

IZAZOVI ODRŽAVANJA VELIKE KOLIČINE GRAFIČKIH PODATAKA

CHALLENGES IN MAINTENANCE OF BIG DATA IN COMPUTER GRAPHICS

Samir Lemeš, v.prof.dr.
Univerzitet u Zenici, Politehnički fakultet
Zenica, Bosna i Hercegovina

REZIME

Pod pojmom "Big Data" podrazumijevaju se one količine podataka koji prevazilaze mogućnosti uobičajeno korištenog softvera za skladištenje, obradu i upravljanje podacima. Posebno ako se radi o grafičkim podacima, upravljanje stotinama terabajta i petabajtima podataka zahtijeva sasvim drugi pristup sa aspekta hardvera, komunikacija, sigurnosti podataka, i stavlja nove izazove pred osoblje koje se bavi informaciono-komunikacionim tehnologijama. U ovom članku dat je osobni pogled na te izazove kad se radi o grafičkim podacima.

Ključne riječi: Velike količine podataka, Računarska grafika, Održavanje

ABSTRACT

The term "Big Data" refers to quantities of data beyond the capabilities of commonly used software for the storage, processing and data management. Especially in the case of graphic data, managing hundreds of terabytes and petabytes of data requires a completely different approach in terms of hardware, communications, information security, and puts challenges in front of the staff involved in information and communication technologies. This article gives a personal view on these challenges when it comes to graphic data.

Key words: Big Data, Computer Graphics, Maintenance

1. UVOD

Pojam "Big Data" (velike količine podataka) predstavlja tehnologiju koja omogućava prikupljanje i obradu velikih količina strukturiranih i nestrukturiranih podataka u realnom vremenu. Jedna od definicija pojma "Big Data" se opisuje sa 3 slova V: **Volume** (velika količina podataka, reda veličine petabajta, koji se prikupljaju, obrađuju i stavljaju na raspolaganje za analizu), **Velocity** (kontinuirano prikupljanje velike količine podataka u realnom vremenu brzinom koja je veća od brzine obrade podataka), **Variety** (podaci su dostupni u različitim oblicima i izvorima: slike, satelitski snimci, atmosferski podaci, podaci s društvenih mreža, logovi sa servera, podaci prikupljeni senzorima, itd.). Pojam je relativno nov, a od svog nastanka, 2008. godine, stalno se šire područja primjene na mnoge oblasti.

Chen i Zhang su u [1] dali prikaz primjena, prilika i izazova koje "Big data" pruža, te pregled vrhunskih tehnika i tehnologija koje se prilagođavaju rješavanju "Big data" problema. Pokazali su kako bi ova tehnologija mogla poslužiti kao katalizator za razvoj mnogih drugih

tehnologija, kao što su granularno računarstvo (*Granular Computing*), računarstvo u oblaku (*Cloud Computing*), bioračunarstvo (*Bio-inspired Computing*), kvantni računari (*Quantum Computing*), od kojih se neke već koriste, dok su neke još uvijek u razvojnoj fazi.

Zheng, Liu i Hsieh su u [2] obradili primjer upotrebe "Big data" tehnologija za obradu ogromnih količina podataka o kvalitetu zraka, odnosno koncentracijama prašine PM2.5 koja predstavlja opasnost za ljudsko zdravlje. Koristili su niz tehnika za obradu i modeliranje podataka o zagađenju zraka u kineskim gradovima Beijing i Shanghai za koje je poznato da su među najzagađenijim na svijetu.

Grupa autora je u [3] primijenila statistički alat R za analizu velikih količina podataka u četiri sloja: infrastrukturni, virtualizacijski, sloj za procesiranje seta podataka i servisni sloj. Posebnu pažnju su posvetili grafičkom prikazu rezultata analize.

Barbierato, Gribaudo i Iacono su u [4] modelirali upotrebu računarstva u oblaku (*Cloud Computing*) za planiranje i upravljanje resursa unutar geografski distribuirane infrastrukture, sa aspekta planiranja i predviđanja održavanja tako distribuiranih resursa.

Lee, Holgado, Kao i Macchi su u [5] diskutovali nove perspektive za inovacije u održavanju i predložili putanje stvaranja vrijednosti za transformaciju održavanja korištenjem novih tehnologija kao što su *Internet of things* (IoT), *Cloud Computing*, *Big Data*, *PHM* i *cyber-fizičke* sisteme.

Byers i Woo su analizirali primjenu "Big data" tehnologija na obradu geoloških prostornih podataka [6]. Razmotrili su neke od prednosti prostorne grafike i korištenje velikih količina podataka u kontekstu 3D vizualizacije kroz studije slučaja, u kojima bi obilježja objekata, kao što su efikasnosti prikaza, dinamične izmjene, presjeci i dr. mogla poboljšati procese i radne tokove interpretacije podataka.

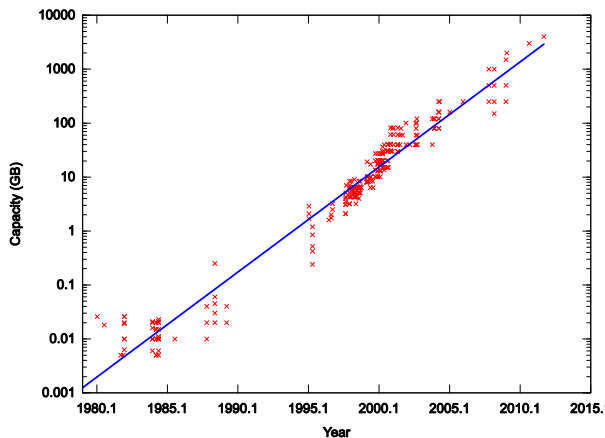
Wang, Aji i Vo su opisali kako se od prvobitne upotrebe "Big data" tehnologije za obradu medicinske računarske grafike došlo do primjene istih tehnologija na obradu podataka u geografskim informacionim sistemima (GIS) [7]. Okazali su kako se različite tehnologije, kao što su *Hadoop-GIS* i paralelni prostorni DBMS, različito ponašaju u smislu efikasnosti i skalabilnosti.

Simmhan je sa grupom autora u uvodniku [8] identifikovao "Big data" tehnologiju kao potencijalni alat za rješavanje vitalnih naučnih i društvenih problema 21. stoljeća, dok su se Yu, Mu, Lu i Ren u uvodniku [9] osvrnuli na sve aktuelniji problem sigurnosti informacionih sistema i privatnosti u velikim računarskim mrežama, koje čine osnovu infrastrukture "Big data" tehnologija.

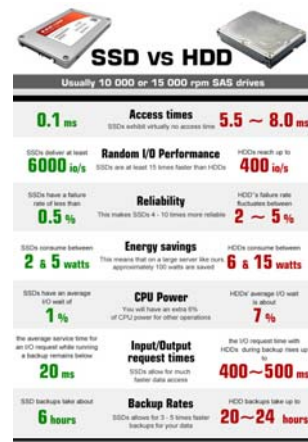
2. HARDVER ZA VELIKE KOLIČINE PODATAKA

Kapacitet hardvera za lokalno pohranjivanje podataka se mijenjao na način da se mora prikazati u logaritamskoj skali. Na slici 1 je prikazana promjena maksimalnog kapaciteta hard diskova, koji se kretao od 10 MB 1985. godine, preko 1 GB 1995. godine, zatim 100 GB 2005. godine do 8 TB 2015. godine. Samsung je za august 2016. najavio prvi disk kapaciteta 16 TB, što je do sada najveći kapacitet diska.

S druge strane, pored kapaciteta, vrlo važan aspekt postaje i brzina medija za trajno pohranjivanje podataka, tako da su sve češće računarske konfiguracije koje standardni HDD (Hard Disk Drive) kombinuju sa SSD (Solid State Drive) diskovima baziranih na Flash tehnologiji bez pokretnih mehaničkih dijelova, ili se računari isporučuju samo sa SSD diskom, a za pohranjivanje većih količina podataka koriste eksterne diskove ili *Cloud Computing*. Na slici 2 prikazana je razlika u brzini pristupa HDD i SSD medija. Kako nema jedinstvene mjerne jedinice za brzinu pristupa, prikazano je više parametara, a po svima je SSD superioran, osim po kapacitetu.



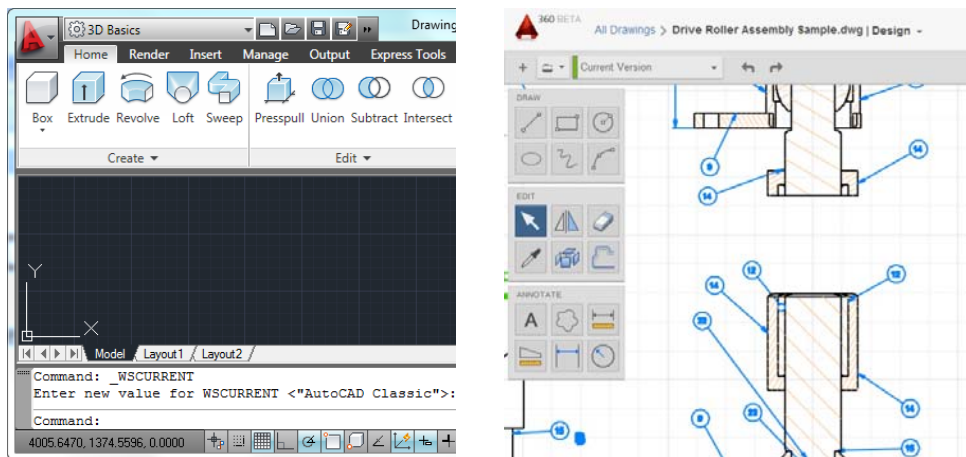
Slika 1. Kapacitet hard diskova tokom vremena [10]



Slika 2. Poređenje brzine pristupa podacima HDD i SSD [11]

S povećanjem brzine pristupa internetu, širokopolasni (*Broadband*) pristup je postao standard, čime su se stvorile mogućnosti za postepeni prelazak s lokalnog pohranjivanja podataka, pa čak i aplikacija, na mrežne medije za pohranjivanje. Tehnologija poznata kao "računarstvo u oblaku" (*Cloud Computing*), postaje svakodnevica i mijenja paradigmu obrade i pohranjivanja podataka. Njene osnovne prednosti su praktično neograničen kapacitet za pohranjivanje podataka, univerzalna pristupačnost podacima širom svijeta, te minimalno održavanje, kako hardvera tako i softvera. Osnovni nedostaci su ovisnost o internet pristupu i otežana zaštita sigurnosti podataka.

Prednosti računarstva u oblaku su davno prepoznali veliki proizvođači softvera, tako da većina njih nudi, uz klasične instalacije na lokalnom računaru, i mogućnost kupovine licenci za pristup kroz *browser*. Dakle, aplikacija i podaci se nalaze u računarskom oblaku, odnosno distribuiranom sistemu servera i podatkovnih centara, a interfejs se ostvaruje putem *browsera* ili kao aplikacija za mobilne uređaje (pametne telefone ili tablete). Na slici 3 prikazan je primjer interfejsa desktop verzije AutoCAD-a i A360, koji ne treba instalirati na lokalni računar, nego se svi programi i podaci nalaze na internetu.

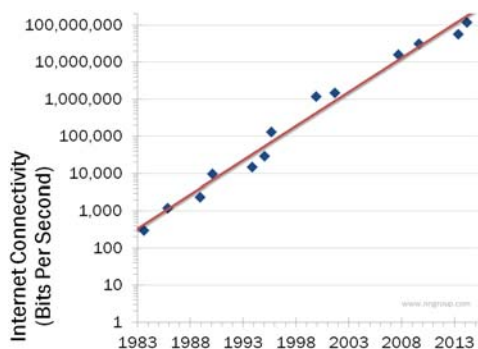


Slika 3. Poređenje AutoCAD interfejsa desktop i 360 verzije koja se koristi kroz browser

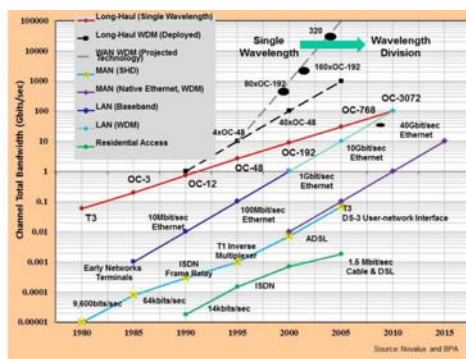
Osnovna prednost računarstva u oblaku za proizvođače softvera se razlikuje od one za krajnje korisnike. Aplikacije koje se pokreću direktno i isključivo sa njihovih servera se ne mogu piratizirati, odnosno nije moguć pristup takvim aplikacijama bez validne licence. To onda mijenja i način licenciranja, tako da se više ne kupuju trajne licence za softver koji se instalira na lokalni disk korisnika, nego se plaća najam softvera, odnosno vremenski ograničeno pravo pristupa. Vlasnik autorskih prava na softver se tako zaštitio od neovlaštene upotrebe intelektualnog vlasništva, ali i vezao korisnike da redovno plaćaju upotrebu softvera, a koja u sebi uključuje i održavanje. Pod održavanjem ovdje se podrazumijeva prvenstveno praćenje uočenih grešaka i njihovo popravljavanje, te automatska nadgradnja na nove verzije.

3. PRENOS VELIKIH KOLIČINA PODATAKA

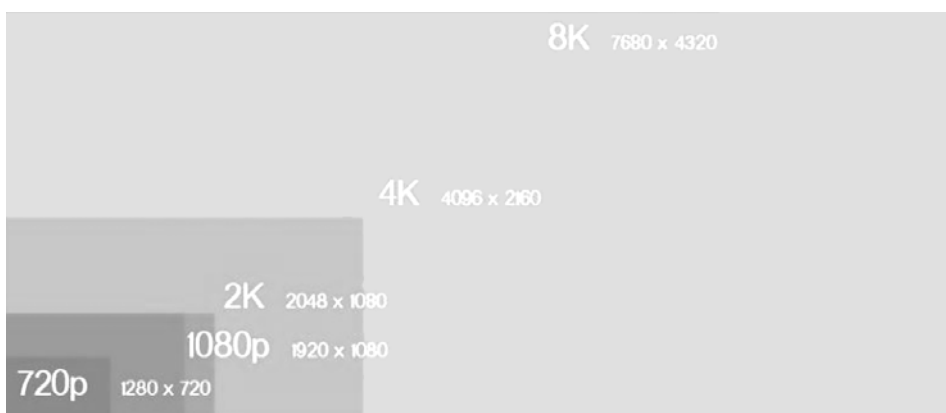
Da bi računarstvo u oblaku, bez kojeg je "Big Data" tehnologija nezamisliva, bilo moguće, veoma je važna propusnost komunikacijskih mreža, koja se također značajno povećala posljednjih godina. Na slici 4 prikazana je promjena standardne propusnosti, odnosno brzine pristupa internetu, također u logartamskoj skali, jer se povećava za 50% svake godine. Treba napomenuti da se ovdje radi o maksimalnoj komercijalno dostupnoj brzini pristupa koju nude internet provajderi u svijetu. Situacija u BiH je nešto drukčija, jer je relativno mali broj korisnika i malo tržište ograničavajući faktor za domaće internet provajdere. Na slici 5 prikazana je promjena brzine pristupa internetu za različite tehnologije. Sa slike se vidi da se brzina udvostručila svakih 5 godina od 1990 do danas.



Slika 4. Brzina pristupa internetu [12]



Slika 5. Brzina pristupa internetu [13]



Slika 6. Standardne rezolucije videa dostupne 2016. godine

Što se tiče grafičkih podataka, posebno videa, zanimljivo je posmatrati i promjenu rezolucije digitalnog videa. Na slici 6 prikazane su različite standardne rezolucije videa koje su do sada bile u upotrebi. Prenos ogromnih količina podataka koje iziskuje stalno povećanje rezolucije slike, zahtijeva razvoj novih algoritama za kompresiju podataka.

Zanimljiv pristup se koristi u prikazu velikih grafičkih datoteka, koje se u postupku zumiranja prikazuju sa sve sitnijom rezolucijom. Za to se koristi poznati *Quadtree* algoritam, po kojem se slike rekurzivno dijele na manje segmente, a uz kolekciju slika visoke rezolucije pohranjuju se i slike krupnije rezolucije, kako bi se izbjegao prenos nepotrebnih količina podataka. Na taj način funkcioniše postupak zumiranja u GIS mapama ili sličnim aplikacijama za prikaz terena kao što je *Google Earth*.

4. SIGURNOST INFORMACIJA

Standard ISO 27000:2016 definiše sigurnost informacija kao "očuvanje povjerljivosti, integriteta i dostupnosti informacija" [14]. Povjerljivost po istom standardu "obežbuje da informacija nije dostupna niti se otkriva neovlaštenim pojedincima, entitetima ili procesima". Integritet je "preciznost i kompletnost informacija", dok se pod dostupnošću smatra da su "podaci dostupni i upotrebljivi na zahtjev ovlaštenog entiteta" [14]. U nastavku je dat osvrt na ta tri aspekta sigurnosti velike količine grafičkih podataka.

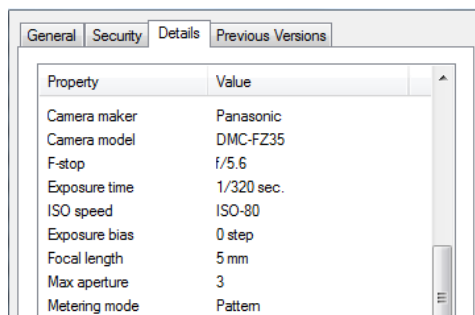
4.1. Povjerljivost velike količine grafičkih podataka

Povjerljivost grafičkih datoteka se ne razlikuje od povjerljivosti drugih vrsta dokumenata, kao što su tekstualni dokumenti, proračunske tablice (*spreadsheet*), web stranice i sl. Najvažniji aspekt povjerljivosti je u stvari zaštita intelektualnog vlasništva, koja se štiti na različite načine. Jedan od najčešćih načina je takozvani vodeni žig (*watermark*), koji se ugradi u sliku, tako da se svaka neovlaštena upotreba lako detektuje. Na slici 7 prikazan je primjer slike s vodenim žigom. Problem s takvom zaštitom je što se originalna slika trajno naruši i izmijeni, čime je njena upotrebljivost ograničena.

Drugi načini zaštite predstavljaju trikove koji ne zaštite sliku u potpunosti, nego samo ograniče pristup informatički manje vještim korisnicima. Takvi primjeri uključuju prekrivanje slike providnim slojem u HTML kodu, rezanjem slike na sitne dijelove i njihov prikaz u HTML tabeli, onemogućavanje desnog klika na web stranici ili pretvaranje slike u *Flash* animaciju. Malo naprednija tehnika, ali skupa kad se radi o većem broju slika, je komercijalni filter *Digimarc* koji je standardno sastavni dio softvera *Adobe Photoshop*, a koji digitalno kodira vodeni žig koji je nevidljiv običnom korisniku, ali se može otkriti pomoću odgovarajućeg softvera. Ta tehnika povećava veličinu datoteke. Na slici 8 prikazani su metapodaci koji se također mogu koristiti za zaštitu intelektualnog vlasništva nad slikom i za otkrivanje više detalja o autoru slike, datumu i načinu nastanka, eventualnim izmjenama originalne slike i sl.



Slika 7. Vodeni žig [15]



Slika 8. Metapodaci u datoteci sa slikom

4.2. Integritet velike količine grafičkih podataka

Pod integritetom grafičke datoteke smatra se njena autentičnost, u smislu da nije došlo do naknadnih izmjena nakon nastanka slike. Ako se radi o rasterskim datotekama koje sadrže fotografije, najslikovitiji primjer predstavljaju fotomontaže. Slika 9 pokazuje primjer fotomontaže kojom se turistkinja željela pohvaliti da je posjetila destinaciju na kojoj nikad nije bila, a na slici 10 je primjer očigledne greške marketinške agencije koja je izmjenama fotografije uklonila drugu osobu s fotografije, ali je zaboravila ukloniti njenu ruku.



Slika 9. Lažna posjeta Kineskom zidu [16]



Slika 10. Photoshop greška na reklami za Snickers čokoladu [17]

Pored ovakvih, bezazlenih primjera, integritet grafičkih datoteka nekad može biti jako važan. Ako se radi o dokaznom materijalu u pravosudnim postupcima, dijagramima koji prikazuju važne podatke (na primjer, količine opasnih zagađujućih materija u zraku ili vodi), medicinskoj grafici kao što su snimci sa CT ili MR uređaja, ali i vojnoj ili industrijskoj tehničkoj dokumentaciji za realizaciju projekata [18], integritet podataka je od izuzetne važnosti. Upravo iz tih razloga, američki FBI je razvio smjernice najboljih praksi očuvanje integriteta digitalne slike i videa [19].

4.3. Dostupnost velike količine grafičkih podataka

Najveći izazov za računarstvo u oblaku predstavlja aspekt dostupnosti svih vrsta podataka, jer su svi podaci dostupni isključivo ako postoji internet konekcija. Svaki prekid te konekcije predstavlja smetnju za dostupnost informacija.

Pružaoци usluga smještaja podataka u oblaku imaju razrađene sisteme pravljenja rezervnih kopija podataka, redundancije servera i medija za pohranjivanje podataka, čime se obezbjeđuje maksimalna dostupnost podataka, ali i pored svega toga postoje moguće situacije kada podaci postaju nedostupni, kao što su hakerski napadi, napadi odbijanjem usluga (DDOS – *Distributed Denial of Service*), ili jednostavno problemi s internet konekcijom klijenta. Jedini način zaštite od takvih slučajeva je izrada rezervnih kopija podataka, s redovnim ažuriranjem, geografskom diverzifikacijom rezervnih kopija i drugim tehnikama, koje se koriste u zavisnosti od važnosti i vrijednosti podataka koji se čuvaju.

Grafički podaci su u ovom smislu posebno ranjivi, jer obično zauzimaju veliki kapacitet memorije, posebno ako se radi o rasterskim slikama visoke rezolucije, koje su nepogodne za algoritme kompresije bez gubitka podataka (*Lossless Compression*), ili ako se radi o velikom broju grafičkih datoteka koje se moraju kreirati u vremenskim intervalima i čuvati određeno vrijeme, kao što su to snimci sa nadzornih kamera i slični grafički sadržaji.

5. ZAKLJUČAK

Kako se informaciono-komunikaciona tehnologija razvija velikom brzinom, tehnike zaštite podataka predstavljaju sve važniji aspekt održavanja velike količine grafičkih podataka. Razvijeni su brojni standardi, smjernice, tehnike i algoritmi koji rješavaju probleme i izazove zaštite podataka, ali se isto tako razvijaju i nove prijetnje po sigurnost informacija.

Sve to pokazuje da su informaciono-komunikacione tehnologije jako dinamična aktivnost, koja zahtijeva stalno usavršavanje, praćenje najnovijih trendova, unapređenje softvera i hardvera, ali i zahtijeva sistematski pristup održavanju sve većih količina podataka, čija količina eksponencijalno raste.

6. LITERATURA

- [1] Chen, C. P., Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, pp. 314-347. doi: 10.1016/j.ins.2014.01.015
- [2] Zheng, Y., Liu, F., Hsieh, H. P. (2013). U-Air: When urban air quality inference meets big data. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1436-1444). ACM.
- [3] Ye, F., Wang, Z., Zhou, F., Wang, Y., Zhou, Y. (2013). Cloud-Based Big Data Mining & Analyzing Services Platform Integrating R. *Advanced Cloud and Big Data (CBD)*, 2013 International Conference on (pp. 147-151). IEEE.
- [4] Barbierato, E., Gribaudo, M., Iacono, M. (2016). Modeling and evaluating the effects of Big Data storage resource allocation in global scale cloud architectures. *International Journal of Data Warehousing and Mining (IJDWM)*, 12(2), 1-20.
- [5] Lee, J., Holgado, M., Kao, H. A., Macchi, M. (2014). New thinking paradigm for maintenance innovation design. *World Congress* (Vol. 19, No. 1, pp. 7104-7109).
- [6] Byers, C., Woo, A. (2015). 3D data visualization: The advantages of volume graphics and big data to support geologic interpretation. *Interpretation*, 3(3), SX29-SX39.
- [7] Wang, F., Aji, A., Vo, H. (2015). High performance spatial queries for spatial big data: from medical imaging to GIS. *SIGSPATIAL Special*, 6(3), 11-18.
- [8] Simmhan, Y., Ramakrishnan, L., Antoniu, G., Goble, C. (2015). Cloud computing for data-driven science and engineering. *Concurrency and Computation: Practice and Experience*.
- [9] Yu, Y., Mu, Y., Lu, R., Ren, J. (2016) Recent advances in security and privacy in large-scale networks. *Concurrency Computat.: Pract. Exper.*, 28: 1080–1082. doi: 10.1002/cpe.3775.
- [10] <https://opencliptart.org/detail/182708/hard-drive-capacity-over-time> (23.3.2016)
- [11] <http://technofaq.org/posts/2014/11/ssd-hosting-or-hdd-hosting-which-one-to-choose/> (23.3.2016)
- [12] <https://www.nngroup.com/articles/law-of-bandwidth/> (23.3.2016)
- [13] <http://ipcarrier.blogspot.ba/2013/05/internet-speeds-have-grown-10x-every-5.html> (23.3.2016)
- [14] ISO/IEC 27000, Information security management systems - Overview and vocabulary
- [15] <http://www.lunapic.com/editor/premade/watermark.gif> (23.3.2016)
- [16] <http://www.dailymail.co.uk/femail/article-3505261/Seve-Gat-s-photoshopping-fail-goes-viral.html> (23.3.2016)
- [17] <http://www.rune.com/wp-content/uploads/2016/02/snickers-pokes-fun-at-photoshop-fails-in-new-terribly-edited-print-ad.jpg> (23.3.2016)
- [18] Reis, G. (2004) *Digital Image Integrity*, Adobe Systems Inc. Whitepaper, http://www.adobe.com/digitalimag/pdfs/digital_image_integrity.pdf (23.3.2016)
- [19] FBI Scientific Working Group on Imaging Technology (SWGIT) (2008) *Best Practices for Maintaining the Integrity of Digital Images and Digital Video*, Forensic Science Communications vol. 10 (2), https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/april2008/index.htm/standards/2008_04_standards01.htm (23.3.2016)