Az internet ökoszisztémája és evolúciója

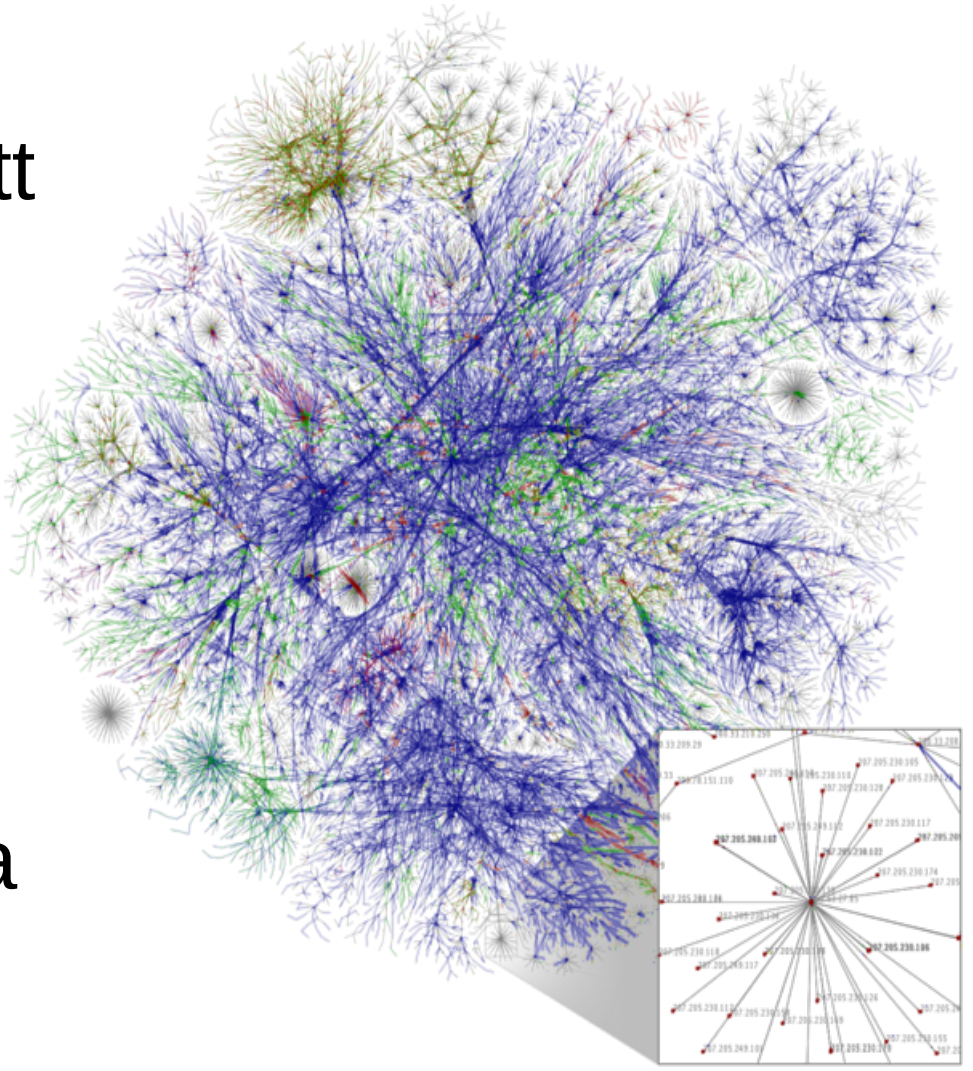
Tartalom

- Rendszerelmélet
- A TCP/IP protokollok
- Az internet architektúrája
  - autonóm rendszerek (AS) és szolgáltatók
  - content/eyeball/tranzit AS
  - access/edge/core AS

Rendszerelmélet

Az internet: „hálózatok hálózata”

- 50 ezer szolgáltató
- 10 milliárd csatlakoztatott eszköz
- 2 milliárd felhasználó
- Több 100 milliárd USD üzleti bevétel
- Megbízhatóság? Biztonság? Magánszféra védelme?...
- Mit hoz a jövő?



Opte Project, Wikipedia

An internet ökoszisztémája

- Hogy megértsük az okokat, meg kell ismernünk a technológiai hátteret...
  - „Hogy működik az internet?”
- … a folyamatot, amely idáig vezetett...
  - a jelenlegi internet 40 éves evolúció eredménye
- … a szolgáltatók motivációit és céljait
  - az internet decentralizált, a szereplők autonóm döntenek, nincs központi kontroll
- … és a lehetséges kiutakat
  - hogy működtethető mégis egy hálózat hatékonyan?

Ökoszisztéma

*Az **ökoszisztéma** […] az ökológiai jelenségek értelmezése, vizsgálata céljából […] létrehozott rendszermodell […] Az ökoszisztéma-modellek a kiválasztott ökológiai rendszer és környezete kapcsolatát vizsgálják. […] Ökológiai rendszermodellt egy akvárium élőlényegyüttesére, egy Petri-csészében szaporodó baktériumok populációjára vagy akár a bioszféra egészére is felírhatunk. A modell azonban mindig egy leegyszerűsítést, lényegkiemelést jelent.*

(Wikipédia)

Rendszerelméleti megközelítés

- Az internetet nem mint mérnöki alkotást, hanem mint **nagy méretű**, **komplex** és **elosztott**, tőlünk függetlenül működő rendszert vizsgáljuk
- Átalakítani, újratervezni nem tudjuk, hiszen nagy számú autonóm szereplő (kormányok, szolgáltatók, stb.) együttes viselkedése alakítja
  - már az internet topológiáját sem ismerjük, közelítőleg sem
- Így célunk inkább megfigyelni, megérteni, és modellezni az interneten zajló folyamatokat
- Hasonló a közgazdaságtan szemléletéhez

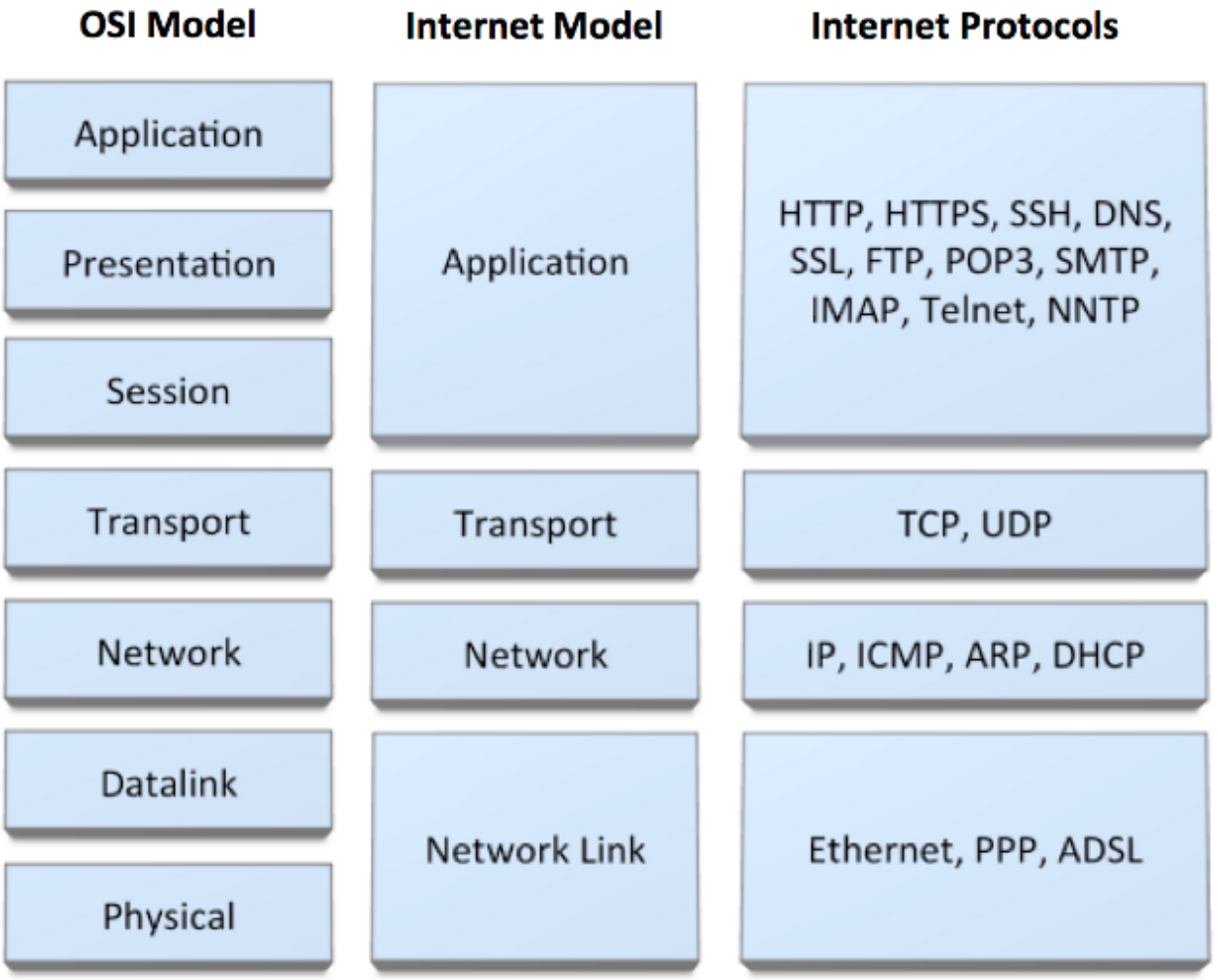
A TCP/IP protokollok

Hálózati rétegek

- Az internet egy rendkívül komplex elosztott rendszer
- A csatlakoztatott eszközök összehangolt működését hálózati **protokollok** biztosítják
- Mérnöki absztrakció: hálózati **rétegek**
  - funkcionális modul a protokoll stack-ben
  - szolgáltatást nyújt a felsőbb rétegek számára, az alsóbb rétegek szolgáltatásait használva
  - az alsóbb és felsőbb rétegek számára egységes interfész (elvileg cserélhető)
  - a rétegek között horizontális kommunikáció

TCP/IP rétegek

- Az ISO/OSI modelltől függetlenül jött létre



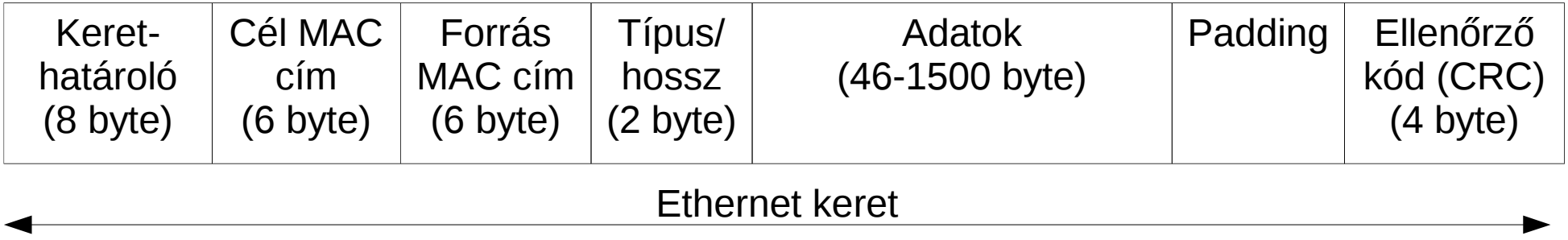
http://vichargrave.com/network-programming-design-patterns-in-c

Adatkapcsolati réteg (Link Layer)

- Helyi hálózati protokoll (Ethernet, PPP, WiFi,...)
- **Funkció: hibamentes csomagátvitel a hálózati réteg számára**
  - közös **linken** levő **hosztok** IP interfészei között
- Szolgáltatásmodell:
  - bemenő adatok tördelése adatkeretekre
  - sorrendhelyes továbbítás, nyugtázás, hibajavítás
  - forgalomszabályozás (vevő adattal való elárasztásának elkerülése céljából)
  - közeghozzáférés (MAC) protokoll: a megosztott csatornához való hozzáférés vezérlése
  - a fizikai réteget is ideértjük

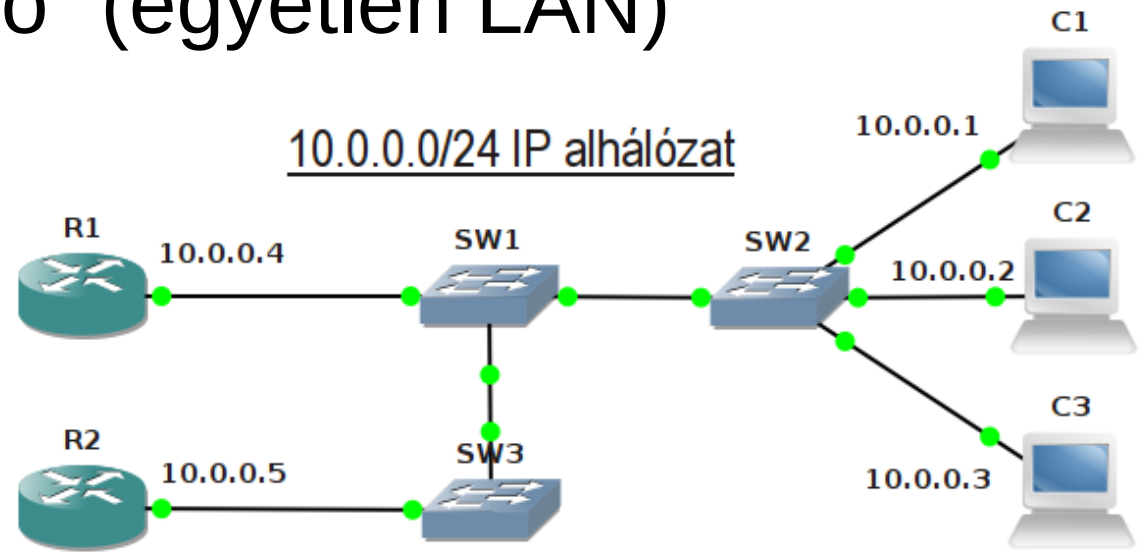
Adatkapcsolati réteg: az Ethernet

- IEEE 802.3: elterjedt helyi hálózati protokoll
  - 3 Mbit/sec ↔ 100 Gbit/sec, 48 bites „flat” címtér
- CSMA/CD közeghozzáférés
- Ethernet hálózatok összekapcsolhatók:
  - **hub/repeater:** az összekapcsolt szegmensek között minden keretet továbbít
  - **switch:** Spanning Tree Protocol vagy SP Bridging



Adatkapcsolati réteg: az Ethernet

- Az `SW1`, `SW2`, és `SW3` switch-ek Ethernet interfészek nem rendelkeznek IP címmel
- Az `R1` és `R2` IP routerek illetve a `C1`, `C2` és `C3` hosztok interfészei IP címmel vannak ellátva
- Számukra a bridge-elt Ethernet hálózat „átlátszó” (egyetlen LAN)



Hálózati réteg: Internet Protokoll

- Az internetre csatlakozó eszközök „nyelve”
- **Funkció: megbízhatatlan, összeköttetés-mentes, best-effort datagram szolgáltatás a szállítási réteg számára**
  - **megbízhatatlan:** nincs hibajavítás (header chksum!)
  - **összeköttetésmentes:** kapcsolatfelépítési fázis nincs az adatátvitel előtt
  - **datagram:** minden csomag tartalmazza a cél azonosítóját és egyenként továbbítódik, akár a kommunikáció során több eltérő útvonalon
  - **best-effort:** „all packets are created equal” (?)

Összeköttetés alapú vs. összeköttetésmentes protokollok

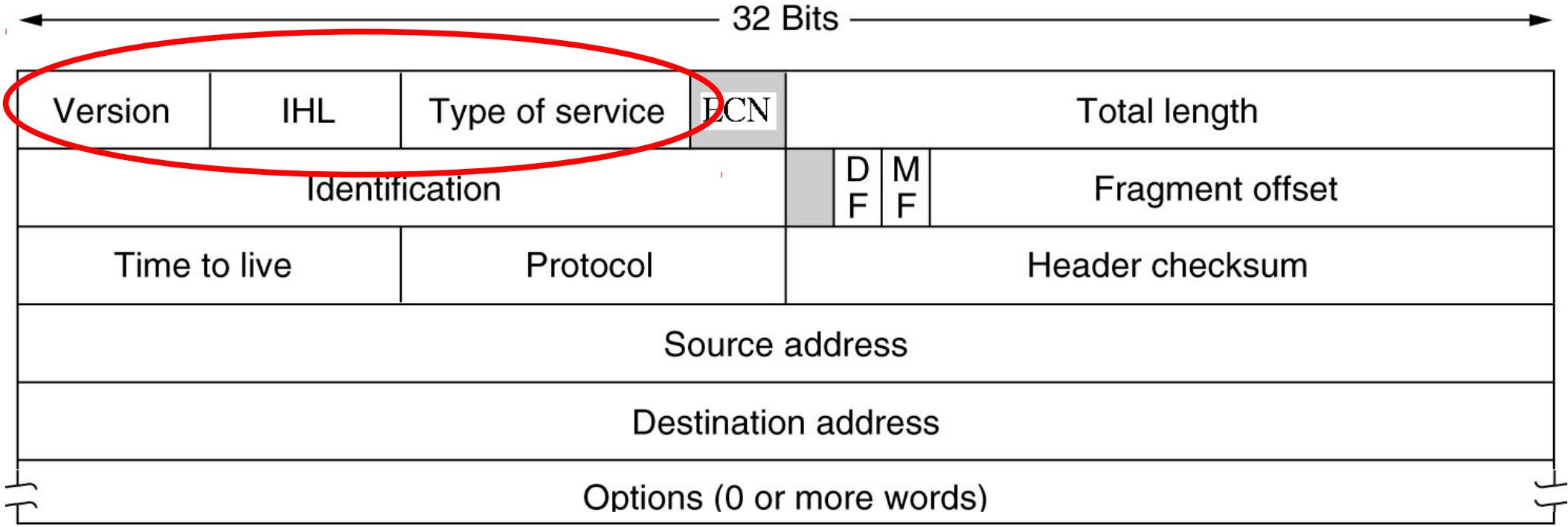
	Összeköttetésmentes	**Összeköttetés alapú**
Áramkörök felépítése	Nincs	Kommunikáció előtt mindig
Címzés	Minden csomag tartalmazza a teljes forrás- és célcímet	Minden csomagban virtuális áramkör azonosító, rövidebb, mint a cím
Csomagtovábbítás	Statikus vagy dinamikus routing	(Circuit-)switching
Állapotinformáció	Routing táblák a routerekben	Virtuális áramköröket nyilvántartó táblázatok a swtich-ekben
Torlódásvédelem, beengedés-szabályozás	Bonyolult	Egyszerű

Hálózati réteg: Szolgáltatások

- Hálózati együttműködés (internetworking)
  - heterogén eszközök, alhálózatok, op. rendszerek, adatkapcsolati protokollok közti átjárás
- Címzés (addressing)
  - eszközök egyedi azonosítóval való ellátása
- Forgalomirányítás (routing)
  - **csomagtovábbítás (forwarding):** egyenként, célcím alapján, a forgalom-továbbítási tábla szerinti következő IP címre
  - **útvonalválasztás (routing):** forgalom-továbbítási táblák kitöltése/fenntartása/frissítése
- Egyéb: fragmentáció (IPv4: R2R, IPv6: E2E), stb.

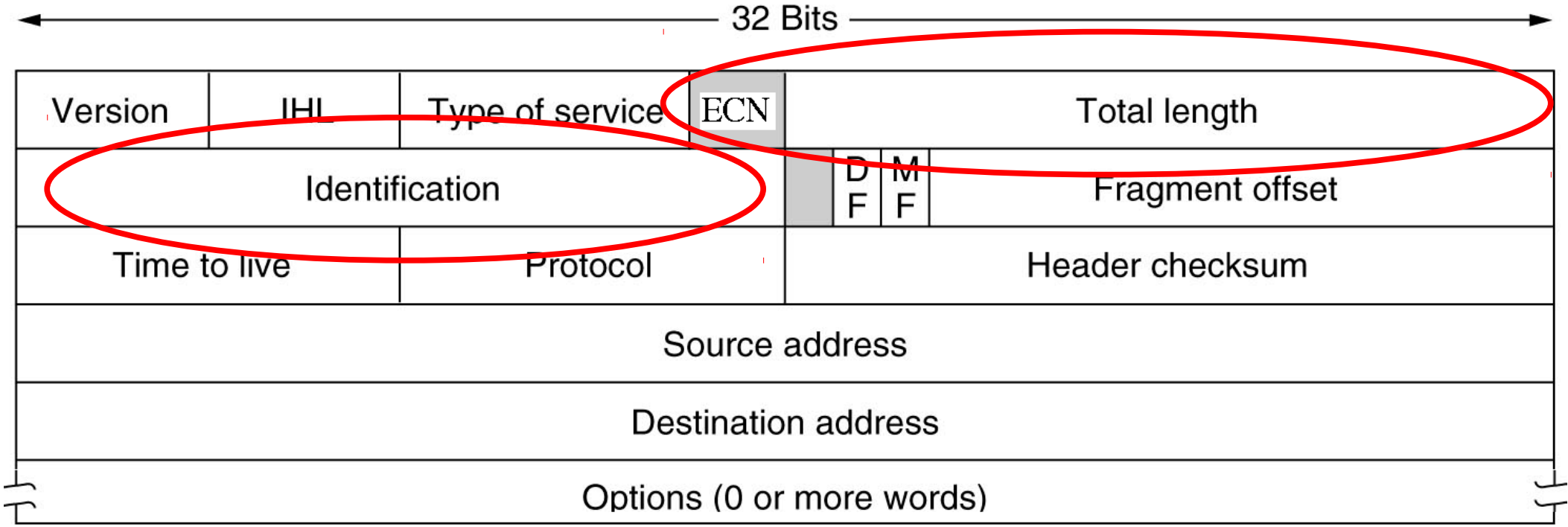
IPv4: fejrész

- Version (4 bit): verzió, mindig `0100` (4)
- Internet Header Length (IHL, 4 bit): fejrész hossz 4 byte-os szavakban (min 20, max 60 byte)
- Type of Service (ToS/DiffServ codepoint, 6 bit)



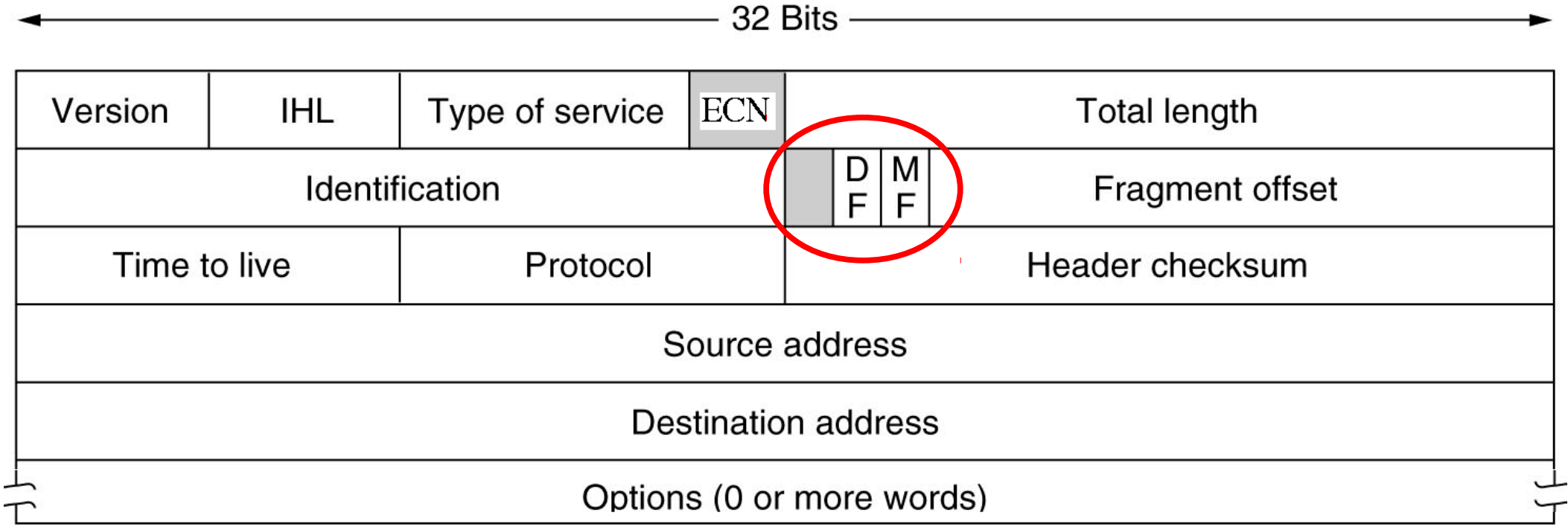
IPv4: fejrész

- Explicit Congestion Notification (ECN, 2 bit): hálózati torlódás jelzése, kevéssé használt
- Hossz (Total length, 16 bit): a csomag hossza
- Identification (16 bit): tördelt csomag azonosítása



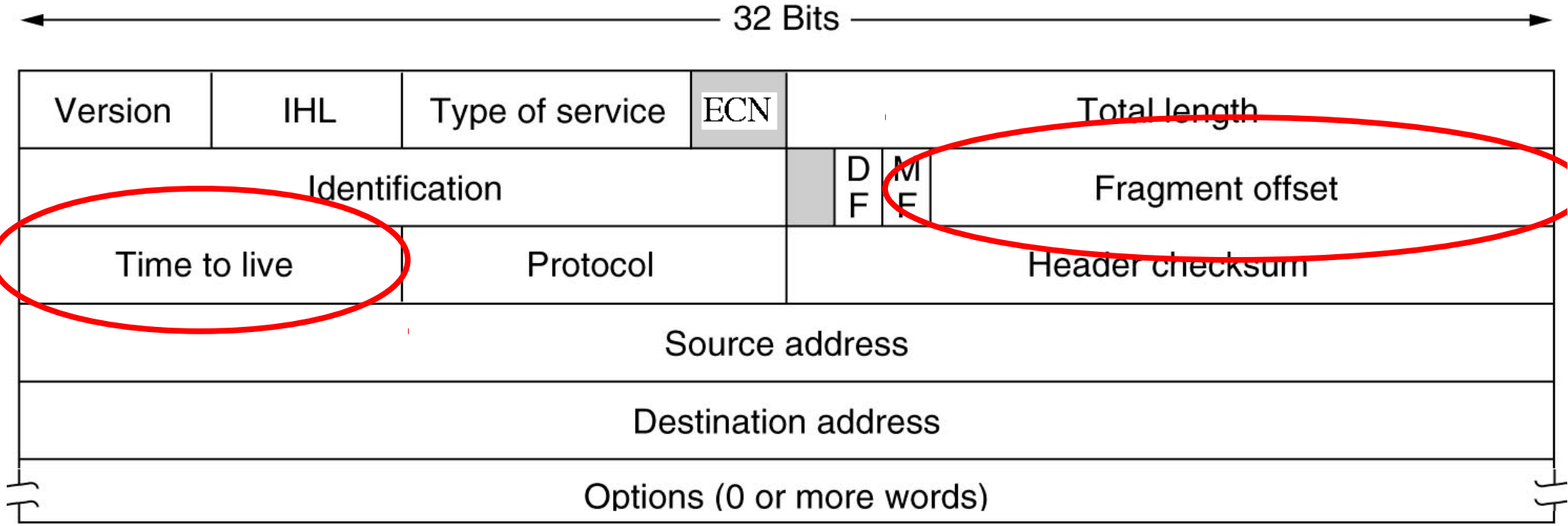
IPv4: fejrész, flag-ek

- bit 0: foglalt, mindig `0`
- bit 1: „Don't fragment” (DF), nem tördelendő
- bit 2: „More fragments (MF), további töredék



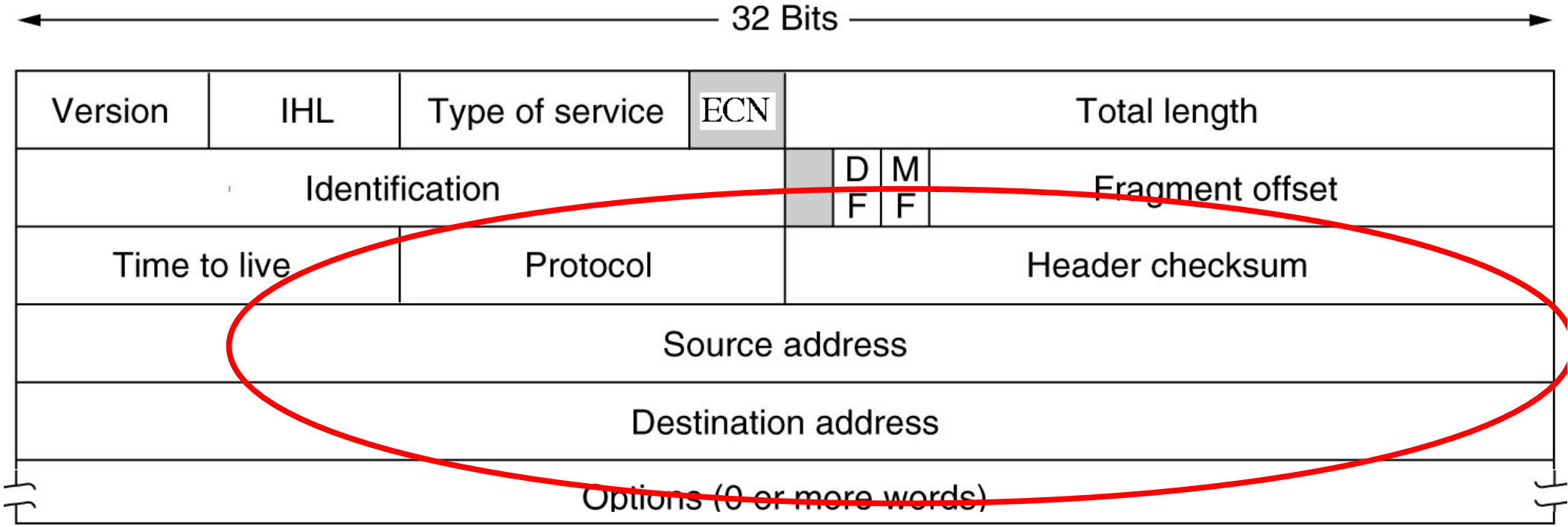
IPv4: fejrész

- Fragment offset (13 bit): a tördelt csomag melyik byte-jától tartalmazza ez a csomag az adatokat
- Time to live (TTL, 8 bit): minden router eggyel csökkenti, a hurokba került csomagok eldobódnak



IPv4: fejrész

- Protocol (8 bit): felsőbb protokoll, amelynek a csomag tartalmát át kell adni (TCP, UDP, ICMP..)
- Header checksum (16 bit): fejrész ellenőrzőösszeg
- Forrás és célcím: 32 bites hoszt azonosítók



IPv4: címzés

- IPv4 cím: 32 bit unsigned integer, 4294967296 (2^{32}) darab egyedi azonosító
- De pl. a `2554524783` cím nehezen olvasható
  - decimális jelölés: `2554524783`
  - bináris:`10011000 01000010 11110100 01101111`
  - „dotted decimal”: 8 bitenként feltördelve négy 1 byte-os számra: `152.66.244.111`

152	66	244	111
10011000	01000010	11110100	01101111
2554524783			

IPv4: alhálózatok

- CIDR (Classless Interdomain Routing): 1993 óta
- IP címek csoportjai **alhálózati prefixbe (subnet)** gyűjthetőek: az alhálózathoz tartozó hosztok
  - közvetlen kommunikálnak
  - a többi IP eszköz számára azonos prefixen „látszanak”
  - így egyetlen routing bejegyzés az összes hosztra
- Az IPv4 cím két részre oszlik:
  - az első X bit az **alhálózat-azonosító**
  - a maradék 32-X bit **hosztazonosító**
  - X-et a **prefix hossz** (pl. `/18`) vagy a **netmask** (pl. `255.255.192.0`) adja meg

IPv4: klasszikus címosztályok

Név	Tartomány	Maszk/CIDR	Példa
Class A	0*******/8 0.0.0.0/8 – 127.0.0.0/8	255.0.0.0 (/8)	17.0.0.0/8 (Apple Inc.)
Class B	10****** ********/16 128.0.0.0/16 – 191.255.0.0/16	255.255.255.0 (/16)	152.66.0.0/16 (BMENET)
Class C	110***** ******** ********/24 192.0.0.0/24 – 223.255.0.0/24	255.255.255.0 (/24)	192.160.172.0/24 (SOTE)
Class D multicast	1110*... 224.0.0.0 – 239.255.255.255	–	224.0.0.5 (All OSPF Routers)
Class E foglalt	1111*... 240.0.0.0 – 255.255.255.255	–	–

- Ma már csak történelmi jelentőségű: pl. a 195.1.0.0/16 papíron Class C, valójában egyben használt (Class B)

IPv4 alhálózatok: CIDR

- CIDR: bármelyik címosztályban bármekkora (> /24) subnet létrehozható
- Variable Length Subnet Masking (VLSM)

CIDR notation	192.168.192.0/18
Prefix hossza	18 bit (az MSB-től)
bináris	11000000 10101000 11000000 00000000
Subnet mask (bináris)	11111111 11111111 11000000 00000000
Subnet mask (dotted)	255.255.192.0
Egyedi IP címek száma	2^{32-18}-2^{14}=16384 (valójában 2-vel kevesebb, a tartomány első és az utolsó IP címe nem használt)
Első IP cím	192.168.192.1
bináris	11000000 10101000 11000000 00000001
Utolsó IP cím	192.168.255.254
bináris	11000000 10101000 11111111 11111110

IPv4 alhálózatok: maszkolás

- Például a `192.168.199.100` cím beletartozik a `192.168.192.0/18` alhálózatba?

A keresett alhálózat	`192.168.192.0/18`
Bináris	`11000000 10101000 11000000 00000000`
Hálózatazonosító 18 bit az MSB-től	`11000000 10101000 11`
Az IP cím	`192.168.199.100`
Bináris	`11000000 10101000 11000111 01100100`
Első 18 bit megegyezik a hálózatazonosítóval	`11000000 10101000 11`

- A hálózatazonosítók (első 18 bit = **prefix**) megegyeznek: a cím része az alhálózatnak
- Nem is mindig olyan egyszerű!

IPv4 forgalomtovábbítás

- An IP-ben a forgalomtovábbítást végző eszközök neve: **router**
  - alhálózatok közötti kommunikáció: interfészenként más alhálózat, (általában) egyedi IP cím
  - a **forgalomtovábbítási tábla** = FIB (Forwarding Information Base) alapján továbbítja a csomagokat
  - FIB karbantartása: statikusan konfigurálva vagy opcionálisan routing protokoll(ok)
  - egyéb szolgáltatások: management (SNMP, CLI), monitoring (SNMP), egyéb protokollok (IGMP, CDP), access control, NAT, stb.

IPv4 forgalomtovábbítás

1) A routing protokollok (routerenként több is futhat, pl. OSPF + BGP) segítségével a routerek topológia-leíró adatokat cserélnek egymás közt
2) A routereken az egyes routing protokollok kitöltik a saját **routing táblájukat** (RIB, Routing Information Base)
3) A router a RIB-ekben található bejegyzésekből egy „legjobb” utat választ minden alhálózathoz
4) A bejegyzéseket letölti a FIB-be: (prefix, prefix-hossz, **next-hop** IP és **next-hop** link-layer cím)

IPv4 forgalomtovábbítás

5) Minden beérkező csomag számára (egyenként!) kikeresi a „legjobb” bejegyzést a FIB-ben

- a csomag IPv4 fejrészében levő célcím alapján
- legjobb bejegyzés: **legspecifikusabb bejegyzés**

6)Kezeli a csomagot (TTL-t csökkenti, stb.)

7)A csomagot továbbítja a FIB-ben talált begyezésben tárolt next-hop számára

- **Hop-by-hop routing:** minden router csak a következő állomást határozza meg, nem az egész utat a célig!

Némi terminológia

- **Routing ≠ Forwarding** (útvonalválasztás ≠ csomagtovábbítás)
  - **forwarding:** csomag eljuttatása a következő hopra
  - **routing:** a továbbítási út meghatározása és a next-hop-ok megtalálása (a FIB kitöltése!)
- Routing table = RIB = „routing tábla”???
  - routing protokollonként egyedi
- Forwarding table = FIB = „routing tábla”???
  - a RIB-ek összefésülése
  - ez alapján továbbítódnak a csomagok!

Legspecifikusabb bejegyzés

- Ha egy célcímre több bejegyzés is illeszkedik
- A legspecifikusabb bejegyzés preferált
  - a legtöbb biten illeszkedő prefix (MSB-től szamítva)
- **Longest Prefix Match (LPM):** az IP routing kulcsa!
  - számtalan hasznos funkció megvalósítható vele
  - ugyanakkor nehezíti a csomagtovábbítást (a FIB-ben való keresés komplexitása miatt)

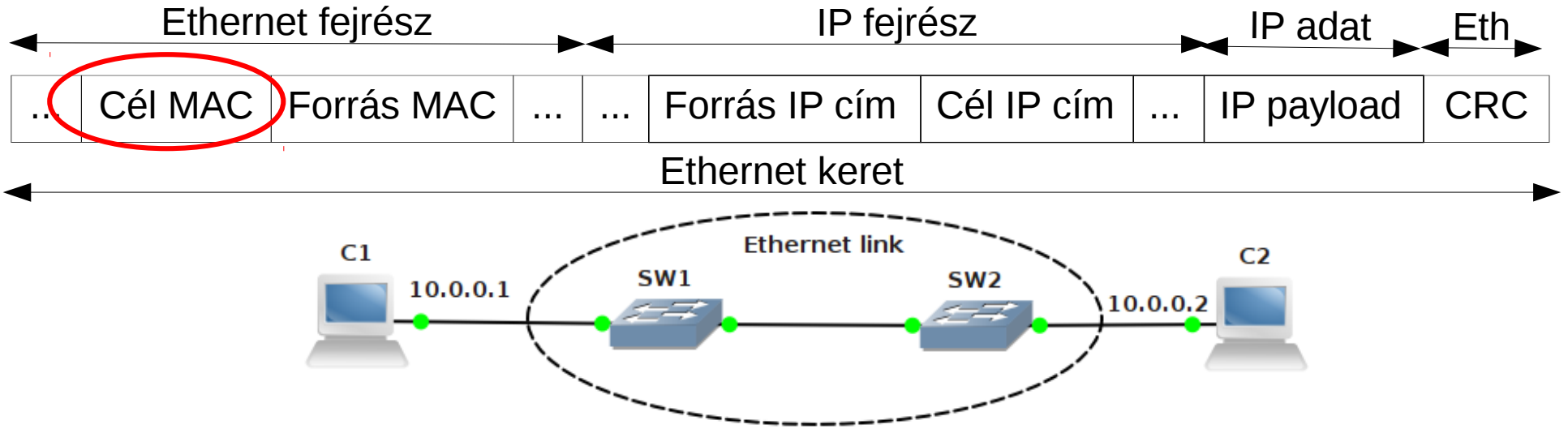
LPM: példa

Egy router FIB-jének részlete		
IP prefix/prefix hossz	**A prefix binárisan**	**Next-hop IP címe**
192.168.0.0/16	11000000 10101000	10.0.0.1
192.168.0.0/17	11000000 10101000 0	10.0.0.2
192.168.64.0/18	11000000 10101000 01	10.0.0.3
192.168.96.0/19	11000000 10101000 011	10.0.0.4

- A 192.168.1.1=x.x.000000001.000000001 címre az első két bejegyzés illik, a 3. és 4. a pirossal jelzett pozíciókban eltér: a 2. preferált
- A 192.168.95.2=x.x.010111111.000000010 címre a 3. bejegyzés, a 192.168.97.3=x.x.011000001.000000011 címre a 4. a LPM

IP over Ethernet

- Az IP legtöbbször Ethernet link layer felett fut



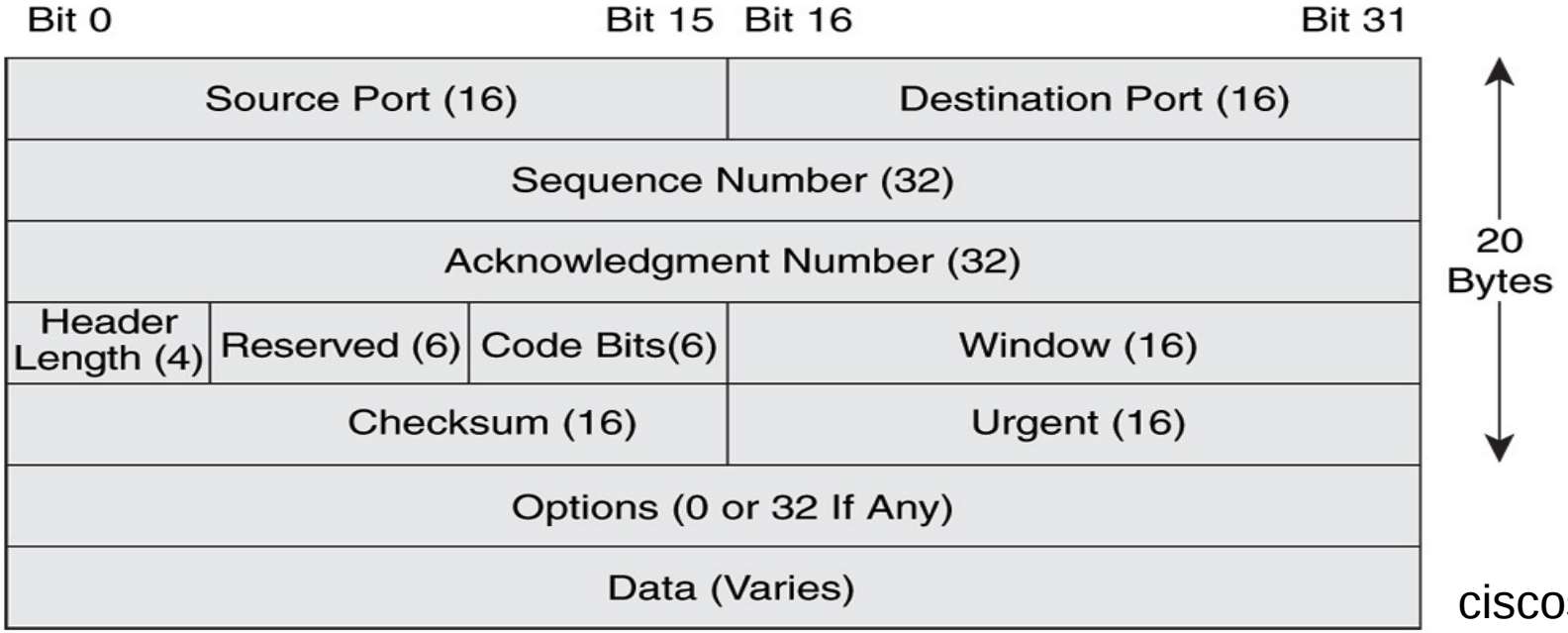
- A `C1` és `C2` hosztok IP szinten szomszédok
- `C1` lekérdezi a linken a `10.0.0.2` címhez tartozó Ethernet MAC címet: **Address Resolution Protocol (ARP)**

Szállítási réteg

- Felhasználók/applikációk közötti adatátvitel
  - egyes applikációk címezhetők (UDP/TCP port)
- A TCP/IP-ben alapvetően két protokoll
- **Transmission Control Protocol (TCP):** összeköttetés alapú megbízható adatfolyam szolgáltatás két hoszt meghatározott TCP portjai között
- **User Datagram Protocol (UDP):** összeköttetésmentes, nem megbízható datagram szolgáltatás UDP portok között

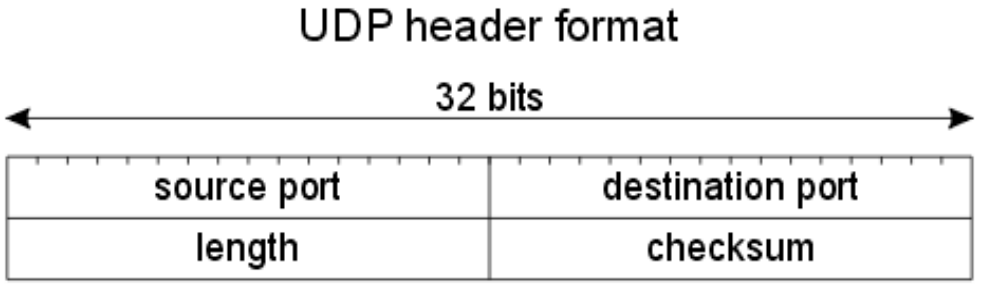
Transmission Control Protocol

- Összeköttetés alapú megbízható adatfolyam
  - sorrendhelyes, csomagvesztés- és hibavédett
  - folyamszabályozás (a vevő elárasztása miatt)
  - torlódásvezérlés (hálózati torlódás elkerülése)
  - több viszony multiplexálása



User Datagram Protocol

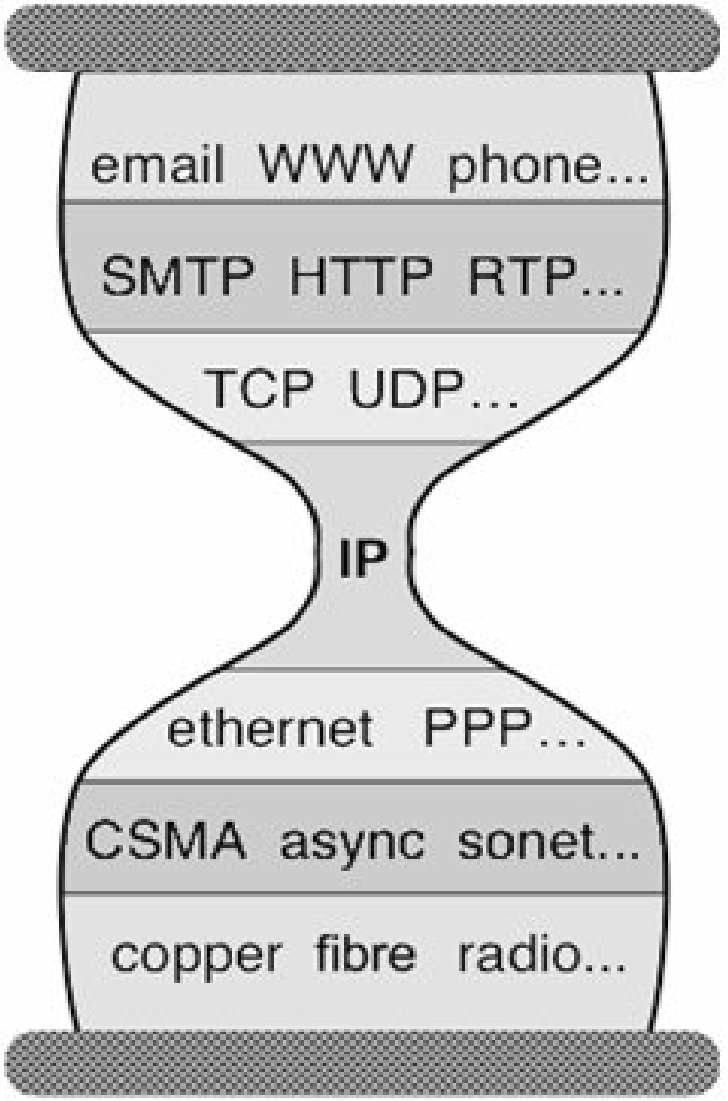
- Összeköttetésmentes datagram szolgáltatás
  - átviteli hiba ellen védett (CRC)
  - de nem megbízható (csomagvesztés), nem sorrendhelyes, nem véd csomagduplikálás ellen
  - nincs összeköttetésvezérlés (handshake, etc.)



wenk.be

IP „homokóra” modell

- IP: a legnagyobb közös osztó
  - minden csomag „áthalad” az IP rétegen
  - minden hoszt érti: igazi „internetworking”
- De pont ezért szinte lehetetlen megváltoztatni
  - IP multicast, IPv6,...
  - „internet ossification”



Trilogy project

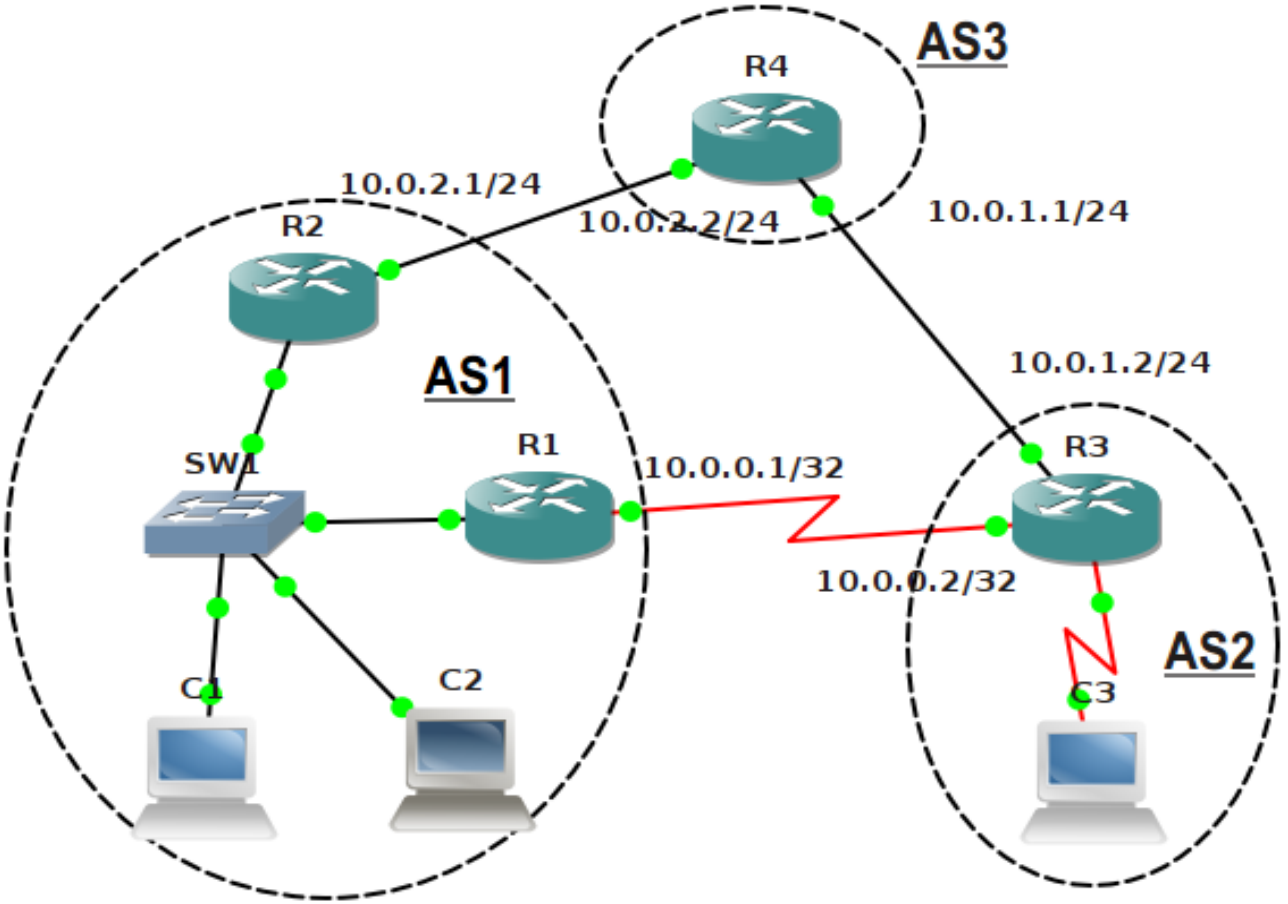
Architektúra

Automóm rendszerek

- Az internet szolgáltatók, egyetemek, cégek, telco-k hálózatait köti egységes rendszerbe
- Mindegyik résztvevő a saját hálózatát egyedileg, a többitől autonóm menedzseli
- **Automóm rendszer (Autonomous System, AS):** hosztok/routerek/hálózatok összessége, melyet valamely szervezet egységesen adminisztrál és amely egységes útválasztási policy-t mutat az internet felé
- Egyedi 32 bites AS azonosítóval van ellátva

Automóm rendszerek

- `AS1` tartalmazza `R1` és `R2` routereket, `SW1` switch-et, és `C1` és `C2` hosztokat



Automóm rendszerek

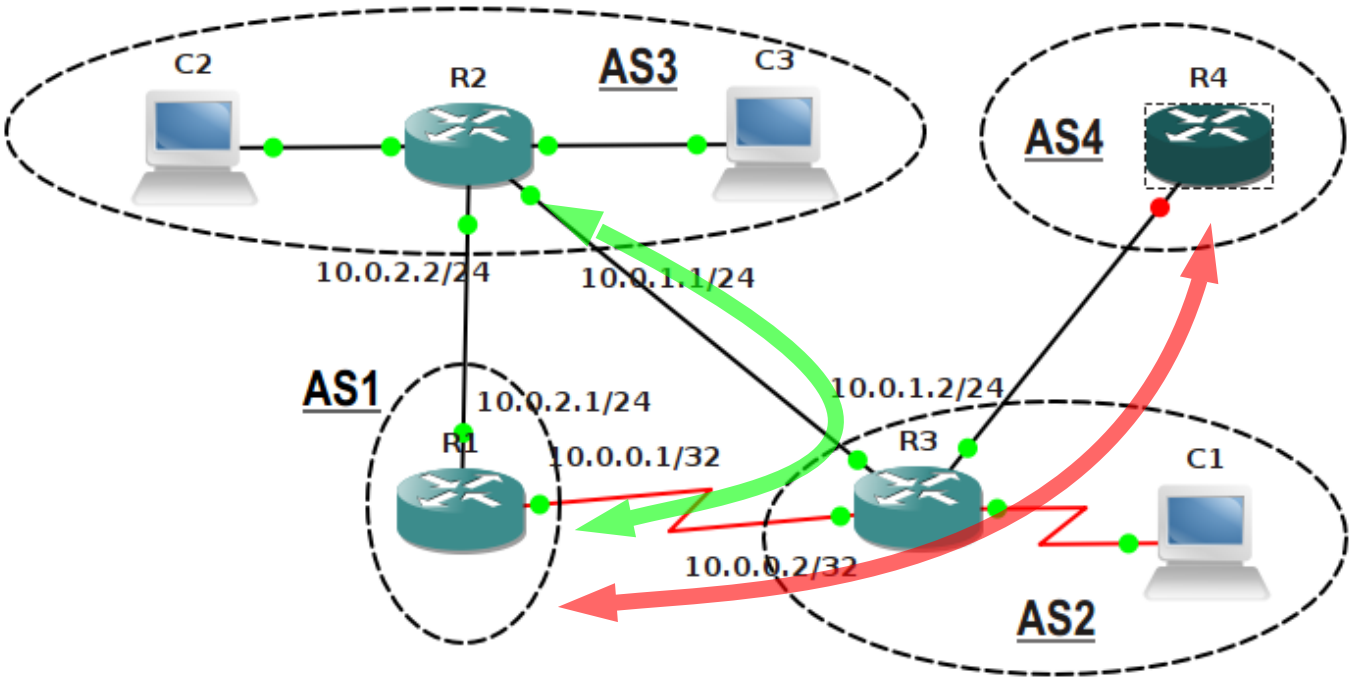
- **ISP (Internet Service Provider):** Magyar Telekom (AS5483), Telenor Hungary (AS8448)
- **Campus:** BME (AS2547), Harvard (AS11), MIT (AS3), UC San Diego (AS7377)
- **Enterprise:** IBM (AS547, AS763, stb., de övék a 9.0.0.0/8 Class A cím is!), Apple (AS714)
- **Globális szolgáltató:** Sprint (AS1239, AS1240, AS6211, AS6242, ...), Cogent (AS174, AS2149, AS6494), TeliaSonera (AS1759)
- Egy ISP-nek több AS száma is lehet!

Útválasztás AS-ben és AS-ek közt

- AS útvonalválasztási szempontból egységes, a többi AS-re AS-enként külön policy
  - milyen forgalmakat enged át a hálózatán
- Ez a routing protokollon keresztül valósul meg
  - milyen útvonalakat exportál és importál
- AS-en belüli útválasztás (intra-domain routing):
  - **Interior-Gateway Protocol (IGP):** OSPF,RIP
- AS-ek közti útválasztás (inter-domain routing):
  - **Exterior Gateway Protocol (EGP):** BGP

AS: útválasztási stratégiák

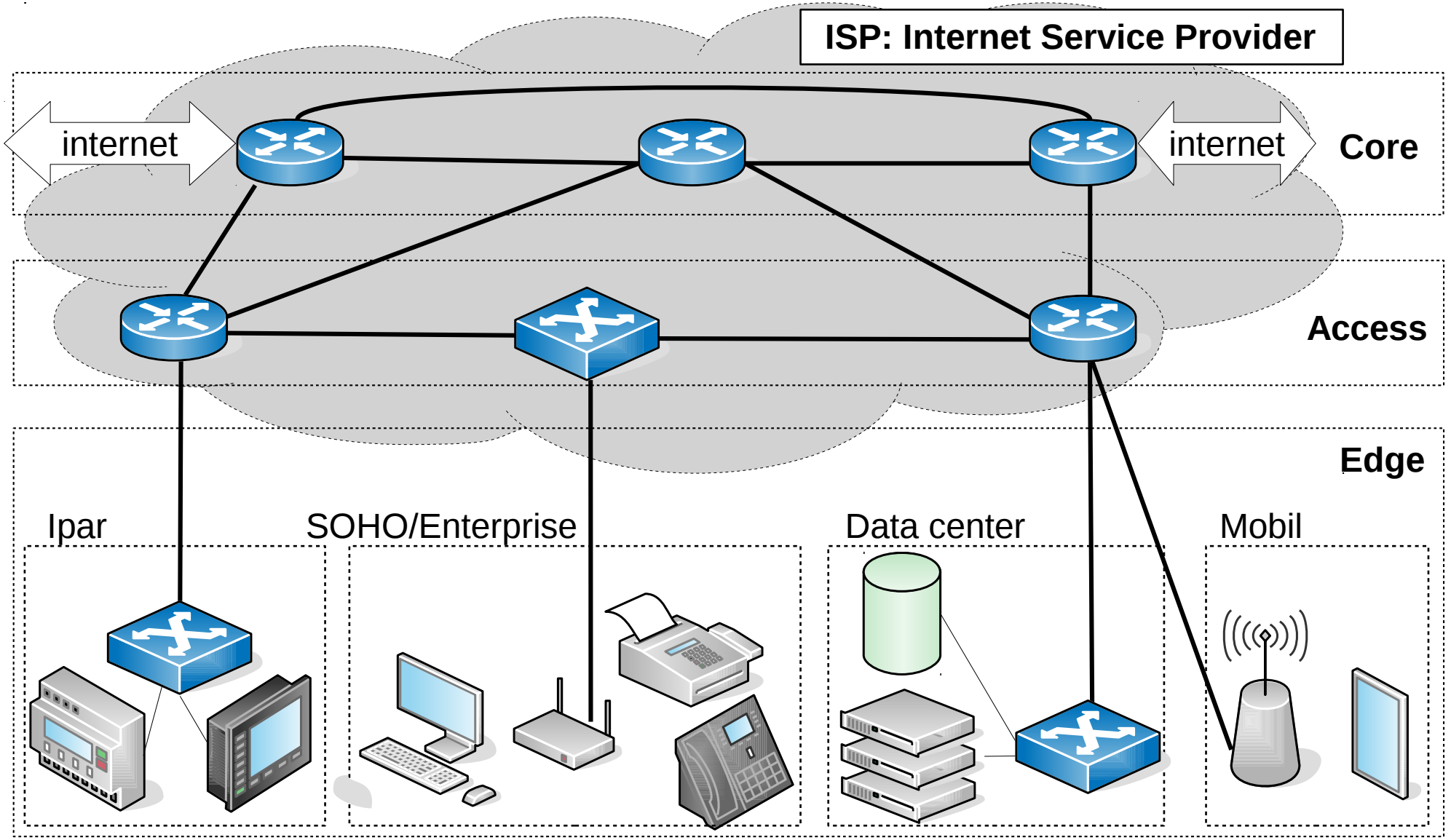
- Példa: `AS2` továbbítja az `AS1`–`AS3` forgalmat
  - **egységes policy:** `AS3` **minden** hosztjának forgalmát (`C2`, `C3`) átviszi `AS1`-be és viszont
- De `AS2` blokkolja `AS1`–`AS4` átmenő forgalmát



AS-ek típusai

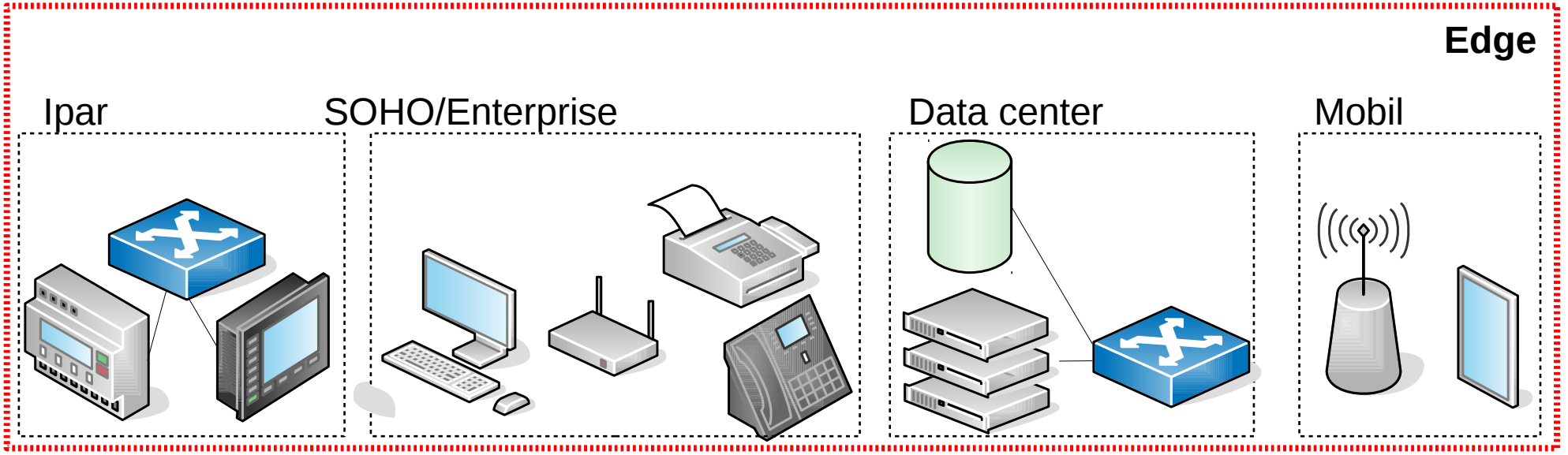
- **Content AS:** tartalomszolgáltató AS
  - kimenő forgalom domináns
  - Youtube, Netflix, HULU
- **Eyeball AS:** a tartalomfogyasztás domináns
  - nagy előfizetői bázist kiszolgáló ISP-k
  - nagy bemenő forgalom, kis késleltetés a content felé!
- **Transit AS:** globális adatátvitel (Cogent,Level3)
- **CDN** (Content Delivery Network): Akamai
  - content + global transit

Egy alternatív felosztás: Edge/Access/Core



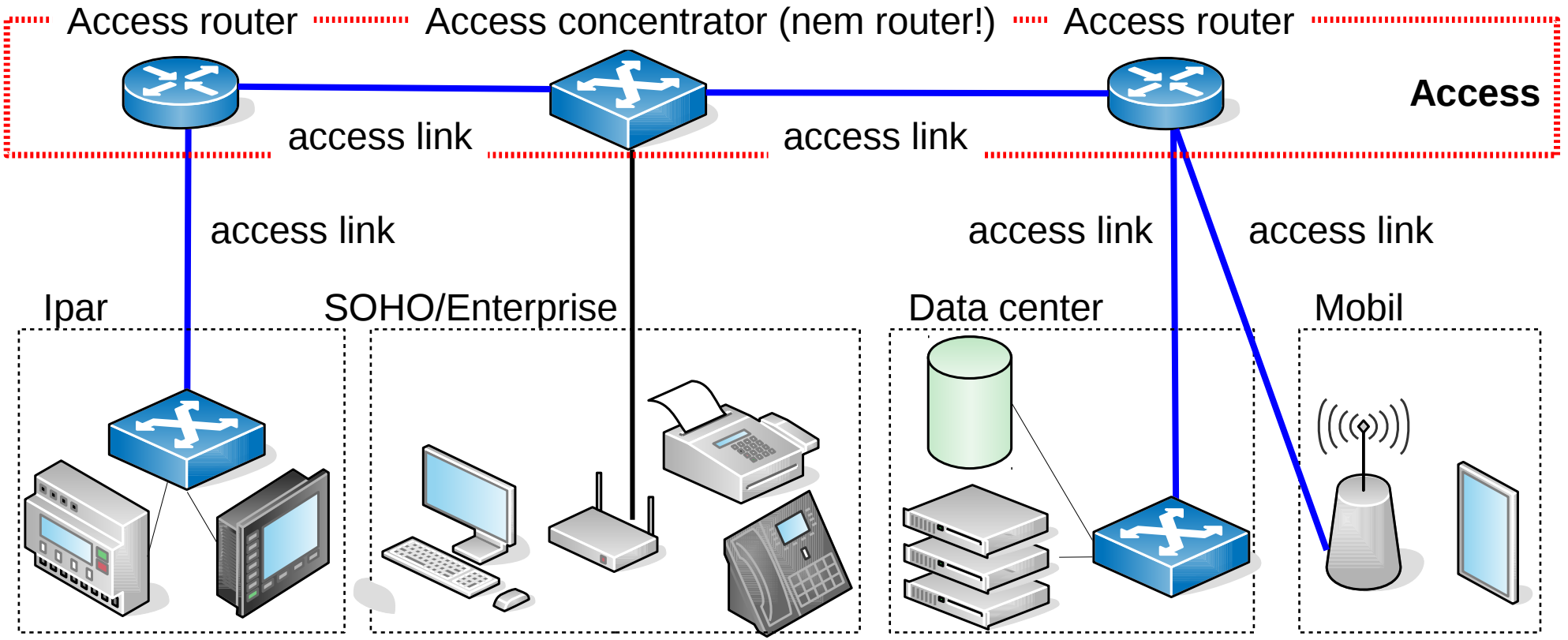
Edge

- Végberendezések, hosztok
  - desktop, laptop, mobil, PDA, tablet
- Egyéni előfizetők, **SOHO** (Small Office/Home Office), mobile, data center, ipari rendszerek
- Lokális hálózatok, egy szolgáltatóhoz bekötve



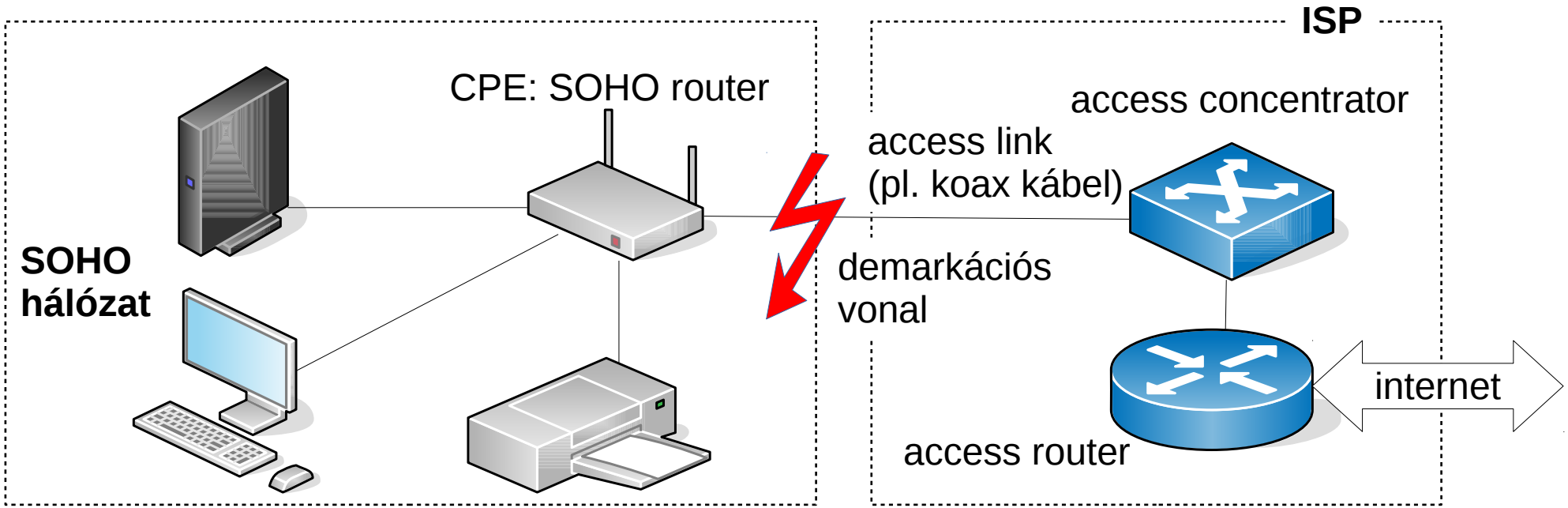
Access

- Edge forgalmának koncentrálása a core felé
- **Access router:** első szolgáltatói router az úton
- **Access link:** átviteli médium az access routerig



Internet access

- **CPE** (Customer-premises Equipment): az előfizetőnél elhelyezett terminál, amely a szolgáltatóhoz van bekötve
- **Demarkációs vonal:** a szolgáltató és az előfizető hálózata közötti határvonal



Előfizetői/SOHO access

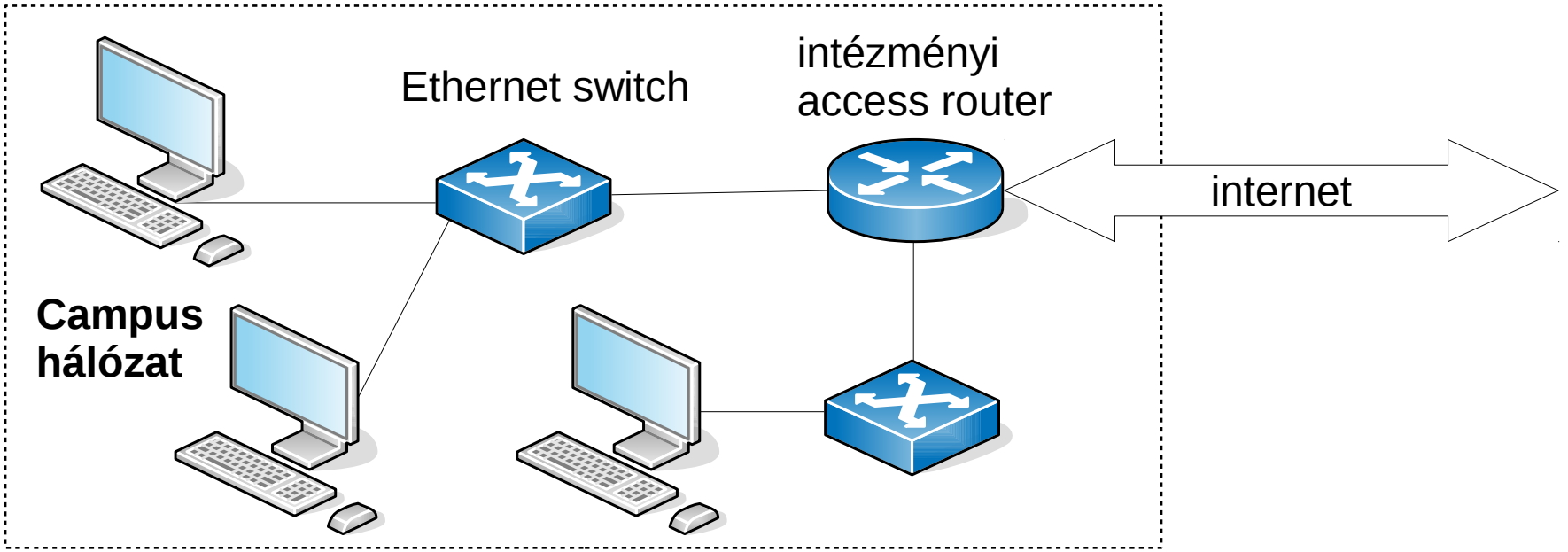
- Dial-up: telefonos betárcsázás
  - CPE: analóg modem
  - access link: telefonvonal
  - ISP/telco access: POTS előfizetői kártya
- DSL (Digital Subscriber Line): ADSL/VDSL
  - CPE: DSL modem
  - access link: telefonvonal, sodort érpár
  - ISP/telco access: DSLAM (Digital Subscriber Line Multiplexer)

Előfizetői/SOHO access

- Kábel(TV):
  - CPE: kábelmodem
  - link: koax vagy HFC (Hybrid Fiber Coax)
  - ISP/telco access: cable headend
- FTTx (Fiber-to-the-X): access üvegszálon
  - FTTH (Fiber-to-the-Home): üvegszál CPE-ig
  - FTTP (Fiber-to-the-Premises): üvegszál az épületig, onnan belső hozzáférési hálózat
  - technológia: PON (Passive Optical Network)

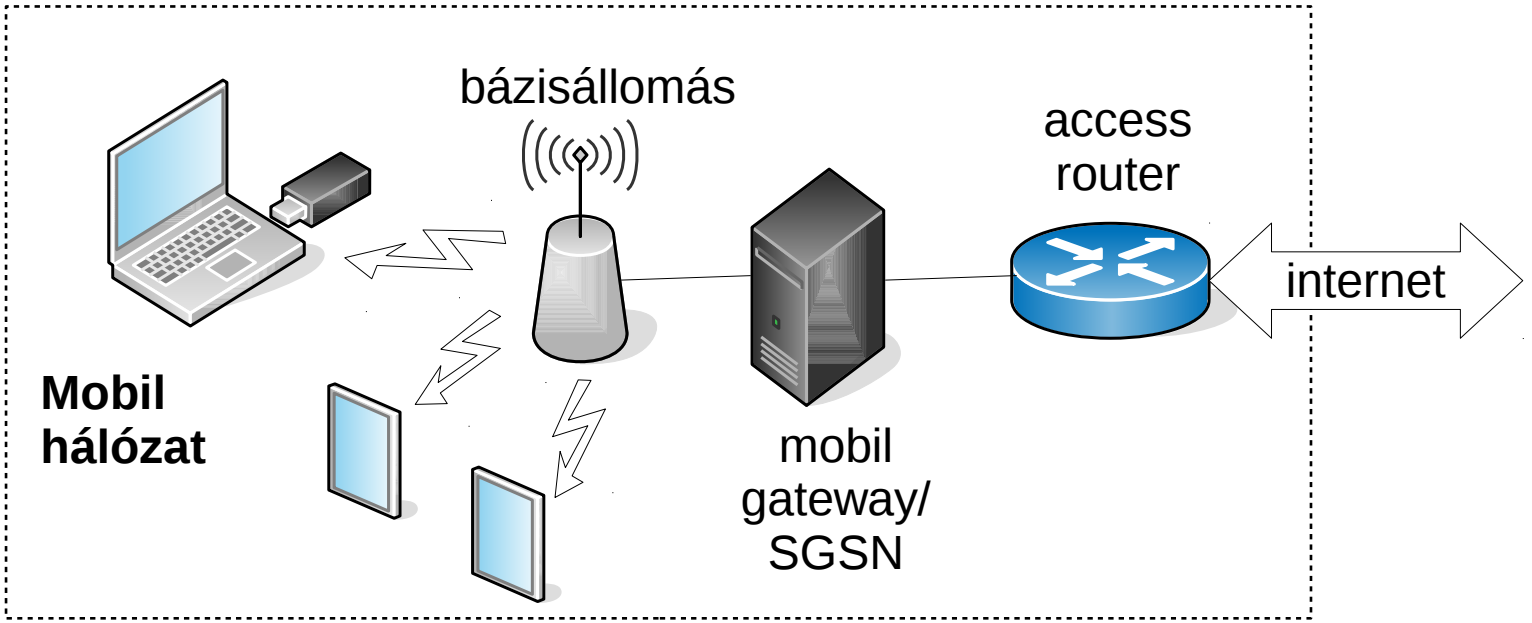
Enterprise/campus access

- **Enterprise:** kereskedelmi egység, cég, multi...
- **Campus:** több épületből álló intézmény
  - tipikusan oktatási, kormányzati, hadsereg
- Egy egyetemi Ethernet hozzáférési hálózat



Mobil access (2G, 3G, 4G, WiMAX)

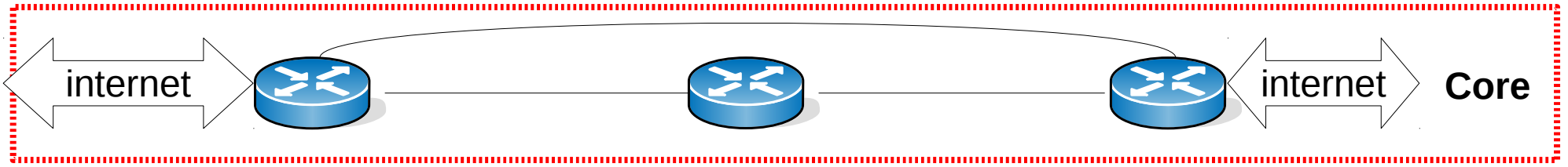
- Internet access celluláris mobilhálózaton
- 2.5G: General Packet Radio Service (GPRS) csomagkapcsolás mobilhálózatokban
- Később: EDGE, UMTS, 4G, stb.



Az internet gerinchálózata: Core

- Az access koncentrálja az internet edge forgalmát és bevezeti az internet core-ba
- AS-en belül: nagy sebességű, de „buta” routerek sűrűn össze-kapcsolt (mesh) hálózatban
- AS-ek között: „okos” routerek a forgalom kicserélésére
- Az internet legizgalmasabb része: fő témánk

Cisco CRS-1
wikipedia



Tipikus AS: az ISP

- **Internet Service Provider:** internetszolgáltató
  - fő profil: kétirányú hozzáférést szolgáltat az előfizető és az internet között
  - egyéb szolgáltatások: DNS regisztráció, email, tárhely, server colocation,...
- Az ISP hálózatok egy vagy több POP-ból állnak
- **POP (Point of Presence):** valamely fizikai lokáción elhelyezett szerverek, switchek, access concentratorok, stb. összessége

Példa: egy globális ISP