

---

Lipanj 2005.

# **WHITE PAPER**

Kompresija kod modema  
Brži prijenos podataka modemom

Ana Stanišić

0036394109

---

## Sadržaj

Uvod.....	3
V.42bis.....	4
V.44.....	5
Usporedba V.42bis i V.44.....	5
Usporedba modema istih karakteristika.....	10
Je li mogući veći stupanj kompresije?.....	11
Sažetak.....	12

## Uvod

Iako je upotreba broadband tehnologija kao DSL ili Kabel (kabelska) sve raširenija, većina ljudi još uvijek imaju pristup samo analognim telefonskim linijama. Zato su dial-up modemi još uvijek osnovno sredstvo pristupa internetu. Brzina pristupa internetu ne ovisi samo o modemu. Značajnu ulogu pri brzini prijenosa podataka modemom ima kompresija podataka. Modem koji šalje podatke, ako je moguće, sažima podatke prije slanja. Podatci se u sažetom obliku šalju telefonskom žicom do drugog modema. Modem koji prima podatke prvo vraća podatke u nesažeti oblik. Nesažete podatke modem pošalje računalu. Stupanj kompresije ovisi o podatku. Različiti podatci se različito sažimaju. Kompresija smanjuje veličinu podatka od za svega nekoliko posto do čak nekoliko puta. Zato je dobro znati kako kompresija funkcionira za određeni modem.

## V.42bis

Najrašireniji standard za kompresiju je ITU (International Telecommunication Union) standard V.42bis. Najveća brzina prijenosa V.42bis je 34 000 bps. Najrašireniji ne znači i optimalan. Najveći stupanj kompresije je 4:1. Iako je V.42bis koristan postoji mnogo prostora za poboljšanja.

V.42bis koristi Lempel-Ziv (LZ) algoritam za kompresiju. Animacija LZ algoritma nalazi se na web-stranici <http://www.data-compression.com/lempelziv.html>.

Učinkovitost kompresije kod V.42bis ovisi o veličini rječnika kompresije i duljini linije. Modem sažima samo po jednu liniju podatka. Svaka linija podatka podijeli se u sekvence koje se zamjenjuju sa kraćim kodnim riječima. Ako u podatku ima veliki broj ponavljajućih sekvenci, kompresija koja koristi duže linije je učinkovitija. Ako nema ponavljajućih sekvenci upotrebom kratkih kodnih riječi povećava se kompresija. Skup svih linija i odgovarajućih nizova kodnih riječi čine rječnik kompresije. Veličina rječnika se izražava u linijama. Struktura rječnika je dinamična. Rječnik je na početku prijenosa prazan. Puni se tokom prijenosa. Prvo se puni svim simbolima koje sadrži linija (sekvencama duljine jedan) kojima pridružuje kodne riječi. Rječnik je tada inicijaliziran. Počinje kompresija linije. Rječnik traži najdulju sekvencu koju sadrži za zamjenu. Zamjenjuje sekvencu kodnom riječi. Upisuje sekvencu koju je zamijenio produženu za sljedeći simbol u liniji i pridružuje joj odgovarajuću kodnu riječ. Proces ponavlja do kraja linije. Pun rječnik se mijenja sa prijenosom podataka ovisno o najčešćim sekvencama. Sekvence koje se najrjeđe pojavljuju zamjenjuju se novima. Veći rječnik sadržava više linija. V.42bis je neučinkovit kod kompresije malih količina podataka npr. e-mail poruka. Rječnik može biti "otrovan" promjenom tipa podatka npr. slika unutar tekstualne datoteke. Fizički parametar koji ograničava V.42bis je brzina RAM-a.

Kompresija ima i prednosti i mane. Pri slanju podataka koji su već sažeti kompresija kod modema će povećati veličinu podatka. Da bi se taj problem izbjegao V.42bis provjerava da li je podatak sažet. Ako je podatak sažet prelazi u

transparentan način rada. Prije prelaska u transparentni način rada modem šalje odgovarajuću kodnu riječ, time obavještava modem koji prima podatke o prijelazu. U transparentnom načinu rada modem šalje nepromijenjene podatke, ali ne prestaje provjeravati da li su podatci sažeti. Ako podatci više nisu sažeti modem izlazi iz transparentnog načina rada. Kao i prije prvo šalje odgovarajuću kodnu riječ. Česti prijelazi iz jednog u drugi način rada smanjuju učinkovitost. Kriterije promjene načina rada određuje proizvođač i mogu biti različiti za različite proizvođače. Prijenos podataka ovisi o učinkovitosti algoritma za promjenu načina rada.

## V.44

ITU standard V.44 koristi Lempel-Ziv-Jeff-Heath (LZJH) algoritam za kompresiju. V.44 je baziran na V.42bis. Standardi su jako slični. Glavna razlika je u stupnju kompresije. LZJH sadrži poboljšani algoritam pridruživanja liniji niz kodnih riječi. Kompresija kod V.44 je znatno učinkovitija nego kod V.42bis, u prosjeku do 25%. Najveći stupanj kompresije je 6:1. Brojke u različitim izvorima variraju od 20% pa sve do 200%.

## Usporedba V.42bis i V.44

Usporedbu V.42bis i V.44 provest ćemo na različitim modelima modema i sa različitim tipovima podataka. Modeli modema koje testiramo su:

- USR Curier 56k (3453 model) sa rječnik/linija veličine 2048/32
- USR Curier 56k (3453 model) sa rječnik/linija veličine 4093/32
- ZyXEL Omni Pro sa rječnik/linija veličine 1024/13
- ZyXEL U-336E sa rječnik/linija veličine 2048/32
- LT Win modem (Genius GM56PCI-L) sa rječnik/linija veličine 2048/32
- LT Win modem (Genius GM56PCI-L) V.44
- USR 56k Faxmodem PCI (OEM-2977) sa rječnik/linija veličine 2048/32.

U testu koristit ćemo pet različitih tipova podataka:

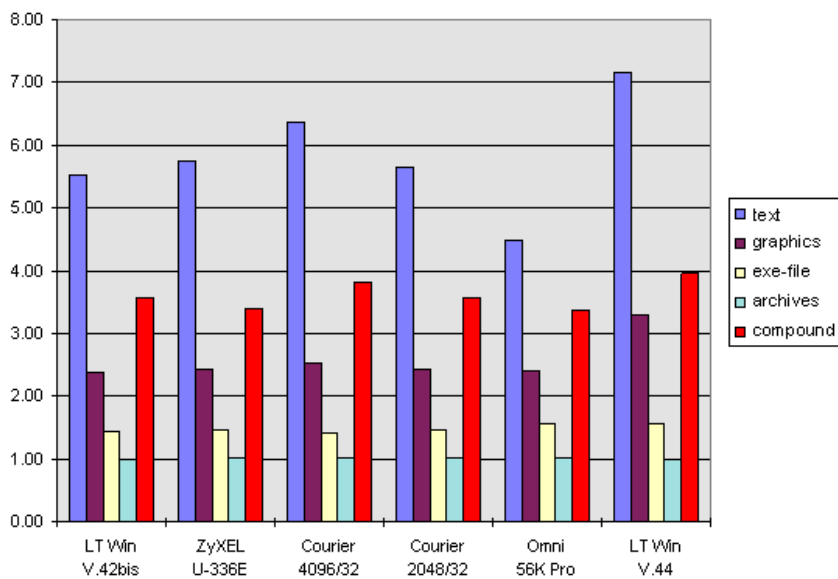
1. text file
2. graphics file (nesažet)
3. exe-file
4. archive file (nesažet)
5. compound file (sažet) koji sadrži fragmente podataka tipa 1, 2 i 4 ,

podatci se nalaze <ftp://ftp.mtu.ru/pub/horgi/test/>. Za svaki modem i svaki tip podatka test je ponovljen 10 puta. Da isključimo utjecaj greške pri prijenosu izbačeni su rezultati sa greškom pri prijenosu i pri slanju i pri primanju.

Relativna učinkovitost kompresije dobivena je omjerom količine nesažetih podataka i količine podataka nakon kompresije modema.

	text	graphics	exe-file	archives	compound
Without compression	1.00	1.00	1.00	1.00	1.00
LT Win V.42bis	5.53	2.39	1.43	1.00	3.56
ZyXEL U-336E	5.76	2.44	1.47	1.02	3.39
Courier 4096/32	6.36	2.52	1.41	1.02	3.82
Courier 2048/32	5.65	2.44	1.47	1.02	3.56
Omni 56K Pro	4.48	2.40	1.56	1.02	3.37
LT Win V.44	7.16	3.29	1.57	1.00	3.97

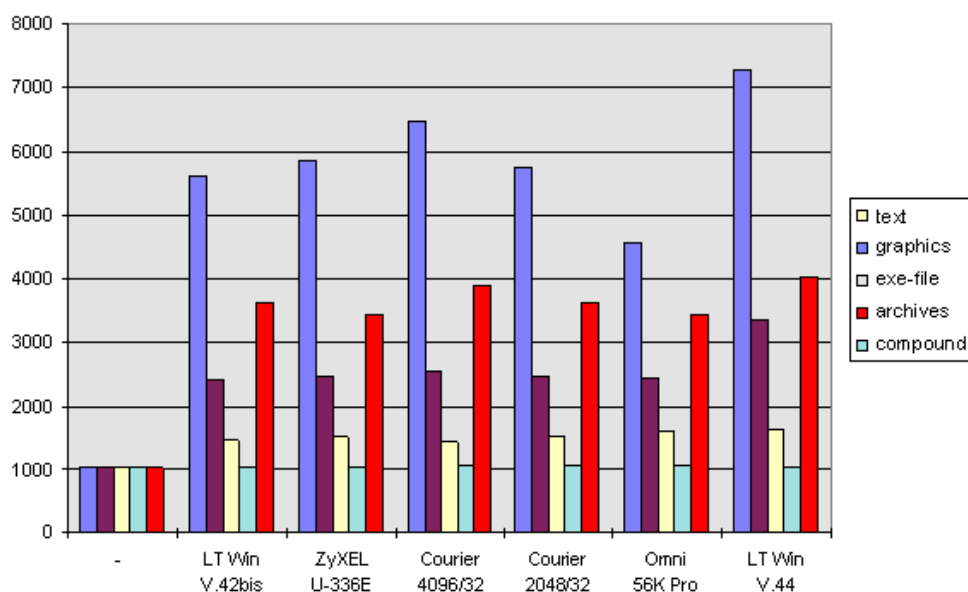
Tablica 1. relativna učinkovitost kompresije  
(relative compression efficiency)



Slika 1. relativna učinkovitost kompresije - rezultati tablice 1. prikazani grafički

	text	graphics	exe-file	archives	compound
Without compression	1017	1018	1019	1016	1016
LT Win V.42bis	5617	2427	1456	1016	3616
ZyXEL U-336E	5851	2482	1501	1032	3449
Courier 4096/32	6467	2560	1435	1040	3884
Courier 2048/32	5749	2482	1501	1040	3616
Omni 56K Pro	4551	2445	1586	1040	3427
LT Win V.44	7282	3344	1598	1016	4033

Tablica 2. apsolutna stopa učinkovitosti primljenih podataka, CPS  
(absolute effective data receive rate, CPS)



Slika 2. apsolutna stopa učinkovitosti primljenih podataka, CPS - rezultati tablice 2. prikazani grafički

Rezultati kompresije V.44 su umjereni. Bolji je od ostalih, ali ne za 25%. Usporedbom standarda V.42bis i V.44 na istom modemu LT Win (GM56PCI-L) vidimo da se učinkovitost kompresije kreće od 12% (tip podatka 5, compound file) do 38% (tip podatka 2, graphics file). Tip



podatka 5 je tipičan za web stranice, dok je tip podatka 2 netipičan za web. Kada usporedimo "maksimum" sheme V.42bis standarda Courier 4096/32 i V.44 LT Win razlika je manja: 31% za tip podataka 2 (graphics file), oko 13% za tip podataka 1 (text file) i manje od 4% za tip podataka 5 (compound file).

Razlike kompresija različitih shema V.42bis od "minimuma" (1024/16) do "maksimum" (4096/32) su veće: 42% za tip podatka 1 (text file) i 13% za tip podatka 5 (compound file). Razlike su manje za tip podataka 3 (exe-files) 9% i samo 5% za tip podataka 2 (graphics file). Od ZyXEL modela Omni 56k Pro je stariji model od U-336E, ali unatoč duplo većem rječniku i liniji U-336E je samo bolji u tip podatka 1 (text file).

## Usporedba modema istih karakteristika

	LT Win	USR 2977	Courier	ZyXEL U- 336E	Omni 56K Pro
CPS	1312	1292	1318	1305	1315
number of switching of modes	300	346	376	50	70
average compression coefficient	1.2	1.2	1.21	1.19	1.2

Tablica 4. V.42bis, veličina rječnika 1024 i veličina linije 16

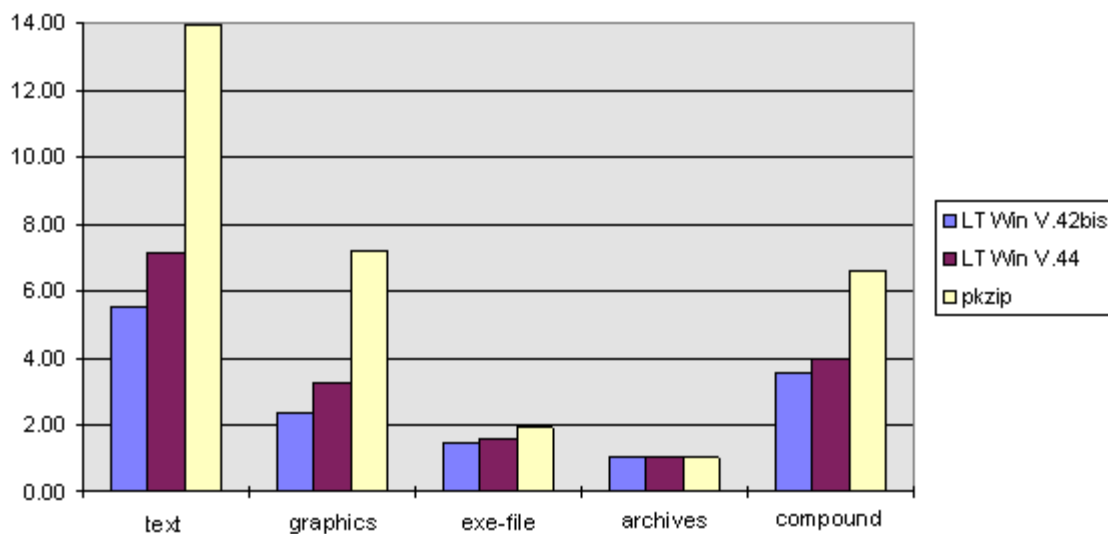
Svi modemi u tablici 4. su istog standarda V.42bis, imaju istu veličinu rječnika i linije (1024/16). Modemi u tablici 4. označeni istom bojom su od istog proizvođača. Unatoč velikoj razlici u algoritmu za promjenu načina rada, razlika prosječnog koeficijenta kompresije (average compression coefficient) je samo 2%. Iznenađujuće je ta razlika prisutna kod modema istog proizvođača.

## Je li mogući veći stupanj kompresije?

Veći stupanj kompresije je mogući ako ne koristimo kompresiju kod modema (npr. V.42bis ili V.44), već koristimo **klasičnu kompresiju** (npr. pkzip).

	text	graphics	exe-file	archives	compound
LT Win V.42bis	5.53	2.39	1.43	1.00	3.56
LT Win V.44	7.16	3.29	1.57	1.00	3.97
pkzip	13.92	7.20	1.88	1.00	6.62

Tablica 3. usporedba V.42bis, V.44 i pkzip-a



Slika 3. usporedba V.42bis, V.44 i pkzip-a - rezultati tablice 3. prikazani grafički

Iz tablice 3. i slike 3. jasno je vidljivo da je pkzip učinkovitiji od bilo koje kompresije modema. Razlika između pkzip-a i V.44 je veća nego između V.42bis-a i V.44

Osnovni koncept slanja podataka modemom se ne mijenja sa novom tehnologijom: svi podaci se prije slanja moraju sažeti. **Klasičnim kompresorom** bolje sažimamo exe-fileove. Kod slanja graphics fileova preporučljivo je slati formate bez ekstenzije gif, jpg i sl.

## Sažetak

Povećanje brzine prijenosa podataka modemom korištenjem novog V.44 standarda ne podudaraju se sa podacima navedenim od proizvođača. Najveća odstupanja su za tipove podataka 1 (text file) i 5 (compound file), a upravo su od takvih tipova podataka web stranice najčešće napravljene. Ali standard V.44 je uvjerljivo bolji od V.42bis. Izjava često dana od proizvođača "algoritam za kompresiju standarda V.44 daje bolju kompresiju podataka do 6:1 u odnosu na algoritam za kompresiju standarda V.42bis koji ima najveću kompresiju 4:1" je netočna. Kod V.42bis standarda korištenje modema sa većim rječnikom i linijom je isplativo u većini slučajeva (npr. USR Courier).

Kompresija kod modema je manja nego kod tradicionalna klasična kompresija. Unatoč tome je jako korisna. Ostavljeno je još mnogo mjesta za poboljšanja. Velika je vjerojatnost da će u budućnosti biti još efikasnije sheme V.44.