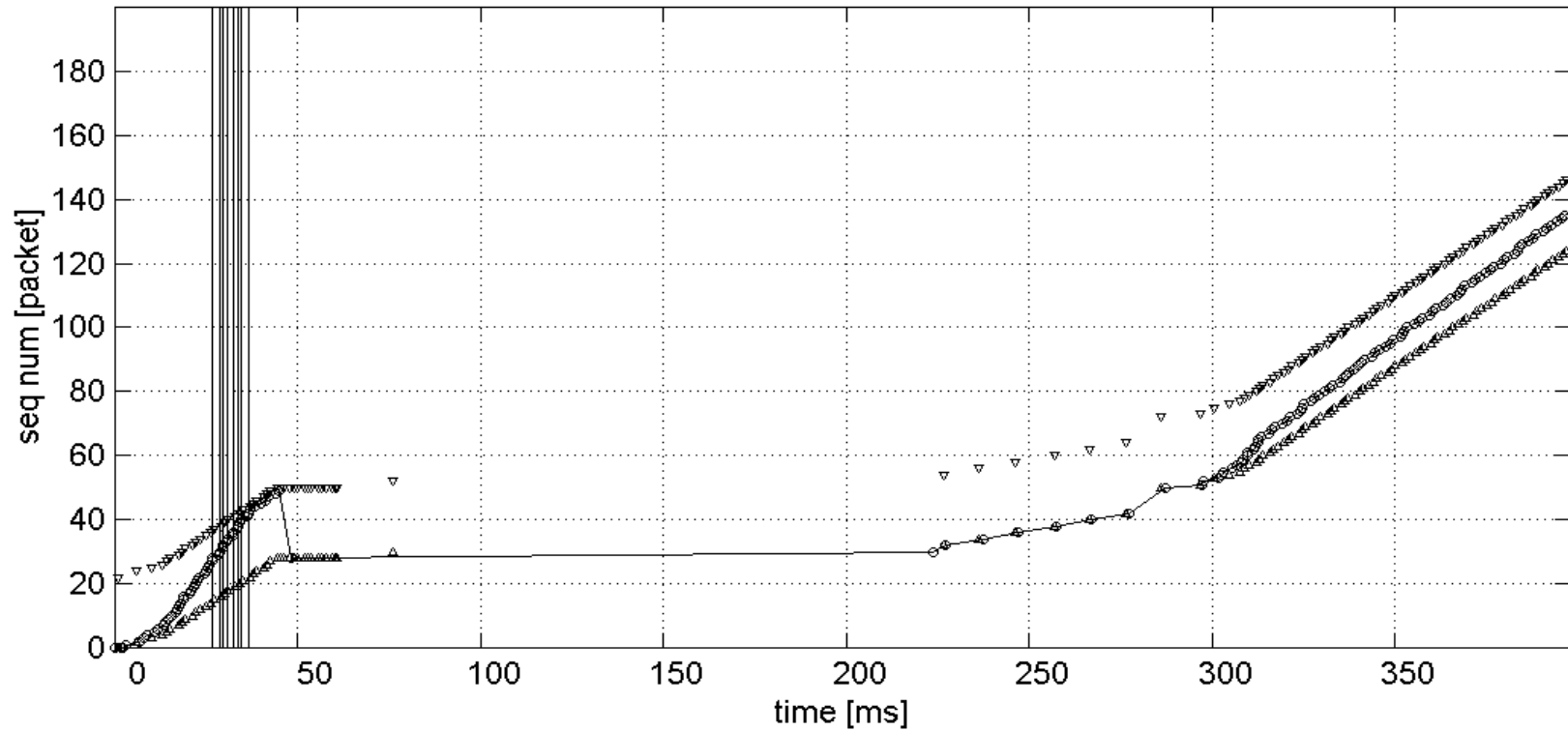


26-36ms	A fogadó felé a sorok túlcsordulnak, a csomagok eldobásra kerülnek
37-45ms	A fogadó felé menő csomagok továbbításra kerülnek, a fogadó fogadja ezeket a csomagokat, de nem nyugtázza őket, mert néhány korábbi csomag is hiányzik
44-47ms	A fogadó duplikált ACK-t küld
48ms	A küldő újraküldi az első nemnyugtázott csomagot (fast retransmit).
49-61ms	A fogadó továbbra is duplikált ACK-kat küld
76ms	A fogadó nyugtázza az újraküldött csomagokat

200 ms környéke



Trace: 443



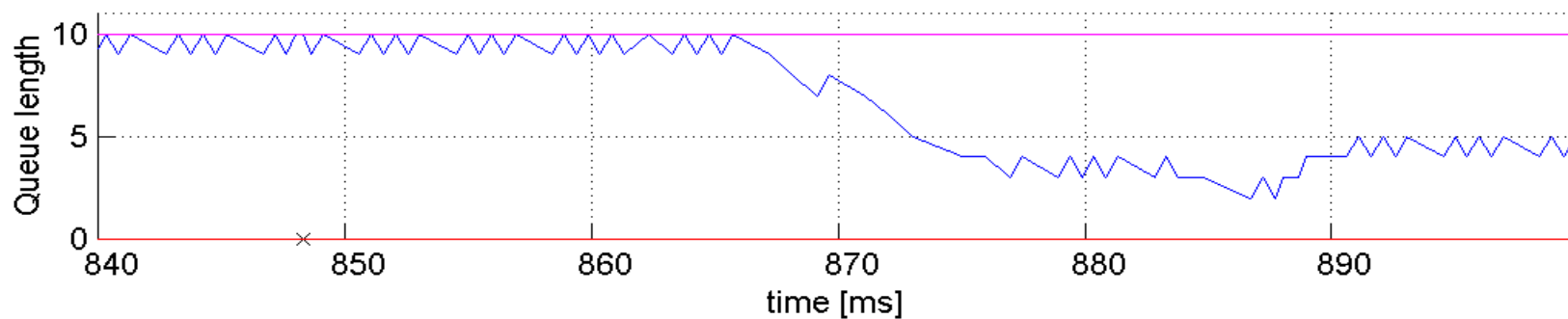
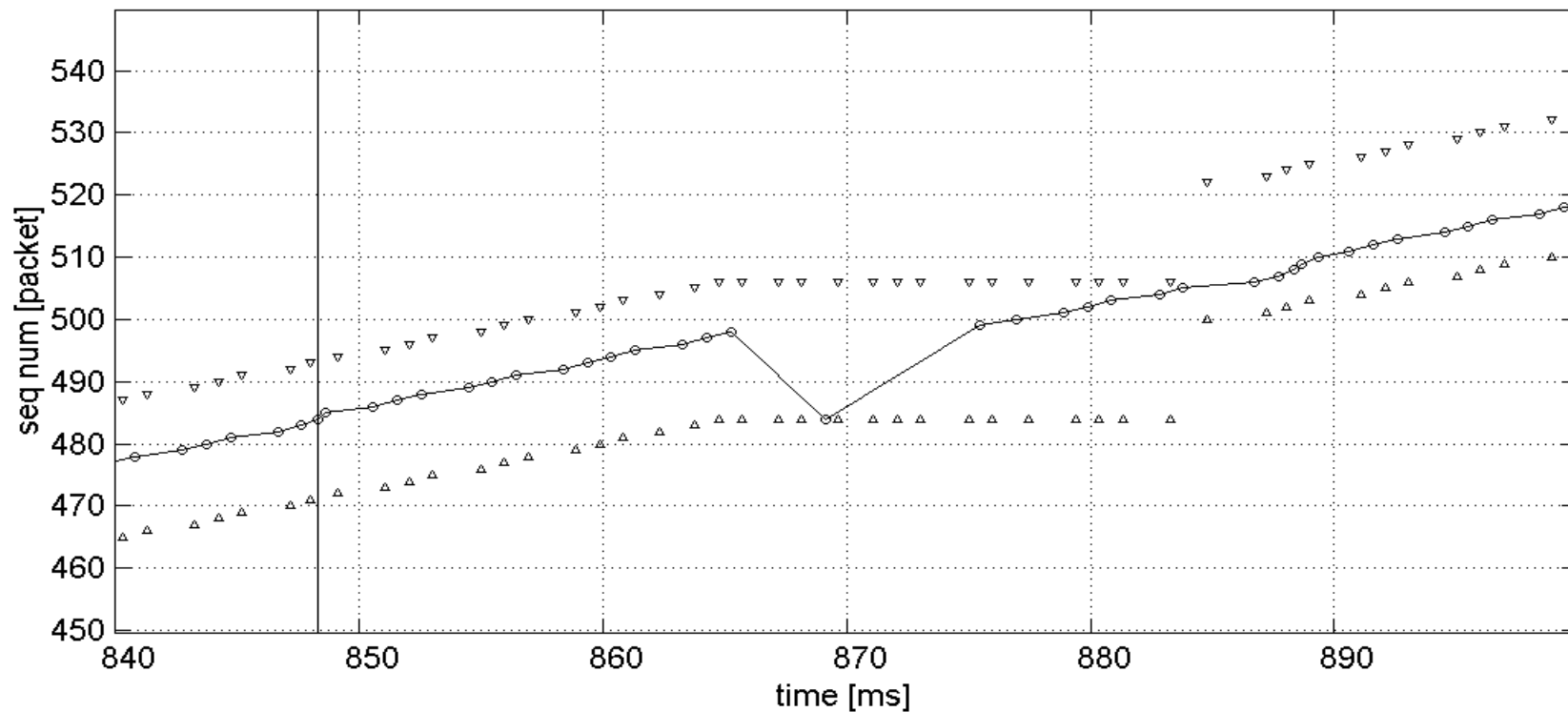
225ms	A küldő időzítője lejár és küld egy következő nemnyugtázott csomagot. Az időztés kb 200 ms
225-275ms	A küldő továbbít egy csomagot rögtön, amint a fogadó nyugtázta az előzőt. A küldő nem növeli a congestion window méretét, amíg újraküldést végez
275ms	A küldő minden elküldött adatára kap nyugtát (Az ACK számok ugrása). Slow start kezdődik.
330ms-	A gyors (exponenciális) növekedése a nemnyugtázott csomagoknak megáll és congestion avoidance szakasz következik – ez látszik a sorhossz lineáris növekedéséből

- Börsztös csomagvesztés történt
- Az első veszteszt a küldő vette észre a duplikált ACK-ból, és gyors újraküldés következett
- A következő veszteséket szintén a küldő vette észre az időzítők lejáratakor – emiatt slow start indult
- Ez a mechanizmus börsztös veszteséknél gyakori

870 ms környéke



Trace: 443



848ms	Egy csomag veszett el a közbenső sorok megtelése miatt
848-864ms	A fogadó a korábbi csomagokat nyugtázza
864-868ms	A fogadó a további csomagokat nem tudja nyugtázni, mert egy hiányzik. Emiatt a fogadó duplikált ACK-t küld
869ms	A küldő 3 duplikált ACK-t kap, majd gyors újraküldést hajt végre
875-884ms	A küldő folytatja a csomagok küldését, mivel nem érte el a fogadó advertised window méretét még.
885ms	Egy ugrás látható az ACK numberek között – a fogadó megkapta a hiányzó csomagot, és az azutániakkal együtt nyugtázta azt.
885ms-	A küldő folytatja a csomagok továbbítását (fast recovery történt: slow start nem következik most). A gyors újraküldés nem okozott nagy teljesítménycsökkenést

Csomagvesztés – jó vagy rossz?



BME-TMIT

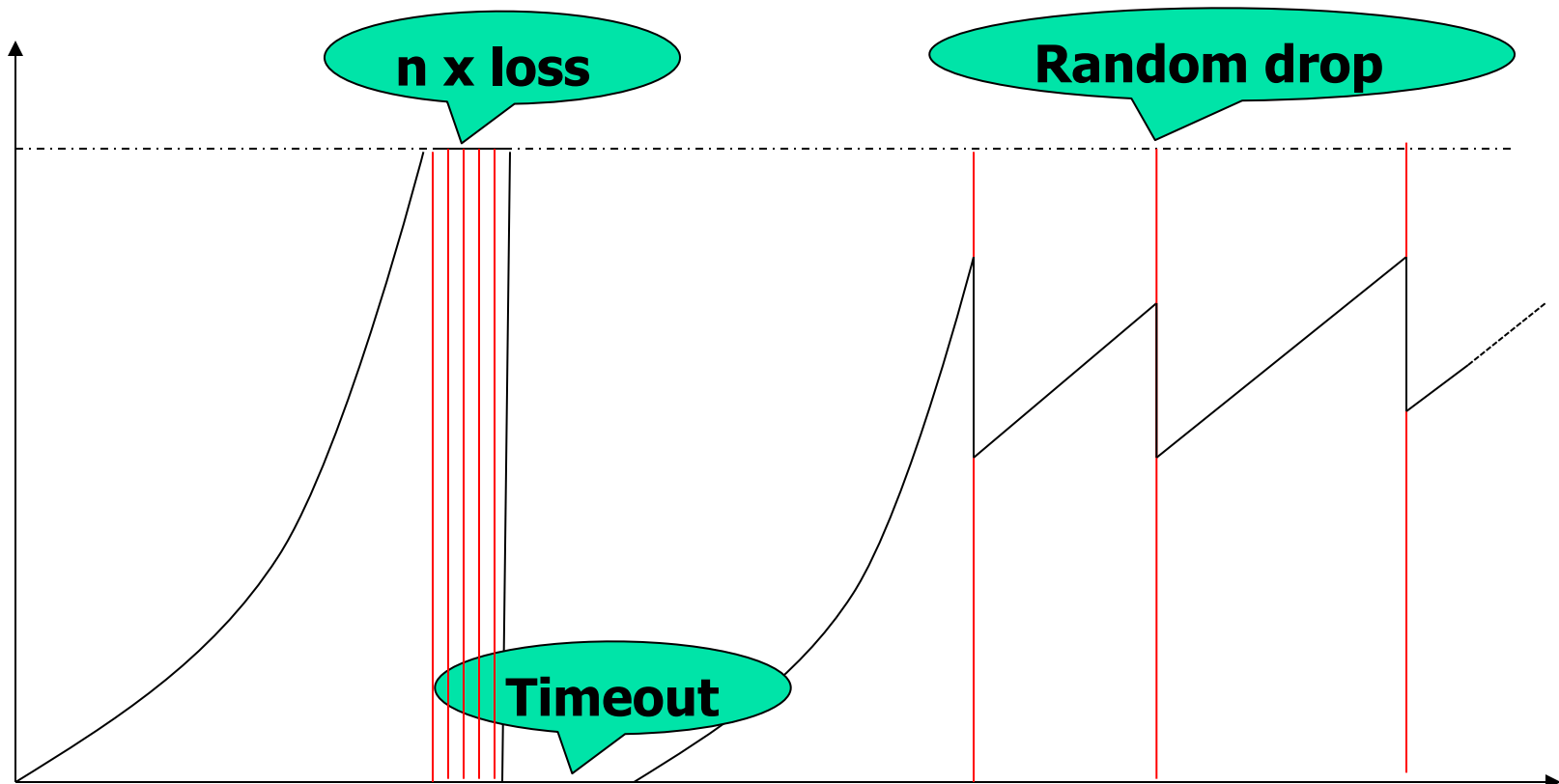
- Példa: DSL 10Mbit/s vonal, 2 lehetőség
 - a) $x\%$ loss vs
 - b) FEC – hibajavító kód, fix 20% adat
- TCP transport, letöltések: melyik jobb?
- a) 10Mbps – $x\%$ loss (pl. $BER=10^{-6}$)
 - ⇒ 1 csomag 1500 byte, 12000bit, minden 83.3 csomag elvész
- b) 8Mbps
 - a hibajavító kód 20% veszteség mindig

RED – Random Early Drop



BME-TMIT

- Buffer menedzsment a routerekben
 - Egy dobás jobb mint egy timeout



- Probléma:
 - Wireless – eleve van loss
 - de az nem túlcsordulás, nem szűk keresztmetszet!
 - TCP rosszul értelmezi, visszaveszi a cwnd-t
- Megoldások - többféle
 - WTCP – proxy
 - SACK

Köszönöm a figyelmet

- Vége -



Budapest University of Technology and Economics

**Department of
Telecommunications and Media Informatics**

