

Digitální stopy v internetové komunikaci

Ondřej Caletka



15. ledna 2018



Uvedené dílo podléhá licenci Creative Commons Uveďte autora 3.0 Česko.

O sdružení CESNET



	n×100 Gb/s		100 Gb/s
	n×10 Gb/s		10 Gb/s
	uzel (PoP)		1-2,5 Gb/s
	uživatel (user)		<1 Gb/s



MetaCentrum



UltraGrid

Jak vlastně funguje internet

Naivní představa

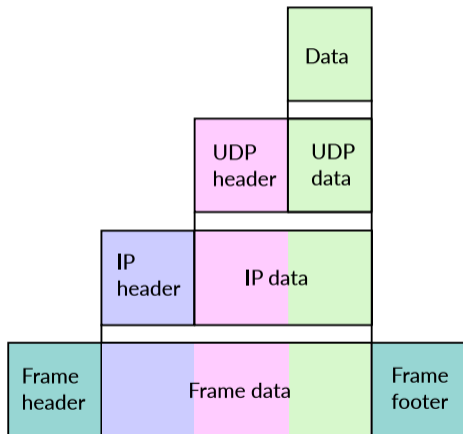
- noviny
- televize
- pošta

Realita

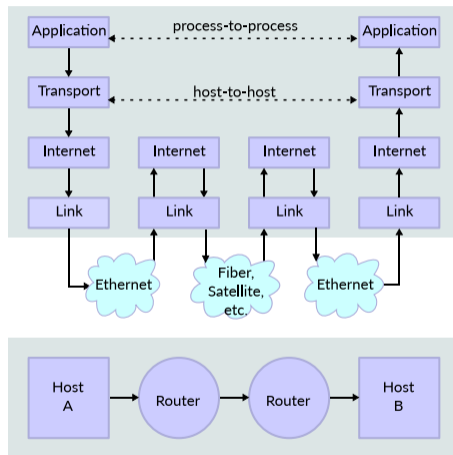
- telefon
- nelze *pasivně* konzumovat

- datová komunikace jeden-na-jednoho
- každé použití zanechává stopy

Vrstvový model

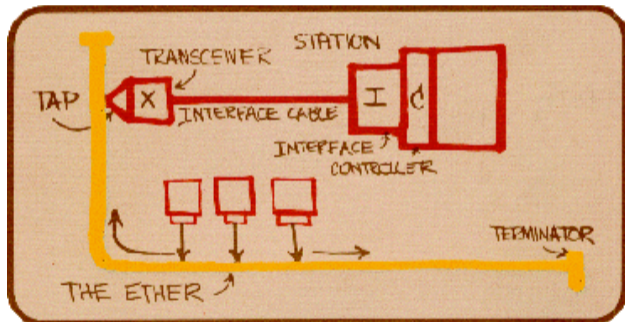


Autor: Cburnet, Kbrose, Wikimedia commons, GFDL



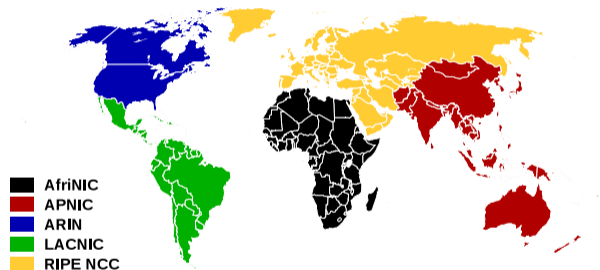
Linková vrstva

- pouze lokální komunikace
- nejčastěji Ethernet
- adresování MAC adresami
- 48bitové identifikátory
- stabilní



Internetová vrstva

- s globálním dosahem
- hierarchické adresování
- identifikátor i lokátor
- 32bitové IPv4
- 128bitové IPv6
- bez pevného vztahu k hardware



- spojení z konce na konec
- přizpůsobuje přenosový kanál
- nejčastěji TCP a UDP
- adresace pomocí portů

Transport Layer Security

- mezivrtstva nad transportní
- šifruje data vyšších vrstev
- brání nižším vrstvám v sledování a **pozměňování** obsahu

`https://www.internetovehazardnihry.cz/casino/?game=blackjack#score`

schéma
protokol

hostitel
internetová adresa

cesta

dotaz

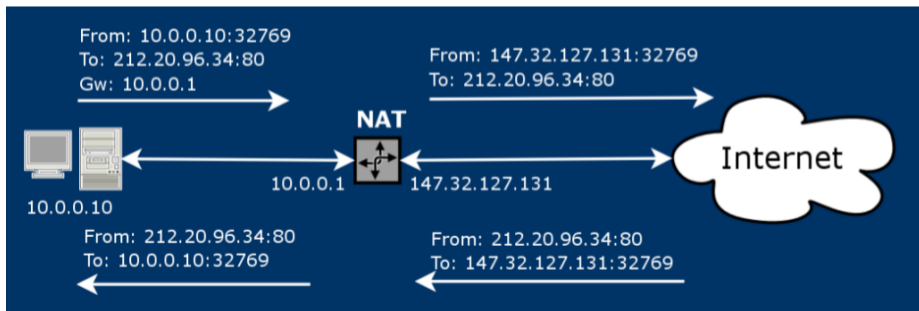
kotva
nepřenáší se

- mnoho protokolů, nejčastěji HTTP/HTTPS
- vlastní adresování
- často utajené nižším vrstvám pomocí šifrování TLS
- spousta **metadat**

- v reálném čase protokolem ARP/NDP
- u IPv4 často *semi*-permanentně pomocí DHCP serveru
- u IPv6 nejčastěji náhodná IP adresa proměnná v čase
- v malých sítích bez trvalých stop
- u větších sítí zaznamenáváno a uchováváno

Sdílení a překlad adres

- běžná praxe v IPv4
- zásah do komunikace porušující vrstevný model
- privátní a veřejné adresy, mapování čísel portů
- proměnlivé v čase
- zaznamenávání historie mapování jen u velkých sítí (Carrier Grade NAT)



- povinnost zaznamenávat provozní a lokalizační údaje
- záznam informací o datových tocích
- data internetové a transportní vrstvy
- *nevidí* za NAT koncového zákazníka
- možnost korelace záznamů z více zdrojů

- drobné odlišnosti v implementacích TCP/IP
- aktivní i pasivní

Příklad

operační systém	hodnota TTL	velikost TCP okna
Linux	64	5840
FreeBSD	64	65535
Windows XP	128	65535
Windows 7	128	8192

Záznamy aplikační vrstvy

- jsou k dispozici pouze na koncových bodech
- obsáhlé HTTP hlavičky
- zejména User-Agent, Referrer, Cookies

▼ Request Headers

```
:authority: en.wikipedia.org
:method: GET
:path: /wiki/P3P
:scheme: https
accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8
accept-encoding: gzip, deflate, br
accept-language: cs,en-US;q=0.8,en;q=0.6,sk;q=0.4
referrer: https://www.google.cz/
upgrade-insecure-requests: 1
user-agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/61.0.3163.100 Safari/537.36
```

- informace, které může webová služba uložit a později přečíst
- nezbytná např. pro funkci přihlášení, stav nákupního košíku, apod.
- zneužívaná pro sledování chování uživatelů
- zajímavé zejména cookies významných třetích stran
 - Google Analytics
 - Facebook Like tlačítko
 - reklamní síť

Je prohlížeč unikátní?

The screenshot shows a web browser window with the URL <https://amiunique.org/fp>. The page title is "Am I Unique?". The main content area displays the question "Are you unique?" followed by the result: "Almost! (You can most certainly be tracked.)". Below this, several statistics are listed:

- 38.51 % of observed browsers are **Chrome**, as yours.
- 1.09 % of observed browsers are **Chrome 61.0**, as yours.
- 14.89 % of observed browsers run **Linux**, as yours.
- 0.43 % of observed browsers have set "**cs**" as their primary language, as yours.
- 22.33 % of observed browsers have **UTC+2** as their timezone, as yours.

At the bottom, it states: "But **only 2** browsers out of the 506182 observed browsers (0.00 %) have exactly the same fingerprint as yours." Two buttons are visible: "View more details" and "View graphs". The left sidebar contains navigation links: Home, My fingerprint, My timeline (New), Global statistics, FAQ, Privacy policy, Privacy tools, and Links (Updated).

- MAC jsou stabilní, ale nepřenáší se
- Vazbu IP adresy na MAC adresu dlouhodobě evidují jen velké sítě
- Překlady adres a náhodné IPv6 adresy dohledání majitelů komplikují
- Webový prohlížeč odesílá spoustu metadat, často unikátních
- Zajímavé informace mají mimo jiné provozovatelé reklamních systémů

Děkuji za pozornost

Ondřej Caletka
Ondrej.Caletka@cesnet.cz
[https://Ondřej.Caletka.cz](https://Ondrej.Caletka.cz)

