

UNIVERZITET U BEOGRADU
MATEMATIČKI FAKULTET

Ozren Demonja

**REALIZACIJA P2P PROTOKOLA ZA
DOSTAVU SINHRONIZOVANOG
SADRŽAJA**

master rad

Beograd, 2018.

Mentor:

dr Aleksandar KARTELJ, docent
Univerzitet u Beogradu, Matematički fakultet

Članovi komisije:

dr Saša MALKOV, profesor
Univerzitet u Beogradu, Matematički fakultet

dr Miroslav MARIĆ, profesor
Univerzitet u Beogradu, Matematički fakultet

Datum odbrane: _____

Mami i tati

Naslov master rada: Realizacija P2P protokola za dostavu sinhronizovanog sadržaja

Rezime:

Ključne reči: P2P mreža, deljenje video sadržaja, sinhronizovano deljenje sadržaja, P2P protokol, deljenje u realnom vremenu, decentralizovana razmena podataka

Sadržaj

1	Uvod	1
2	P2P mreža	2
2.1	Osnovni koncepti	3
2.2	Klasifikacija P2P mreža	4
2.3	Topologija	5
2.4	Protok video sadržaja	9
2.5	Modelovanje mreže	9
3	Zaključak	10
	Bibliography	11

Glava 1

Uvod

Glava 2

P2P mreža

Osnovna svrha interneta kao globalne mreže u današnje vreme je razmena digitalnog sadržaja. Arhitektura koja se koristi za podršku razmeni često je zasnovana na modelu klijent-server (eng. *Client-Server architecture*). Model klijent-server deli zadatke između servera koji obezbeđuju resurse ili usluge i klijenata koji zahtevaju usluge [6].

Kako svaki pristup sistemu od strane klijenta zahteva veću upotrebu resursa i bolje performanse, jasno se uočava da je server usko grlo sistema. Tokom 1990-ih godina internet se najvećim delom zasnivao na klijent-server modelu. Poslednjih desetak godina dolazi do velikog proboja novih tehnologija usled želje korisnika za interaktivnijim sadržajem, porastom brzine pristupa internetu (eng. *broadband*), rasprostranjenosti pristupa, boljom stabilnosti veze i jačim računarima korisnika.

Navedene činjenice dovode do promene načina korišćenja interneta i pravljenja mreže ravnopravnih članova (eng. *peer-to-peer (P2P) network*) u kojoj korisnici dele svoje resurse [20]. Primena ovakvog pristupa obećavajuće je rešenje za mnoge oblasti.

U poglavlju 2.1 opisani su osnovni koncepti mreže ravnopravnih članova. U poglavlju 2.2 opisane su struktuirane i nestruktuirane mreže ravnopravnih članova. U poglavlju 2.3 opisane su topologija stabla (eng. *tree topology*), mrežna (eng. *mesh topology*) i hibridna (eng. *hybrid topology*) topologija. Način pripreme i slanja video sadržaja opisan je u poglavlju 2.4. U poglavlju 2.5 opisani su osnovni problemi prilikom pravljenja mreže ravnopravnih članova.

2.1 Osnovni koncepti

Korišćenje aplikativnog sloja u formiranju pokrivača (eng. *overlays*) za pristup internet servisima ima dugu istoriju. I pored duge tradicije ovakav pristup se do nedavno koristio samo za dizajn specifičnih protokola i interno povezivanje servisa infrastrukture, bez mogućnosti korišćenja između dva računara koja nisu susedi. Popularizacija ideje za formiranje pokrivača na aplikativnom sloju počinje 1999. godine sa Napster-om [1] i nastavlja se kroz druge sisteme za razmenu digitalnog sadržaja kao što su Gnutella [4], FastTrack [8], KaZaa [13] i BitTorrent [3].

Arhitektura ravnopravnih članova (eng. *peer-to-peer*, *P2P*) definiše se kao distribuirani vid arhitekture u kojoj svaki parnjak (eng. *peer*) može samostalno da obavlja poslove, ali omogućava drugim parnjacima da koriste njegove resurse kao što i sam može po potrebi koristiti resurse drugih parnjaka [17]. Drugim rečima, svaki parnjak komunicira sa ostalim parnjacima u mreži koristeći softver koji se ponaša i kao klijent i kao server. Na osnovu navedenih činjenica zaključuje se da arhitektura ravnopravnih članova omogućava veću autonomiju članova mreže koji sami definišu pravila o deljenju resursa. Glavna primena arhitekture ravnopravnih članova prisutna je u sistemima sa većom tolerancijom greške kod distribuirane obrade. Pored toga, ovaj vid arhitekture koristi se prilikom razmene fajlova, direktne komunikacije, obrade velikih količina podataka i softvera za zabavu.

Prilikom deljenja resursa svaki parnjak doprinosi resursima mreže ravnopravnih članova. Idealno, deljenje resursa je proporcionalno broju parnjaka koji koriste mrežu ravnopravnih članova. Međutim, usled različitih problema u praksi navedeno svojstvo većinom nije ispunjeno.

P2P sistem čini skup međusobno povezanih parnjaka, direktno ili preko drugih parnjaka formirajući povezani graf. Kako u grafu ne postoji centralna tačka kontrole i kolektivno se obavlja posao, P2P sistemi su decentralizovani. U nekim situacijama definiše se centralni server za poslove kao što su identifikacija, autentikacija i različite sigurnosne provere. U ovakvim uslovima, po pravilu server obavlja što je moguće manji deo posla. Pored toga, ponekad se u P2P sistemima iz različitih razloga uvode parnjaci zaduženi za dodatni deo posla. Ovakvi parnjaci se nazivaju super parnjaci (eng. *super peers*).

Učešće parnjaka u obavljanju posla u P2P sistemima određeno je lokalno. Iz ove činjenice proizilazi bolja organizacija P2P sistema vremenom usled korišćenja lokalnih znanja i operacija na svakom parnjaku. Ovim se takođe sprečava neplanirana

dominacija nekog parnjaka sistemom.

Dva važna svojstva P2P sistema su samoskalabilnost i stabilnost. Samoskalabilnost, kao glavna prednost P2P sistema, ispoljava se povećanjem ukupne snage sistema prilikom pristupanja parnjaka. Stabilnost mreže se obično definiše kao njena otpornost na veliku stopu promene strukture mreže usled čestog dolaska i odlaska parnjaka. U literaturi ove promene se nazivaju talasanja (eng. *churn*). P2P sistem bi trebao biti stabilan do određenog nivoa talasanja, u kontekstu održavanja grafa konekcije i upućivanja (eng. *route*) sadržaja deterministički sa prihvatljivim granicama broja skokova (eng. *hop-count bounds*).

P2P sistem, kao tehnologija u zamahu, privlači veliku pažnju običnih i korporativnih korisnika. Takođe, pored pažnje korisnika, značajno interesovanje beleži se i u istraživačkim krugovima [6]. U poslednjih nekoliko godina, veliki deo istraživanja bavi se P2P deljenjem resursa i prenosom podataka.

2.2 Klasifikacija P2P mreža

Usled velikog broja kriterijuma dizajniranja P2P mreža, ne postoji jedinstven način klasifikacije [17]. Na primer, sistemi za deljenje fajlova često se klasifikuju po generacijama. Prvu generaciju karakteriše hibridni dizajn. Glavnu karakteristiku ove generacije predstavlja kombinacija servera sa P2P upućivanjem sadržaja. Druga generacija je zasnovana na decentralizovanoj arhitekturi. Trećom generacijom smatraju se anonimni P2P sistemi kao što su Freenet [5] i I2P [11]. Ovakav način kategorizacije prema generacijama ima značajnih nedostataka. Glavni nedostatak ogleda se u nedefinisanim važnim dimenzijama problema i nejasnoći dodatnih mogućnosti sledeće generacije. Takođe, sistemi sve tri generacije su korišćeni u isto vreme što se ne može zaključiti na osnovu njihovih imena. Zbog navedenih nedostataka, u literaturi se češće sreće podela prema P2P pokrivaču. Prema ovoj podeli, razlikujemo struktuiranu i nestruktuiranu P2P mrežu [9].

U struktuiranim P2P mrežama topologija je precizno definisana. Takođe, neophodna je striktna kontrola protoka sadržaja. Sadržaj se ne sme nalaziti kod proizvoljnih parnjaka već mora biti na strogo definisanim lokacijama koje će omogućiti efikasniju pretragu. Većina struktuiranih P2P mreža je bazirana na distribuiranim heš tabelama (eng. *Distributed Hash Table (DHT)*). Kao primere struktuiranih P2P mreža navodimo Tapestry [21], Chord [18], Pastry [15] i Kademlia [14].

Ne struktuirani P2P sistem je sačinjen od parnjaka koji se pridružuju mreži sa

nekim labavim skupom pravila, bez ikakvog predznanja o njenoj topologiji. Freenet, Gnutella, FastTrack, KaZaA i BitTorrent su primeri ovakvih sistema. Ovakve mreže se zbog svojih svojstava tipično koriste za udruživanje snage parnjaka nebi li se neki posao brže obavio ili razmenu fajlova kod kojih nije od presudnog značaja brzina pronalaska fajla.

U nestruktuiranim P2P mrežama, pretraga je bazirana na prosleđivanju upita. Svaki parnjak upit prosleđuje komšijama, osim u slučaju kada parnjak može odgovoriti na traženi upit ili se brojač skokova upita smanji na 0 [17]. Sa ovakvim načinom prosleđivanja postoje različita pravila kontrole prosleđivanja upita. Način prosleđivanja može se kontrolisati tako što se upit neće proslediti svim komšijama nego samo određenim. Izbor kome će upit biti prosleđen se vrši prema informacijama koje parnjak čuva o komšijama. Te informacije predstavljene su istorijom ili indeksima sadržaja koji komšija može obezbediti. Takođe, često se prosleđivanje vrši slučajno bez ikakvog znanja. Naravno, ovo svojstvo smanjuje šansu da se pronađe sadržaj. Opravdanje nestruktuiranog pristupa zasniva se na promenljivosti broja članova mreže i ispravnosti prvog izbora komšije.

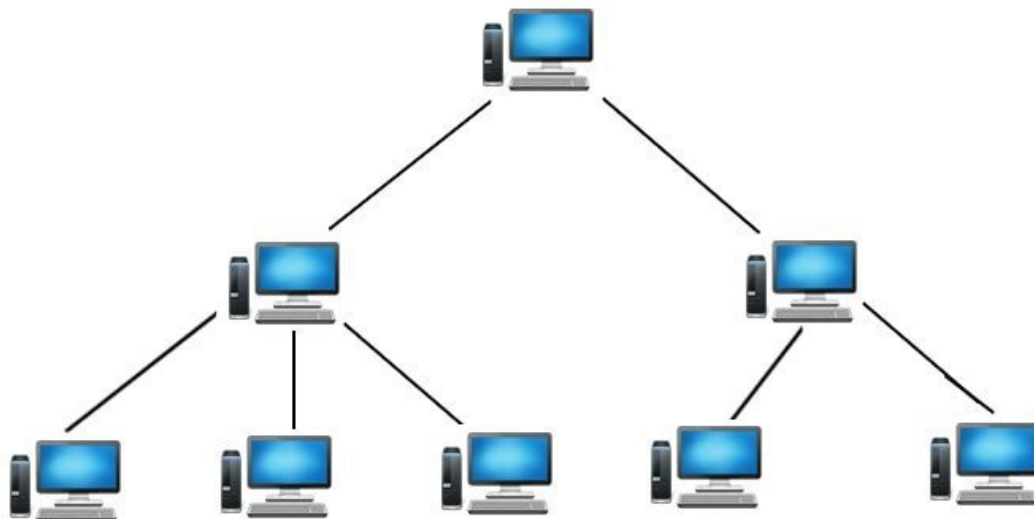
Nestruktuirane mreže ravnopravnih računara uglavnom su sastavljene od parnjaka koji su kućni PC računari. Oni su najčešće povezani sa mrežom slabim propusnim opsegom (eng. *bandwidth*). Takođe, većinom je propusni opseg nesrazmeran u odnosu na dolazni i odlazni saobraćaj u korist dolaznog saobraćaja. Navedena činjenica blago narušava samoskalabilnost mreže. Postoje različiti načini da se spreči narušavanje samoskalabilnosti. Neki od njih su zasnovani na kažnjavanju parnjaka koji ne doprinose ili jako malo doprinose resursima sistema.

2.3 Topologija

P2P sistem je sačinjen od dva podsistema koji se izvršavaju paralelno. To su konstrukcija i održavanje izabrane topologije kao i prosleđivanje bloka (eng. *chunk*) podataka. Algoritmi za konstrukciju i prosleđivanje zajedno formiraju P2P protokol. U zavisnosti od odluka donešenih pri dizajnu sistema tj. od izabrane topologije i algoritma prosleđivanja može se napraviti P2P sistem sa različitim karakteristikama.

Topologija drveta

Među prvim predloženim topologijama P2P sistema za razmenu sadržaja predstavljena je topologija drveta. U okviru ove topologije, svi parnjaci su organizovani tako da čine strukturu drveta. Primer topologije drveta prikazan je na slici 2.1.



Slika 2.1: Primer topologije drveta

U korenu drveta nalazi se parnjak koji deli sadržaj-izvor. Svaki parnjak u sistemu može imati tačno određen broj dece koji je ili unapred definisan ili određen kapacitetom parnjaka. Deca su komšije koje preuzimaju sadržaj od roditelja. Konstrukcija drveta i relacija roditelj-dete može biti određena različitim faktorima kao što su ukupno vreme potrebno da podatak stigne od izvora do svih parnjaka u mreži, dostupan propusni opseg i fizička topologija same mreže. Podaci koji se šalju na ovakav način se najpre podele na manje delove koji se nazivaju paketi. Paket se zatim šalje od izvora do parnjaka koji se nalaze na prvom nivou drveta. Nakon toga parnjaci sa prvog nivoa drveta šalju podatke parnjacima na drugom nivou drveta i tako redom sve dok paket ne dođe do svih parnjaka u mreži [6].

Iako jednostavna za konstrukciju, topologija drveta ima dosta nedostataka. Najveći nedostatak je neravnomerna raspodela tereta prosleđivanja koja se javlja kao posledica činjenice da veliki deo parnjaka čine listovi koji ne daju nikakav doprinos sistemu. Dodatno, parnjak koji nema odlazni (eng. *upload*) kapacitet proporcionalan proizvodu broja dece i veličine bloka koji prenosi, konstantno će nagomilavati

blokove podataka i usporavati ceo sistem. Ovakav parnjak treba da bude lociran jedino u listovima stabla ili veoma blizu listova. Pored navedenih osobina, važno je napomenuti da je ovakva struktura ranjiva na napuštanje parnjaka. Prilikom napuštanja parnjaka na višem nivou, njegova podstabla privremeno ostaju odsečena od ostatka mreže.

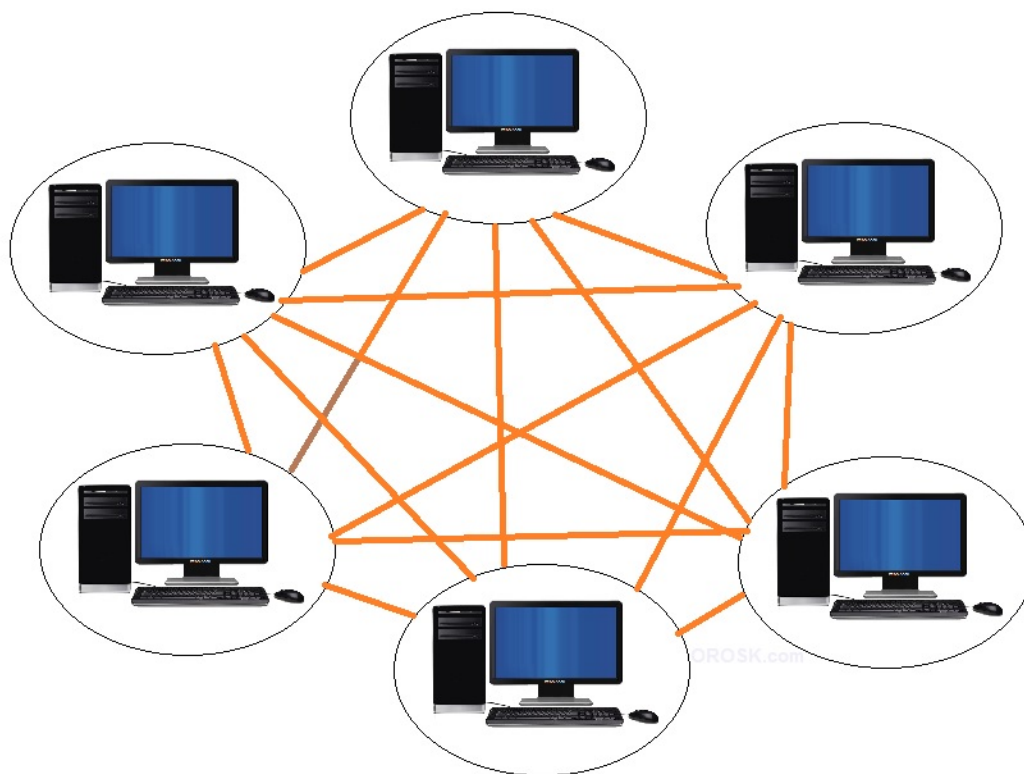
Da bi se povećala otpornost na napuštanja parnjaka i povećao kapacitet mreže koristi se rešenje bazirano na višestrukim stablima. U ovakvim sistemima parnjaci se nalaze u jednom ili više stabala. Svaki parnjak prima sadržaj od svih stabala čiji je član. Izvorni čvor je izvor za sva stabla. Dobijen sadržaj parnjak prosleđuje unapred određenim stablima. Sistem se na ovaj način deli u podsisteme u kojima svaki od podsistema predstavlja jedno stablo. Radi bržeg distribuiranja sadržaja svakom podsistemu se dodeljuje jedinstven paket.

Topologija zasnovana na višestrukim stablima rešava neke od problema upotrebe samo jednog stabla, ali uvodi dodatne probleme [6]. Prvo, na ovakav način postoji više putanja od izvora do lista pa je potreban značajno složeniji algoritam prosleđivanja. Drugo, kako stabla imaju različite podatke potrebno je obezbediti da svi podaci dođu do svakog parnjaka. Treće, prilikom konstrukcije mreže treba kreirati stabla koja su što je moguće manje dubine čime bi se obezbedio brži protok podataka. Dodatno, javlja se povremena potreba prebacivanja parnjaka iz jednog stabla u drugo uz dodatni zadatak balansiranja dubina svih stabala. Četvrti problem leži u balansiranju kapaciteta stabala.

Topologija mreže

Sistem zasnovan na mreži se fokusira na kreiranje topologije koja se brzo adaptira na promene. Glavna ideja pri pravljenju ove topologije bila je da parnjaci provode što je moguće manje vremena za održavanje pokrivača i brzi oporavak mreže pri njihovom napuštanju. Prednost ove topologije je brža i jednostavnija konstrukcija pokrivača. Sa druge strane, u ovakvoj konstrukciji potrebno je uložiti značajno vreme za implementaciju algoritma za prosleđivanje sadržaja između parnjaka. Primer topologije mreže nalazi se na slici 2.2.

Značajan kontrast se može primetiti ukoliko se posmatraju razlike između topologije mreže i topologije drveta 2.3. Kod topologije drveta kreiranje same topologije je složenije i zahteva više vremena, ali jednom kad se postigne željeni dizajn mreže prosleđivanje podataka je jednostavno. Kod topologije mreže situacija je potpuno



Slika 2.2: Primer topologije mreže

drugačija. Sama izrada pokrivača je krajnje jednostavna, jeftina i brza, ali je prosleđivanje podataka značajno kompleksnije.

U mrežnom pokrivaču glavni cilj parnjaka je da održavaju veliki broj dolaznih konekcija [17]. U slučaju izlaska komšije iz mreže, njegov uticaj na parnjake je minimalan. Pokrivač se obično gradi na distribuirani način i svaki parnjak je svestan samo malog dela učesnika, njegovih komšija. Način biranja komšija se razlikuje od politike (eng. *policy*) do politike, pri čemu je u svim slučajevima ovaj proces brz i jednostavan. Primer brze i jednostavne politike je slučajno biranje komšija.

U odnosu na topologiju drвета, prosleđivanje podataka u mreži ne odvija se pravolinijski. Nedostatak konkretne strukture koja bi definisala put kojim paket treba ići čini nemogućim određivanje unapred putanje kojom paket treba proći. Zbog navedenog svojstva prosleđivanje paketa je bazirano na lokalnim odlukama koje svaki parnjak donese. Parnjaci donose odluke prema informacijama koje imaju o komšijama. Paket se može razmeniti primenom guranja (eng. *push*) ili povlačenja (eng. *pull*). Kod guranja parnjak određuje koji paket šalje svakom komšiji, a kod povlačenja određuje koji paket traži od svakog komšije. Oba pristupa imaju dobre

i loše strane.

Pristup sa guranjem je efikasniji u mreži sa ograničenim odlaznim kapacitetom jer izbegava višestruke zahteve za pakete. Pristup sa povlačenjem je dobar izbor za mreže sa ograničenim dolaznim kapacitetom jer parnjak može da kontroliše koliko brzo će primati pakete od komšija. Kod obe šeme zbog lokalne koordinacije propagacije paketa dolazi do blage neefikasnosti u prosleđivanju. Kod sistema guranja ova činjenica se ogleda u višestrukim kopijama koje parnjak može da dobije ako mu više komšija šalje isti paket. Kod sistema povlačenja može se desiti da komšije zaguše parnjaka sa previše zahteva za paketima. Takođe sistemi povlačenja unose dodatni sadržaj u mrežu (eng. *overhead*) s obzirom da se paket mora tražiti.

Hibridna topologija

Hibridna topologija kombinuje prednosti topologija drveta i mreže. Preciznije, predstavlja kombinaciju robusnosti mrežne topologije sa jednostavnošću i efikasnošću prosleđivanja paketa koju nudi topologija drveta.

U hibridnoj topologiji, mreža je podeljena na više podmreža slično kao kod topologije drveta [17]. Da bi parnjak mogao da prima pakete, mora pronaći roditelje koji će mu prosleđivati pakete iz svih podmreža. Kada se novi parnjak pridružuje mreži, on dobija listu svih parnjaka na koje može da se poveže. Zatim parnjak među njima pokušava da odabere komšije tako da pokrije sve podmreže. Nakon što je pronašao takve parnjake povezuje se na njih. Glavna mana ovakvog sistema je ta što je dobijanje pravilne topologije težak zadatak. Često se može desiti da mala greška dovodi do situacije da parnjak ili grupa parnjaka bude izolovana od ostatka mreže.

2.4 Protok video sadržaja

2.5 Modelovanje mreže

Glava 3

Zaključak

Bibliography

- [1] K. Aberer, A. Datta, and M. Hauswirth. „Efficient, Self-Contained Handling of Identity in Peer-to-Peer Systems”. In: *IEEE Transactions on Knowledge and Data Engineering* 16.07 (July 2004), pp. 858–869. DOI: 10.1109/tkde.2004.1318567. URL: <https://doi.org/10.1109%2Ftkde.2004.1318567>.
- [2] Josh Beggs and Dylan Thede. *Designing Web Audio*. Ed. by Richard Koman. 1st. Sebastopol, CA, USA: O’Reilly & Associates, Inc., 1999. ISBN: 1565923537.
- [3] *Bittorrent Opis*. <http://www.bittorrent.com/>.
- [4] Fernando Bordignon and Gabriel Tolosa. „Gnutella: Distributed System for Information Storage and Searching Model Description”. In: (Aug. 2001).
- [5] Ian Clarke et al. „Freenet: A Distributed Anonymous Information Storage and Retrieval System”. In: *Designing Privacy Enhancing Technologies*. Springer Berlin Heidelberg, 2001, pp. 46–66. DOI: 10.1007/3-540-44702-4_4. URL: https://doi.org/10.1007%2F3-540-44702-4_4.
- [6] Jorn De Boever. „Peer-to-Peer Networks as a Distribution and Publishing Model”. In: (Jan. 2007).
- [7] Stephen E. Deering and David R. Cheriton. „Multicast Routing in Datagram Internetworks and Extended LANs”. In: *ACM Trans. Comput. Syst.* 8.2 (May 1990), pp. 85–110. ISSN: 0734-2071. DOI: 10.1145/78952.78953. URL: <http://doi.acm.org/10.1145/78952.78953>.
- [8] *FastTrack Opis*. <http://ntrg.cs.tcd.ie/undergrad/4ba2.02/p2p/index.html>.
- [9] Prasanna Ganesan, Krishna Gummadi, and H. Garcia-Molina. „Canon in G major: designing DHTs with hierarchical structure”. In: *24th International Conference on Distributed Computing Systems, 2004. Proceedings*. IEEE,

2004. DOI: 10.1109/icdcs.2004.1281591. URL: <https://doi.org/10.1109%2Ficdcs.2004.1281591>.
- [10] Yuri Gurevich and Saharon Shelah. “Expected computation time for Hamiltonian path problem”. English. In: *SIAM Journal on Computing* 16 (1987), pp. 486–502.
- [11] *I2P Dokumentacija*. <https://geti2p.net/en/docs>.
- [12] „INTERNAP The next big wave in convergence – Building the right foundation for Internet TV”. In: *J. Exp. Algorithmics* (May 2010). URL: www.internap.com.
- [13] Jian Liang, Rakesh Kumar, and Keith W Ross. „The kazaa overlay: A measurement study”. In: (Oct. 2004).
- [14] Petar Maymounkov and David Mazières. „Kademlia: A Peer-to-Peer Information System Based on the XOR Metric”. In: *Peer-to-Peer Systems*. Springer Berlin Heidelberg, 2002, pp. 53–65. DOI: 10.1007/3-540-45748-8_5. URL: https://doi.org/10.1007%2F3-540-45748-8_5.
- [15] Antony Rowstron and Peter Druschel. „Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems”. In: *Middleware 2001*. Springer Berlin Heidelberg, 2001, pp. 329–350. DOI: 10.1007/3-540-45518-3_18. URL: https://doi.org/10.1007%2F3-540-45518-3_18.
- [16] Robert R. Schaller. *Moore’s law: Past, present, and future*. IEEE Spectr.,34(6):52–59, 1997.
- [17] Xuemin Shen et al. *Handbook of Peer-to-Peer Networking*. 1st. Springer Publishing Company, Incorporated, 2009. ISBN: 0387097503, 9780387097503.
- [18] I. Stoica et al. „Chord: a scalable peer-to-peer lookup protocol for internet applications”. In: *IEEE/ACM Transactions on Networking* 11.1 (Feb. 2003), pp. 17–32. DOI: 10.1109/tnet.2002.808407. URL: <https://doi.org/10.1109%2Ftnet.2002.808407>.
- [19] „Telecom, ATIS Telecom Glossary 2007”. In: *American National Standard for Telecommunications* (2010). URL: <http://www.atis.org>.
- [20] David J. Wetherall and Andrew S. Tanenbaum. *Computer Networks*. Pearson Education, 2013. ISBN: 1292024224, 9781292024226.

BIBLIOGRAPHY

- [21] B.Y. Zhao et al. „Tapestry: A Resilient Global-Scale Overlay for Service Deployment”. In: *IEEE Journal on Selected Areas in Communications* 22.1 (Jan. 2004), pp. 41–53. DOI: 10.1109/jsac.2003.818784. URL: <https://doi.org/10.1109%2Fjsac.2003.818784>.