



Komprimácia



Otázky na rozmýšľanie

- Ako je možné zmenšiť veľkosť súboru bez toho, aby sa z neho stratili informácie?
- Môžeme komprimovať dáta tak, že stratíme nejaké informácie?
- Ako veľmi vieme skomprimovať dáta?

Komprimácia

- **Komprimácia** (kompresia) je proces, pri ktorom dochádza k zmenšovaniu objemu dát na dátovom médiu
- Dáta pred komprimáciou nazývame **hrubé dáta**
- Dáta po komprimácii nazývame **komprimované dáta**
- Postup, ktorým vieme zmenšiť objem dát a vytvoriť z hrubých dát komprimované, nazývame **kompresný algoritmus**

Komprimačný pomer (kp)

- Je pomer medzi hrubými (h) a komprimovanými (k) dátami

$$kp = \frac{V_k}{V_h} * 100$$

- Udáva sa v percentách
- Ak je väčší ako 100%, hovoríme o negatívnej komprimácii, kedy vznikne väčší súbor ako pôvodný (napríklad keď sa snažíme skomprimovať už skomprimované dáta)

Rozdelenie komprimácia

- Podľa zachovania pôvodného obsahu:
 - **Stratová komprimácia** – dáta po dekomprimácii ***nie sú tie isté***, ako boli pred komprimáciou
 - **Bezstratová komprimácia** – dáta po dekomprimácii ***sú tie isté***, ako boli pred komprimáciou
- Podľa potreby dekomprimačného programu:
 - **Offline komprimácia** – ***je potrebný program*** na komprimáciu a dekomprimáciu
 - **Online komprimácia** – dekomprimáciu ***vykonáva priamo operačný systém*** pri pokuse o otvorenie súboru

Bezstratová komprimácia

- Komprimujeme ňou **binárne súbory programov** (spustiteľné súbory, knižnice, uložené súbory...) alebo **čitateľné textové súbory** (konfiguračné súbory, zdrojové kódy programov ...)
- Vykonáva sa zhusteným zápisom sekvenčných dát alebo slovníkovými metódami, či zápisom adres opakujúcich sa dát
- Súbory s bezstratovou komprimáciou majú prípony **zip, 7zip, arj, rar, bzip, gzip**, ale napríklad aj **png**, či **docx, pptx, jar**

Bezstratová | Zhustený zápis sekvenčných dát

- Predstavme si, že máme súbory, v ktorých sa opakujú dlhé sekvencie dát (napríklad súbor, v ktorom máme informácie o prejdení 30 levelov zapísané ako y=yes, n=no, s-skipped pre každý level)

yyyyyyssssyyyyyysssyynnnnnnnnnnn (30 bajtov)

- Súbor môžeme skomprimovať tak, že zapíšeme jednotlivé sekvencie ako dvojice [číslo,znak] kde číslo udáva počet výskytov znaku

5y4s7y3s2y9n (12 bajtov)

Bezstratová | Zápisy adries opakujúcich sa dát

- Ak dáta nie sú sekvencie rovnakých znakov, ale obsahujú opakujúce sa slová, môžeme tieto slová nahradiť len odkazom na prvý výskyt slova alebo reťazca:

V elektrotechnike skúmame elektromagnetické javy, ktoré sú vyvolané a prejavujú sa elektrickým a magnetickým poľom. (117 bajtov)

- Môžeme hneď vidieť reťazce, ktoré sa často opakujú a nahradiť ich len zápisom adresy a dĺžky opakujúcich sa dát
- Napríklad reťazec elektr (začína sa 3. písmenom a je dlhý 6 znakov) alebo reťazec magnetick (začína sa 34. písmenom a má dĺžku 9 znakov)
- Namiesto 6B reťazca elektr tak vo výsledku uvedieme len dve jednobajtové čísla

V elektrotechnike skúmame [3;6]omagnetické javy, ktoré sú vyvolané a prejavujú sa [3;6]ickým a [34;9]ým poľom. (99 bajtov)

Bezstratová | Slovníkové metody

- Vytvára sa slovník pevnej dĺžky (napr 4096 záznamov), ktorý v prvom bajte obsahuje štandardné ASCII kódy a zvyšok sú adresy a k nim prislúchajúce opakujúce sa sekvencie dát
- Následne sa potom vo výslednom súbore zapíše pri danej sekvencii len odkaz do slovníka a na konci samotný slovník

Tab.1 Príklad Slovníka	
kód	reťazce
0000	0
0001	1
:	:
0254	254
0255	255
0256	145 201 4
0257	243 245
:	:
4096	XXX XXX XXX

originál dátový tok: 123 145 201 4 119 89 243 245 59 11 206 145 201 145 201 4 243 245,...

dáta zakódovane pomocou: 123 256 119 89 257 59 11 206 256 257 ...

slovníka

Stratová kompresia

- Komprimujeme ňou vo väčšine prípadov **multimediálne súbory** (video, audio, obrázky)
- Prvým krokom je odstránenie „nepotrebných“ dát, teda tých dát, ktoré sme ochotní obetovať pre lepšiu kompresiu a následne sa použije algoritmus bezstratovej kompresie
- Súbory so stratovou kompresiou majú prípony **jpg, gif, mp3, wma, mp4, mpeg**

Stratová | Príklad kompresie s pevnou paletou

- Namiesto 24-bitovej farebnej palety sa z obrázku vyberie 256 najviac používaných farieb a tie sa zakódujú pomocou 8 bitov
- Takto vieme zmenšiť súbor o tretinu a navyše opakujúce sa znaky vieme ďalej komprimovať



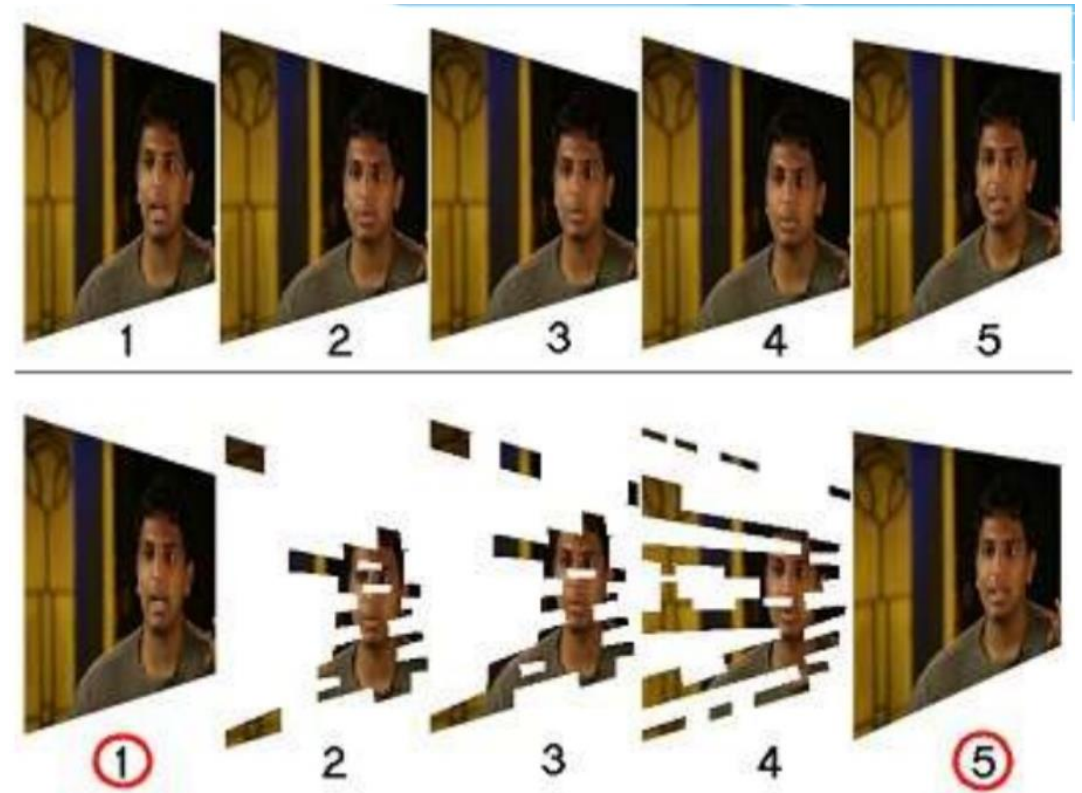
Stratová | Príklad kompresie s adaptívnou paletou

- Umožňuje nastaviť úroveň kompresie, v rámci ktorej sa zjednocujú príbuzné farby pixelov na jednu a tým vznikajú dlhé sekvencie rovnakých znakov, ktoré je možné následne efektívne skomprimovať
- Na obrázku môžeme vidieť rôzne úrovne kompresie od maximálnej (vľavo) po minimálnu (vpravo)



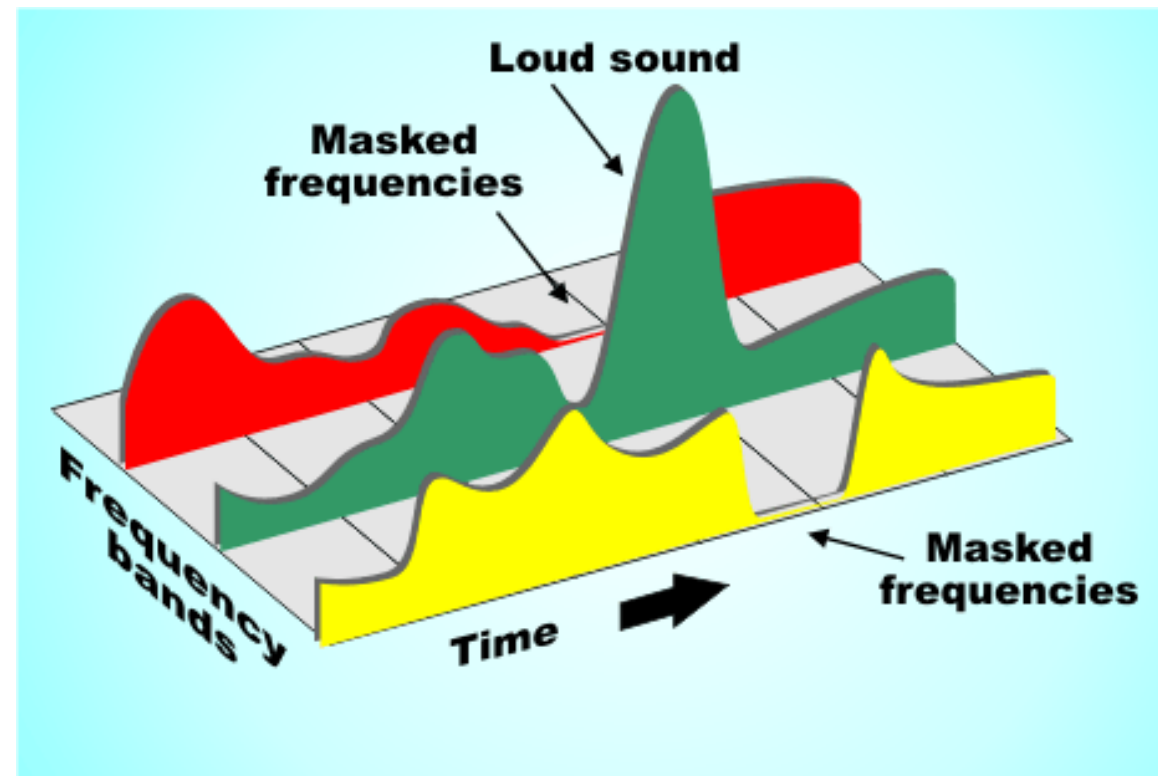
Stratová | Komprimácia videa

- Keďže video sa skladá z viacerých po sebe idúcich obrazoch, kde sa často mení len niekoľko oblastí v scéne, nie je potrebné ukladať každý jeden obraz, iba každý n -tý a následne rozdiely medzi nimi



Stratová | Komprimácia zvuku

- Programy odrežú nepočuteľné zvuky, ktoré môžu byť spôsobené napríklad neschopnosťou ucha počuť vysoké frekvencie (nad 20kHz) alebo príliš hlasným zvukom v jednom frekvenčnom pásme (viď obr.)
- Výsledok sa následne spracuje bezstratovou kompresiou

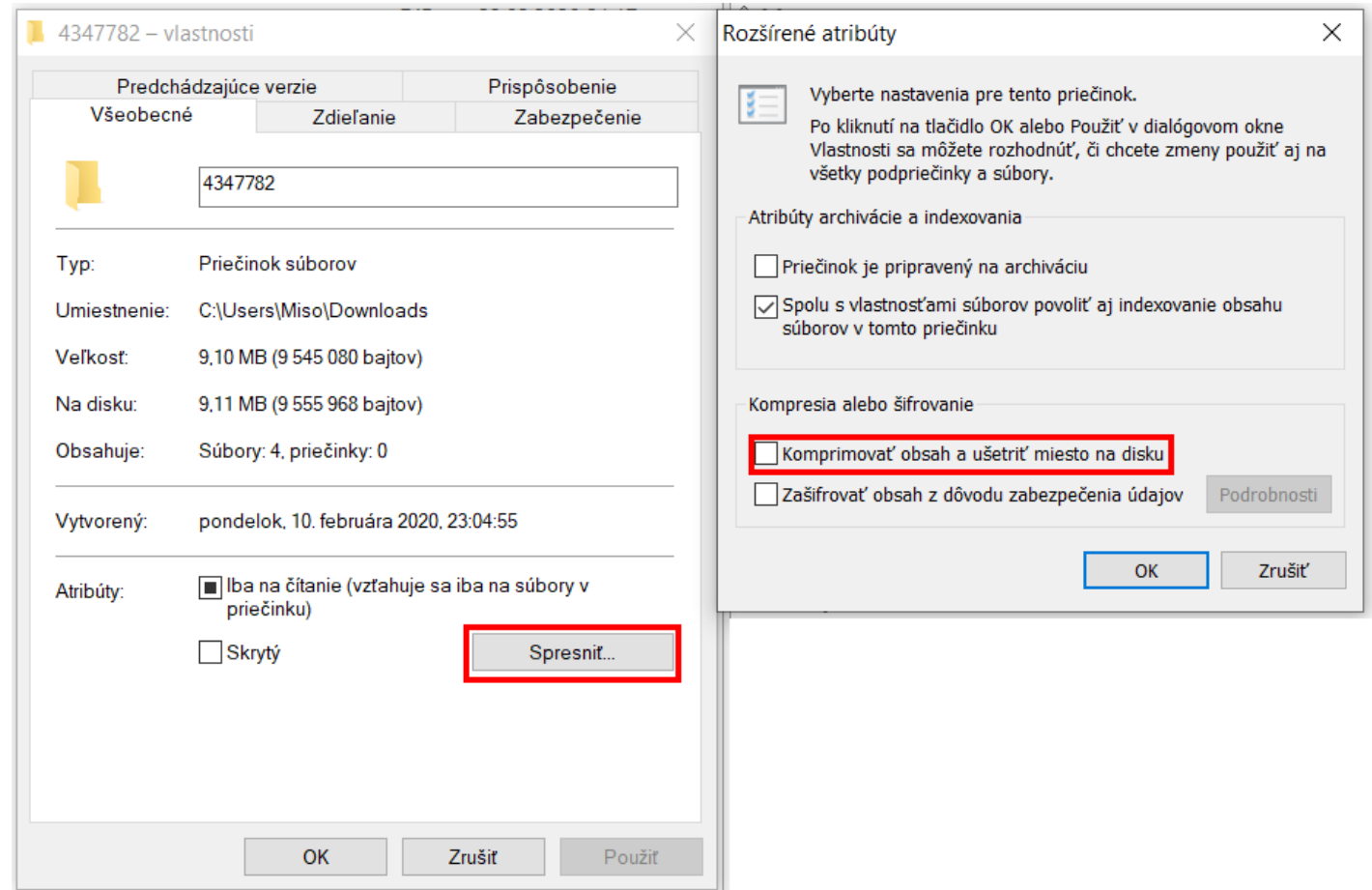


Offline komprimácia

- Pre offline komprimáciu je potrebné spustiť špeciálny program
- Používateľ teda komprimuje a dekomprimuje dáta vedome
- Programy, ktoré slúžia na komprimáciu a dekomprimáciu sú napríklad 7-Zip, WinZip, WinRAR, Tar, Gzip, Bzip2, ale napríklad aj Total Commander, ktorý má integrovaný komprimačný program

Online komprimácia

- Robí ju operačný systém automaticky, bez vedomia používateľa
- Windows napríklad dokáže nad súborovým systémom NTFS robiť online komprimáciu



File archivers ⇄	Windows ⇄	DOS ⇄	Mac OS X ⇄	Linux ⇄	BSD ⇄	Other Unix ⇄	AmigaOS ⇄	Android ⇄	Windows Mobile ⇄	Windows Phone ⇄
7-Zip	Yes ^[1]	Command-line interface ^[2]	Yes ^[2]	Yes ^[2]	Command-line interface ^[2]	Command-line interface ^[2]	Command-line interface	No	Yes	No
ALZip	Yes	Command-line interface ^[17]	No	No	No	No	No	No	No	No
ALZip for Mac	No	No	Yes	No	No	No	No	No	No	No
Archive Manager	No	No	No	Yes	Yes	Yes	No	No	No	No
Archive Utility	No	No	Yes	No	No	No	No	No	No	No
Ark	No	No	No	Yes	Yes	Yes	No	No	No	No
B1 Free Archiver	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	No
BetterZip	No	No	Yes	No	No	No	No	No	No	No
bsdtar/libarchive	Yes	No	Yes, bundled	Yes	Yes, bundled	Yes	Partial	Unknown	Unknown	No
Commander One	No	No	Yes	No	No	No	No	No	No	No
CRAX Commander	No	No	Yes	No	No	No	No	No	No	No
Disk ARchiver	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
Filzip	Yes	No	No	No	No	No	No	No	No	No
FreeArc	Yes	No	No	Yes	Yes	Yes	No	No	No	No
GNU tar	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No
iArchiver	No	No	Yes	No	No	No	No	Unknown	Unknown	?
Info-ZIP	Yes	Yes	Yes	Yes	Yes	Yes ^[9]	Yes	No	Yes ^[8]	No
KGB Archiver	Yes ^[7]	Partial	Command-line interface	Yes	No	No	No	Yes	Unknown	?
PeaZip	Yes ^[1]	Unknown	Under Development ^[18]	Yes	Yes	No	No	Unknown	Unknown	?
PKZIP	Yes	Yes	No	Yes	No	Yes	Yes	Unknown	Unknown	?
PowerArchiver	Yes	Unknown	Command-line interface	No	No	No	No	Unknown	Unknown	?
Star	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Unknown	?
Stuftit	Yes	Unknown	Yes	X86 only, older version (5.2.0?)	No	older version (5.2.0?)	No	Unknown	Unknown	?
The Unarchiver	Command-line interface	No	Yes	Command-line interface	No	No	No	Unknown	No	?
TUGZip	Yes	Unknown	No	No	No	No	No	Unknown	Unknown	?
WinAce	Yes	Command-line interface ^[5]	Command-line interface expander only ^[5]	X86, Command-line interface expander only ^[5]	No	No	Command-line interface expander only	Unknown	Unknown	?
WinRAR and RAR	Yes ^[1]	Yes	Command-line interface ^[6]	X86 and x86-64, Command-line interface ^[6]	Command-line interface ^[6]	Command-line interface ^[6]	Command-line interface expander only	Yes ^[19]	Yes	?
WinZip	Yes ^[1]	Command-line interface	Yes	No	No	No	Yes	Yes	Unknown	?

Komprimáčné programy