

## 2. kapitola – správy

### Jednotky

Neistotu vzbudila otázka, v akých jednotkách sú vlastne dostupné údaje – či ide o kilobity alebo kilobajty alebo ešte niečo iné. Pre veci neznalých – v počítačoch sa zvykne všetko prekódovať na nuly a jednotky. Jeden bit (značka je [b]) je práve jedna nula alebo jednotka. Keď sa dá takýchto núl alebo jednotiek veľa seba osem, je to jeden bajt (po anglicky byte, značka [B]). Do ôsmich bitov – teda do jedného bajtu – viete zakódovať číslo od 0 do 255 a do toho sa dá už celkom dobre schovať abeceda. Keď vám hovoria, aký veľký disk vám idú predať, údaj je väčšinou v gigabajtoch alebo terabajtoch, lebo bajt je najmenšia jednotka, ktorá sa dá na disk zapísať. Keď vám poskytovateľ pripojenia hovorí, že akú rýchlosť pripojenia vám ide dať, údaje väčšinou udáva v kilobitoch prípadne megabitoch za sekundu, pretože to číslo je osemkrát väčšie, ako ten istý údaj v kilobajtoch či megabajtoch za sekundu, vyzerá to lepšie a lepšie sa to predáva. Preto si nejakí ľudia najprv mysleli, že keď sa bavíme o prenosovej rýchlosti, údaj je v kilobitoch za sekundu a chceli to prepočítavať na bajty. Nie je to ale tak, program, ktorý som použil, pracuje s bajtami.

Druhý problém je, že ani kilobajty za sekundu [kB/s] nie sú presne tá jednotka, ktorú graf používa. Namiesto toho používa kibibajty za sekundu [KiB/s]. Kilobajt je tisíc bajtov tak, ako sme navyknutí z metrickej sústavy. Keďže informatici pracujú s nulami a jednotkami, tisíc pre nich ale často nie je ten optimálny násobok, s ktorým by chceli pracovať. Neďaleko čísla 1000 sa však nachádza číslo 1024, čo je  $2^{10}$  (v sústave núl a jednotiek – teda v dvojkovej sústave – sa zapisuje ako 10 000 000 000, čo je krásne okrúhle číslo), ktoré informatikom vyhovuje niekedy viac. Kibibajt je teda 1024 bajtov. Informatikom sa to hodí a ak to náhodou niekto pochopí ako kilobajt, až tak veľmi sa nepomýli. Ale je dobré mať na pamäti, že náš výsledok bude v kibibajtoch, lebo údaje o veľkosti tých súborov na webe viedenskej univerzity sú v kilobajtoch a ak si chce človek porovnať svoj výsledok so skutočnou veľkosťou, treba si to prepočítať.

Niekoľko sa ma pýtal, či označenie jednotky naozaj má byť KiB a nie kiB. Tvrdil som, že možno to druhé, ale že ako program napísal, tak som urobil snímku obrazovky. Overoval som to ale na wikipedii a zistil som, že to majú dobre a kibibajt má naozaj značku KiB. Podobne mebibajt, čo je  $1024^2$  teda 1 048 576 teda 100 000 000 000 000 000 000<sub>(2)</sub> bajtov<sup>1</sup> má značku MiB. Neplieť si s MB, čo je  $10^6$  B ani s mužmi v čiernom.<sup>2</sup>

### Úloha 1

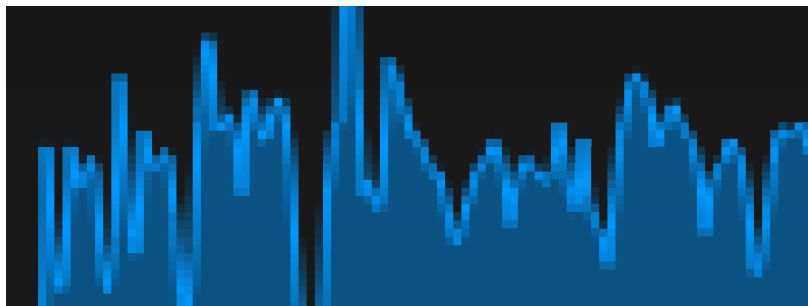
Pri tejto úlohe bolo treba graf merať, aby sa dalo zistiť, v akých časoch dochádzalo v sieťovej prevádzke k zásadným zmenám. Vzhľadom na katastrofálny nedostatok pravítok v triede sa siahalo po rôznych improvizovaných nástrojoch, zvlášť populárna sa ukázala byť jednotka jeden Kelbel, používaná Dávidom a Maťom. (Išlo o vzdialenosť dvoch nemeckých slovíčok na pracovnom liste, jednotka bola nazvaná na počesť profesorky nemčiny.) Na každom úseku sa odhadla priemerná rýchlosť sťahovania, vynásobila sa trvaním úseku a získané údaje sa sčítali. Aj napriek improvizovaným prostriedkom boli dosiahnuté relatívne dobré výsledky. Bolo prekvapivé, že všetky výsledky, ktoré boli dosiahnuté touto metódou, boli väčšie, než skutočná veľkosť súboru. Celý súbor má 157 MB a dohady boli väčšinou asi o 10 MB väčšie. Úplne najbližšie z tejto skupiny sa dostal Rišo, ktorému vyšlo 158,6 MB.

1 Zápís  $1010_{(2)}$  znamená, že číslo je zapísané v dvojkovej sústave. Teda že je tam 1 osmička ( $2^3$ ), 0 štvoriek ( $2^2$ ), 1 dvojka ( $2^1$ ) a 0 jednotiek ( $2^0$ ). Teda po našom  $8+2$  čiže 10.

2 Men in Black je klasická sci-fi akčná komédia. [http://meninblack.wikia.com/wiki/Men\\_in\\_Black\\_Series](http://meninblack.wikia.com/wiki/Men_in_Black_Series)

Sú ľudia (Konkrétne jednak Kubo, jednak Maťo s Bashou, ktorá bola na návšteve. Peťo tiež sľuboval, ale zatiaľ nič nedodal.), ktorým sa takéto odhadovanie zdalo málo presné, stiahli si pôvodný obrázok s grafom a povedali si, že sa do grafu pustia stĺpec po stĺpci, pixel po pixeli. Keďže údaje na grafe pribúdajú rýchlosťou 1 pixel za sekundu, stačí spočítať výšky jednotlivých stĺpcov prepočítané na kibibajty a máme presné údaje, koľko sme stiahli.

Kubo použil Photoshop, graf orezal vpravo, vľavo aj zhora (zhora to orezal na rýchlosti 450 KiB/s), zlikvidoval vodorovné čiary na pozadí a prehnal to farebným filtrom tak, aby bolo pozadie čierne a celý zvyšok obrázka – čiže graf – biely. Potom si nechal spočítať, koľko má obrázok bielych pixelov a koľko všetkých a v rovnakom pomere rozdelil údaj  $450 \text{ KiB/s} \cdot 724 \text{ s}$  – toľko by sa stiahlo, keby bol obrázok čisto biely. Výsledná veľkosť súboru mu vyšla 154 MB, čo je asi o 3 MB menej, než skutočnosť.



Obrázok 1: Detail grafu

Maťo s Bashou sa do toho pustili programátorsky. Nevšimli si, že majú k dispozícii originálny screenshot, spravili si screenshot z pdf-ka, čím prišli o informáciu, že 1 pixel na šírku je jedna sekunda, ale to nevadilo. Tiež najprv v GIMPe<sup>3</sup> upravili obrázok na dvojfarebný (čierne pozadie vs. zvyšok), vyexportovali vo formáte .ppm vhodnom na spracovanie a v programe, ktorý si napísali, spočítali pixely a dostali horný odhad. Potom pre každý stĺpec pridali do čiernej oblasti jeden pixel a dostali dolný odhad.<sup>4</sup> Vyšlo im to nejak rozumne, zabudli žiaľ zaznamenať, koľko. Tak som Maťu poprosil, aby to doma zrátal ešte raz a poslal mi výsledok. Maťo sa do toho pustil a odpísal, že to nevyšlo tak pekne, ako na hodine. Že to vychádza divne, dolný odhad že je 160 000 KiB a horný 173 000 KiB. (163,8 MB a 177,1 MB). Neskôr výsledky trochu upravil. Ja som sa pustil do roboty tiež, orezal som v GIMPe graf a vymazal som z neho nepodstatné detaily, aby ostalo iba vnútro, vonkajšok a hranica. (Ak chcete robiť experimenty, upravený obrázok visí na stránke.) Dolný odhad určený modrou oblasťou, ktorá je zaručene celá vo vnútri grafu mi vyšiel 161 638 KiB, horný – teda všetko mimo čiernej oblasti – vyšiel 173 067 KiB (165,5 MB a 177,2 MB). Oba naše výsledky mali dve spoločné črty – boli relatívne blízko údajov, ktoré namerali ostatní ľudia priamo z papiera a správna hodnota 157 MB bola o dosť menšia, než naše dolné odhady. Chvíľu sme s Maťom špekulovali, čím by to mohlo byť, ale potom sme to nechali tak.

A potom som na radu môjho syna Tomiho nainštaloval Whireshark<sup>5</sup>. To je program, ktorý vie zobrazíť, čo sa deje pri sieťovej komunikácii do úplných detailov. Dal som znovu sťahovať Almagest a pozrel som sa, ako vyzerajú jednotlivé pakety, ktoré mi do počítača chodia. A dozvedel som sa toto:

Každý paket, ktorý dorazil z Viedenskej univerzity s kúskom Almagestu do môjho počítača, mal 1514 bajtov. Obsahoval jednak hlavičky, v ktorých boli údaje o tom, odkiaľ a kam ide, kedy bol poslaný, aký je dlhý atď. Okrem hlavičiek obsahoval samotné dáta. Tých bolo v každom pakete 1448 bajtov. Ak teda chcem stiahnuť súbor, ktorý má 157 MB, reálne musím kvôli tým hlavičkám stiahnuť toho viac. Konkrétne  $157 / 1448 \cdot 1514 = 164,2 \text{ MB}$ . Okrem toho dorazia niektoré pakety

<sup>3</sup> Voľne prístupný softvér na prácu s obrázkami <http://www.gimp.org>

<sup>4</sup> Keď sa pozriete na obrázok, pridať jeden pixel nemusí stačiť, pretože tá hranica je niekedy vyššia.

<sup>5</sup> Voľne prístupný na <https://www.wireshark.org>

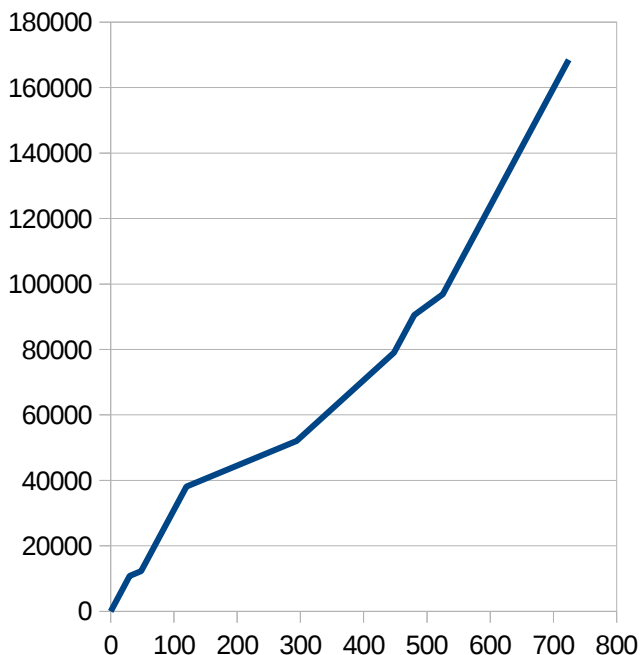
duplicitne, pretože potvrdenka od klienta neprišla včas a niektoré dorazia poškodené, takže si ich klient vyžiada znovu, čo ešte zvýši množstvo prenesených dát. V každom prípade ľudia, ktorým to vyšlo viac, než očakávaných 157 MB, mali naprostú pravdu.

## Úloha 6

Pôvab tejto úlohy je v tom, že jej riešenie v sebe skrýva riešenie všetkých ostatných úloh z tejto kapitoly. Pri konštrukcii grafu treba najprv zistiť, v akom čase došlo k zmenám v rýchlosti sieťovej prevádzky, potom si v týchto bodoch vypočítať, koľko sa toho stiahlo, naniesť na graf a pospájať. V ďalšom budú použité Dušanove dáta:

Trvanie	Rýchlosť	Koniec	Stiahnuté	Stiahnuté spolu
		<b>0</b>		<b>0</b>
30	360	<b>30</b>	10800	<b>10800</b>
18	80	<b>48</b>	1440	<b>12240</b>
72	360	<b>120</b>	25920	<b>38160</b>
174	80	<b>294</b>	13920	<b>52080</b>
154	175	<b>448</b>	26950	<b>79030</b>
32	360	<b>480</b>	11520	<b>90550</b>
45	140	<b>525</b>	6300	<b>96850</b>
199	360	<b>724</b>	71640	<b>168490</b>

V prvých dvoch stĺpcoch má Dušan údaje, ktoré nameral – koľko trvajú jednotlivé úseky a ako rýchlo sa súbor sťahuje. V treťom stĺpci si vypočítal, v akom čase jednotlivé úseky končia, v štvrtom koľko dát počas jednotlivých úsekov stiahol a v poslednom stĺpci sčítal čísla zo štvrtého, aby zistil, koľko toho bolo stiahnutého na konci jednotlivých úsekov. Z tretieho a piateho stĺpca nakoniec urobil graf:

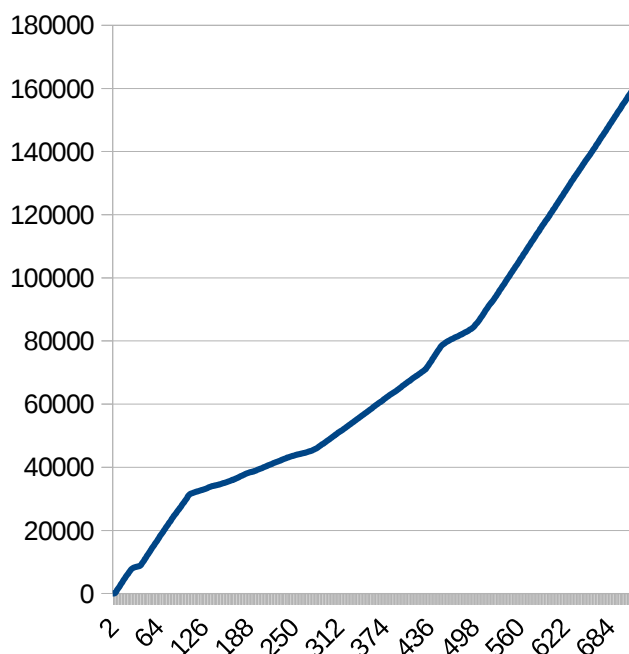


Graf stále stúpa (pretože zo súboru neubúda a sieť nepadla, takže vždy niečo pribúdalo). Čím sa sťahovalo rýchlejšie, tým je graf strmější. Tiež si môžete všimnúť, že sklon grafu na každom úseku, kde išla sieť naplno, je rovnaký.

Ak by sme pomocou tohto grafu chceli riešiť napríklad úlohu 5, teda koľko dát sa stiahlo medzi 430. a 500. sekundou, pozrieme sa, aké hodnoty má graf v čase 430 a v čase 500 a spravíme rozdiel.

Vedeli by ste v tomto grafe nájsť riešenie úlohy 7?

Na hodine sme riešili otázku, ako by graf vyzeral, keby sa pripočítavali prírastky po sekunde. Keď som počítal tú veľkosť, tak som si napísal program v jazyku Python a pri tej príležitosti som si nechal spočítať pixely v jednotlivých stĺpcoch a taký graf urobil.<sup>6</sup> Celkom sa na ten Dušanov podobá:



<sup>6</sup> Použil som dolné odhady – bral som teda iba tie pixely, ktoré sa zaručene nachádzali vo vnútri grafu.