



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

Kompiuterių tinklų katedra

Rokas Zakarevičius

DUOMENŲ CENTRŲ PASLAUGŲ KOKYBĖS ĮVERTINIMAS

Magistro baigiamasis darbas

IFM-1/1 gr. studentas Rokas Zakarevičius

Vadovas: prof. dr. Rimantas Plėštys

Recenzentas: doc. dr. Pranas Kanapeckas

Kaunas 2007

TURINYS

1.	IVADAS	3
2.	DUOMENŲ CENTRŲ PASLAUGOS	5
2.1.	Tarnybinių stočių nuomos paslauga	5
2.2.	Duomenų saugyklų nuomos paslauga	5
2.3.	Duomenų perdavimo tinklo ir Interneto prieigos paslaugos	6
2.4.	„Hosting“ paslaugos	6
2.5.	Taikomųjų programų ir resursų nuoma – informacijos „outsourcing“	7
2.6.	Išskirtinės kokybės paslaugos	8
3.	DUOMENŲ CENTRŲ SLA (Service Level Agreement) PARAMETRŲ ĮVERTINIMAS	9
3.1.	Paslaugų pateikiamumo įvertinimas	13
3.2.	Duomenų perdavimo tinklo vėlinimo įvertinimas	17
3.3.	Duomenų perdavimo tinklo vėlinimo sklaidos įvertinimas	18
3.4.	Duomenų perdavimo tinklo paketų pristatymo įvertinimas	19
3.5.	Informacijos perdavimo sparta duomenų perdavimo tinklu	19
3.6.	Informacijos saugumas	20
3.7.	Duomenų centro paslaugos vėlinimas	21
4.	SLA (Service Level Agreement) PARAMETRŲ MATAVIMO METODIKOS	23
4.1.	CISCO SAA (Service Assurance Agent) technologija	23
4.2.	RIPE TTM (Test Traffic Measurement) technologija	28
4.3.	Išvados	30
5.	DUOMENŲ CENTRŲ DIDELIO PATEIKIAMUMO SERVERIŲ SISTEMOS	31
5.1.	Didelio pateikiamumo failų serveriai duomenų centruose	31
5.2.	Didelio pateikiamumo Web serveriai duomenų centruose	35
5.3.	Tinklo apkrovos balansavimas	36
6.	„PLONO KLIENTO“ (Thin-client) PASLAUGOS MODELIS	39
6.1.	„Plono kliento“ paslaugos panaudojimo galimybės	40
6.2.	Windows Remote Desktop Protocol (RDP) įvertinimas	44
6.3.	„Plono kliento“ paslaugos kokybė	45
7.	WEB PASLAUGOS KOKYBĖS ĮVERTINIMAS	46
7.1.	Web paslaugos vėlinimo įvertinimas iš serverio ir kliento perspektyvos	49
7.1.1.	Matavimų metodika	49
7.1.2.	Vėlinimo įvertinimo rezultatai	51
7.2.	Web paslaugos vėlinimas dėl interneto naršyklės laikinosios atminties (cache) įtakos	53
7.2.1.	Tyrimų tikslas	53
7.2.2.	Matavimų metodika	54
7.2.3.	„Cache“ įtakos vėlinimui matavimo rezultatai	54
7.3.	WWW paslaugos kokybė	55
8.	MAGISTRINIO DARBO IŠVADOS	57
9.	LITERATŪRA	59
10.	QUALITY EVALUATION OF DATACENTER SERVICES - SUMMARY	61
11.	PRIEDAI	62
11.1.	Web paslaugos vėlinimo įvertinimų išsamūs rezultatai	62
11.2.	Straipsnis konferencijoje „Informacinės Technologijos 2006“	65
11.3.	Straipsnis konferencijoje „Informacinės Technologijos 2007“	69

1. IVADAS

Šiuolaikinėse didelės spartos duomenų perdavimo tinklų sistemose duomenų centrai tapo viena esminių tinklinės infrastruktūros sudedamųjų dalių. Juose centralizuotai įrengiama duomenų perdavimo tinklo komutacinė bei tarnybinių stočių aparatūrinė ir programinė įranga, talpinami didelės apimties duomenys.

Duomenų centrų paslaugos - tai informacijos saugojimo, apdorojimo, valdymo ir persiuntimo duomenų perdavimo tinklu paslaugos.

Paslaugų teikėjai siūlo vartotojams plataus spektro duomenų centrų paslaugas: didelio našumo ir pateikiamumo duomenų saugyklos kritinėms taikymo sritims, duomenų perdavimo tinklo bei prieigos prie interneto paslaugas, Web svetainių bei failų talpinimo tarnybinėse stotyse paslaugas, terminalines vartotojo aplinkų sistemas taikomųjų programų naudojimui, bei kitas.

Duomenų centrų paslaugos – tai dar nauja ir labai perspektyvi sritis, kuri vis plačiau pradedama naudoti ne tik didelės apimties verslo, bet ir mažesnės apimties verslo bei namų vartotojų.

Tam, kad vartotojas galėtų pasirinkti vieną ar kitą duomenų centrą reikia nustatyti paslaugų lygio kokybinius parametrus ir juos pateikti paslaugų lygio sutartyje (SLA – Service Level Agreement), kuri sudaroma tarp paslaugų vartotojo ir duomenų centro administracijos. Sutartyje duomenų centro paslaugų teikėjas įsipareigoja teikti tam tikras nustatytos kokybės paslaugas.

Kartu paslaugos kokybė yra labai subjektyvus parametras, t.y. priklauso nuo kiekvieno vartotojo poreikių. Iš čia kyla problema, kaip vartotojui vertinti skirtingų duomenų centrų teikiamas paslaugas, palyginti jų kokybę, bei išsirinkti tinkamą paslaugos teikėją, o duomenų centrų paslaugų teikėjui reikalinga metodika, padedanti įvertinti vartotojo norimą paslaugų kokybę ir už ją užtikrinti.

Šio darbo tikslas sudaryti duomenų centrų paslaugų kokybės vertinimo metodiką, kuria naudojantis paslaugos vartotojas galėtų pasirinkti geriausiai poreikius atitinkančią paslaugą, o paslaugos teikėjas galėtų užtikrinti reikalaujamą tos duomenų centro paslaugos kokybę. Tikslui pasiekti turi būti išspręsti tokie uždaviniai:

- atlikta duomenų centrų paslaugų klasifikacija,
- atlikti dažniausiai vartojamų duomenų centro paslaugų kokybės įvertinimai,
- išanalizuoti paslaugų lygio sutartyse pateikiami parametrai, sudarytos jų paskaičiavimo bei matavimo metodikos,
- įvertinta duomenų centrų aparatūrinės bei programinės įrangos įtaka paslaugos kokybei, ir nustatytos kritinės vietos,

- atlikti dažniausiai naudojamų paslaugų - terminalinių serverių bei web paslaugų kokybinių parametrų matavimai,
- pasiūlytos skirtingų duomenų centrų paslaugų kokybės palyginimo metodikos.

2. DUOMENŲ CENTRŲ PASLAUGOS

Šiuolaikinėse informacinių technologijų sistemose duomenų centrai, kaip apibendrintas terminas, vartojamas labai dažnai. Duomenų centrai - tai tarnybinės stotys bei duomenų perdavimo tinklo įranga, patalpinta specialiai tam pritaikytose patalpose.

Duomenų centrų paslaugos - tai informacijos saugojimo, apdorojimo, valdymo ir persiuntimo duomenų perdavimo tinklu paslaugos.

Įvairūs duomenų centrai siūlo įvairias paslaugas. Tačiau daugumos duomenų centrų paslaugos apima:

- tarnybinių stočių nuoma,
- duomenų saugyklų nuoma,
- prieigą prie duomenų centrų,
- „Hosting“ paslaugas,
- taikomųjų programų ir resursų nuomą – informacijos „outsourcing“ paslaugas,
- išskirtinės kokybės paslaugas.

Toliau apžvelgsime šias paslaugas.

2.1. Tarnybinių stočių nuomos paslauga

Šia paslauga duomenų centras siūlo vartotojams nuomotis tarnybines stotis ir duomenų perdavimo tinklą nuo jų iki duomenų centro. Yra įmanoma nuomotis paslaugos teikėjo dedikuotas tarnybines stotis, arba tiesiog pačią vietą nuosavai tarnybinei stotčiai laikyti [1][2].

Paskutiniu metu pradedama plačiai taikyti paslaugų virtualizacija duomenų centruose, t.y. kuriamos virtualios tarnybinės stotys ar kiti vartotojams reikalingi resursai, kuriuos vartotojai mato kaip tik jiems skirtus. Pačių duomenų centrų turimų resursų panaudojimas tampa gerokai lankstesnis - vienoje tarnybiniėje stotyje arba jų klasteryje (pagal poreikius) gali veikti daugelio skirtingų vartotojų aplikacijos. Duomenų centrų techniniam personalui tampa gerokai lengviau kontroliuoti ir efektyviai išnaudoti turimus resursus, kuriuos kiekvienas vartotojas mato kaip tik jam skirtus.

2.2. Duomenų saugyklų nuomos paslauga

Duomenų saugyklos – tai specializuota aparatūrinė įranga, skirta centralizuotai, efektyviai ir saugiai laikyti didelius kiekius duomenų. Tai dideli diskų masyvai (informacijos saugojimui) ar

juostinių įrenginių masyvai (atsarginėms kopijoms daryti). Kaip ir tarnybinėms stotims, plačiai taikoma duomenų saugyklų virtualizacija. Jos pagalba ne tik efektyviau valdomi sistemų resursai, bet suteikiama daugiau lankstumo didelių duomenų saugyklų architektūrų kūrimui.

Detalesnis šių technologijų aprašymas pateiktas 5-ame skyriuje.

2.3. Duomenų perdavimo tinklo ir Interneto prieigos paslaugos

Dažnai duomenų centrų paslaugų teikėjai teikia tuo pačiu ir duomenų perdavimo tinklo ir Interneto prieigos paslaugas. [3] Suprantama – juk duomenys turi būti greitai ir kokybiškai pasiekiami tiek klientui, tiek ir kitiems vartotojams (priklausomai nuo paslaugos), todėl duomenų perdavimo tinklo ryšio linijos yra vienas esminių paslaugų elementų. Dažnai siūlomos ne tik IP tinklo paslaugos, bet ir papildomos saugumą užtikrinančios paslaugos, tokios kaip VPN (*angl.* Virtual Private Network). VPN – tai paslauga, leidžianti sukurti saugius virtualius duomenų kanalus per nesaugų tinklą (tokį, kaip Internetas). Tuo pačiu teikiamos visa eilė kitų su duomenų perdavimo tinklais susijusių paslaugų: tinklo saugumo paslaugos (ugniasienės, įsilaužimų aptikimo sistemos), bei pastovi duomenų perdavimo tinklo priežiūra, ryšio linijų rezervavimai ir t.t.

2.4. „Hosting“ paslaugos

Tai paslauga, kai duomenų centre yra talpinama vartotojo informacija, kuri vėliau gali būti pasiekama duomenų perdavimo tinklu, naudojantis skirtingais protokolais. Šis apibrėžimas apibūdina daugiausiai siūlomų duomenų centrų paslaugas. Tai dažniausiai vartojamos interneto ar intraneto paslaugos, kuriomis remiasi visos šiuolaikinės interneto technologijos. Web „hosting“ [2],[3] yra dažniausiai naudojama duomenų centrų paslauga. Pagal Web technologiją veikia didžioji dalis šiuolaikinių taikomųjų programų. Klientų interneto portalai ir web aplikacijos talpinamos duomenų centrų tarnybinėse stotyse (arba dedikuotose arba virtualizuotose). Paslaugos veikimui vartojama visa eilė įvairios paskirties įrenginių: duomenų perdavimo tinklo įrenginiai, pirminiai užklausas aptarnaujantys serveriai, duomenų prieigos serveriai, duomenų saugyklos (diskų masyvai). Duomenų centrų architektūros aprašymas pateiktas 3-ame skyriuje.

Didžiuosiuose duomenų centruose „Web hosting“ terminu apibūdinamas ne tik WWW svetainės talpinimas, bet ir visa eilė kitų su ja susijusių paslaugų:

- Viena iš jų yra DNS vardo (domeno) palaikymas [3]. Domeno vardai registruojami šalies ar regiono akredituoto domeno registratoriaus duomenų bazėse. Visa domeno vardų zonos konfigūracija saugoma duomenų centro DNS serveriuose.

- E-mail paslauga [3] leidžia vartotojui susikurti, talpinti ir vartoti elektroninį paštą. E-pašto transportą aptarnauja pašto serveriai, o patys duomenys (laiškai ir t.t.) laikomi duomenų saugyklose.
- Informacinių sistemų (duomenų bazių) paslauga įgalina kurti Web svetaines ir taikomas programas, kuriose naudojamos duomenų bazių priemonės. Duomenų centruose veikia visa eilė duomenų bazių valdymo sistemų (DBVS, *angl. DBMS*), kurių pagalba kuriamos duomenų struktūros, o duomenys galiausiai talpinami duomenų saugyklose. Dažniausiai naudojamos DBVS yra Oracle, Microsoft SQL, MySQL ir t.t.
- Įvairių failų talpinimo duomenų centruose paslauga – leidžia talpinti ir naudoti informaciją duomenų centruose. Duomenų pasiekiamumas gali būti realizuojamas įvairiomis technologijomis: HTTP, HTTPS, FTP, VPN ir t.t.

2.5. Taikomųjų programų ir resursų nuoma – informacijos „outsourcing“

Iki šiol minėtos „hosting“ paslaugos vienaip ar kitaip apėmė tik informacijos talpinimą duomenų centruose. T.y. vartotojai duomenų centrą naudodavo kaip didelę duomenų saugyklą, palaikančią įvairias duomenų prieigos technologijas.

Taikomųjų programų paslaugų teikėjas – (*angl. Application Service Provider - ASP*) tai tam tikra duomenų centrų paslaugų teikėjų atmaina. Jie teikia programinės įrangos nuomos paslaugas, leisdami naudotis duomenų centrų serverių resursais (programine įranga, duomenų laikmenomis, duomenimis, ryšio linijomis ir t.t.). Šios paslaugos pasiekiamos naudojantis Web technologijomis ar per specializuotą programinę įrangą. Tokiu atveju klientas apmoka paslaugas teikėjui tik už tą programinę įrangą, kuria naudojami tam tikrą laiką. Priklausomai nuo situacijos, tai gali būti labai ekonomiškai naudinga.

„Outsourcing“ - tai programinės įrangos nuoma, duomenų talpinimas, tiek ir platesnės IT infrastruktūros perkėlimas. Šiuo atveju klientas patiki visą ar dalinį savo IT infrastruktūros (programinę įrangą, aparatūrinius resursus, tinklo įrangą, duomenis, sistemų priežiūrą ir aptarnavimą) duomenų centrui. Paslaugų teikėjas (duomenų centras) savo turimus resursus panaudoja kliento IT infrastruktūros patalpinimui savo sistemose. Tokiu atveju klientas moka už „outsourcing“ paslaugą, kurios pagalba teikėjas užtikrina maksimalios kokybės IT sistemų darbą. Detalesnis ASP panaudojimo aprašymas pateiktas 6-ame skyriuje.

2.6. Išskirtinės kokybės paslaugos

Dažnai duomenų centrų paslaugų teikėjai siūlo tiek pigias ir paprastiems vartotojams prieinamas paslaugas, tiek ir brangias dideliems verslo klientams skirtas paslaugas. Kainą lemia paslaugų kokybė, resursų sunaudojimas ir visa eilė kitų reiškinių. Aukštos kokybės paslaugos pasižymi tuo, kad jos yra prižiūrimos ir valdomos realiu laiku, o gedimai šalinami kaip įmanoma greičiau [2]. Veikiančių sistemų saugumas yra užtikrinamas tiek duomenų perdavimo tinklo stebėjimu ir apsaugos priemonėmis, tiek ir serverių atnaujinimais bei stebėjimais. Duomenų saugumas gali būti realizuojamas kodavimu, o patys duomenys ne tik laikomi ypač aukšto patikimumo duomenų masyvuose, bet ir periodiškai daromos jų atsarginės kopijos, laikomos duomenų archyvuose, esant reikalui – kaip įmanoma greičiau atstatomi duomenys.

Tarp kliento ir teikėjo pasirašomos paslaugų lygio sutartys (*angl. Service Level Agreement – SLA*), kuriose paslaugos teikėjas formaliai įsipareigoja teikti konkrečiai specifiikuotos kokybės paslaugas, kurių kokybė turėtų būti aprašoma visa eile skirtingų metrikų. Detalesnis paslaugos kokybės bei jos užtikrinimo aprašymas pateiktas 3-ame skyriuje.

Čia buvo pateikta apibendrinta duomenų centrų paslaugų klasifikacija. Skirtingi teikėjai siūlo labai įvairų spektrą IT paslaugų, todėl kiekvienas atvejis gali būti skirtingas. Toliau šiame darbe atlikta tam tikrų atskirų paslaugų analizė bei su jomis susiję eksperimentai, jų rezultatai, išvados.

3. DUOMENŲ CENTRŲ SLA (*Service Level Agreement*) PARAMETRŲ ĮVERTINIMAS

Kaip minėta praėjusiame skyrelyje, duomenų centrai susideda iš visos eilės paslaugų. Tarp kliento ir duomenų centro paslaugų teikėjo pasirašomos paslaugų lygio sutartys (*angl. Service Level Agreement – SLA*), kuriose teikėjas formaliai įsipareigoja teikti konkrečiai specifiikuotos kokybės paslaugas, kurių kokybė turėtų būti aprašoma visa eile skirtingų parametrų (kai kur vadinamų metrikomis).

Šie įsipareigojimai nėra vienodi skirtingiems duomenų centrams. Atlikus didžiausių duomenų centrų [1][2][3] paslaugų SLA apžvalgą, išskyrėme šiuos pagrindinius kokybinius parametrus:

- Paslaugos įrengimo trukmė;
- Elektros energijos pateikiamumas;
- Duomenų perdavimo tinklo pateikiamumas;
- Duomenų perdavimo tinklo vėlinimas;
- Duomenų perdavimo tinklo vėlinimo sklaida (*angl. jitter*);
- Duomenų perdavimo tinklo paketų pristatymas;
- Informacijos perdavimo sparta duomenų perdavimo tinklu;
- Informacijos saugumas;

Paskutiniu metu yra siūloma įtraukti [14][16] ir šiuos kokybinius parametrus:

- Duomenų centro paslaugos pateikiamumas;
- Duomenų centro paslaugos vėlinimas;

Jų įvertinimas pateiktas toliau šiame skyrelyje bei 6-ame ir 7-ame skyriuose.

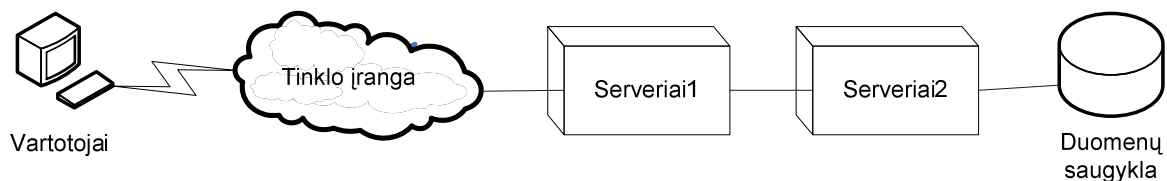
Minėtus atskirus parametrus galima pažymėti: *I* - paslaugos įrengimo trukmė, *E* - elektros energijos pateikiamumas (*angl. availability*), *NA* - duomenų perdavimo tinklo pateikiamumas, *NL* - duomenų perdavimo tinklo vėlinimas (*angl. latency*), *J* - duomenų perdavimo tinklo vėlinimo sklaida (*angl. jitter*), *D* - duomenų perdavimo tinklo paketų pristatymas (*angl. packet delivery*), *G* - informacijos perdavimo sparta duomenų perdavimo tinklu (*angl. throughput*), *S* - informacijos saugumas, *A* - duomenų centro paslaugos pateikiamumas (ne tik duomenų perdavimo tinklo), *L* - konkrečios duomenų centro paslaugos vėlinimas.

Iš jų galima spręsti apie paslaugos kokybę (*angl. QoS – Quality of Service*). Duomenų centro paslaugos kokybė nėra vienareikšmiškas parametras, nes skirtingiems vartotojams ji gali atrodyti labai skirtinga. Tai yra bedimensis dydis, kuris gali būti išreiškiamas naudojantis pritaikyta [20] formule:

$$Q = \frac{K_1 \cdot I + K_2 \cdot E + K_3 \cdot NA + K_4 \cdot NL + K_5 \cdot J + K_6 \cdot D + K_7 \cdot G + K_8 \cdot S + K_9 \cdot A + K_{10} \cdot L}{\sum_{i=1}^N K_i} (1),$$

kur minėti parametrai yra kokybinių parametrų įverčiai (normuoti [0;1] intervale). N yra parametrų skaičius. K_i yra atitinkamų parametrų svorių koeficientas, kurį paslaugos vartotojas nustato skaičiumi duotame intervale (pvz. [1;10]), priklausomai nuo to, kiek kiekvienas parametras yra svarbus bendrai konkrečios paslaugos kokybei. Duomenų centro paslaugos kokybę apibūdinantis dydis Q gali būti naudojamas kelių duomenų centrų paslaugų teikėjų tos pačios paslaugos palyginimams. Detalesni parametrų ir jų įverčių apibūdinimai pateikti toliau.

Duomenų centro paslaugų teikimo struktūrinė schema pateikta 1 pav. Ji sudaryta iš tokių elementų: duomenų perdavimo tinklo, pirminių serverių („Serveriai1“), antrinių serverių („Serveriai2“) ir duomenų saugyklos. Elementų vykdomos funkcijos priklauso nuo paslaugos tipo.



1 pav. Duomenų centro paslaugos teikimo struktūrinė schema

Duomenų perdavimo tinklą sudaro ryšio linijos, maršrutizatoriai (*angl. router*), komutatoriai (*angl. switch*) bei ugniasienės (*angl. firewall*) įrenginiai. Duomenys serveriams ir iš serverių pristatomi IP protokolu (OSI Layer 3), nepriklausomai nuo fizinių (OSI Layer 2) protokolų.

Blokeliu „Serveriai1“ pažymėtas duomenų centro struktūros elementas, kurį sudaro pirminiai vartotojus aptarnaujantys serveriai. Jie būna subalansuoti pagal apkrovimą (5 skyrius), priima prisijungimo prie paslaugos užklausas, juose veikia taikomosios programos, tačiau patys vartotojų duomenys juose nėra saugomi. Failų sistemos užklausas aptarnauja antriniai serveriai („Serveriai2“) - jie tiekia duomenis pirminiems serveriams. Patys duomenys saugomi duomenų saugyklose, kurios yra atskiri didelio pateikiamumo diskiniai įrenginiai (5 skyrius).

Duomenų centrų paslaugos struktūrinių elementų įtaka bendrai paslaugos kokybei pateikta 1-oje lentelėje.

1 lentelė. Duomenų centro paslaugos struktūrinių elementų įtaka kokybiniams parametrams

	Vartotojas	Tinklas	Serveriai 1	Serveriai 2	Duomenų saugykla
Paslaugos įrengimo trukmė		+	+	+	+
Elektros energijos pateikiamumas		+	+	+	+
Duomenų perdavimo tinklo pateikiamumas		+			
Duomenų perdavimo tinklo vėlinimas		+			
Duomenų perdavimo tinklo vėlinimo sklaida		+			
Duomenų perdavimo tinklo paketų pristatymas		+			
Informacijos perdavimo sparta		+			
Informacijos saugumas		+	+	+	+
Duomenų centro paslaugos pateikiamumas	+	+	+	+	+
Duomenų centro paslaugos vėlinimas	+	+	+	+	+

- *Paslaugos įrengimo trukmė* labai priklauso nuo paslaugos tipo. Priklausomai nuo to, kurie minėti elementai yra reikalingi paslaugai įrengti, gali labai skirtis įrengimo trukmė. Pvz. jeigu vartotojas užsisako tik duomenų perdavimo tinklo paslaugą, tai duomenų centro paslaugos teikėjui reikės sukonfigūruoti ar paruošti tik reikiamą duomenų perdavimo tinklo įrangą. Jeigu vartotojas užsisako web svetainės talpinimo paslaugą, tai paprastai duomenų perdavimo tinklo struktūrinio elemento įtakos nereikia įvertinti įrengimo trukmei (nes tinklo įrenginių specialiai tam paruošti nereikia). Tačiau kita elementų grandinė (1 lentelė) įtakoja bendrą įrengimo trukmę.
- *Elektros energijos pateikiamumas* užtikrinamas visai duomenų centruose esančiai įrangai. Čia nevertinamas duomenų perdavimo tinklo įrangos maitinimas, nes tinklas nėra pačiame duomenų centre, o yra naudojamas vartotojo prieigai prie jo. Elektros energijos pateikiamumas gali apimti visą struktūrinių elementų grandinę (1 lentelė), tiek ir atskirus elementus - jeigu vartotojas talpina savo įrangą duomenų centre.
- *Duomenų perdavimo tinklo pateikiamumas* - tai dydis, parodantis kiek laiko (per nagrinėjamą laikotarpį) duomenų centras yra pasiekimas duomenų perdavimo tinklu. Jis apima tik duomenų perdavimo tinklo įrenginių ir ryšio linijų komponentų pateikiamumus. Detalesnis aprašymas pateiktas 3.1 skyrelyje.
- *Duomenų perdavimo tinklo vėlinimas, vėlinimo sklaida ir paketų pristatymas* - tai parametrai, apibūdinantys duomenų perdavimo tinklo savybes (vėlinimą, vėlinimo sklaidą, informacijos paketų pristatymą), kurios pasireiškia, perduodant informacijos paketą tinklu. Jie apima tik tinklo įrenginių ir ryšio linijų komponentų parametrus. Jų reikšmės suprantamos kaip informacijos paketų sulaikymas (vėlinimas, vėlinimo sklaida) ar atmetimas (priešingai nei

paketų pristatymas), kai jie keliauja tinklu. Patys tinklo įrenginiai neturi šių parametrų, jį turi duomenų perdavimo tinklu keliaujantys paketai. Detalesnis aprašymas pateiktas 3.2, 3.3 ir 3.4 skyreliuose.

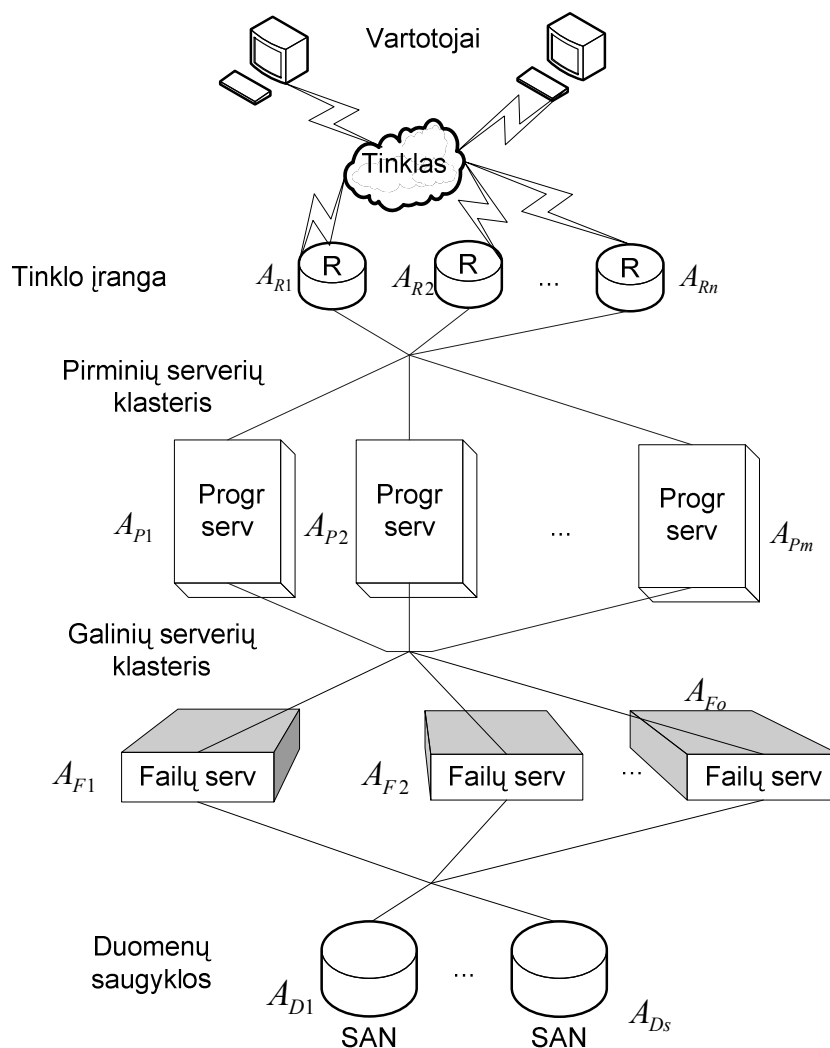
- *Informacijos perdavimo sparta* - tai parametras, parodantis kiek informacijos duomenų perdavimo tinklas gali persiųsti per laiko vienetą. Naudojami matavimo vienetai bitai per sekundę: bps, Kbps, Mbps. Detalesnis aprašymas pateiktas 3.5 skyrelyje.
- *Informacijos saugumas* gali būti suprantamas labai skirtingai. Jeigu kalbama tik apie duomenų perdavimo tinkle keliaujančių duomenų saugumą (saugius VPN šifruojamus tunelius, autentifikavimo sertifikatus ir pan.), tai reikia įvertinti tik tinklo įrenginių ir ryšio linijų saugumą šiuo požiūriu. Tačiau jeigu kalbama ir apie serveriais keliaujančių bei laikomų duomenų saugumą (pvz. šifravimą), tai saugumas gali apimti visą elementų grandinę (1 lentelė). Duomenų atsarginių kopijų darymą taip pat galima sieti su informacijos saugumu, nes kopijos reikalingos, norint atkurti duomenis sistemų gedimo, duomenų sugadinimo, stichinių nelaimių atvejais. Detalesnis aprašymas pateiktas 3.6 skyrelyje.
- *Duomenų centro paslaugos pateikiamumas* – tai dydis, parodantis kiek laiko (per nagrinėjamą laikotarpį) konkreti duomenų centro paslauga yra pasiekama vartotojui duomenų perdavimo tinklu. Jis apima ne tik duomenų perdavimo tinklo įrenginių ir ryšio linijų komponentų pateikiamumus (kaip duomenų perdavimo tinklo pateikiamumas), bet ir, priklausomai nuo paslaugos, gali apimti arba dalinių, arba visos struktūrinių elementų grandinės (1 pav.) pateikiamumus. Detalesnis aprašymas pateiktas 3.1 skyrelyje.
- *Duomenų centro paslaugos vėlinimas*, priklausomai nuo paslaugos, gali apimti arba dalinių, arba visos struktūrinių elementų grandinės (1 pav.) pateikiamumus. Šį dydį siūloma įvesti į SLA tam, kad konkrečios paslaugos vėlinimą įtakoja ne vien tik duomenų perdavimo tinklo vėlinimas, bet ir atskirų sistemos elementų bendras vėlinimas. T.y. kiekvienas iš serverių ir programinės įrangos sulaiko informacijos paketus, nes pačios informacijos apdorojimas tampa gerokai sudėtingesnis – ne tik jų perdavimas duomenų perdavimo tinklu (kaip tinklo vėlinimo atveju). Detalesnis aprašymas pateiktas 3.7 skyrelyje.

3.1. Paslaugų pateikiamumo įvertinimas

Pateikiamumas - tai dydis, parodantis kurią dalį laiko per nagrinėjamą laikotarpį (dažniausiai per mėnesį) paslauga (pvz. elektros energija, duomenų perdavimo tinklas, interneto prieiga, prieiga prie Web svetainės) buvo pateikiama vartotojui (vartotojas galėjo paslauga pasinaudoti). Dažnai jis išreiškiamas procentais. Kadangi yra norima pateikiamumo parametą naudoti (1) formulėje, reikia jį normuoti $[0;1]$ intervale - t.y. apibrėžti kokybinį įverčiu A . Tada jį galima skaičiuoti pagal formulę:

$$A = \frac{\sum_{i=1}^N T_i}{T} \quad (2)$$

Čia T_i - laiko tarpas, kai paslauga buvo teikiama, T - pateikiamumo vertinimo laikotarpis (pvz. mėnuo).



2 pav. Duomenų centro paslaugos didelio pateikiamumo struktūrinė schema

Laiko dalis per nagrinėjamą laikotarpį, kai paslauga nebuvo teikiama – nepateikiamumas:

$$\bar{A} = 1 - \frac{\sum_{i=1}^N T_i}{T} \quad (3)$$

Bendras sistemos pateikiamumas gali būti nustatytas laipsniškai suskaidant sistemą į atskirus elementus (1 pav.). Struktūriniai elementai gali būti sujungti nuosekliai ir lygiagrečiai, o iš tokių sujungimų formuojamos sudėtingesnės architektūros - rezervuoti ir paskirstomi elementai (2 pav.). Jų sujungimo tvarka turi didelę įtaką bendram sistemos pateikiamumui.

2 pav. struktūriniai elementai:

- A_{Ri} , kur $i = \overline{1, n}$, - duomenų perdavimo tinklo elementų (čia maršrutizatorių) pateikiamumai.
- A_{Pi} , kur $j = \overline{1, m}$, - pirminių serverių klasterio elementų (čia taikomųjų programų ir serverinių programų) pateikiamumai.

Serverių klasteris - tai serverių, atliekančių tas pačias funkcijas, rinkinys, skirtas padidinti sistemos našumą. Vartotojui matomas kaip vienas užklausas aptarnaujantis serveris.

- A_{Fk} , kur $k = \overline{1, o}$, - galinių serverių klasterio elementų (čia failų serverių) pateikiamumai.
- A_{Ds} , kur $l = \overline{1, s}$, - duomenų saugyklų elementų (čia diskų masyvų) pateikiamumai.

Duomenų centro struktūrinių elementų nuoseklus sujungimas – kai vienam veiksmui atlikti naudojami keli nuosekliai sujungti elementai. Iš n nuosekliai sujungtų elementų grandinės bendras pateikiamumas skaičiuojamas pagal formulę:

$$A = A_1 \cdot A_2 \cdot \dots \cdot A_n \quad (4),$$

kur A_n - atskirų elementų pateikiamumo įverčiai. Šiuo atveju bendras sistemos pateikiamumas yra žemesnis nei silpniausiojo struktūrinio elemento pateikiamumas.

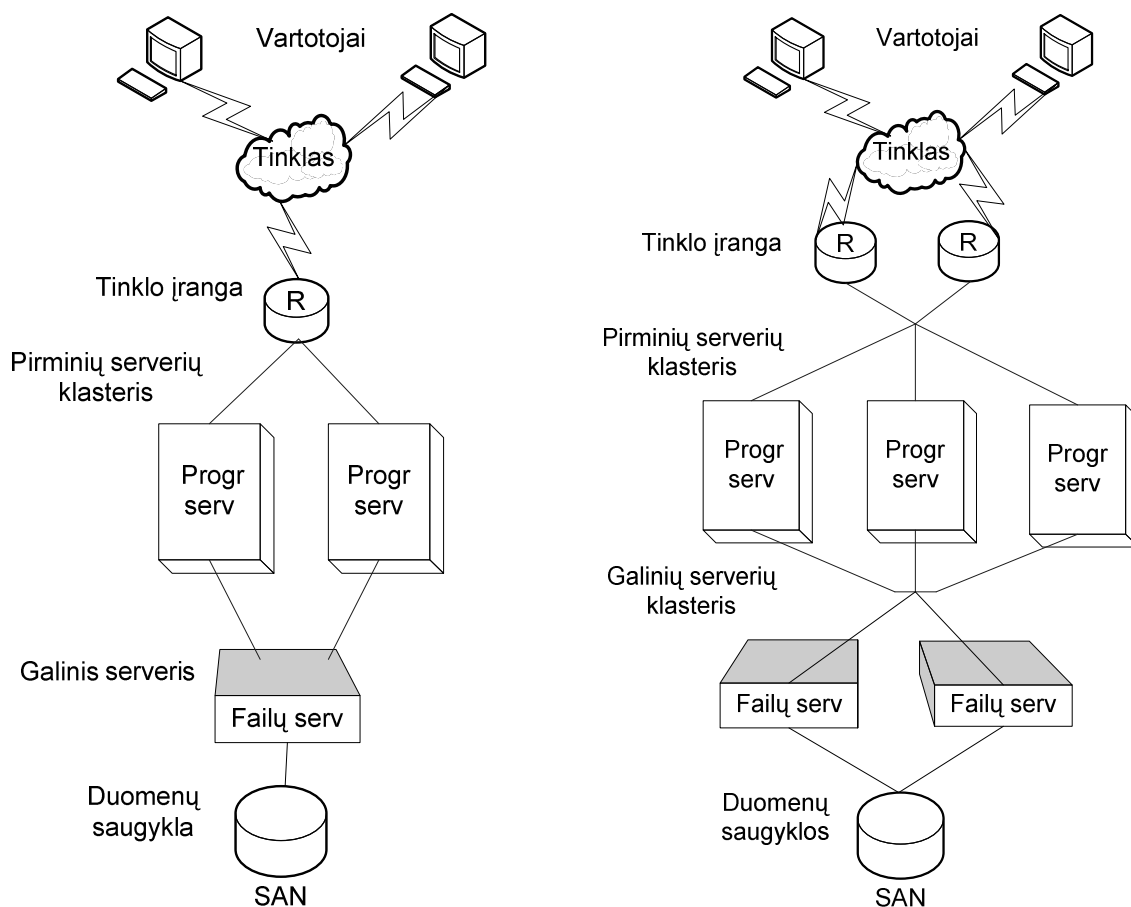
Duomenų centro struktūrinių elementų lygiagretus sujungimas – atitinka elementų rezervavimą. Iš m lygiagrečiai sujungtų elementų grandinės bendras pateikiamumas skaičiuojamas pagal formulę:

$$A = 1 - (\bar{A}_1 \cdot \bar{A}_2 \cdot \dots \cdot \bar{A}_m) \quad (5),$$

kur $\bar{A}_m$ - atskirų elementų pateikiamumų atvirkštiniai įverčiai ($\bar{A} = 1 - A$). Šiuo atveju bendras pateikiamumas gali būti daug didesnis nei atskirų komponentų pateikiamumai. Taigi toks komponentų diegimas turėtų būti naudojamas pateikiamumo padidinimui. Jis labai patogus dar ir

dėl to, kad galima diegti palaipsniui: iš pradžių nedubliuoti, po to dubliuoti kritiškiausias sistemos veiklas, taip didinant bendrą duomenų centro paslaugos pateikiamumą.

Panagrinėkime dviejų skirtingų struktūrų duomenų centrų, pateiktų 3 pav., pateikiamumą palyginimą pavyzdį. Duomenų centro sistemos architektūra su taikomųjų serverių rezervavimu pateikta 3.a pav. Sudėtingesnė architektūra (su tinklo maršrutų, taikomųjų programų serverių ir failų serverių rezervavimais) parodyta 3.b pav. Pavyzdyje abiejų struktūrų kiekvieno elemento pateikiamumas yra 0,9 (Realiai elementų pateikiamumas yra daug artimesnis vienetui).



3 pav. Duomenų centro paslaugos pateikiamumo įvertinimo atvejai: a) – mažai rezervuotų įrenginių; b) – daug rezervuotų įrenginių

Skaičiuojame atskirai lygiagrečius sujungimus naudodamiesi (5) formule, ir paskaičiuojame bendrą paslaugos pateikiamumo įvertį naudodamiesi (4) formule.

Mažai rezervuotos struktūros pateikiamumas:

Pirminių serverio klasterio pateikiamumas:

$$A_p = 1 - (\overline{A_{p1}} \cdot \overline{A_{p2}}) = 1 - (0,9 \cdot 0,9) = 0,99 \quad (6)$$

Bendras sistemos pateikiamumas:

$$A_1 = A_R \cdot A_P \cdot A_F \cdot A_D = 0,9 \cdot 0,99 \cdot 0,9 \cdot 0,9 = 0,72171 \quad (7)$$

Daug rezervuotos struktūros pateikiamumas:

Duomenų perdavimo tinklo įrangos pateikiamumas:

$$A_R = 1 - (\overline{A_{R1}} \cdot \overline{A_{R2}}) = 1 - (0,9 \cdot 0,9) = 0,99 \quad (8)$$

Pirminių serverio klasterio pateikiamumas:

$$A_P = 1 - (\overline{A_{P1}} \cdot \overline{A_{P2}} \cdot \overline{A_{P3}}) = 1 - (0,9 \cdot 0,9 \cdot 0,9) = 0,999 \quad (9)$$

Galinių serverio klasterio pateikiamumas:

$$A_F = 1 - (\overline{A_{F1}} \cdot \overline{A_{F2}}) = 1 - (0,9 \cdot 0,9) = 0,99 \quad (10)$$

Bendras paslaugos pateikiamumas:

$$A_2 = A_R \cdot A_P \cdot A_F \cdot A_D = 0,99 \cdot 0,999 \cdot 0,99 \cdot 0,9 \approx 0,88121 \quad (11)$$

Pateikiamumas padidėjo nuo 0,72 iki 0,88. Antru atveju bendrą pateikiamumą lemia duomenų saugyklos pateikiamumas. Tokiu atveju tikslinga rezervuoti duomenų saugyklą.

Skirtingų paslaugų pateikiamumai gali būti interpretuojami skirtingai. T.y. jis priklauso nuo to, kokie įrenginiai privalo veikti teikiant paslaugą. Pvz. jeigu kalbama apie duomenų perdavimo tinklo pateikiamumą – jis priklauso nuo tinklo įrenginių veikimo; interneto prieigos pateikiamumą – tiek nuo reikiamų duomenų perdavimo tinklo įrenginių, tiek nuo išorinių tinklų operatorių veikimo; tam tikrų serverių pateikiamumą – nuo visos eilės įrenginių (pradedant nuo duomenų perdavimo tinklo lygio), leidžiantis žemyn, iki reikiamo serverio aparatūrinės, o kai kuriais atvejais ir programinės įrangos. Taigi tuo pačiu gali apimti ir duomenų centro vietinį atsarginių kopijų darymą ir atkūrimą iš jų. T.y. būtina tiksliai suprasti, ką paslaugos teikėjas konkrečiu atveju išsipareigoja.

Kitas pateikiamumo skaičiavimo metodas remiasi duomenų centro paslaugos gedimų skaičiavimu. Produktų specifikacijose naudojamas dydis MTBF (*angl. Mean Time Between Failure*) – vidutinis laiko tarpas tarp gedimų. Jis apibūdina tam tikro įrenginio valandų skaičių tarp gedimų. MTTR (*angl. Mean Time to Repair*) – vidutinis laiko tarpas tarp sistemos sugedimo ir jo atstatymo. Galima kalbėti apie sistemą ar tam tikrą jos dalį apibendrintai, arba detalesniu lygiu.

Žinant tam tikro įtaiso MTBF ir MTTR, galima tiesiogiai paskaičiuoti pateikiamumą:

$$A = \frac{MTBF}{MTBF + MTTR} \quad (12)$$

Sistemos pateikiamumas stipriai surištas su ją sudarančių komponentų pateikiamumu. Mažinant MTTR (sistemų atstatymo laiką) didėja pateikiamumas. Nagrinėjant duomenų centro paslaugos struktūrą, parodytą 2 pav., kuri susideda iš visos eilės architektūrinių lygių, galima sakyti, kad į tarnybinės stoties (ar stočių) MTTR įeina [10]:

1. Diagnostikos fazė – problemos diagnozavimas ir jos sprendimo veiksmų nustatymas.
2. Pirminė aprūpinimo fazė – nustatyti, surasti, transportuoti ir fiziškai instaliuoti reikiamą aparatūrinę, programinę įrangą ir/ar atsarginių kopijų laikmenas.
3. Pagrindinė aprūpinimo fazė – sistemos aparatūrinės dalies konfigūracija ir bazinės operacinės sistemos instaliavimas.
4. Atkūrimo fazė – atkurti visą sistemą iš atsarginių kopijų laikmenų, įskaitant sisteminius failus ir vartotojų duomenis.
5. Patikrinimo fazė – patikrinti visos sistemos funkcionalumą ir vartotojų duomenų vientisumą.

Nepaisant to, kas įrašyta SLA, reikia žinoti, kiek laiko truks kiekviena fazė, t.y. per kiek laiko realiai galima atstatyti pilną sistemos darbą. Norint sumažinti bendrą laiką, yra siekiama sumažinti problemos diagnozavimo laiką, t.y. jeigu matoma, kad jis bus didesnis nei pvz. 15 minučių, yra pereinama prie antros fazės. Ji taip pat gali trukti ilgai, jeigu pvz. atsarginių kopijų laikmenos yra toli nuo sistemos, tada jų pristatymo laikas pailgėja.

Nagrinėjant pavyzdį, kai turima 3 valandos 43 minutės [10] (nuo gedimo pradžios) pilnai atstatyti sistemą, iš tiesų yra tik apie 2 valandas įvykdyti realią sistemos atstatymo procedūrą, nes likusį laiką užima diagnozavimo, aprūpinimo bei patikrinimo fazės.

Realiose duomenų centrų paslaugų sistemose pateikiamumas turi būti matuojamas periodiškai renkant atitinkamos paslaugos veikimo statistiką. Tai gali būti daroma tiek iš paslaugos teikėjo, tiek ir iš kliento perspektyvų. Tiek matavimai, tiek sistemų skaičiavimai (įvertinant atskirus komponentus ir jų parametrus) paslaugos teikėjui suteikia informaciją, kuriuos sistemos komponentus reikėtų keisti kitais, norint laikytis visų SLA įsipareigojimų. Yra visa eilė matavimo technologijų ir realizacijų, apie kurias kalbama 4-ame skyriuje.

3.2. Duomenų perdavimo tinklo vėlinimo įvertinimas

Duomenų perdavimo tinklo vėlinimas – tai dydis (dažniausiai išreiškiamas milisekundėmis), parodantis, kiek laiko truko informacijos paketo persiuntimas reikiamam gavėjui. Tipiniu atveju vartojamas parametras yra dvikryptis vėlinimas (*angl. RTT Round Trip Time*), parodantis trukmę

tarp išsiųsto paketo ir gauto atsakymo į jį. Kadangi yra norima vėlinimo parametą naudoti (1) formulėje, reikia jį normuoti santykinu dydžiu [0;1] intervale - t.y. apibrėžti kokybiniu įverčiu NL . Tada jį galima skaičiuoti pagal formulę [20]:

$$NL = \frac{NL_{SLA} - NL'}{NL_{SLA} - NL_{\min}}, \quad (13)$$

kur NL' - vidurkinė duomenų perdavimo tinklo vėlinimo reikšmė (skaičiuojama apibrėžtam laiko momentui), NL_{SLA} - duomenų perdavimo tinklo vėlinimo reikšmė SLA sutartyje, $NL_{\min}$ - minimali duomenų perdavimo tinklo vėlinimo reikšmė. Čia visada $NL_{\min} < NL_{SLA}$.

3.3. Duomenų perdavimo tinklo vėlinimo sklaidos įvertinimas

Vėlinimo sklaida (*angl. jitter, delay variation*) – tai dydis (dažniausiai išreiškiamas milisekundėmis), parodantis vėlinimo pokyčius laike. Nors šis parametras kai kurioms paslaugoms nėra labai aktualus, tačiau tai ypač aktualu VoIP (balsas per IP) paslaugai, kuri veikia UDP paketais – paketų srautas turi eiti tolygiai ir nuosekliai, norint, kad balso kokybė būtų aukšta.

Vėlinimo sklaida gali būti teigiama (vėlinimas išaugo, lyginant su prieš tai buvusio paketo vėlinimu), neigiama (vėlinimas sumažėjo, lyginant su prieš tai buvusio paketo vėlinimu), nulinė (vėlinimas nepasikeitė, lyginant su prieš tai buvusio paketo vėlinimu).

Vėlinimo sklaida išmatuojama abejoms kryptims atskirai, ir šiems matavimams nereikalinga tiksli įrenginių laiko sinchronizacija, nes yra skaičiuojamas vėlinimo pokytis tarp gretimų paketų porų. Dėl tokio skaičiavimo principo statistiniai rezultatai gali gautis skirtingi, lyginant su vėlinimo sklaidos statistiniais skaičiavimais, įvertinant sumines paketų emisijos vėlinimų sklaidos reikšmes.

Kadangi yra norima vėlinimo sklaidos parametą naudoti (1) formulėje, reikia jį normuoti santykinu dydžiu [0;1] intervale - t.y. apibrėžti kokybiniu įverčiu J . Tada jį galima skaičiuoti pagal formulę [20]:

$$J = \frac{J_{SLA} - J'}{J_{SLA} - J_{\min}}, \quad (14)$$

kur J' - vidurkinė duomenų perdavimo tinklo vėlinimo sklaidos reikšmė milisekundėmis (skaičiuojama apibrėžtam laiko momentui), J_{SLA} - duomenų perdavimo tinklo vėlinimo sklaidos reikšmė SLA sutartyje, $J_{\min}$ - minimali duomenų perdavimo tinklo vėlinimo sklaidos reikšmė. Čia visada $J_{\min} < J_{SLA}$.

3.4. Duomenų perdavimo tinklo paketų pristatymo įvertinimas

Duomenų perdavimo tinklo paketų pristatymas (*angl. packet delivery*) – tai parametras, parodantis kuri dalis paketų per nagrinėjamą laikotarpį (dažniausiai per mėnesį) yra sėkmingai persiunčiama. Dažnai jis išreiškiamas procentais. Kadangi yra norima paketų pristatymo parametą naudoti (1) formulėje, reikia jį normuoti [0;1] intervale - t.y. apibrėžti kokybiniu įverčiu D . Tada jį galima skaičiuoti pagal formulę [20]:

$$D = \frac{\sum_{i=1}^N D_i}{D_{sum}} \quad (15)$$

Čia D_i - sėkmingai persiųstų paketų skaičius per tam tikrą laiko intervalą, D_{sum} - bendras per visą laikotarpį (pvz. per mėnesį) persiųstų paketų skaičius.

3.5. Informacijos perdavimo sparta duomenų perdavimo tinklu

Informacijos perdavimo sparta (dažnai dar vadinama pralaidumu) – tai dydis (dažniausiai išreiškiamas bitais per sekundę: bps, Kbps, Mbps), parodantis kiek duomenų tinklas gali persiųsti per laiko vienetą. Kadangi yra norima informacijos perdavimo spartos parametą naudoti (1) formulėje, reikia jį normuoti [0;1] intervale - t.y. apibrėžti kokybiniu įverčiu G . Tada jį galima skaičiuoti pagal formulę [20]:

$$G = \frac{G'}{G_{SLA}} \quad (16)$$

kur G' - vidurkinė informacijos perdavimo spartos reikšmė bitais per sekundę (skaičiuojama apibrėžtam laiko momentui), G_{SLA} - informacijos perdavimo spartos reikšmė, kuri nurodyta SLA sutartyje. Norint stebėti informacijos perdavimo spartą, turėtų būti periodiškai atliekami jos matavimai (siunčiami duomenys ir matuojamas greitis), po to reikšmės vidurkinamos apibrėžtam laiko momentui.

Idealiu atveju maksimali informacijos perdavimo sparta turėtų būti lygi reikšmei, kuri įrašyta SLA sutartyje - duomenų perdavimo tinklo ryšio linija pilnai išnaudojama, tačiau realiu atveju skirtingose situacijose ši sparta būna mažesnė. Tai iš dalies priklauso nuo tinklu siunčiamų duomenų, o taip pat nuo paslaugos tipo. T.y. ryšio linija gali būti rezervuota specialiai tik vienam klientui (tada niekas kitas negali naudotis jos resursais), arba nerezervuota (keli vartotojai gali naudoti tą pačią liniją). Antruoju atveju paprastai informacijos perdavimo spartos išipareigojimas

SLA būna mažiau apibrėžtas, t.y. nurodoma maksimali galima ryšio linijos informacijos perdavimo sparta, tačiau ji nėra garantuojama.

3.6. Informacijos saugumas

Informacijos saugumas gali būti suprantamas labai skirtingai. Jeigu kalbama tik apie duomenų perdavimo tinklu keliaujančių duomenų saugumą, tai reikia įvertinti tik tinklo įrenginių ir ryšio linijų saugumą šiuo požiūriu. Tačiau jeigu kalbama ir apie serveriais keliaujančių bei laikomų duomenų saugumą, tai saugumas gali apimti visą minėtą elementų grandinę.

Paprastai saugumo parametras parodo, kaip sistemoje yra užtikrinamas vartotojų informacijos konfidencialumas ir saugojimas. Tai užtikrinama šifravimo bei didelio patikimumo autentifikacijos mechanizmais. Tinklo lygyje naudojama VPN (*angl. Virtual Private Network*) - virtualaus privataus tinklo technologija, pagrįsta šifruotų tunelių kūrimu, kurių pagalba užtikrinamas saugus duomenų perdavimas net ir per viešo naudojimo duomenų perdavimo tinklus, kaip pvz. internetas. Persiunčiamos informacijos kodavimui (pvz. el. paštas) naudojama viešo-privataus rakto technologija ir skaitmeninis parašas. Vartotojų autentifikacijai naudojami slaptažodžiai, skaitmeniniai sertifikatai ir ID kortelės.

Duomenų atsarginių kopijų darymą taip pat galima sieti su informacijos saugumu, nes kopijos reikalingos, norint atkurti duomenis sistemų gedimo, duomenų sugadinimo, stichinių nelaimių atvejais. Paprastai duomenys kiekvieną naktį archyvuojami naudojantis juostiniais įrenginiais, taip pat atliekamos ir savaitinės bei mėnesio kopijos, priklausomai nuo vartotojo pageidaujamos paslaugos. Kai kurioms paslaugoms dar vykdomos ir papildomų kopijų kūrimas, kurios išvežamos iš duomenų centro ir saugomos pas vartotoją, arba tam skirtose saugyklose.

Informacijos saugumą konkrečiai duomenų centro paslaugai galima įvertinti, apibrėžus ir įvertintus minėtas priemones saugumo parametrais, ir naudojantis formule [20]:

$$S = \frac{\sum_{i=1}^N (I_i \cdot Z_i)}{\sum_{i=1}^N I_i} \quad (17)$$

Čia Z_i - i -ojo saugumo parametro tenkinimas: $Z_i = 1$ - jeigu tenkinamas, ir $Z_i = 0$, jeigu netenkinamas. I_i yra atitinkamų parametrų svoris (pvz. [0;10] intervale), kuris yra nustatomas pagal vartotojo poreikius). N yra vertinamų saugumo parametrų skaičius.

3.7. Duomenų centro paslaugos vėlinimas

Duomenų centro paslaugos vėlinimas - tai dydis (dažniausiai išreiškiamas milisekundėmis), parodantis, kiek laiko truko tam tikros duomenų centro paslaugos informacijos gavimas, išsiuntus užklausą duomenų centrui. Tai yra vienas naujausių kokybės parametrų - konkrečių duomenų centro paslaugų vėlinimas. Šį dydį taip pat siūloma pateikti SLA sutartyse [14],[16]. Tradicinėse sutartyse (daugiausia duomenų perdavimo teikėjų sutartyse) šis parametras nebuvo minimas. Konkretios paslaugos vėlinimą įtakoja ne vien tik duomenų perdavimo tinklo vėlinimas, bet ir atskirų sistemos komponentų bendras vėlinimas. T.y. kiekvienas serveris ir programinė įranga sulaiko informacijos paketus, nes pačios informacijos apdorojimas tampa gerokai sudėtingesnis – ne tik jų perdavimas tinklu (kaip duomenų perdavimo tinklo vėlinimo atveju). Kadangi plečiantis duomenų centrams iškyla poreikis vertinti paslaugos vėlinimą ne tik pagal netiesioginius parametrus (duomenų perdavimo tinklo vėlinimas), tai pastaruoju metu nagrinėjama galimybė įvertinti ir šią metriką, teikiant aukštos kokybės IT paslaugas.

Kadangi yra norima duomenų centro vėlinimo parametą naudoti (1) formulėje, reikia jį normuoti santykinu dydžiu $[0;1]$ intervale - t.y. apibrėžti kokybinį įverčiu L . Tada jį galima skaičiuoti pagal formulę [20]:

$$L = \frac{L_{SLA} - L'}{L_{SLA} - L_{\min}}, \quad (18)$$

kur L' - vidurkinė paslaugos vėlinimo reikšmė milisekundėmis (skaičiuojama apibrėžtam laiko momentui), L_{SLA} - paslaugos vėlinimo reikšmė SLA sutartyje, $L_{\min}$ - minimali paslaugos vėlinimo reikšmė. Čia visada $L_{\min} < L_{SLA}$.

Apie tai plačiau kalbama 6-ame ir 7-ame šio darbo skyriuose.

Tiek minėti duomenų perdavimo tinklo kokybiniai parametrai, tiek ir bendras tam tikros duomenų centro paslaugos vėlinimas SLA sutartyse paprastai specifikuojami tam tikram duomenų perdavimo tinklo segmentui. T.y. jeigu duomenų centrų paslaugų teikėjas turi savo duomenų perdavimo tinklą, arba turi susitarimus su kitais duomenų perdavimo tinklų paslaugų teikėjais, tada jis gali prisiimti įsipareigojimą užtikrinti reikiamas šių parametrų reikšmes tam tikruose (tiksliai apibrėžtuose) tinklo segmentuose. Kitu atveju paslaugos teikėjas negali įsipareigoti, nes tokie parametrai kituose tinkluose (pvz. viešajame internete) nėra pastovūs ir užtikrinti.

Norėdamas laikytis SLA nurodytų reikalavimų, paslaugos teikėjas privalo turėti šių parametrų užtikrinimo metodikas, ir naudotis atitinkamomis priemonėmis, norint teikti maksimalios galimos kokybės paslaugas. Realiose duomenų centrų paslaugų sistemose minėti parametrai turi būti matuojami periodiškai renkant paslaugos veikimo statistiką. Tai gali būti daroma tiek iš duomenų centro paslaugos teikėjo, tiek ir iš kliento perspektyvų. Šie matavimai suteikia paslaugos teikėjui reikiamą informaciją, kuriuos sistemos komponentus reikėtų keisti kitais, norint laikytis visų SLA įsipareigojimų. Yra visa eilė matavimo technologijų ir realizacijų, apie kurias kalbama 4-ame skyriuje.

4. SLA (Service Level Agreement) PARAMETRŲ MATAVIMO METODIKOS

Duomenų perdavimo tinklo parametrų reikšmės galima stebėti naudojant tiesioginius (pasyvius) arba netiesioginius (aktyvius) matavimus. Tiesioginiai matavimai - reikalingas duomenų perdavimo tinklo įrenginiais keliaujančių informacijos paketų pastovus stebėjimas. Netiesioginiai matavimai - veikia siunčiant kontrolinius paketus ir iš gautų parodymų sprendžiant apie tinklo būklę.

Jeigu reikia matuoti tam tikrą duomenų perdavimo tinklo ryšio maršrutą (nuo A maršrutizatoriaus iki B), kurio tarpe gali būti visa eilė kitų tinklo įrenginių, tai tiesioginių matavimų (pasyvių) atveju reikėtų atlikti kiekvieno iš tarpinių tinklo įrenginių matavimus - nuskaityti sukaupią informaciją nuo kiekvieno tinklo įrenginio interfeiso, ją palyginti, atmesti netinkamus paketus ir t.t. Taip pat matavimų periodiškumas ir rezultatų išbaigtumas labai priklausytų nuo realiu laiku duomenų perdavimo tinklu keliaujančių vartotojų paketų ir jų parametrų vertinimo. Tą pačią užduotį atliekant netiesioginiais matavimais reikėtų įvertinti tik galinių nagrinėjamos ryšio linijos tinklo įrenginių gautus rezultatus, o ir pats matavimo sistemos įrengimas bei valdymas būtų paprastesnis. Tačiau netiesioginiai matavimai įveda paklaidas dėl pačių kontrolinių paketų siuntimo. T.y., pvz. lėtose ryšio linijose atsiranda pastebimos paklaidos, nes kontroliniai paketai papildomai apkrauna tinklą. Tuo tarpu pasyvūs matavimo metodai remiasi tinkle keliaujančių paketų stebėjimu ir ryšio linijų neapkrauna.

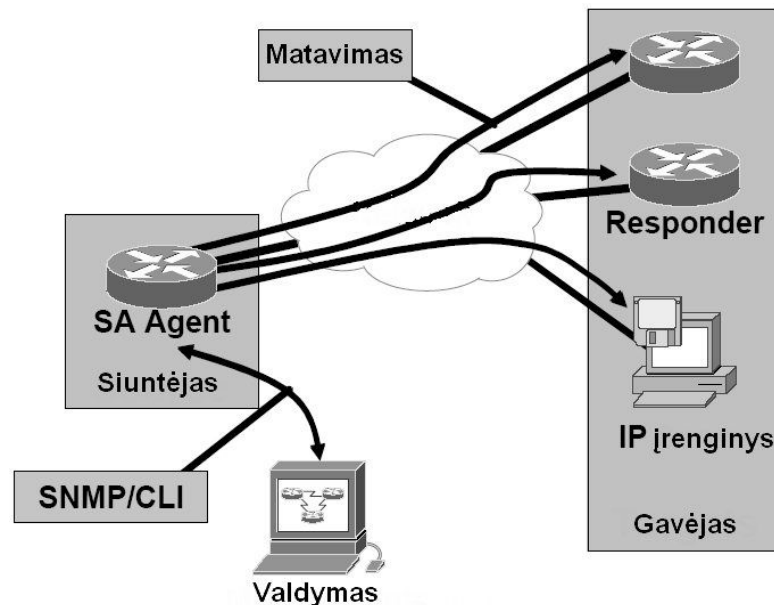
Visgi paskutiniu metu, dėl minėtų priežasčių, SLA parametrų matavimams daugiausia naudojami netiesioginiai (aktyvūs) metodai, todėl čia ir išnagrinėtos dvi plačiausiai naudojamos pagal šią metodiką skirtingai veikiančios tinklo parametrų matavimų technologijos.

4.1. CISCO SAA (Service Assurance Agent) technologija

Cisco „Service Assurance Agent“ (SAA), kuris ankstesnėse versijose buvo žinomas kaip „Response Time Reporter“ (RTR), yra Cisco įrenginių operacijų sistemos IOS (Internetwork Operating System) dalis [5]. Kadangi „Cisco Systems“ yra didžiausias duomenų perdavimo tinklo įrenginių gamintojas pasaulyje, tai ši technologija yra labiausiai paplitusi - jos pagrindu sukurta daug programinės įrangos. SAA yra nuolat tobulinama, o su kiekviena Cisco IOS programinės įrangos versija pridedama vis daugiau naujų funkcijų. Ji leidžia stebėti duomenų perdavimo tinklo našumo parametrus tarp Cisco įrenginio ir kito nutolusio įrenginio (kito Cisco įrenginio ar bet kokio IP įrenginio). Matavimams nereikalingi skirtiniai matavimų įrenginiai, galima išnaudoti

esamus Cisco įrenginius, todėl ši technologija yra labai plačiai naudojama įvairių duomenų perdavimo paslaugų teikėjų.

Tai netiesioginis (aktyvus) matavimų metodas, kurio metu įrenginys „siuntėjas“ siunčia dirbtinius kontrolinius paketus (tam tikrą skaičių paketų – paketų emisiją), kurie įterpiami į vartotojų srautą. Pasirinkus tinkamas matavimų sąlygas, pagal šių paketų matavimų parodymus galima spręsti apie viso likusio srauto parametrus.



4 pav. SAA architektūra

4 pav. pateiktoje SAA architektūros diagramoje matome tinklo įrenginį „Responder“ (atsakiklis). Tai taip pat Cisco įrenginys, dirbantis atsakiklio režime. Jis yra reikalingas kai kuriems specifiniams SAA matavimams, norint padidinti matavimų tikslumą. Šiuo atveju abudu tinklo segmento įrenginiai turi turėti Cisco SAA mechanizmą, bet matavimai yra kuriami tik įrenginyje „siuntėjas“, o gavėjas veikia „SAA Responder“ režimu, t.y. atsakinėja į tam tikras siuntėjo užklausas.

Ši technologija yra pripažįstama kaip moderni, efektyvi ir nebrangi duomenų perdavimo tinklo parametrų stebėjimo priemonė, kuri gali būti naudojama paslaugų lygio sutarčių (SLA – Service Level Agreement) sąlygoms užtikrinti. Cisco SAA technologija leidžia atlikti tokių duomenų perdavimo tinklo parametrų matavimus [5]:

- UDP paketų atsako laikas, vienakryptis duomenų perdavimo tinklo vėlinimas ir vėlinimo sklaida (jitter)
- „Žingsnis-po-žingsnio“ (hop-by-hop) ICMP paketų atsako laikas ir vėlinimo sklaida
- Interneto puslapio našumo metrikos: DNS paieškos laikas, TCP prisijungimo ir HTTP transakcijų laikas

- Paketų praradimo statistika
- „Frame Relay“ grandinių statistika

Kaip jau buvo minėta, tinklo administratorius turi atlikti matavimų parametrų nustatymus kiekviename duomenų tinklo segmento Cisco įrenginyje atskirai. Tai gali būti atliekama, naudojant Cisco IOS CLI komandinės eilutės sąsają arba naudojant SNMP (Simple Network Management Protocol) protokolą per CISCO-RTTMON-MIB [4] valdymo informacinę bazę. Matavimų rezultatai taip pat gali būti peržiūrimi abiem minėtais būdais, tačiau priimtinesnis yra SNMP skaitiklių nuskaitymas, nes šis procesas gali būti automatizuotas, bei pateikiami išsamesni rezultatai.

Atliekant matavimų nustatymus per komandinės eilutės sąsają, reikalingas prisijungimas prie įrenginio ir privilegijuoto vartotojo slaptažodis, o pakeitimai daromi pagrindiniame įrenginio konfigūraciniame faile, todėl kyla grėsmė duomenų tinklo saugumui, nes tai padidina klaidų tikimybę. Kitas būdas atlikti konfigūraciją yra naudojant SNMP protokolą, tačiau čia yra tam tikrų apribojimų – ne visų tipų matavimai gali būti nustatyti per SNMP protokolą.

Tinklo įrenginių konfigūravimas rankiniu būdu yra sudėtingas ir neefektyvus – tinklo administratorius turi atlikti nustatymus kiekviename duomenų tinklo segmento įrenginyje atskirai, kas padidina klaidų tikimybę, sumažina tinklo saugumą ir padidina laiko sąnaudas. Rezultatai taip pat yra pateikiami neapdoroti, todėl neparodo konkrečių duomenų perdavimo tinklo parametrų reikšmių, kurias galima būtų analizuoti. Taigi turi būti naudojami papildomi programiniai paketai, kurie leistų atlikti matavimus ir gauti prasmingus rezultatus šio metodo pagalba. Yra visa eilė komercinių programinių produktų, veikiančių remiantis Cisco SAA:

- Infovista VistaView [6] – visa eilė duomenų perdavimo tinklo parametrų stebėjimo programų, skirtų atskiroms konkrečioms užduotims spręsti. Infovista VistaView for VoIP SAA – skirta VoIP (balsas per IP) paslaugos kokybės stebėjimui. Infovista VistaView for IP Service Level Agreements – skirta paslaugų lygio sutarčių (SLA – Service Level Agreement) sąlygoms užtikrinti.
- CiscoWorks Internetwork Performance Monitor (IPM) [7] - CiscoWorks programų paketo dalis, universali programa, leidžianti išnaudoti didžiąją dalį Cisco SAA metodo funkcijų. IPM sprendimas susideda iš IPM serverio, IPM klientinės programos ir Cisco SAA savybės tinklo įrenginiuose.
- Fluke Networks Response Watch [8] - dar viena SLA parametrų stebėjimo programa, galinti atlikti matavimus skirtingų lygių įrenginiams (pradedant serveriu, tinklo įrenginiais, baigiant kliento PC). Rezultatus kaupia duomenų bazėje, pateikia grafiniu ir tekstiniu pavidalais. Pranešimus apie SLA pažeidimus pateikia syslog pranešimų pavidalu.

SAA technologija suteikia galimybę apibrėžti kylančius ir krentančius parametrų slenksčius, siekiant stebėti duomenų perdavimo tinklo parametrus, specifikuotus paslaugų lygio sutartyse (SLA). Juos viršijus įrenginys gali pranešti apie tai tinklo valdymo sistemai (NMS – Network Management System) [5].

Skirtingi tinklo parametrai matuojami atliekant skirtingas operacijas. Žemiau esančioje lentelėje pateiktos pagrindinės SAA operacijos ir paminėtos jų pritaikymo galimybės.

2 lentelė. Pagrindinės Cisco SAA operacijos [5]

SAA operacija	Operacijos funkcionalumas	Pagrindiniai pritaikymai
IP tinklo operacijos		
UDP Echo	Abipusis vėlinimas	Tikslus UDP paketų atsako laiko matavimas
UDP Jitter	Abipusis vėlinimas, vienakryptis vėlinimas, vėlinimo sklaida (jitter), paketų praradimas Pastaba: vienakrypčio vėlinimo matavimui reikalinga tiksli įrenginių laiko sinchronizacija	Didžioji dalis operacijų tinklams, kurie perneša UDP srautą, pvz. VoIP (balsas per IP). VoIP yra ypatingai jautrus tinklo vėlinimo parametrų mokyčiams, todėl yra svarbu stebėti šiuos parametrus, teikiant VoIP paslaugą tinkle.
ICMP Echo	Abipusis vėlinimas	Gedimų paieška ir pateikiamumo (availability) stebėjimas
Paslaugų operacijos - aukštesnių OSI lygių		
TCP Connect	Prisijungimo laikas	Internetinio puslapio našumo stebėjimas
DNS	DNS paieškos laikas	Internetinio puslapio našumo stebėjimas
DHCP	Laikas, sugaištas gaunant IP adresą	Paslaugos našumo stebėjimas
FTP	Failo siuntimo abipusis vėlinimas	Paslaugos našumo stebėjimas
HTTP	Puslapio siuntimo abipusis vėlinimas	Paslaugos našumo stebėjimas

„UDP Jitter“ operacija gali būti tinkamai naudojama siekiant įvertinti paties duomenų perdavimo tinklo kokybinius parametrus (pateikti 2-oje lentelėje). Šie parametrai specifikuojami SLA sutartyse, apibrėžiant duomenų perdavimo tinklo veikimo išsipareigojimus. Duomenų centrų paslaugos veikia duomenų perdavimo tinklu, todėl tai yra viena iš esminių paslaugos kokybės dedamųjų.

2-oje lentelėje pateiktos paslaugų operacijos, kurios veikia aukštesniuose nei IP (Layer 3) OSI lygiuose, suteikia galimybę atlikti dažniausiai naudojamų IT paslaugų pateikiamumo ir našumo matavimus. Šios operacijos apima HTTP, DNS, FTP ir kitų TCP ir UDP protokolais veikiančių paslaugų testavimus.

Naudojantis šia į Cisco tinklinius įrenginius integruota technologija galima sukurti tokią duomenų centrų paslaugų teikėjo duomenų perdavimo tinklo architektūrą, kad pats tinklas (t.y. jo tiek galiniai, tiek tarpiniai įrenginiai) periodiškai atliktų tinklo paslaugų pateikiamumo ir našumo matavimus. Gauti rezultatai gali būti nuskaityti ir centralizuotai saugomi bei apdorojami paslaugų teikėjo. Tokiu būdu paslaugų teikėjas gali pastoviai stebėti vartotojų gaunamą paslaugų kokybę net jeigu vartotojas konkrečiu metu tos paslaugos ir nenaudoja.

Cisco SAA technologijos privalumai:

- Leidžia atlikti duomenų perdavimo tinklo parametrų matavimus kontroliniais paketais tiesiai iš komutacinių įrenginių (maršrutizatorių, komutatorių), nenaudojant specialių matavimo stočių.
- Gali atlikti visų pagrindinių duomenų perdavimo tinklo parametrų matavimus - tiek dvikrypčių, tiek ir vienakrypčių. Tuo pačiu atlieka ne tik tinklo (OSI Layer 3), bet ir aukštesniuose OSI sluoksniuose veikiančių protokolų kokybinius matavimus.
- Tinka matavimams tiek tarp pagrindinių (backbone) tinklo įrenginių, tiek ir iki vartotojo galinio įrenginio ar kitos tinklo aparatūrinės įrangos.

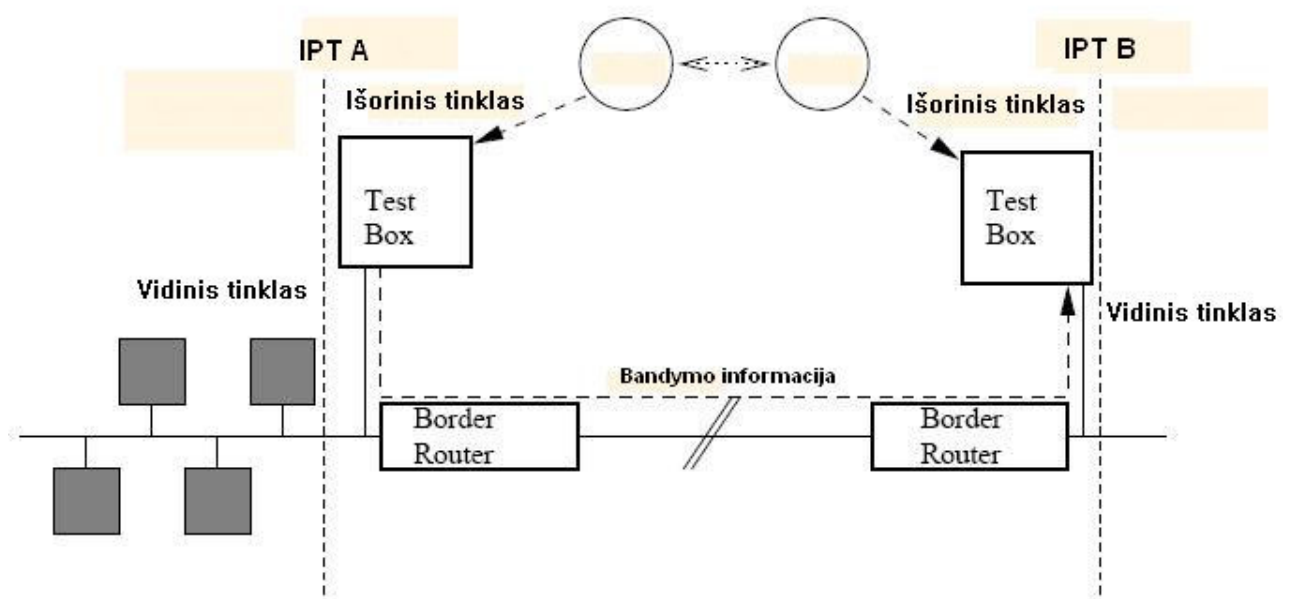
Cisco SAA technologijos trūkumai:

- Tai išskirtinai Cisco technologija, veikiantis tik su šio gamintojo įranga.
- Pati Cisco SAA yra tik technologija, bet nėra pilnai naudojimui paruoštas produktas. Jis vienas gali būti naudojamas tik testavimui. SLA parametrų stebėjimams realiame duomenų perdavimo tinkle turi būti naudojama papildoma programinė įranga, leidžianti patogiai konfigūruoti, planuoti, vykdyti matavimus, rinkti, apdoroti ir vizualizuoti rezultatus.
- Tie patys įrenginiai turi atlikti savo tiesioginę (maršrutizavimo ar komutavimo) užduotį, o taip pat atlikti ir matavimus, todėl kartais neužtenka turimų procesoriaus ir atminties resursų – taigi esant didesniai tinklo įrenginių apkrovimui, matavimuose atsiranda nemažos paklaidos.
- Pats SAA metodas neturi laiko sinchronizavimo galimybės, o tai yra reikalinga, norint tiksliai išmatuoti vienakrypčio vėlinimo (one-way delay) parametro reikšmes. Tinklinės įrangos laikas turi būti tiksliai sinchronizuotas naudojant GPS sistemą, šios problemos

sprendimas paliekamas duomenų perdavimo tinklo parametrų matavimo programinės įrangos kūrėjams bei tinklų projektuotojams.

4.2. RIPE TTM (Test Traffic Measurement) technologija

RIPE NCC sukurta technologija TTM (Test Traffic Measurements) [9], kuri kaip ir Cisco SAA, remiasi netiesioginiais (aktyviais) vienakrypčiais matavimais. TTM yra skirta atlikti matavimus tarp vieno paslaugos teikėjo duomenų perdavimo tinklo iki kito tinklo (5 pav.). Čia naudojami gamintojo (RIPE NCC) sukurti atskiri testiniai įrenginiai (test-box), susidedantys iš nepriklausomų personalinių kompiuterių. Jie yra pilnai paruošti darbui, ir jungiami tiesiai į testuojamo tinklo segmentą, norint atlikti matavimus. Dažniausiai jie jungiami tiesiai už tinklo ribinių maršrutizatorių (*angl. border routers*), norint leisti ir priimti testinius informacijos paketus tiksliai iš tos vietos, kur nuosavas duomenų perdavimo tinklas jungiasi su kitų paslaugų teikėjų tinklais (5 pav.).



5 pav. RIPE TTM architektūra

Ši testavimo technologija veikia su visų tipų IP tinklais ir nepriklauso nuo įvairių gamintojų tinklo įrenginių specifikos. RIPE TTM įrenginiai atlieka matavimus OSI tinklo lygyje (Layer 3), taigi nepriklauso nuo aukštesnio lygio taikomųjų programų naudojimo. Šis metodas atitinka RFC 2330, 2678-2681 standartus, kuriuos išleido IETF organizacijos „IP Performance Metrics“ darbo grupė, kuri testuoja bei standartizuoja IP tinklų kokybės parametrų metodikas.

RIPE TTM yra pilnai išbaigtas ir realiam naudojimui paruoštas produktas, galintis atlikti matavimus [9]:

- Duomenų perdavimo tinklo vėlinimo matavimai (dvikryptis (RTT) vėlinimas, vienakryptis (one-way) vėlinimas).
- Tinklo paketų praradimo procento matavimai.
- Tinklo vėlinimo sklaidos parametrų matavimai.
- Tinklo ryšio kanalo apkrovų (bandwidth) matavimai.

Matuojant vienakryptį vėlinimą testiniai paketai yra išsiunčiami iš IPT A tinklo IPT B tinklui. Prieš išsiuntimą jiems yra uždedamos laiko atžymos (*angl. timestamps*). Gavėjo tinkle pagal gauto paketo išsiuntimo laiką gali būti nesunkiai paskaičiuojamas vienos krypties vėlinimas.

Tiksliam testinių įrenginių laiko nustatymui naudojami GPS imtuvai. RIPE TTM įrenginiai taip pat gali veikti kaip NTP (Network Time Protocol) tarnybinės stotys, tiekiančios tikslas (10 mikrosekundžių tikslumo) „stratum 1“ tikslumo lygio laiko reikšmes visiems tinklo įrenginiams.

Panašiu principu atliekami ir vėlinimo sklaidos, paketų praradimų procento, kanalo apkrovų matavimai.

Duomenys yra siunčiami ir kaupiami centralizuotai RIPE NCC (Network Coordination Centre) serveriuose. RIPE NCC yra Europos regioninis interneto koordinatorius. Yra suformuojami ir kaupiami dienos, savaitės ir mėnesio grafikai, informatyviai pavaizduojantys duomenų perdavimo tinklo būklę bei jos kitimą. Tinklo administratorius taip pat turi galimybę pasiekti pilnus statistinius duomenis, gauti maršrutizavimo informaciją, bei analizuoti visus neapdorotus duomenis, kai ieškoma sudėtingų problemų sprendimų.

RIPE TTM technologijos privalumai:

- Veikia su visų tipų IP tinklais ir nepriklauso nuo įvairių gamintojų tinklo įrenginių specifikos
- Gali atlikti visų pagrindinių duomenų perdavimo tinklo parametrų matavimus - tiek dvikrypčių, tiek ir vienakrypčių.
- Tinka parametrų matavimams tarp skirtingų duomenų perdavimo paslaugų teikėjų tinklų.
- Pilnai paruoštas produktas - testiniai įrenginiai (*angl. test-boxes*) tiesiog jungiami į duomenų perdavimo tinklą reikiamoje vietoje, joks konfigūravimas nereikalingas - tai atlieka RIPE NCC (Network Coordination Centre).

RIPE TTM technologijos trūkumai:

- Netinka matavimams vidiniame duomenų perdavimo tinklo paslaugų teikėjo tinkle: tiek tarp pagrindinių (backbone) tinklo įrenginių, tiek ir iki vartotojo galinio įrenginio ar kitos tinklo aparatūrinės įrangos.
- Matavimų valdymas yra centralizuotas, taigi tinklo paslaugų teikėjas neturi galimybės valdyti testinius įrenginius.
- Neatlieka aukštesniuose OSI sluoksniuose veikiančių protokolų kokybinius matavimus (pvz. HTTP, FTP, DNS ir t.t.).

4.3. Išvados

Skirtingi duomenų perdavimo tinklo paslaugų teikėjai teikia skirtingo lygio paslaugas, todėl negalima vienareikšmiškai teigti, kuri technologija yra geresnė ar blogesnė - viskas priklauso nuo poreikių. Įvertinus prieš tai pateiktuose skyreliuose suformuluotus Cisco SAA ir RIPE TTM technologijų privalumus bei trūkumus, buvo pateiktos tokios išvados:

- Cisco SAA labiausiai tinka duomenų perdavimo tinklams, veikiantiems tik Cisco įrenginių pagrindu. Šiuo atveju nereikia diegti jokios kitos tinklo aparatūrinės įrangos. Tačiau kadangi Cisco SAA nėra paruošta realiam naudojimui, reikalinga įrengti tinklo valdymo sistemas (NMS - Network Management System), t.y. tam skirtus serverius, kuriuose veiktų specializuota programinė įranga, paremta Cisco SAA technologija matavimams atlikti. Šios programinės įrangos pasirinkimas yra labai didelis.
- RIPE TTM veikia su visų tipų IP tinklais ir nepriklauso nuo įvairių gamintojų tinklo įrenginių specifikos, tačiau kiekvienam tinklo taškui, kur norima atlikinėti matavimus, reikalinga įdiegti specializuotą aparatūrinę įrangą.
- Kadangi Cisco SAA matavimai atliekami iš realius informacinius paketus pernešančių tinklo įrenginių (maršrutizatorių ir komutatorių), didelio apkrovimo arba mažo pralaidumo ryšio linijose matavimų atlikimas įvestų dideles paklaidas arba net duomenų perdavimo tinklo sutrikimus. Tokiu atveju gali tekti tam tikruose tinklo segmentuose jungti papildomą tik matavimams skirtą įrangą. RIPE TTM šios problemos neturi, nes ir taip vartojami atskiri įrenginiai.
- Cisco SAA tinka matavimams tiek tarp pagrindinių (backbone) duomenų perdavimo tinklo įrenginių, tiek ir iki vartotojo galinio įrenginio ar kitos tinklo aparatūrinės įrangos, kai tuo tarpu RIPE TTM skirta matavimams tik tarp skirtingų duomenų perdavimo paslaugų teikėjų tinklų.

5. DUOMENŲ CENTRŲ DIDELIO PATEIKIAMUMO SERVERIŲ SISTEMOS

3-ame skyriuje buvo pateiktas pateikiamumo apibrėžimas ir jo skaičiavimo metodikos. Praktikoje skirtingoms sistemoms aukštas pateikiamumas užtikrinamas skirtingais būdais. Šiame skyrelyje apžvelgtos metodikos atskirų duomenų centrų paslaugų pateikiamumo užtikrinimui.

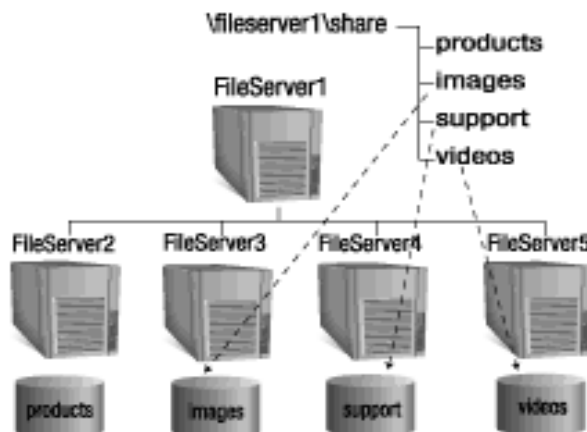
5.1. Didelio pateikiamumo failų serveriai duomenų centruose

Failų serveriai nereikalauja didelių procesoriaus ir operatyviosios atminties resursų. Pvz. tam, kad palaikyti 500 vartotojų ir 200GB duomenų, galima naudoti vieną mažą serverį, pvz. su dviem procesoriais ir 512MB RAM. Tai yra pigus sprendimas, kai yra norima teikti duomenų saugyklų paslaugas. Tarkime, kad serveris sugedo, reikia jį pakeisti nauju ir atstatyti visus 200GB duomenų. Jeigu kopijoms daryti naudosisime juostinį įrenginį, kurio vidutinis duomenų siuntimo greitis yra 5MB per sekundę, tai 200GB duomenų atkūrimas užtruks 666 minutes (11 valandų ir 6 minutes). Dar pridėjus papildomo laiko problemos diagnozavimui, aprūpinimui ir tikrinimui, gausime daugiau nei 12 valandų. Jeigu manysime, kad gali įvykti tik vienas toks gedimas per mėnesį laiko, tai šio failų serverio pateikiamumas SLA bus 98,37%.

Norint padidinti failų serverių pateikiamumą, galima siekti sumažinti sistemos atstatymo laiką gedimo atveju, o taip pat sumažinti pačių gedimų atsiradimų dažnį. Yra visa eilė skirtingų strategijų, kurios naudojamos įrengiant duomenų centrų failų serverius [10].

Duomenų padalijimas

Viena iš paprasčiausių strategijų – duomenų padalijimas (data partitioning).



6 pav. Pavyzdinis duomenų padalijimo scenarijus [10]

6 pav. pateiktoje pavyzdinėje schemoje, skirtinguose serveriuose yra laikomi skirtingi duomenys, t.y. jie yra išdalinti per visą eilę serverių. Vartotojas gali netgi nežinoti apie šią architektūrą, nes skirtingų serverių katalogai gali jam būti matomi kaip vieno serverio duomenys. Tai yra įgyvendinama sukuriant virtualią failų sistemą. Šios architektūros privalumas yra tas, kad yra sumažinamas duomenų atstatymo laikas, nes tas pats duomenų kiekis yra išdalintas per visą eilę serverių, t.y. kiekviename serveryje kaupiama gerokai mažiau duomenų. Todėl, sugedus vienam serveriui, jo atstatymo laikas bus gerokai trumpesnis, nes reikės atstatyti kur kas mažiau duomenų.

3 lentelė. SLA kaina ir duomenų padalijimas [10]

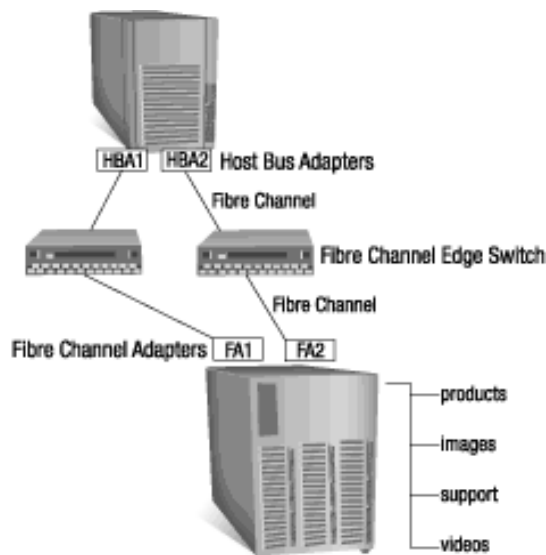
Serverių skaičius	Serverio duomenų talpa	Serverio reikiamas atstatymo laikas (min)	Neveikimo laikas (pagal SLA) (min)	SLA	Aparatūros kaina (pavyzdinė)
1	200GB	666	726	98.37%	\$9,700
2	100GB	333	393	99.12%	\$14,500
3	66GB	220	280	99.37%	\$20,100
4	50GB	166	226	99.49%	\$26,800
5	40GB	133	193	99.57%	\$33,500

3-oje lentelėje pateiktas SLA pateikiamumo palyginimas, vykdant duomenų padalijimą. Čia yra laikoma, kad vienu metu neveiks tik vienas serveris. Kiekvieno serverio atstatymo laikas mažėja, kai imama daugiau serverių ir kai serverių kietųjų diskų talpa mažėja. Savaimė suprantama, kad ši architektūra yra brangi aparatūrinės įrangos atžvilgiu, nes reikia daugiau atskirų serverių.

Momentinis duomenų saugojimas ir atkūrimas (Snapshot backup and restore)

Kita strategija yra mažinti duomenų atstatymo laiką, naudojant kitokio pobūdžio duomenų saugojimą, kad būtų įmanoma labai greitai atstatyti duomenis gedimo atveju.

7 pav. pateikta duomenų saugyklų tinklo (SAN) architektūra. Duomenų saugyklos yra jungiamos prie serverio per optinio kanalo sąsają (pageidautina dubliuotą) ir failų sistema vartotojui pasiekama kaip esanti pačiame serveryje. Tada gali būti naudojami specialūs momentinio duomenų saugojimo metodai, kurie labai greitai padaro SAN esančių duomenų atsargines kopijas, o, esant reikalui, taip pat greitai juos atkuria [10].



7 pav. SAN (Storage Area Network) – duomenų saugyklų tinklo architektūra [10]

Jeigu momentiniai duomenų nuskaitymai (snapshots) kuriami pakankamai dažnai, tai gedimo atveju atkurti duomenis iš atsarginių diskų įmanoma labai greitai, ir taip patenkinti netgi labai griežtus SLA reikalavimus. Jeigu norima išvengti SAN duomenų saugyklos gedimų, tai galima įrengti keletą duomenų saugyklų, kuriose duomenys būtų dubliuojami, o duomenų kanalus perjunginėtų specialūs optinio kanalo komutatoriai (fibre channel switch).

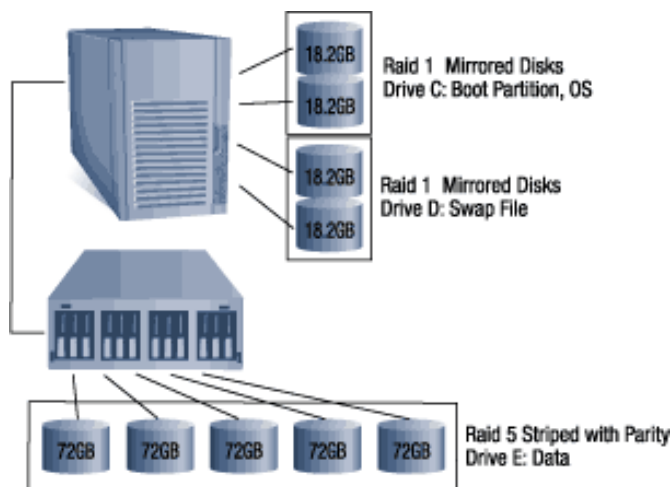
Dfs (distributed file system) technologija

Ši technologija įgyvendinama serverių aplinkoje, kuri buvo aprašyta kalbant apie duomenų padalijimą (6 pav.). Tačiau čia duomenis ne išdalinami pagal tam tikrus kriterijus, o yra dvigubinami skirtinguose serveriuose. Dfs failų sistema leidžia sukurti vientisą loginę failų sistemą, kurioje vartotojas mato tik duomenis, kurie fiziškai gali būti arba viename, arba kitame serveryje. Pagal nustatytą tvarkaraštį duomenys gali būti sinchronizuojami tarp serverių. Tokiu atveju, jeigu nėra pasiekiamas vienas serveris, kitas gali sėkmingai aptarnauti klientus. Tačiau ši technologija nėra pritaikyta greitai kintantiems duomenims, todėl kai kuriais atvejais nėra tinkama.

RAID (Redundant Array of Independent disks) diskų masyvai

Šios technologijos pagalba duomenys yra replikuojami tarp skirtingų diskų tame pačiame serveryje. Tyrimai rodo, kad kietieji diskai yra vieni iš labiausiai gendančių serverių komponentų. Nors ir diskų gamintojai pateikia labai dideles MTBF (Mean Time Between Failures) reikšmes, tačiau tai neatspindi realios situacijos [10]. Jų gedimai yra dažni ir juos sukelia visa eilė išorinių reiškinių.

Kuriant didelio pateikiamumo sistemas, naudojama gedimams atspari RAID technologija tam, kad diskų gedimai nesukeltų visos sistemos gedimo.



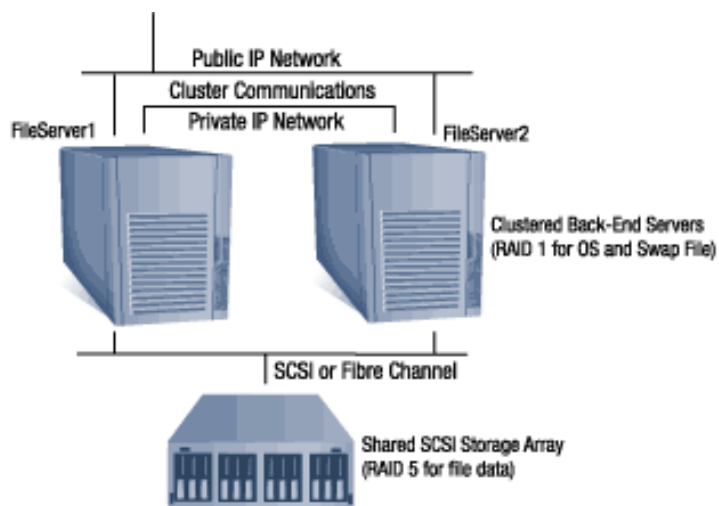
8 pav. Paprastos gedimams atsparios RAID sistemos konfigūracija [10]

8 pav. pateikta paprastos gedimams atsparios RAID sistemos konfigūracija. RAID 1 technologija dubliuoja (mirror) diskus realiu laiku, taigi jeigu vienas diskas sugenda ar duomenys susigadina, kitas diskas ir toliau dirba kaip įprastai ir sistemos našumas nesumažėja, nors tokiu atveju ji jau dirba ne kaip perteklinė sistema. 8 pav. pateiktoje sistemoje RAID 1 dubliavimas naudojamas serverio sisteminiam diskui ir virtualios diskinės atminties failų diskui (swap). Sugedus bet kuriam diskui, jis gali būti išimamas ir pakeičiamas nauju be sistemos perkrovimo. Įdėjus naują diską vietoje sugedusio, į naująjį įrašomas sisteminio disko atvaizdas ir sistema vėl tampa perteklinė (redundant).

Tačiau dubliuoti diskų laikmenas visais atvejais yra per brangu, nes kaina išauga dvigubai. Kita technologija RAID 5 turi kitokį veikimo principą: kiekviename diske yra išskiriama vietos pertekliniams duomenis (parity data), kurie gali būti naudojami sugadintiems ar trūkstamiems duomenims atstatyti. Tokia architektūra sumažina perteklinių diskų skaičių sistemoje iki vieno papildomo disko. RAID 5 sistemoje tik vieno disko vieta sistemoje nėra naudojama, taigi ši technologija gerokai pigesnė. 8 pav. pavaizduotas RAID 5 technologijos naudojimas duomenų saugyklai. Vieno disko gedimo atveju klientas tikriausiai netgi nepastebės diskų našumo pablogėjimo. 8 pav. pavaizduota architektūra yra tipinė duomenų centrų sistemoms.

Serverių klasteriai (Server clusters)

Priklausomai nuo duomenų centrų paskirties ir jo struktūros bei aparatūrinės įrangos, sekanti silpniausia grandis gali būti duomenų perdavimo tinklo įranga arba serveriai. Norint užtikrinti atskirų serverių sistemų perteklišumą, yra kuriami serverių klasteriai.



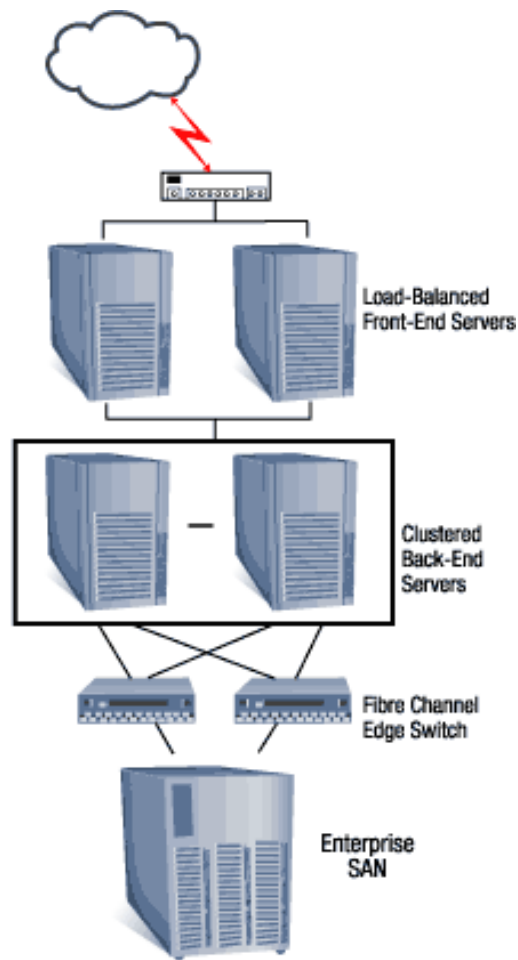
9 pav. RAID pagrindu sutvirtintas serverių klasteris [10]

9 pav. pavaizduotas serverių klasteris, kuriame yra panaudota RAID technologija. Serverių klasteris gali būti sukonfigūruotas įvairiai, bet pagrindinė mintis yra ta, kad jeigu vienas serveris sugenda, kitas serveris perima sugedusio funkcijas. Jeigu laikysime, kad 9 pav. yra pavaizduotas failų serverių klasteris, tai, esant serverio gedimui, vartotojas gali ir toliau dirbti su duomenų saugykloje saugomais duomenimis, nes ši saugykla yra pasiekama iš abiejų serverių. Gali būti pastebimas trumpas vėlinimas taikomosioms programoms persijungiant prie kito serverio. Tuo metu neveikiantis serveris gali būti atjungiamas ir taisomas nepažeidžiant vartotojų darbo ir SLA reikalavimų.

Serverių klasteriavimas nesumažina duomenų atkūrimo laiko, o taip pat ir nesukuria perteklinių duomenų saugyklų (taip, kaip RAID technologija). Klasteriai tiesiog padidina tų serverių pateikiamumą, kurie yra naudojami duomenų pristatymui iš duomenų saugyklos tam tikram vartotojui.

5.2. Didelio pateikiamumo Web serveriai duomenų centruose

Didelės ir visiems plačiai prieidamos interneto svetainės talpinamos į ypatingai didelio pateikiamumo sistemas, kuriose naudojamos įvairios technologijos, siekiant užtikrinti pateikiamumą.



10 pav. Didelio pateikiamumo Web serverių architektūra [10]

10 pav. pavaizduota didelio pateikiamumo Web svetainės architektūra, kurioje įdiegtos įvairios pateikiamumą užtikrinančios technologijos [10].

Schemos pradžioje matomi pertekliniai (redundant) pirminiai (front-end) Web serveriai, kurie yra subalansuoti pagal apkrovimą. Taigi vartotojas visada galės prisijungti prie svetainės.

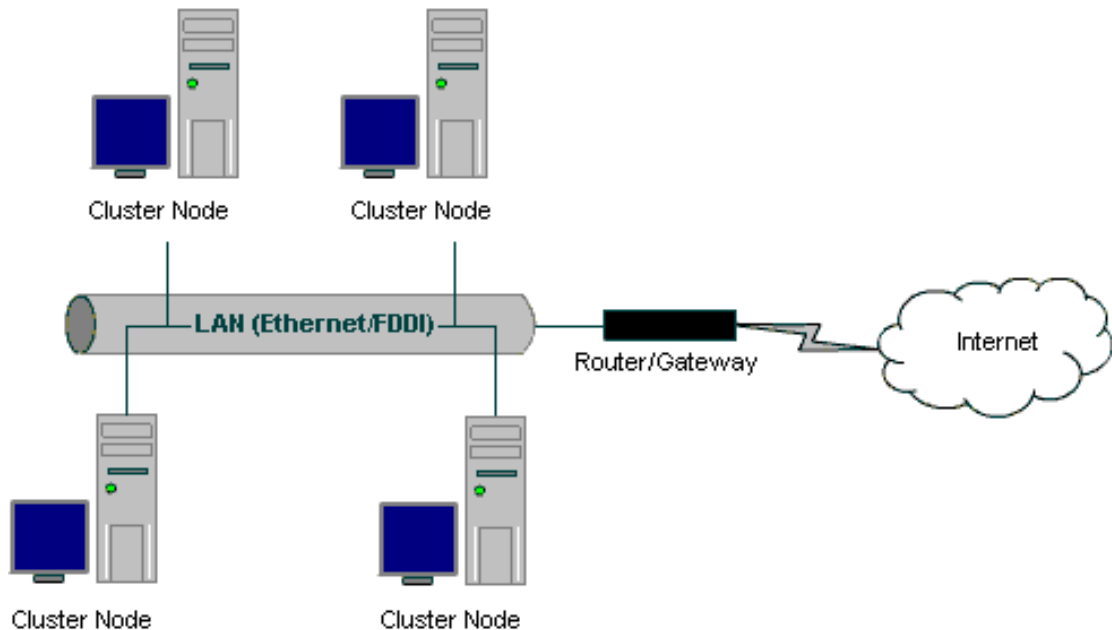
Žemiau esantis serverių klasteris užtikrina, kad serveris visada aptarnaus failų sistemos užklausas. Galiausiai, norint užtikrinti duomenų pateikiamumą, sistema naudoja perteklinę optinio kanalo (fibre channel) prieigą prie duomenų saugyklų tinklo (SAN).

Šioje architektūroje tiksliai pirminių (front-end) serverių ir antrinių (back-end) klasterių skaičius, o taip pat ir komunikacinių komponentų skaičius riboja šios architektūros pateikiamumą.

5.3. Tinklo apkrovos balansavimas

11 paveiksle pavaizduotas tinklo apkrovimo balansavimo (Network Load Balancing) klasteris, susidedantis iš keturių serverių. Network Load Balancing yra viena iš MS Windows serverio operacinės sistemos funkcijų, leidžiančių sukurti serverių klasterį programiškai ir be

jokios specializuotos aparatūrinės įrangos ir taip paskirstyti tinklo apkrovimą tolygiai visame klasteryje [11].



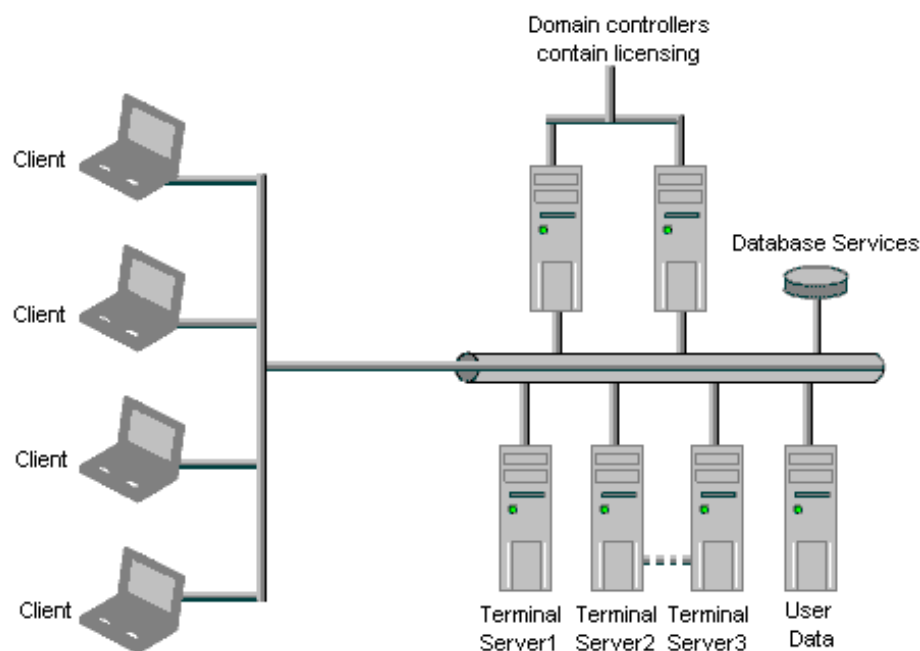
11 pav. MS Windows Network Load Balancing architektūros pavyzdys [11]

Kiekviename serveryje veikia atskiros reikalingų serverio programų kopijos, tokios kaip Web serveris, FTP, Telnet servisas. Kai kuriems servisams, tokiems kaip Web serveris, programos kopija veikia visuose klasterio serveriuose, o tinklo apkrovimo balansavimo paslauga paskirsto apkrovimą tarp serverių. Kitoms paslaugoms tik vienas serveris aptarnauja klientų užklausas, tačiau šiam serveriui sugedus, užklausos automatiškai nukreipiamos į kitą serverį. Vieno serverio gedimo atveju šis persijungimas atliekamas greičiau nei per 10 sekundžių [11].

Network Load Balancing naudojimas terminalinių serverių klasterizavimui

Balansuojant tinklo apkrovas galima paskirstyti didelį klientų skaičių visai grupei terminalinių serverių (6-skyrius, [13]). Kuriant tokio pobūdžio klasterį, kiekvienas serveris turi būti paruoštas, kad galėtų aptarnauti visus vartotojus. T.y. kiekvieno vartotojo informacija, jų sisteminė informacija ir reikalingi duomenys turi būti saugomi tam skirtame antro lygio (ne tiesiogiai vartotojui pasiekiamame) failų serveryje [11].

12 paveikslėlyje pavaizduoti duomenų bazių serveriai ir vartotojų duomenų serveriai, kurie turi būti suprojektuoti kaip atskiri serverių klasteriai su duomenų saugyklomis ar diskų masyvais.



12 pav. Terminalių serverių klasterizavimo pavyzdys [11]

Terminaliniai serveriai tik aptarnauja klientus, t.y. juose veikia tam tikro kliento programos, tačiau visi duomenys yra saugomi atskirai tam skirtuose serveriuose. Toks terminalinių serverių klasterizavimas leidžia paskirstyti apkrovimus, taigi kiekvienas serveris gali tinkamai ir kokybiškai aptarnauti tam tikrą skaičių klientų, tačiau klientai nėra susieti su tam tikru terminaliniu serveriu.

Minėtos technologijos leidžia užtikrinti didelį duomenų centrų paslaugų pateikiamumą. Pateikiamumo įvertinimo skaičiavimas pateiktas 3-ame skyriuje. Dažnai jų pagalba užtikrinamas ne tik aukštas pateikiamumas, bet ir aukštesnės kokybės duomenų centrų paslaugos. T.y. rezervuotos sistemos, serverių klasteriai, diskiniai masyvai suteikia galimybę sumažinti paslaugų vėlinimą ir kitus kokybinius parametrus - tuo pačiu aptarnauti daugiau vartotojų ir tiekti jiems kokybiškesnes paslaugas.

6. „PLONO KLIENTO“ (Thin-client) PASLAUGOS MODELIS

“Plonas klientas” (Thin-client) — tai aparatūrinė ar/ir programinė įranga, veikianti kliento pusėje, skirta prisijungti prie nutolusio paslaugų serverio ir naudotis jo resursais realiu laiku. “Plonu klientu” paprastai vadinama aparatūrinė įranga: kompiuteris (darbo stotis) su minimaliais aparatūriniais resursais (dažniausiai be ilgalaikės atminties – kietojo disko), kuris yra skirtas tik prisijungimui prie nutolusio terminalinio serverio. Visi arba didžioji dalis duomenų apdorojimo atliekama serveryje, o “plonas klientas” yra tik tarpininkas tarp terminalinio serverio ir vartotojo. Taip pat dar „plonu klientu“ gali būti vadinama šioms funkcijoms atlikti skirta programinė įranga, veikianti kliento kompiuteryje. Ji naudojama tam, kad bet kurį personalinį kompiuterį būtų galima pritaikyti darbui nutolusiame darbastalyje (remote desktop) tiekėjo terminaliniame serveryje.

Priešingybė „plonam klientui“ yra “storas klientas” (Thick-client). Šis klientas tik naudoja serverių resursais duomenų saugojimui, tačiau programos veikia pas klientą, taigi visas duomenų apdorojimas vyksta kliento kompiuteryje. Šiam modeliui galima priskirti didžiąją dalį vartotojų kompiuterių, veikiančių tinkle, tačiau tinklą naudojančių tik duomenų perdavimui, o ne įprastiniams veiksams atlikti – jie atliekami kliento kompiuteryje.

„Plono kliento“ modelio privalumai:

- Mažesnės IT administravimo išlaidos – „ploni klientai“ yra beveik pilnai valdomi serveryje. Kvalifikuotas duomenų centro IT personalas yra atsakingas už duomenų ir programinės įrangos valdymą, todėl vartotojui nereikia turėti savo sistemų administratoriaus IT ūkiui prižiūrėti.
- Lengviau apsaugoti – saugumo užtikrinimas centralizuotas (atliekamas tiekėjo serveriuose), o vartotojų duomenys niekada nesaugomi pas klientą. Aparatūrinė įranga, esanti pas klientą, yra patikimesnė, taip pat sukuriama saugesnė darbo aplinka – apsaugoma nuo kompiuterių virusų, kenkėjiškų programų ar kitų saugumo pažeidimų.
- Mažesnės išlaidos aparatūriniai kompiuteriniai įrangai. „Plono kliento“ aparatūrinė įranga yra pigesnė už įprastą vartotojo aparatūrinę įrangą, nes susideda iš mažiau komponentų bei jos yra paprastesnės ir pigesnės (nėra kietojo disko, mažiau atminties, nereikalingas galingas procesorius ir t.t.). Taip pat tokia įranga paprastai skirta ilgesniam funkcionavimo laikui, taigi patikimesnė, todėl reikia rečiau atnaujinti ar taisyti.

„Storo kliento“ modelio privalumai:

- Mažesnis duomenų perdavimo tinklo resursų naudojimas, mažesni tinklo pralaidumo reikalavimai.

- Mažesni serverių reikalavimai – nereikalingi tokie galingi serveriai, kaip „plonam klientui“, kadangi saugomi tik duomenys.
- Geresnės kokybės vaizdo ir garso (multimedia) duomenų naudojimas. Geros kokybės vaizdai ir garsai sunaudoja labai daug duomenų perdavimo tinklo resursų „plono kliento“ modelyje, todėl dažnai ta kokybė mažinama norint sutaupyti tinklo resursų.
- Daugiau lankstumo – daugiau programinės įrangos pritaikymo galimybių kliento kompiuteriuose.

6.1. „Plono kliento“ paslaugos panaudojimo galimybės

Taikomųjų programų paslaugų teikėjas – (*angl. Application Service Provider - ASP*) tai tam tikra duomenų centrų paslaugų teikėjų atmaina. Jie teikia programinės įrangos nuomos paslaugas, leisdami naudotis duomenų centrų serverių resursais (programine įranga, duomenų laikmenomis, duomenimis, ryšio linijomis ir t.t.). Šios paslaugos pasiekiamos naudojantis Web technologijomis arba naudojantis specialios paskirties programine įranga ar aparatiniais įrenginiais. Tokiu atveju klientas apmoka paslaugas teikėjui tik už tą programinę įrangą, kuria naudojasi tam tikrą laiką. Priklausomai nuo situacijos, tai gali būti labai ekonomiškai naudinga. Tiek programinės įrangos nuoma, tiek duomenų talpinimas, tiek platesnės IT infrastruktūros perkėlimas vadinamas „outsourcing“. Šiuo atveju klientas patiki visą ar dalinę savo IT infrastruktūrą duomenų centrui (programinę įrangą, aparatūrinius resursus, tinklo įrangą, duomenis, sistemų priežiūrą ir aptarnavimą). Paslaugų teikėjas savo turimus resursus panaudoja kliento IT infrastruktūros patalpinimui savo sistemose. Tokiu atveju klientas moka už „outsourcing“ paslaugą, kurios pagalba teikėjas užtikrina maksimalios kokybės IT sistemų darbą.

Yra visa eilė šio sprendimo privalumų:

- Programinės įrangos suderinamumo problemos yra pašalinamos kliento pusėje.
- Programinės įrangos kaina yra išdalinama visam skaičiui klientų, taigi vienam klientui kainuoja pigiau.
- Teikėjai paprastai turi aukštesnio lygio IT specialistus, kurie sprendžia programinės įrangos gedimus, rūpinasi atnaujinimais, suderinamumu ir t.t.
- Kadangi klientų duomenys taip pat saugomi teikėjo serveriuose, tai yra teikiamos ir duomenų atsarginių kopijų darymo paslaugos – teikėjas išipareigoja saugoti duomenis ir kaip galima greičiau juos atkurti, įvykus sistemos gedimui ar duomenų pažeidimui.

Tai pat yra ir trūkumų:

- Klientas turi labai pasitikėti teikėju, ir tikėtis, kad teikėjas užtikrins kuo didesnę pateikiamumą, ypač kai tokiu būtu yra naudojamos verslui kritinės taikomosios programos. Teikėjui nesilaikant reikalavimų, gali labai nukentėti kliento verslas, taigi klientas yra labai priklausomas nuo paslaugos teikėjo.
- Integracija su kitomis kliento taikomosiomis programomis gali būti problematiška.

Spartus duomenų perdavimo tinklų pralaidumų augimas bei kainų kritimas sukūrė naują rinką taikomųjų programų paslaugų teikėjams. Anksčiau teikę daugiausia tik interneto svetainių talpinimo (*angl. Web hosting*) paslaugas, teikėjai dabar suteikia galimybę klientams gauti „outsourcing“ paslaugas, naudojantis profesionalių duomenų centrų galimybėmis – visa eile taikomųjų programų, jų palaikymo paslaugomis, duomenų saugumo užtikrinimu ir t.t. Taigi įmonės, organizacijos, o taip pat ir privatūs asmenys gali sumažinti išlaidas tiek aparatūrinei, tiek programinei įrangai, o ypač jų priežiūrai, atnaujinimui bei duomenų saugojimui. Taikomųjų programų paslaugų teikėjai, norėdami kurti naujas paslaugas ir teikti centralizuoto duomenų apdorojimo paslaugas per globalųjį tinklą naudoja „plono kliento“ modeliu pagrįstas sistemas: virtualaus nutolusio darbaltio (*angl. Remote desktop*) paslaugas, tuo pačiu taikomųjų programų nuomą, duomenų saugyklų paslaugas ir t.t.

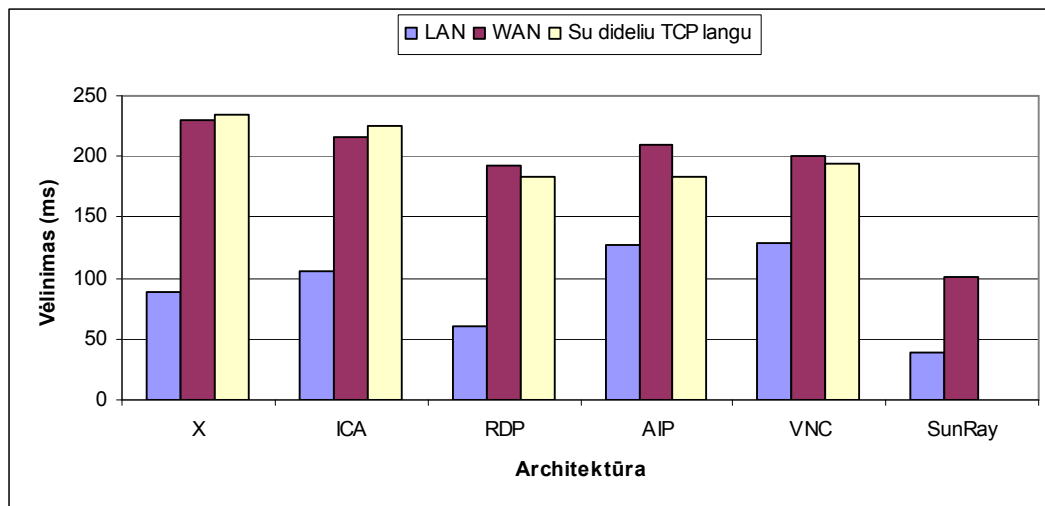
„Plono kliento“ modelio sistema susideda iš terminalinio serverio ir kliento, kuris komunikuoja per duomenų perdavimo tinklą, naudodamas tam skirtus protokolus. Klientas naudoja grafiniu ekranu, klaviatūra, pele, bei kartais papildomais įvesties įrenginiais, o visos programų atliekamos operacijos yra vykdomos serveryje, o rezultatai bei vartotojo komandos siunčiami tinklu. Didelė dalis šių nutolusio atvaizdavimo (*angl. Remote display*) protokolų leidžia efektyviai naudoti jau esamas taikomas programas be jokio jų modifikavimo. Klientas gali naudoti specializuotą aparatūrinę įrangą arba nutolusiam prisijungimui skirtą specialią programą, kuri veikia visose populiariose operacinių sistemų architektūrose.

„Plono kliento“ modelio efektyvumo ir galimybių tyrimai daugiausia atliekami šešiose populiariausiose „plono kliento“ pagrindu veikiančiose architektūrose [12]: Citrix MetaFrame, Microsoft Windows 2000 Terminal Services, AT&T VNC, Tarantella, Sun Ray ir X. Jos pasirinktos, atsižvelgiant į jų populiarumą, veikimo efektyvumą bei veikimo skirtumus. Reikalinga identifikuoti globalaus tinklo (WAN) poveikį sistemų veikimui, todėl tuo tikslu lyginamas sistemų veikimas tiek globaliajame tinkle, tiek vietiniame tinkle (LAN).

„Plono kliento“ sistemų efektyvumas nustatomas trijų rūšių testavimus [12][14]: vėlinimo - matuojant atsako laiką, web testavimą - matuojant naršymo efektyvumą, bei vaizdo duomenų

testavimą - matuojant video peržiūros efektyvumą. Vėlinimo matavimai atliekami, norint išmatuoti paprastų elementarių operacijų trukmę, o web ir vaizdo matavimai – norint sudaryti realistiškesnį bendrą sistemos veikimo vaizdą [14].

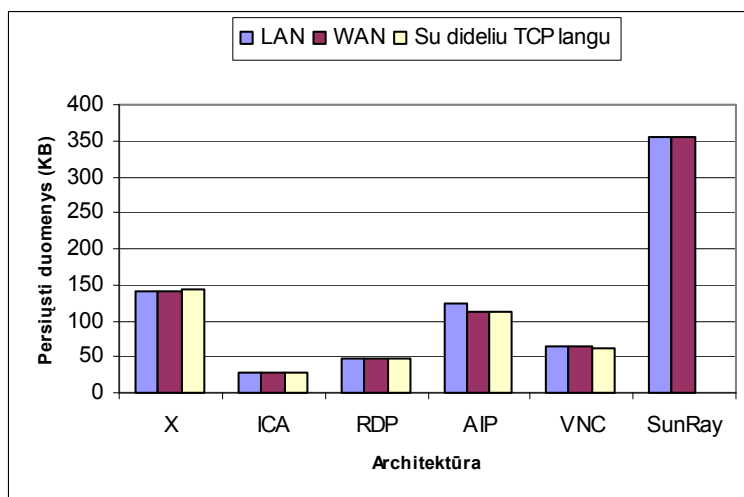
Trijų elementarių operacijų rezultatai: vieno simbolio atspausdinimo, teksto vartymo (angl. scrolling) ir ekrano pildymo tam tikro dydžio stačiakampiu (angl. fill), o taip pat dviejų sudėtingesnių operacijų, kurių metu persiunčiami paveikslėliai, rezultatai pateikti 13 pav.



13 pav. Letter, Scroll, Fill, Red Bitmap Image ir Image operacijų apibendrinti rezultatai

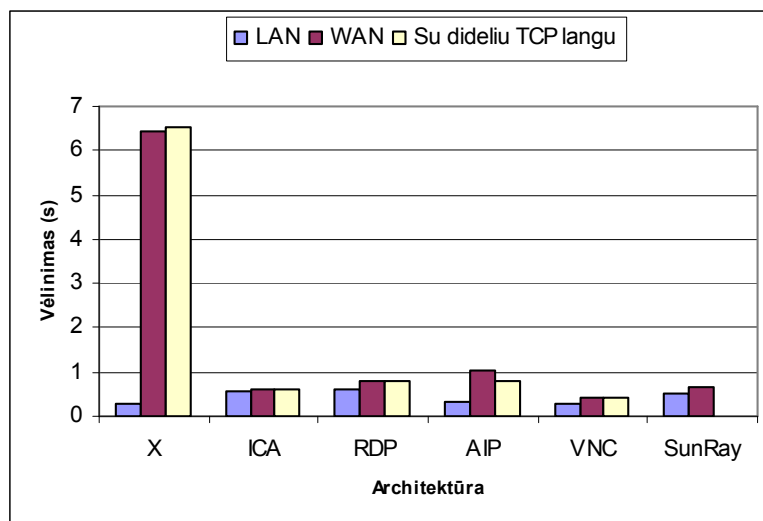
Bendru atveju paprastoms operacijoms – tokioms, kaip spausdinimas klaviatūra, kursoriaus judinimas ar pelės judesys, sistemos atsako vėlinimas turėtų neviršyti 50-150 ms ribos, kad esant tokiam vėlinimui žmogus nepastebėtų sistemos darbingumo sulėtėjimo [12].

Iš pateiktų rezultatų galime padaryti išvadą, kad dalis sistemų pasirodė geriau nei reikalaujama 150 ms ribinė reikšmė didžiajai daliai operacijų. Konkrečiai Sun Ray sistema veikia su mažiau nei 100 ms vėlinimu tiek vietinio tinklo (LAN), tiek globalaus tinklo aplinkose.



14 pav. Persiūsti duomenys web puslapiui

14 ir 15 pav. pateikti web aplikacijų testavimo rezultatai visoms šešioms testuojamoms „plono kliento“ sistemoms. 14 pav. pavaizduotas vidutinis persiųstų duomenų kiekis vienam web puslapiui skirtingoms tinklo architektūroms: vietiniame tinkle (LAN), globaliajame tinkle bei naudojant didelį 1 MB TCP langą.



15 pav. Vėlinimas web puslapiui

15 pav. pavaizduotas vidutinis sistemos vėlinimas vienam web puslapiui atvaizduoti. Reikalaujama, kad vieno puslapio atsiuntimas turėtų užtrukti ne ilgiau nei vieną sekundę. Didžioji dalis sistemų užtikrina web puslapio persiuntimą ir atvaizdavimą iki vienos sekundės [12].

Matavimai rodo, kad Sun Ray architektūros sistema veikė geriausiai visuose bandymuose. Lyginant sistemas tarpusavyje pastebėta, kad skirtingų architektūrų veikimo kokybė labai skiriasi.

Testuojant video aplikacijas, didžioji dalis „plono kliento“ sistemų pasirodė neefektyvios tiek vietiniame, tiek globaliajame tinkle [14]. Tik Sun Ray sistema užtikrina 70% norimos kokybės. Kitose sistemose kokybė nesiekė 35% ir tai buvo nepriimtina.

Galima išskirti penkis principinius aspektus, į kuriuos reikėtų atsižvelgti projektuojant „plono kliento“ modeliu paremtą sistemą, naudojančią globalaus kompiuterių tinklo technologijas:

- Optimizuoti vėlinimą, lyginant su duomenų perdavimo tinklo pralaidumu.
- Minimizuoti kliento ir serverio sinchronizaciją, t.y. kad būtų atnaujinama tik pasikeitusi reikalinga informacija.
- Naudoti paprastesnes atvaizdavimo operacijas, norint padidinti veikimo greitį.
- Kompresuoti ekrano atnaujinimo informaciją.

Gauti rezultatai remiasi atliktais matavimais, kai nebuvo atsižvelgta į paketų praradimą bei linijų persipildymą, kurie pasitaiko realiose globalaus tinklo (WAN) situacijose.

6.2. Windows Remote Desktop Protocol (RDP) įvertinimas

Šiuo protokolu veikia MS Windows Terminal Services ir Remote Desktop Connection paslaugos, kurių pagalba vartotojas gali prisijungti prie nutolusio duomenų centro ir per virtualų darbstaalį dirbti su juo tinkle taip, lyg dirbtų darbo vietoje. Kadangi Windows operacijų sistema yra populiariausia kompiuteriams darbo stotims, todėl tikslinga įvertinti RDP protokolo tinkamumą „plono kliento“ sistemoms.

Šio protokolo resursų sunaudojimas tirtas atliekant įvairius elementarius veiksmus, kurie paprastai vykdomi dirbant kompiuteriu Windows aplinkoje [13].

Tyrimo rezultatai pateikti 4 lentelėje.

4 lentelė. Remote Desktop Protocol elementarių veiksmų tinklo resursų naudojimas

	Valdymo duomenų kanalas	Atvaizdavimo duomenų kanalas		MS Windows
		MS Word programa	WordPerfect programa	
Tekstų spausdinimas (75 žpm)	1,56 KBps	6,26 KBps	11,66 KBps	-
Pelės judesiai	1,95 KBps	-	-	-
Meniu navigacija	-	-	-	3,12 KBps
Meniu navigacija su "Alt" klavišu	-	39,82 KBps	17,13 KBps	-
Dokumento vartymas (angl. scrolling)	-	59,24 KBps	192,54 KBps	-

Kadangi RDP protokolas yra sukurtas Microsoft kompanijos ir jo veikimo specifikacija viešai neskelbiama, tai RDP veikimo kokybę yra įmanoma nustatyti tik netiesiogiai – stebint duomenų perdavimo tinklo apkrovimus bei kliento ir serverio kompiuterių resursų naudojimą. Testuojant reikalinga emuluoti tipinius vartotojo veiksmus Windows sistemoje, ir matuoti tinklo apkrovimus bei atsako laikus. Galime daryti išvadą (4 lentelė), kad pikinis tinklo resursų panaudojimas gali siekti iki 200KBps, tačiau dažniausiai neviršija 60 KBps. Klientų gaunamos paslaugos kokybė labai priklauso nuo serverio procesoriaus apkrovimo, o taip pat ir nuo kitų parametrų – atminties, tinklo apkrovimo. Matome, kad RDP protokolas nėra skirtas dirbti labai lėtose komunikacinio tinklo linijose. Ši technologija labiau skirta naudojimui vietiniame tinkle, bet gali būti naudojama ir didelės spartos platesnės apimties tinkle.

6.3. „Plono kliento“ paslaugos kokybė

„Plono kliento“ (terminalinės programinės įrangos) architektūrų analizė buvo atlikta 6.1 ir 6.2 skyreliuose. Buvo pateikti paslaugų vėlinimo bei naudojamų tinklo resursų rezultatai

Paslaugos vėlinimo parametrai skirtingoms „plono kliento“ architektūroms yra ypač svarbūs vartotojui, nes jie yra tiesiogiai pastebimi. Vartotojas, pirkdamas „outsourcing“ terminalinių serverių paslaugą, tikisi, kad galės naudotis serveryje veikiančia programine įranga taip pat, lyg ji veiktų jo asmeniniame kompiuteryje. Per didelis vėlinimas sukelia informacijos užlaikymą, kartais net visai sutrikdydamas darbą su nutolusiu serveriu.

Todėl „plono kliento“ architektūroje šis parametras ypatingai įtakoja bendrą paslaugos kokybę, kuri išreiškiama formule (1). Šioje formulėje yra reikalingas minėtos paslaugos vėlinimo įvertis L intervale $[0;1]$, kuris gali būti paskaičiuojamas pagal (18) formulę. Vidurkinė paslaugos vėlinimo reikšmė L' (milisekundėmis) gali būti paskaičiuojama apibrėžtam laiko momentui (pvz. dienai, savaitei ar mėnesiui). SLA sutartyje turėtų būti apibrėžtos terminalinių serverių paslaugos SLA nurodoma reikalaujama L_{SLA} ir minimali L_{min} vėlinimo reikšmės. Vidutiniam vėlinimui viršijus reikalaujamą reikšmę, SLA terminalinių serverių paslaugos vėlinimo metrikos įsipareigojimas pažeidžiamas - paslaugos teikėjas privalo prisiimti atsakomybę už tai.

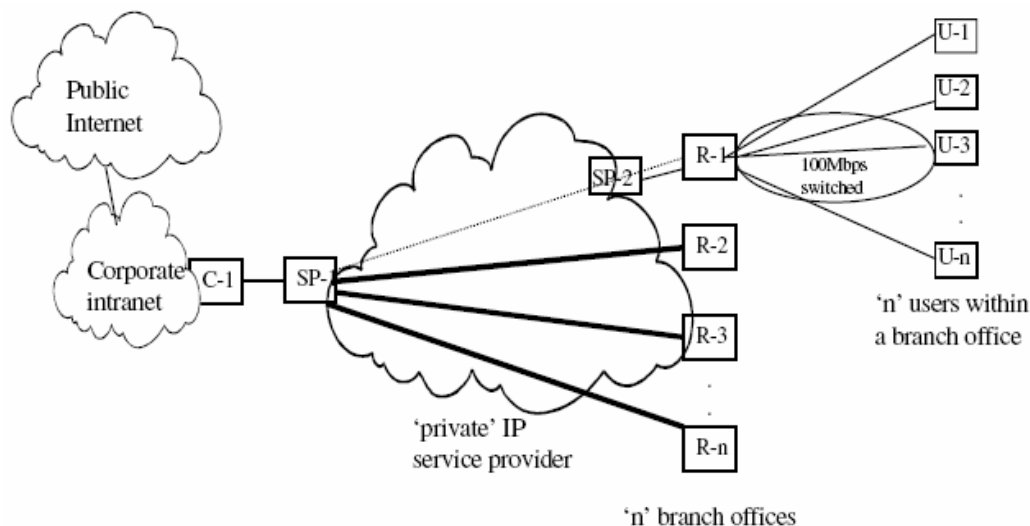
Naudojamų duomenų perdavimo tinklo resursų (tinklo apkrovimo - matuojamo bps, Kbps, Mbps) apimtis skirtingoms „plono kliento“ architektūroms yra taip pat svarbūs vartotojui, nes nuo jų priklauso, kokio ryšio linijos pralaidumo reikia, norint turėti aukštos kokybės paslaugą. Esant per mažam pralaidumui, atsiranda vėlinimo padidėjimas, taip pat duomenų paketų praradimas, o tai sukelia klientui tiesiogiai pastebimus sutrikimus. Duomenų perdavimo tinklo resursų sunaudojimą sunku apibrėžti kokybiniu parametru - tinklo apkrovimas pastoviai kinta, priklausomai nuo kliento atliekamų veiksmų. Taigi gali būti išvestos tik abstrakčios vidutinės reikšmės, pagal kurias reikėtų rinktis panašaus pralaidumo (ir spartesnes) ryšio linijas kaip ir resursų sunaudojimas. Todėl „plono kliento“ architektūros kokybę resursų sunaudojimo atžvilgiu siūlyčiau vertinti pagal paslaugos vėlinimą ir paketų pristatymą (trūkstant resursų paketai pradedami atmetinėti).

Bendras paslaugos kokybės įvertis gaunamas apskaičiavus paslaugos vėlinimo įvertį L , ir įstačius jį į kokybės formulę (1). Tuo pačiu turi būti įvertinami ir paskaičiuojami kiti (1) formulėje atsispindintys kokybiniai parametrai (priklausomai nuo kiekvieno duomenų teikėjo SLA specifiкуotų parametrų ir vartotojo poreikių): paslaugos pateikiamumas, duomenų perdavimo tinklo parametrai (vėlinimas, vėlinimo sklaida, informacijos perdavimo sparta) ir t.t.

7. WEB PASLAUGOS KOKYBĖS ĮVERTINIMAS

Pastaruoju metu, kuriant naujas IP paslaugų SLA užtikrinimo metodikas, yra pastebima, kad tradicinių SLA metrikų garantijos neužtikrina galinio vartotojo gaunamų paslaugų kokybės. Todėl yra siekiama įvesti naujas metrikas, kurios tiksliau atspindėtų konkrečios teikiamos paslaugos kokybę [17].

16 pav. pateiktas duomenų perdavimo tinklo paslaugos scenarijus, kur įmonės vidinis tinklas (tuo pačiu vidinio tinklo resursai – web serveriai) duomenų perdavimo paslaugų teikėjo tinklu yra pasiekiamas iš visos eilės nutolusių taškų (įmonės ofiso padalinių). T.y. duomenų perdavimo paslaugų teikėjo tinklas yra naudojamas įmonės informacijos persiuntimui (dažniausiai iš pagrindinio biuro tinklo į mažesnius padalinius).



16 pav. Duomenų perdavimo tinklo paslaugos infrastruktūra [15]

Remiantis pavyzdine WorldCom paslaugų teikėjo SLA, buvo pasiūlytos tokios tariamoje sutartyje nurodomos kokybės metrikos [15]:

- Pateikiamumas (*angl. availability*) – duomenų perdavimo tinklas bus pasiekiamas 99,8% laiko per mėnesį. Tinklas yra skaitomas nepasiekiamu, jeigu 5 minutes nėra gaunamo sėkmingo *ping* atsakymo. *Ping* matavimų dažnis yra 20 sekundžių, kai atliekant matavimus laukiama 2 sekundes.
- Vėlinimas (*angl. latency*) – *ping* procesas pastoviai monitoruoja vėlinimą tarp kiekvieno duomenų perdavimo tinklo galinio taško (tarp R-n ir C-1 taškų). *Ping* matavimų dažnis yra 20 sekundžių, ir vidutinis *ping* rezultatas (abipusis atsako laikas) turi būti mažiau nei 120 milisekundžių (vidurkinama mėnesiui laiko).

- SLA apribojimas: SLA sąlygos yra galiojančios, jeigu kiekvieno ryšio kanalo prieigos apkrovimas (*angl. utilization*) yra mažiau nei 50% (vidurkinama mėnesiui laiko).

Iš karto galima pastebėti nagrinėjamos SLA trūkumus [15]:

- Pateikiamumo metrika yra orientuota į duomenų perdavimo tinklą, o ne į galinį vartotoją. Kad ši metrika būtų pažeista, tinklas turi būti fiziškai neveikiantis apie 14 valandų per duotą mėnesį. Tačiau gali būti, kad nors ir tinklas bus laikomas pasiekiamu (*angl. available*), tačiau kokia nors kita duomenų centro paslauga nebus galima naudotis. Taigi, į vartotoją orientuota pateikiamumo metrika parodytų, kad paslauga jau yra nepasiekiamą (*angl. unavailable*), jeigu vartotojas negali ja naudotis.
- Per piko valandas tinklas gali būti gerokai apkrautas. Bet kadangi vėlinimo ir tinklo apkrovimo metrikos yra vidurkinamos ir skaičiuojamos per ilgą laiko tarpą (mėnesio laiko), SLA įsipareigojimai gali būti nepažeisti.

Taigi, gana aišku, kad SLA daugiau skirta norint apsaugoti duomenų perdavimo tinklo paslaugų teikėjo interesus, prisiimant minimalius kokybės įsipareigojimus galiniam vartotojui.

Grįžtant prie pavyzdinės tinklo schemos (16 pav.), galima laikyti, kad pagrindiniame įmonės tinkle (*angl. corporate network*) yra visa eilė serverių, kurie turi aptarnauti tiek vietinių vartotojų, tiek ir įmonės padaliniuose esančių vartotojų užklausas. Nagrinėjant kuo realesnius pavyzdžius, galime teigti, kad didžioji dalis tokių paslaugų bus HTTP duomenys, taigi turėsime Web serverius, kurių pagalba nutolusiems įmonės darbuotojams bus pasiekiamą reikiama informacija. Kadangi tada didžioji dalis tolumo ryšio linijomis keliaujančio srauto bus HTTP duomenys, buvo nuspręsta pasiūlyti į web paslaugą orientuotą našumo metriką, siekiant kuo tiksliau įvertinti vartotojui matoma paslaugos kokybę [15].

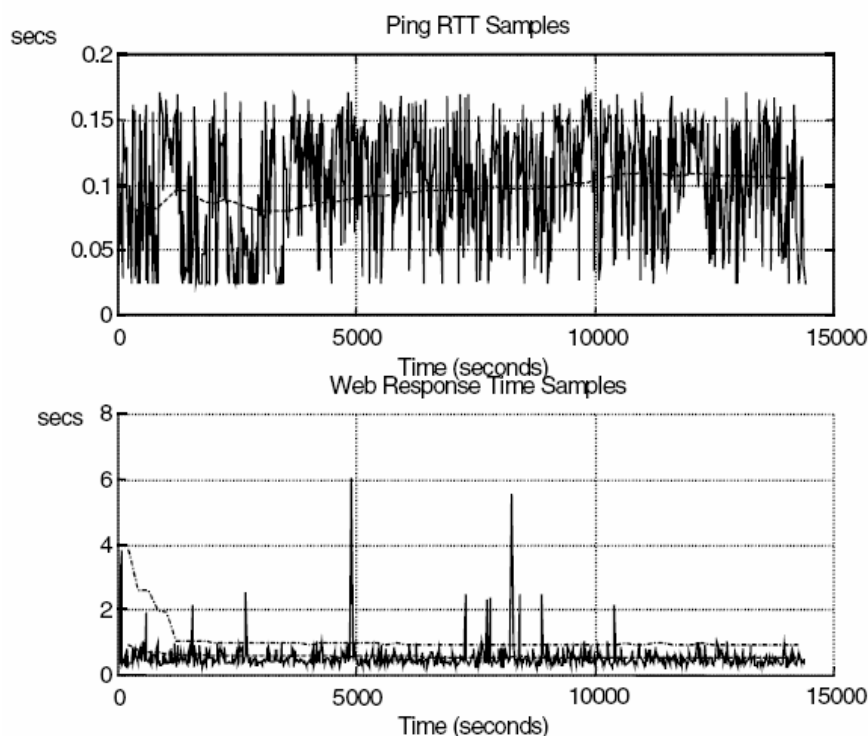
Šią naują metriką „web atsako laikas“ (*angl. Web Response Time – WRT*) galima būtų apibūdinti taip: klientas periodiškai inicijuoja HTTP objekto užklausą iš tam tikro web serverio, o šios užklauskos atsiuntimo laikas ir būtų WRT rezultatas – realus laikas, kurio reikėjo, kad klientas gautų reikiamą informaciją iš serverio. Šiuo atveju web serveris yra pagrindiniame įmonės tinkle, o vartotojas – tam tikro įmonės padalinio tinkle. Šios papildomos serverio protokolų lygio metrikos naudojimas turi kelis privalumus, lyginant su tradicinėmis *ping* pagrindu veikiančiomis metrikomis [15]:

- WRT metrikos rezultatai įvertina duomenų perdavimo tinklo vėlinimo ir paketų praradimo įtaką konkrečios paslaugos (HTTP) kokybei.

- Yra daug lengviau interpretuoti šios metrikos rezultatus, ir vertinti tai, kaip duomenų perdavimo tinklas veikia galinių vartotojų naudojamų paslaugų kokybę.

Šią metriką galima vadinti „į vartotoją orientuota“, kurią pažeidus paslauga gali būti laikoma nepasiekiamą.

17 pav. viršutiniame paveiksle matomas momentinis (stora linija) ir vidutinis (brūkšninė linija) *ping* dvipusio atsako laiko (RTT) rezultatas (vidutinis apie 100 milisekundžių).



17 pav. Skirtingų metrikų rezultatai (tradicinio RTT ir naujo WRT) [15]

Yra sunku pagal šią statistiką spręsti, kaip duomenų perdavimo tinklas veikia galinio vartotojo vartojamų paslaugų kokybę. 17 pav. apatinis paveikslas iliustruoja web atsako laiko metrikos skaičiavimo rezultatus. Brūkšninė kreivė atvaizduoja 95 percentilio statistinę informaciją, kuri realiu laiku buvo skaičiuojama, modeliuojant HTTP klientų darbą. Ši statistika vidurkinama ir skaičiuojama 2 valandų laikotarpiui [15].

Galima sakyti, kad SLA RTT slenkstis (120 milisekundžių) yra per daug griežtas įsipareigojimas, jeigu tinkle dominuoja web aplikacijos [15][17]. Taip yra todėl, kad web technologija iš prigimties yra pakankamai elastinga ir nėra labai jautri tinklo RTT vėlinimui. Išeina, kad paslaugų teikėjui yra lengviau užtikrinti tokias SLA vėlinimo parametrų reikšmes, kurios remiasi konkrečiai WRT (Web Response Time) metrika. Realiam duomenų perdavimo tinkle optimalios parametrų slenkstinės reikšmės pastoviai kinta, o į web paslaugą orientuota

metrika ne tik sprendžia šias problemas, bet ir vartotojui suprantamiau parodo konkrečios paslaugos kokybę duomenų perdavimo tinkle.

Taigi reikėtų vengti apie duomenų perdavimo tinklo kokybę spręsti tik pagal vienos dimensijos metrikas (tokias kaip vidutinis duomenų perdavimo tinklo apkrovimas, vidutinis *ping* vėlinimas ar pan.), nes galinio vartotojo atžvilgiu šie vertinimai gali būti neteisingi. Kelių dimensijų vertinimas (pagrįstas aukštesnio lygio metrikos – šiuo atveju WRT) pateikia gerokai tikslesnius, į vartotoją orientuotus paslaugų kokybės vertinimo rezultatus.

7.1. Web paslaugos vėlinimo įvertinimas iš serverio ir kliento perspektyvos

Web paslaugai, kuri standartiškai veikia HTTP protokolu, vienas iš kokybinių parametrų yra web paslaugos vėlinimas. Jį galima apibūdinti kaip internetinio puslapio, esančio duomenų centro paslaugos teikėjo serveryje, HTTP operacijos (ar operacijų) atlikimo trukmę. Šiame eksperimente buvo siekiama išsiaiškinti, kas iš tikrųjų lemia web paslaugos vėlinimą, o po to ir iš šios kokybinės metrikos sekančią apibendrintą Web paslaugos kokybę. Visa tai reikalinga, norint suprasti, kaip užtikrinti aukštos kokybės paslaugą.

7.1.1. Matavimų metodika

Buvo atliekami KTU ITPI LITNET valdymo centre esančio WWW serverio dviragis.ktu.lt paslaugos vėlinimo matavimai. Serveryje veikia www.itpi.ktu.lt interneto svetainė, o taip pat visa eilė kitų interneto svetainių. Minėta interneto svetainė periodiškai buvo pasiekama iš dviejų skirtingai nutolusių klientų kompiuterių – buvo vykdomos internetinio puslapio parsisiuntimo užklausos (HTML puslapio, bei reikiamų grafinių failų siuntimas) – kurių vykdymo metu buvo matuojamas paslaugos vykdymo laikas iš dviejų skirtingų perspektyvų (serverio ir kliento). Du skirtingai nutolę klientai pasirinkti norint vertinti paslaugos vėlinimo padidėjimą dėl papildomo duomenų perdavimo tinklo vėlinimo. Arti esantis klientinis kompiuteris (193.219.184.93) veikė tame pačiame vietiniame tinkle (LAN), o nutolęs klientinis kompiuteris (85.206.232.6) – kito interneto paslaugų teikėjo duomenų perdavimo tinkle.

Serveryje (LINUX OS) matavimai buvo atliekami abiejų skirtingų klientų užklausoms. Buvo matuojama naudojantis paketų sekimo/fiksavimo programine įranga „tcpdump“ [18], kuri suteikia galimybę stebėti paketus juos patogiai filtruojant, pateikiama informacija apie transakcijų trukmę (kada paketas gautas ar išsiųstas, koks yra laiko skirtumas tarp gretimų paketų ir t.t.).

Klientų pusėje (LINUX OS) automatizuotai tekstinio failų siuntimo įrankio „wget“ [19] pagalba buvo parsisiunčiamas interneto svetainės puslapis su reikalingais paveikslėliais (t.y. buvo vykdomas rekursyvus 1 lygio failų parsisiuntimas), ir buvo matuojama, kiek laiko klientas užtruko vykdydamas šią operaciją. Taigi ši operacija buvo laikoma elementaria HTTP puslapio parsisiuntimo operacija – transakcija. Visi matavimai buvo atliekami periodiškai kas 5 minutes. Programos „tcpdump“ pateikiama informacija buvo saugoma atskiruose failuose, o vėliau automatizuotų programinių scenarijų (scripts) pagalba buvo atliekama rezultatų analizė ir sisteminimas. Toliau pateikta pavyzdinė „tcpdump“ programos pateikiama informacija apie vienos nagrinėjamos operacijos paketus:

```
root@dviragis:/disks/disk1/home/rokas/060503# tcpdump -r test_local_001
reading from file test_local_001, link-type EN10MB (Ethernet)
03:10:03.634367 IP 193.219.184.93.1317 > dviragis.ktu.lt.www: S 3120683314:3120683314(0) win 5840
<mss 1460,sackOK,timestamp 125339320 0,nop,wscale 0>
03:10:03.634473 IP dviragis.ktu.lt.www > 193.219.184.93.1317: S 73855227:73855227(0) ack
3120683315 win 5840 <mss 1460,nop,nop,sackOK,nop,wscale 0>
03:10:03.634804 IP 193.219.184.93.1317 > dviragis.ktu.lt.www: . ack 1 win 5840
03:10:03.635342 IP 193.219.184.93.1317 > dviragis.ktu.lt.www: P 1:103(102) ack 1 win 5840
03:10:03.635375 IP dviragis.ktu.lt.www > 193.219.184.93.1317: . ack 103 win 5840
<...>
03:10:03.808731 IP dviragis.ktu.lt.www > 193.219.184.93.1318: P 81772:82052(280) ack 2130 win
18224
03:10:03.809457 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: . ack 80312 win 62780
03:10:03.809689 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: . ack 82052 win 62780
03:10:03.812378 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: F 2130:2130(0) ack 82052 win 62780
03:10:03.812523 IP dviragis.ktu.lt.www > 193.219.184.93.1318: F 82052:82052(0) ack 2131 win 18224
03:10:03.812810 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: . ack 82053 win 62780
```

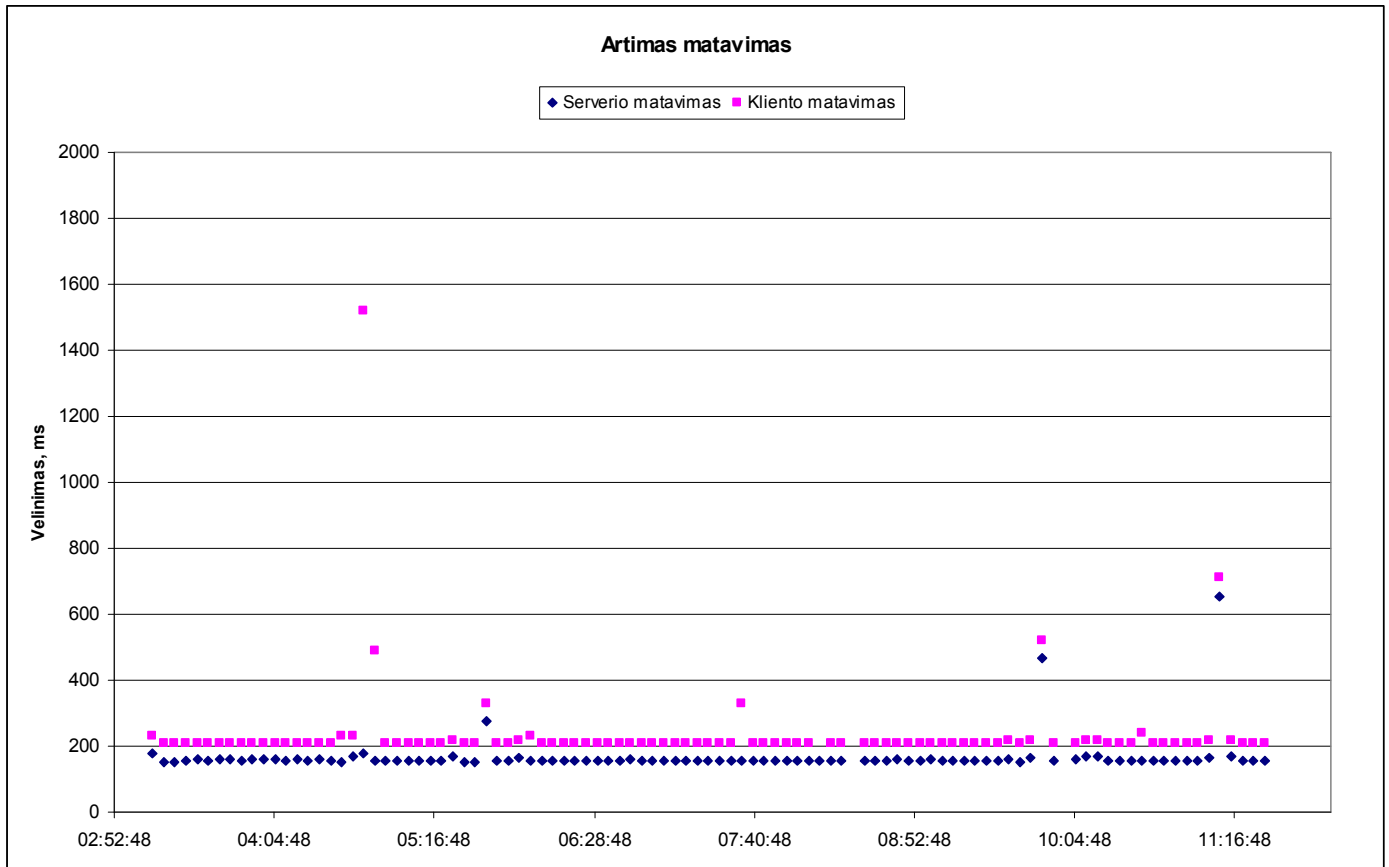
Analogiškai pateikiamos trukmės tarp gretimų paketų, jas galima sumuoti ir taip įvertinti bendrą paslaugos vėlinimą (skaičiuojant serverio pusėje):

```
root@dviragis:/disks/disk1/home/rokas/060503# tcpdump -r test_local_001 -ttt
reading from file test_local_001, link-type EN10MB (Ethernet)
000000 IP 193.219.184.93.1317 > dviragis.ktu.lt.www: S 3120683314:3120683314(0) win 5840 <mss
1460,sackOK,timestamp 125339320 0,nop,wscale 0>
000106 IP dviragis.ktu.lt.www > 193.219.184.93.1317: S 73855227:73855227(0) ack 3120683315 win
5840 <mss 1460,nop,nop,sackOK,nop,wscale 0>
000331 IP 193.219.184.93.1317 > dviragis.ktu.lt.www: . ack 1 win 5840
000538 IP 193.219.184.93.1317 > dviragis.ktu.lt.www: P 1:103(102) ack 1 win 5840
000033 IP dviragis.ktu.lt.www > 193.219.184.93.1317: . ack 103 win 5840
<...>
000015 IP dviragis.ktu.lt.www > 193.219.184.93.1318: P 81772:82052(280) ack 2130 win 18224
000726 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: . ack 80312 win 62780
000232 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: . ack 82052 win 62780
002689 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: F 2130:2130(0) ack 82052 win 62780
000145 IP dviragis.ktu.lt.www > 193.219.184.93.1318: F 82052:82052(0) ack 2131 win 18224
000287 IP 193.219.184.93.1318 > dviragis.ktu.lt.www: . ack 82053 win 62780
```

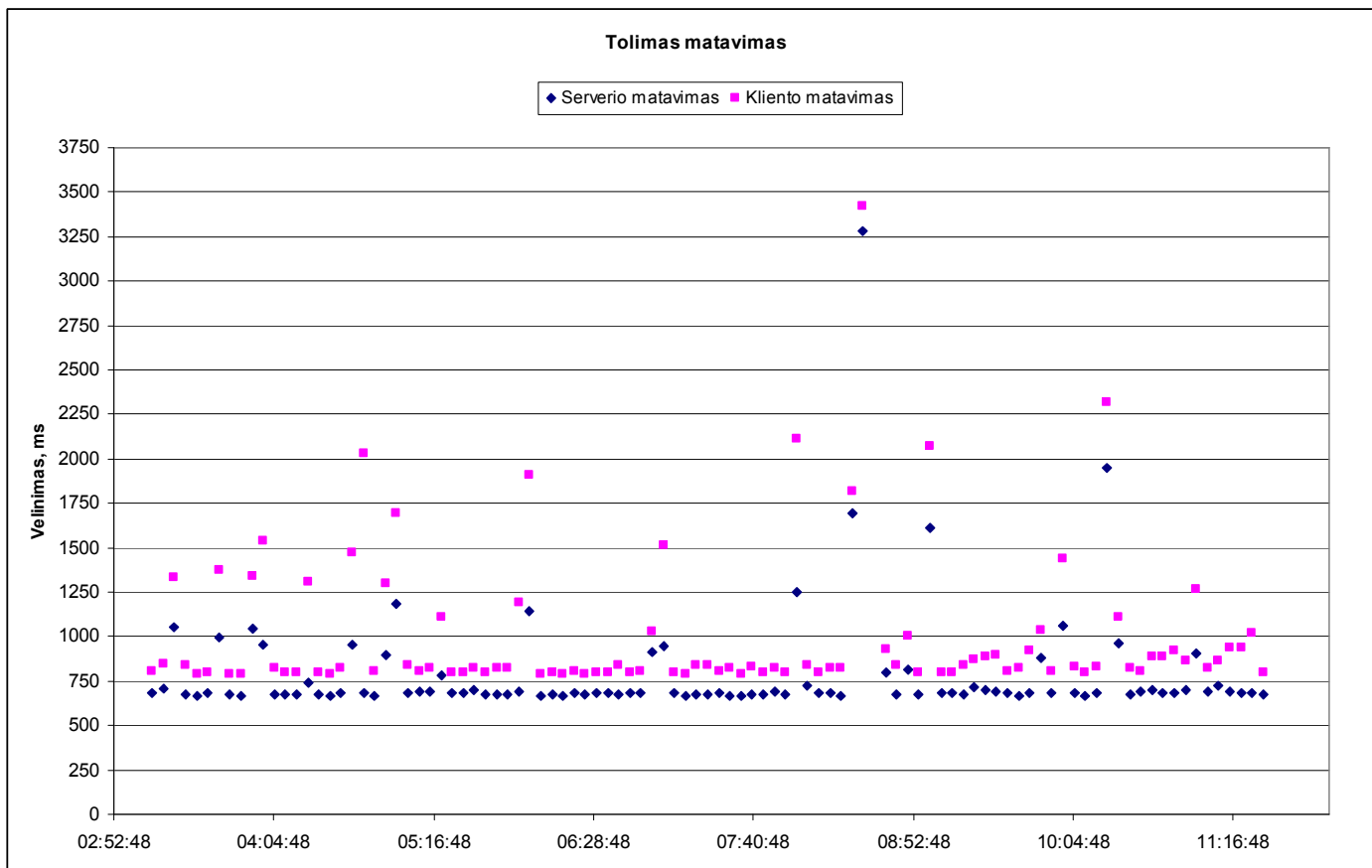
Buvo siekiama serverį užklausinėti tuo pačiu momentu iš abiejų klientų, tačiau labai tiksliai pradėti matavimus yra sunku dėl serverių laiko sinchronizacijos paklaidų.

7.1.2. Vėlinimo įvertinimo rezultatai

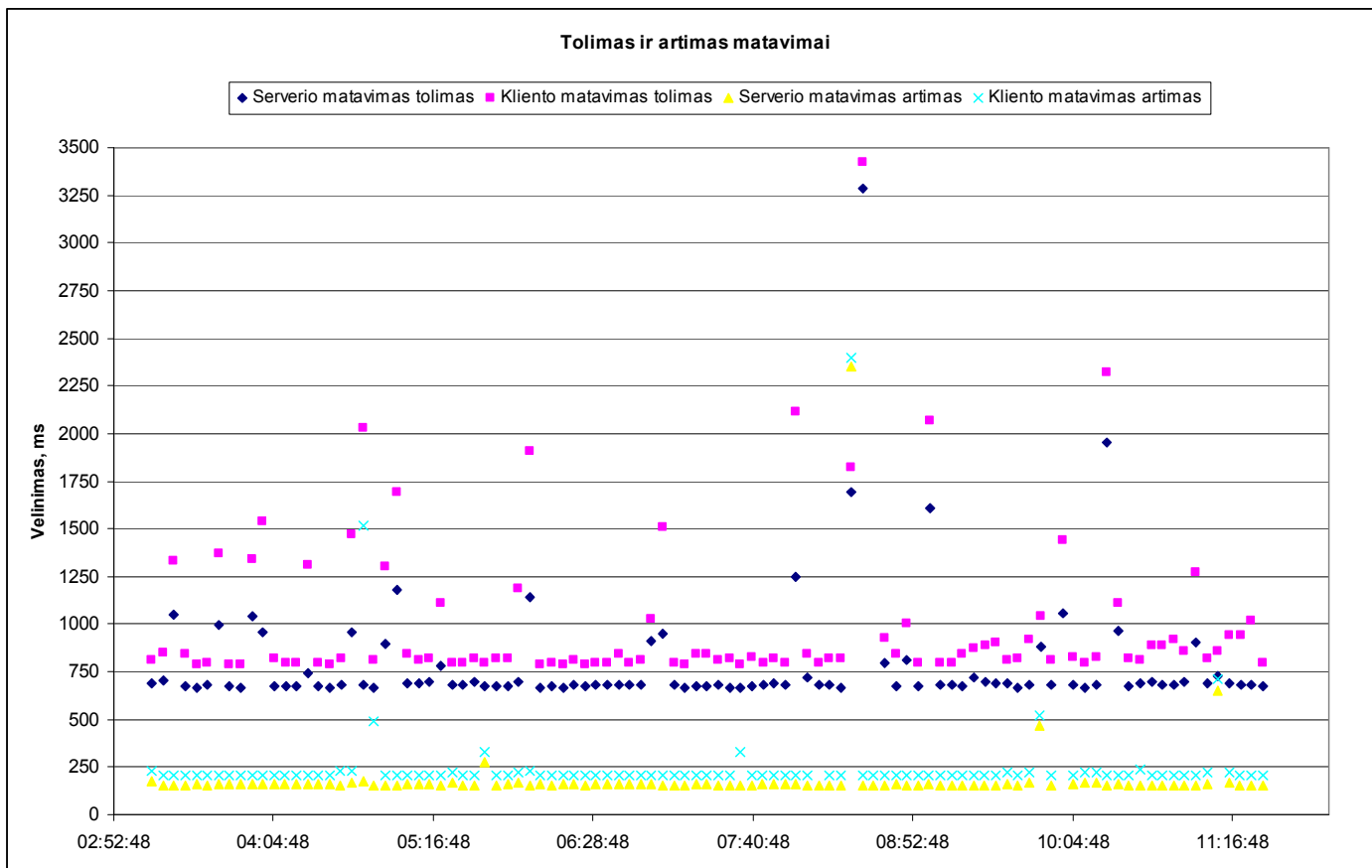
Atlikus visą eilę matavimų bei apdorojus rezultatus, buvo sudaryta priede (10.1 skyrelyje) pateikta lentelė. Matavimo rezultatų grafikai pateikti 18-20 pav.



18 pav. Puslapio užkrovimo vėlinimas iš serverio ir kliento perspektyvų (artimas matavimas)



19 pav. Puslapio užkrovimo vėlinimas iš serverio ir kliento perspektyvų (tolimas matavimas)



20 pav. Puslapio užkrovimo vėlinimas iš serverio ir kliento perspektyvų (artimas ir tolimas matavimai kartu)

Rezultatai rodo, kad iš kliento perspektyvos gauti web paslaugos vėlinimai visada buvo didesni nei skaičiuojant serverio pusėje. Tačiau serverio ir kliento pusėse gauti vėlinimai koreliuoja tiesiškai. Natūralu, kad kliento pusėje yra užfiksuotas didesnis vėlinimas, nes papildomas laikas turi būti sugaištamasis TCP sesijos užmezgimui, o vėliau ir paskutinės informacijos paketo gavimui, informacijos surinkimui į vientisus failus iš atskirų paketų, bei jos įrašymui į kliento kompiuterio kietąjį diską. Trumpai tariant, papildomai laiko sugaištama kliento kompiuteryje, ir visa tai labai priklauso nuo kliento kompiuterio aparatūrinės įrangos (procesoriaus taktinio dažnio, atminties kiekio, kietojo disko greičio, bei konkrečiu laiko momentu turimų laisvų resursų), bei programinės įrangos: operacinės sistemos, taikomųjų programų ir t.t.

Kadangi HTTP protokolas veikia viršutiniame OSI modelio lygyje, jo veikimui įtakos turi visa eilė žemesniuose lygiuose esančių protokolų: TCP (4 OSI lygis), IP (3 OSI lygis), bei žemesniuose lygiuose veikiantys fiziniai protokolai. Tai duomenų perdavimo tinklo įtaka web paslaugos vėlinimui. Kadangi serveris turi sugaišti laiką užklausų apdorojimui (surasdamas ir paimdamas reikiamą failą), tai bendram paslaugos vėlinimui didelę reikšmę turi ir serverio aparatūrinė bei programinė įranga, o taip pat turimų resursų kiekis konkrečiu laiko momentu. Tai nėra elementarios HTTP užklausų operacijos, o visa eilė informacinių paketų, kurie turi būti tvarkingai ir iš eilės persiųsti iš serverio klientui, ir iš kliento serveriui (patvirtinimo paketai), taigi vieno paketų apdorojimo trukmė lemia tolimesnių paketų išsiuntimo ir gavimo trukmę.

Matavimai buvo vykdomi apie 12 val. ir buvo pastebėti matavimų rezultatų nukrypimai nuo vidurkio. Manoma, kad tai įvyko dėl serverio laisvų resursų sumažėjimo.

Iš šių matavimų galime suformuluoti išvadą, kad ne visada vien duomenų perdavimo tinklo resursai įtakoja paslaugos kokybinius parametrus, tačiau nemažą įtaką turi ir serverio turimi resursai, t.y. jo sugebėjimai aptarnauti klientines užklausas konkrečiu laiko momentu. Taip pat įtaką turi ir kliento vartojama aparatūrinė bei programinė įranga, tačiau šią įtaką yra labai sunku įvertinti.

7.2. Web paslaugos vėlinimas dėl interneto naršyklės laikinosios atminties (cache) įtakos

7.2.1. Tyrimų tikslas

WWW puslapiai paprastai yra sudėtingos kompleksinės struktūros dokumentai, susidedantys iš HTML kalba rašytų tekstinių duomenų, grafinės, video ar garso informacijos failų. Norint juos atvaizduoti vartotojui, visa informacija turi būti parsiončiama į kliento kompiuterį bei

atvaizduojama grafinėje naršyklėje (*angl. browser*). Kiekviena grafinė naršyklė šių puslapių saugojimui „paskiria“ tam tikrą kiekį vietos kompiuterio kietajame diske, kuri vadinama laikinąja atmintimi (*angl. cache*). Kadangi dažniausiai interneto puslapiai nekinta labai dažnai, o jeigu kinta, tai ne visa juose esanti informacija, pakartotinis WWW puslapio parsisiuntimas turėtų užtrukti trumpiau – nes ne visa informacija bus siunčiama duomenų perdavimo tinklu, o tik tai, kuri yra pasikeitusi, t.y. nepakitusi informacija tiesiog paimama iš kliento kompiuterio laikinosios atminties (cache). Šio darbo tikslas buvo atlikti eksperimentus ir ištirti interneto naršyklės laikinosios atminties (cache) poveikį bendrai (vartotojui realiai pastebimai) tam tikros WWW paslaugos vėlinimui.

7.2.2. Matavimų metodika

Tyrimas atliktas trejose skirtingose duomenų perdavimo tinklų architektūrose: KTU ITPI LITNET valdymo centre esančiame vietiniame tinkle – nuo klientinio kompiuterio iki WWW serverio dviragis.ktu.lt, kuriame veikia visa eilė interneto svetainių (www.litnet.lt, tinklas.ktu.lt ir pan.), iš namų kompiuterio (prieiga prie interneto per ADSL liniją) iki populiaraus WWW serverio www.delfi.lt, bei prieiga prie interneto per testinį 64kbps greičio prisijungimą iki www.delfi.lt ir www.cnn.com svetainių.

Klientiniame kompiuteryje veikė LINUX operacinė sistema su KDE grafine sąsaja ir Mozilla Firefox interneto naršykle. Buvo sukurti matavimų automatizavimui reikalingi komandų scenarijai (*angl. scripts*), kurių pagalba minėtos svetainės buvo periodiškai parsisiunčiamos. Laikas stebėtas naudojantis paketų sekimo/fiksavimo programine įranga „tcpdump“ [18]. Buvo atliekama po 3, 5 arba 10 parsisiuntimų/matavimų atitinkamai kas 2, 1 ir 0,5 minutės, periodiškai tai atliekant kas kelias valandas. Programos „tcpdump“ pagalba gauta informacija saugota atskiruose failuose.

Taigi, buvo siekiama nustatyti, kaip keičiasi tam tikro WWW puslapio užkrovimo trukmė nuo užkrovimo periodiškumo.

7.2.3. „Cache“ įtakos vėlinimui matavimo rezultatai

Atlikus visą eilę matavimų su skirtingais laiko periodiškumo parametrais trejose duomenų perdavimo tinklo aplinkose, buvo gauti tokie galutiniai rezultatai:

4 lentelė. Interneto naršyklės atminties (cache) įtaka WEB paslaugos kokybei – apibendrinti rezultatai.

	Testas	Pirmi puslapių atvertimai (s)	Paskutiniai puslapių atvertimai (s)	Santykinis skirtumas (%)
Artimas matavimas	1	15,83	15,65	1,14
	2	15,67	14,41	8,04
Tolimas matavimas	3	18,35	15,89	13,41
	4	15,22	14,16	6,96
Tolimas matavimas – lėtas kanalas 64kbps	5	77,42	68,66	11,32
	6	144,99	125,39	13,51

Pirmi puslapių atvertimai – tai laikas, sugaištamasis parsisiųsti WWW puslapį, kai šis siuntimas vykdomas periodiškai, praėjus tam tikram ilgesniam laikui po paskutinių puslapių atvertimų. Visa eilė matavimų buvo vykdoma po 12 – 24 valandas, ir visuose jų rezultatuose matėsi, jog po ilgesnės pertraukos puslapiai buvo parsisiunčiami su didesniu vėlinimu. Santykinis skirtumas pateiktas 4-oje lentelėje procentų pavidalu. Matoma, kad vartotojo kompiuterio laikinosios atminties įtaka web paslaugų vėlinimui labiausiai pastebima esant mažos spartos duomenų perdavimo kanalui tarp WWW serverio ir kliento.

7.3. WWW paslaugos kokybė

Tradicinėse SLA sutartyse web paslaugos vėlinimas nėra apibrėžiamas, tačiau paskutiniu metu pradedama svarstoma galimybė naudoti minėta WRT (*angl. Web Response Time*) metrika paslaugų kokybei vertinti [15].

Web vėlinimo reikšmės turi būti periodiškai matuojamos naudojantis tam skirtais aparatūriniais ir programiniais produktais. Kadangi web paslaugos kokybę pastebi paslaugos vartotojas savo personaliniame kompiuteryje, svarbu parinkti reikiama matavimo metodika, kuri kuo tiksliau atspindėtų vartotojo išsivaizduojamą paslaugos kokybę. Iš 7.1 skyrelio išvadų matome, kad ne tik duomenų perdavimo tinklo resursai įtakoja paslaugos kokybinius parametrus. Nemažą įtaką turi serverio (ar serverių sistemų) turimi resursai (t.y. sugebėjimas aptarnauti klientines užklausas konkrečiu laiko momentu). Vartotojo personalinio kompiuterio aparatūrinė bei programinė įranga taip pat gali turėti nežymią įtaką paslaugos kokybei (7.2. skyrelis), tačiau paslaugos teikėjui dažniausiai ją vertinti yra labai sunku.

Taigi matavimai neturėtų apsiriboti tik magistralinėse paslaugos teikėjo duomenų perdavimo tinklo ryšio linijose vykdomais tinklo vėlinimo matavimais. Kai kurios matavimo technologijos (pvz. Cisco SAA) turi galimybę atlikti ne tik duomenų perdavimo tinklo parametrų, bet ir atskirų paslaugų parametrų matavimus. Paslaugos atsako laikas (vėlinimas iki paslaugas teikiančių serverių) turėtų būti periodiškai matuojamas iš vartotojo tinklo prieigos, t.y. kuo arčiau vartotojo: tinklo įrangoje arba specializuotoje programinėje įrangoje, įdiegtoje vartotojo kompiuteryje. Tokiu būdu (bet dažniausiai skirtingomis priemonėmis) paslaugos kokybės parametrus galėtų stebėti tiek paslaugos teikėjas, tiek ir paslaugos vartotojas.

Minėtu sprendimu būtų įvertinamas tiek paslaugos vėlinimas (į kurią tuo pačiu įeina ir tinklo dvikryptis vėlinimas - RTT), tiek paslaugos pateikiamumas. Tuo pačiu būtų galimybė vertinti tiek paslaugos vėlinimo sklaidą (analogija duomenų perdavimo tinklo vėlinimo sklaidai), tiek ir nepavykusių HTTP sesijų skaičių (analogija duomenų perdavimo tinklo paketų praradimui). T.y. konkrečios duomenų centrų paslaugos atveju tinklo parametrai galėtų būti keičiami aukštesnio lygio (apimančius daugiau sistemos dalių) paslaugų kokybiniais parametrais.

7.1 ir 7.2 skyreliuose buvo skaičiuojamas web paslaugos kokybinis parametras: paslaugos vėlinimas. Šis parametras įtakoja bendrą paslaugos kokybę, kuri išreiškiama formule (1). Šioje formulėje yra reikalingas paslaugos vėlinimo įvertis L intervale $[0;1]$, kuris gali būti paskaičiuojamas pagal (18) formulę. Vidurkinė paslaugos vėlinimo reikšmė L' (milisekundėmis) gali būti paskaičiuojama apibrėžtam laiko momentui (pvz. dienai, savaitei ar mėnesiui). SLA sutartyje gali būti apibrėžtos web paslaugos SLA nurodoma reikalaujama L_{SLA} ir minimali L_{min} vėlinimo reikšmės. Vidutiniam vėlinimui viršijus reikalaujamą reikšmę, SLA web paslaugos vėlinimo metrikos įsipareigojimas pažeidžiamas - paslaugos teikėjas privalo prisiimti atsakomybę už tai.

Bendras web paslaugos kokybės įvertis gaunamas apskaičiavus paslaugos vėlinimo įvertį L , įstačius jį į kokybės formulę (1). Tuo pačiu turi būti įvertinami ir paskaičiuojami kiti (1) formulėje atsispindintys kokybiniai parametrai (priklausomai nuo kiekvieno duomenų teikėjo SLA specifiкуotų parametrų ir vartotojo poreikių): paslaugos pateikiamumas, duomenų perdavimo tinklo parametrai (vėlinimas, vėlinimo sklaida, informacijos perdavimo sparta) ir t.t.

8. MAGISTRINIO DARBO IŠVADOS

Sudaryta duomenų centrų paslaugų kokybės vertinimo metodika, kuria gali vadovautis paslaugos vartotojas, rinkdamasis duomenų centrų paslaugų teikėją bei geriausiai poreikius atitinkančią paslaugą. Taip pat ją galėtų naudotis duomenų centrų paslaugų teikėjas, projektuodamas savo duomenų centro aukštos paslaugų kokybės užtikrinimo sistemas.

Šią duomenų centrų paslaugų kokybės vertinimo metodiką apibendrė išvados:

- Atlikus paslaugų lygio sutarčių (SLA) pateikiamų parametrų analizę, nuspręsta, kad skirtingoms paslaugoms šių parametrų įtaka bendrai paslaugos kokybei gali labai skirtis. Taip yra todėl, kad kokybė nėra vienareikšmiškas parametras – skirtingiems vartotojams ji gali atrodyti skirtinga (pagal poreikius). Pasiūlyta duomenų centro paslaugos kokybę skaičiuoti bedimensiniu dydžiu, į kurį įtrauktos SLA parametrų reikšmės bei svorių koeficientai.
- Duomenų centro paslaugos teikėjas, norėdamas laikytis SLA nurodytų reikalavimų, privalo turėti kokybinių parametrų užtikrinimo metodikas – naudotis atitinkamomis priemonėmis periodiniams matavimams atlikti. Buvo atliktas SLA parametrų matavimo technologijų Cisco SAA ir RIPE TTM įvertinimas bei palyginimas, ir nustatyta, kad:
 - Cisco SAA tinka duomenų centrams, kurie kuria savo duomenų perdavimo tinklo infrastruktūrą Cisco įrenginių pagrindu. Ji turi daugiau matavimo funkcijų: tiek tarp magistralinių tinklo įrenginių, tiek ir iki vartotojo galinių įrenginių.
 - RIPE TTM nepriklauso nuo duomenų perdavimo tinklo įrenginių – veikia su visų tipų IP tinklais, tačiau tinka matavimams tik tarp magistralinių tinklo įrenginių.
- Atlikus „plono-kliento“ – terminalinių serverių – architektūrų bei reikiamų resursų analizę, prieita išvados, kad išaugus tinklo pralaidumui, tokia sistema techniniu požiūriu gali būti ekvivalentiška įprastai personalinių kompiuterių sistemai, o ekonominiu požiūriu – naudingesnė, nes vartotojas turi mokėti tik už tą programinę įrangą, ir tik tuo metu, kai ją naudoja.
- Atlikus duomenų centrų paslaugų bei duomenų centrų serverių sistemų klasifikaciją, pastebėta, kad tiek „hostingo“, tiek ir terminalinių serverių paslaugų pateikiamumas

užtikrinamas naudojant struktūrinių elementų (serverių, ryšio linijų, duomenų saugyklų) rezervavimus.

- Buvo prieita išvados, kad apie Web paslaugos kokybę negalima spręsti vien tik pagal įprastus duomenų perdavimo tinklo kokybinius parametrus (vidutinis tinklo apkrovimas, vidutinis vėlinimas, paketų praradimo procentas ar pan.). Kelių dimensijų vertinimas (pagrįstas aukštesnio lygio parametrais – pvz. WRT – *Web Response Time*) įgalina gauti tikslesnius ir į vartotoją orientuotus paslaugų kokybės vertinimo rezultatus. Tą parodė ir atlikti eksperimentai:
 - web paslaugos vėlinimo matavimų rezultatai rodo, kad web paslaugos kokybė priklauso tiek nuo duomenų perdavimo tinklo parametrų, tiek ir nuo serverio bei kliento personalinio kompiuterio aparatūrinės ir programinės įrangos resursų.
 - matavimai vartotojo interneto naršyklės laikinosios atminties (cache) įtakai web paslaugoms ištirti parodė, kad jos įtaka web paslaugų kokybei labiausiai pastebima esant mažos spartos duomenų perdavimo kanalui tarp WWW serverio ir kliento.
- Pristatytas straipsnis konferencijoje „Informacinės Technologijos 2006“. Tema: „Plono Kliento Modelio Taikymas Duomenų Centrų Paslaugoms“ (bendra autoriai prof. dr. R. Plėštys – KTU, L. Gustauskas – KTU)
- Pristatytas straipsnis konferencijoje „Informacinės Technologijos 2007“. Tema: „WEB paslaugos kokybės įvertinimas“ (bendra autorius prof. dr. R. Plėštys – KTU)

9. LITERATŪRA

- [1]. Primus Networks: Data Center Services. [Žiūrėta 2007-03-21]. Prieiga per Internetą:
<http://www.primusnetworks.com/services/data/>
- [2]. Samsung e-DataCenter Services. [Žiūrėta 2007-03-21]. Prieiga per Internetą:
<http://www.samsungsdsa.com/product/services/brochure.pdf>
- [3]. Web Hosting Services. [Žiūrėta 2007-03-25]. Prieiga per Internetą:
<http://www.dmiwebhosting.com/web-hosting/webhost.php>
- [4]. Cisco RTT-MON MIB. [Žiūrėta 2005-03-25]. Prieiga per Internetą:
<ftp://ftp.cisco.com/pub/mibs/v2/CISCO-RTTMON-MIB.my>
- [5]. Service Level Monitoring with Cisco IOS Service Assurance Agent. „Cisco systems“ white paper. [Žiūrėta 2005-05-15]. Prieiga per Internetą:
http://www.cisco.com/en/US/tech/tk648/tk362/technologies_white_paper09186a00800e9012.shtml
- [6]. Infovista Vistaview products. [Žiūrėta 2006-04-15]. Prieiga per Internetą:
<http://www.infovista.com/products/vistaviews/>
- [7]. User Guide for Internetwork Performance Monitor 2.5. [Žiūrėta 2006-04-15]. Prieiga per Internetą:
http://www.cisco.com/en/US/products/sw/cscowork/ps1008/products_user_guide_chapter09186a008017ae7f.html
- [8]. Fluke Networks Response Watch. [Žiūrėta 2007-05-15]. Prieiga per Internetą:
<http://www.flukenetworks.ca/FNet/en-ca/findit?Document=2826280>
- [9]. RIPE Test Traffic Measurements. [Žiūrėta 2006-05-20]. Prieiga per Internetą:
<http://www.ripe.net/ripe/docs/ripe-158.html>
- [10]. High-Availability System Architecture. [Žiūrėta 2005-11-09]. Prieiga per Internetą:
<http://www.microsoft.com/technet/archive/windows2000serv/plan/hiavsys.mspx>

- [11]. System Redundancy. [Žiūrėta 2005-11-09]. Prieiga per Internetą: <http://www.microsoft.com/technet/archive/itsolutions/ecommerce/plan/sysredun.msp>
- [12]. Limits of Wide-Area Thin-Client Computing. [Žiūrėta 2005-10-21]. Prieiga per Internetą: <http://portal.acm.org/citation.cfm?id=511363>
- [13]. Evaluating Windows NT Terminal Server Performance. [Žiūrėta 2005-10-21]. Prieiga per Internetą: <http://www.usenix.org/publications/library/proceedings/usenix-nt99/wong.html>
- [14]. Measuring Thin-Client Performance Using Slow-Motion Benchmarking. [Žiūrėta 2005-10-23]. Prieiga per Internetą: http://www.ncl.cs.columbia.edu/publications/tocs2001_slowmotion.pdf
- [15]. J. Martin, A. Nilsson, „On Service Level Agreements for IP Networks“, INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE.
- [16]. J. Padhye, et. Al., “Modeling TCP Throughput: A Simple Model and its Empirical Validation”, ACM SIGCOMM98, 1998.
- [17]. A. Schmietendorf, R. Dumke, D. Reitz, „SLA management - challenges in the context of Web-service-based infrastructures“, Proceedings of the IEEE International Conference on Web Services (ICWS'04)
- [18]. <http://www.tcpdump.org>
- [19]. <http://www.gnu.org/software/wget/>
- [20]. Rimantas Plėštys, Gytis Vilutis, Donatas Sandonavičius, Rasa Vaškevičiūtė, Rimantas Kavaliūnas: “The Measurement of Grid QoS Parameters” – straipsnis priimtas į IEEEExplore organizuojamą konferenciją “ITI 2007” Kroatijoje;

10. QUALITY EVALUATION OF DATACENTER SERVICES

SUMMARY

Nowadays datacenters provide different network, hosting, terminal desktop and other services, and it has become very important to be able to evaluate the quality of these services. SLA (Service Level Agreement) metrics define the quality of service, but it also depends on the needs of the particular user. So the evaluation methodology was created in order to compare the quality of services, both: for the user, and for the datacenter service provider.

A service provider has to implement the server architecture with network connection and server redundancy in order to assure high service availability.

Two most popular datacenter services were analyzed: “thin-client” architecture and web services. The web service quality evaluation experiments have been made to show that web oriented performance metric Web Response Time (WRT) is more informative for web service quality assurance, comparing to the traditional latency and loss metrics in the network environment. “Thin-client” architecture analysis show that it can be used both in LAN and WAN networks, and it can be economically useful to outsource IT systems to datacenters by using this technology.

Network and server resource performance monitoring has to be done in order to supply high quality services - so datacenter service providers have to use monitoring hardware and software to make periodical measurements and collect, store and process the results.

11. PRIEDAI

11.1. Web paslaugos vėlinimo įvertinimų išsamūs rezultatai

Artimas matavimas: 193.219.184.93 - www.itpi.ktu.lt			Tolimas matavimas: 85.206.232.6 - www.itpi.ktu.lt		
Laikas	Serverio trukme - ms	Kliento trukme - ms	Laikas	Serverio trukme - ms	Kliento trukme - ms
03:10:03	178,443	230	03:10:03	685,536	810
03:15:03	152,659	210	03:15:04	708,176	850
03:20:03	151,802	210	03:20:04	1049,748	1330
03:25:03	155,852	210	03:25:04	677,722	840
03:30:03	157,956	210	03:30:04	669,917	790
03:35:03	155,83	210	03:35:04	683,175	800
03:40:03	158,077	210	03:40:04	998,808	1370
03:45:03	158,016	210	03:45:04	670,97	790
03:50:03	157,497	210	03:50:04	669,191	790
03:55:03	158,102	210	03:55:04	1043,354	1340
04:00:03	158,112	210	04:00:05	954,153	1540
04:05:03	158,224	210	04:05:04	673,402	820
04:10:03	157,455	210	04:10:05	677,754	800
04:15:04	157,943	210	04:15:05	675,975	800
04:20:04	157,369	210	04:20:05	743,728	1310
04:25:04	157,965	210	04:25:05	676,116	800
04:30:04	157,681	210	04:30:05	666,354	790
04:35:04	151,426	230	04:35:05	681,869	820
04:40:04	167,341	230	04:40:06	956,304	1470
04:45:06	177,58	1520	04:45:07	684,737	2030
04:50:05	154,93	490	04:50:06	669,064	810
04:55:05	154,892	210	04:55:06	892,287	1300
05:00:05	156,486	210	05:00:06	1182,168	1690
05:05:05	157,46	210	05:05:06	685,528	840
05:10:05	157,426	210	05:10:06	688,395	810
05:15:05	157,45	210	05:15:06	693,965	820
05:20:05	156,427	210	05:20:06	778,703	1110
05:25:05	166,913	220	05:25:06	683,617	800
05:30:05	153,302	210	05:30:06	682,053	800
05:35:05	151,341	210	05:35:06	696,5	820
05:40:05	277,347	330	05:40:07	670,763	800
05:45:05	156,914	210	05:45:07	677,193	820
05:50:05	157,416	210	05:50:07	675,198	820
05:55:05	165,526	220	05:55:07	693,584	1190

06:00:05	155,061	230		06:00:08	1143,959	1910
06:05:05	157,458	210		06:05:08	670,013	790
06:10:05	156,787	210		06:10:08	677,483	800
06:15:05	157,544	210		06:15:08	664,617	790
06:20:05	157,297	210		06:20:08	683,492	810
06:25:05	156,567	210		06:25:08	671,285	790
06:30:05	157,27	210		06:30:08	679,877	800
06:35:05	157,635	210		06:35:08	678,861	800
06:40:05	157,575	210		06:40:08	678,071	840
06:45:05	157,879	210		06:45:08	678,795	800
06:50:05	157,559	210		06:50:08	679,148	810
06:55:05	157,475	210		06:55:08	910,51	1030
07:00:05	155,036	210		07:00:09	946,566	1510
07:05:05	155,514	210		07:05:09	680,016	800
07:10:05	156,948	210		07:10:09	667,929	790
07:15:05	157,214	210		07:15:09	674,888	840
07:20:05	157,268	210		07:20:09	676,024	840
07:25:05	156,689	210		07:25:09	683,188	810
07:30:06	156,609	210		07:30:09	662,771	820
07:35:06	154,758	330		07:35:09	667,039	790
07:40:06	156,671	210		07:40:09	676,364	830
07:45:06	157,122	210		07:45:09	678,253	800
07:50:06	157,294	210		07:50:09	687,228	820
07:55:06	157,099	210		07:55:09	678,148	800
08:00:06	157,268	210		08:00:10	1249,015	2110
08:05:06	156,433	210		08:05:10	722,395	840
08:10:13	156,696	7950		08:10:10	679,025	800
08:15:06	156,89	210		08:15:10	682,093	820
08:20:06	156,304	210		08:20:10	667,506	820
08:25:06	2348,284	2400		08:25:11	1693,169	1820
08:30:06	156,903	210		08:30:11	3283,39	3420
08:35:06	156,717	210				
08:40:06	156,919	210		08:40:10	799,641	930
08:45:06	158,049	210		08:45:10	671,224	840
08:50:06	156,736	210		08:50:11	812,91	1000
08:55:06	156,913	210		08:55:11	676,069	800
09:00:06	157,967	210		09:00:11	1608,359	2070
09:05:06	155,54	210		09:05:11	681,195	800
09:10:06	156,526	210		09:10:11	681,266	800
09:15:06	156,82	210		09:15:11	676,224	840

09:20:06	156,766	210		09:20:11	717,097	870
09:25:06	156,587	210		09:25:11	696,099	890
09:30:06	156,471	210		09:30:11	690,846	900
09:35:06	160,075	220		09:35:11	685,59	810
09:40:06	152,777	210		09:40:11	666,188	820
09:45:06	166,663	220		09:45:11	682,67	920
09:50:06	467,226	520		09:50:12	879,232	1040
09:55:06	156,647	210		09:55:12	679,178	810
10:05:06	158,315	210		10:00:12	1060,664	1440
				10:05:12	682,223	830
10:10:06	167,367	220		10:10:12	668,896	800
10:15:06	167,396	220		10:15:12	682,562	830
10:20:06	156,586	210		10:20:13	1949,782	2320
10:25:06	157,024	210		10:25:12	965,104	1110
10:30:07	156,079	210		10:30:12	676,708	820
10:35:07	155,201	240		10:35:12	687,394	810
10:40:07	155,413	210		10:40:13	697,195	890
10:45:07	155,827	210		10:45:13	684,04	890
10:50:07	156,454	210		10:50:13	681,31	920
10:55:07	156,056	210		10:55:13	696,138	860
11:00:07	156,314	210		11:00:13	907,164	1270
11:05:07	163,847	220		11:05:13	689,654	820
11:10:07	654,506	710		11:10:13	727,726	860
11:15:07	167,715	220		11:15:13	687,407	940
11:20:07	156,091	210		11:20:13	683,798	940

11.2. Straipsnis konferencijoje „Informacinės Technologijos 2006“

PLONO KLIENTO MODELIO TAIKYMAS DUOMENŲ CENTRŲ PASLAUGOMS

Rimantas Plėštys, Rokas Zakarevičius, Linas Gustauskas

Kauno Technologijos Universitetas, Kompiuterių tinklų katedra, Studentų g. 50

Vis dažniau taikomųjų programų paslaugų teikėjai teikia programinės įrangos nuomos paslaugas, leisdami naudotis duomenų centrų serverių resursais (programine įranga, duomenų laikmenomis, duomenimis, ryšio linijomis ir t.t.) per interneto naršyklę HTTP protokolu arba naudojant specializuotą programinę įrangą. Jomis naudojasi taip vadinami „ploni klientai“ – aparatinė ar/ir programinė įranga, veikianti kliento pusėje, skirta prisijungti prie nutolusio paslaugų serverio. Tokiu atveju klientas apmoka paslaugas teikėjui tik už tą programinę įrangą, kuria naudojasi tam tikrą laiką.

Šešių populiariausių „plono kliento“ sistemų efektyvumas nustatytas atlikus trijų rūšių testavimus: vėlinimo – matuojant atsako laiką, web testavimą – matuojant naršymo efektyvumą, bei vaizdo duomenų testavimą – matuojant video peržiūros efektyvumą.

Norint užtikrinti didelį paslaugos pateikiamumą, paslaugos teikėjas turi įdiegti serverių architektūrą su tinklo linijų bei serverių rezervavimais. Didelis klientų skaičius gali būti paskirstytas terminalinių serverių grupei. Kiekvieno vartotojo informacija saugoma centralizuotose duomenų saugyklose, pasiekiamose per tam skirtus antro lygio failų serverius.

1 Įžanga

Taikomųjų programų paslaugų teikėjas – (angl. Application Service Provider) teikia programinės įrangos nuomos paslaugas leisdamas naudotis duomenų centrų serverių resursais (programine įranga, duomenų laikmenomis, duomenimis, ryšio linijomis ir t.t.). Šios paslaugos pasiekiamos per naršyklę HTTP protokolu arba naudojant specializuotą programinę įrangą. Jomis naudojasi taip vadinami „ploni klientai“ – aparatinė ar/ir programinė įranga, veikianti kliento pusėje, skirta prisijungti prie nutolusio paslaugų serverio. Tokiu atveju klientas apmoka paslaugas teikėjui tik už tą programinę įrangą, kuria naudojasi tam tikrą laiką.

Darbe parodyta, kad modelis gali būti naudojamas net ir tada, kai klientas ir serveris yra nutolę per kelis tūkstančius kilometrų. Rezultatai rodo, kad paprasta taškų perdavimu (angl. Pixel-based) pagrįsta „plono kliento“ technologija gali veikti žymiai geriau nei sudėtinga „plono kliento“ technologija, kur perduodamos aukštesnio lygio komandos.

2 „Plono kliento“ modelio galimybių įvertinimas

„Plono kliento“ modelio sistema sudaryta iš serverio ir kliento, kurie komunikuoja per duomenų perdavimo tinklą. Klientas naudojasi grafiniu ekranu, klaviatūra, pele, bei kartais papildomais įvesties įrenginiais, o visos programų atliekamos operacijos yra vykdomos serveryje, o rezultatai bei vartotojo komandos siunčiami tinklu.

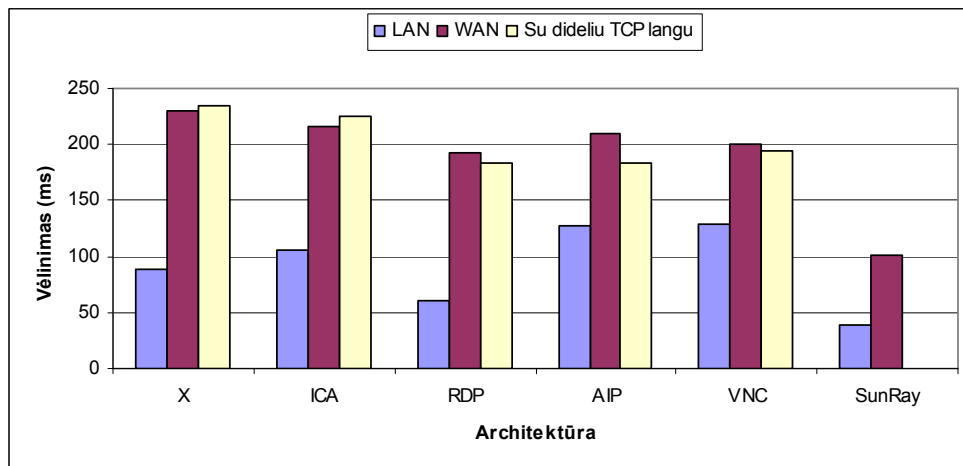
„Plono kliento“ modelio efektyvumo ir galimybių tyrimai buvo atlikti pasirinkus šešias populiariausias „plono kliento“ pagrindu veikiančias architektūras: Citrix MetaFrame, Microsoft Windows 2000 Terminal Services, AT&T VNC, Tarantella, Sun Ray ir X. Jos pasirinktos, atsižvelgiant į jų populiarumą, veikimo efektyvumą bei veikimo skirtumus [1].

„Plono kliento“ sistemų efektyvumas nustatytas atlikus trijų rūšių testavimus: vėlinimo – matuojant atsako laiką, web testavimą – matuojant naršymo efektyvumą, bei vaizdo duomenų testavimą – matuojant video peržiūros efektyvumą. Vėlinimo matavimai buvo atlikti, norint išmatuoti paprastų elementarių operacijų trukmę, o web ir vaizdo matavimai – norint sudaryti realistiškesnę bendrą sistemos veikimo vaizdą.

Trijų elementarių operacijų rezultatai: vieno simbolio atspausdinimo, teksto vartymo (angl. scrolling) ir ekrano pildymo tam tikro dydžio stačiakampiu (angl. fill), o taip pat dviejų sudėtingesnių operacijų, kurių metu persiunčiami paveikslėliai, rezultatai pateikti 1 pav.

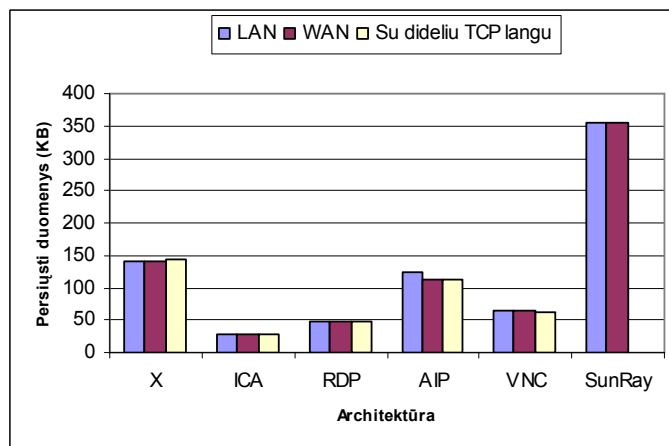
Bendru atveju paprastoms operacijoms – tokioms, kaip spausdinimas klaviatūra, kursoriaus judinimas ar pelės judesys, sistemos priimtinas atsako vėlinimas turėtų neviršyti 50-150 ms ribos, kad esant tokiam vėlinimui žmogus nepastebėtų sistemos darbingumo sulėtėjimo.

Iš pateiktų rezultatų galime padaryti išvadą, kad dalis sistemų pasirodė geriau nei reikalaujama 150 ms ribinė reikšmė didžiajai daliai operacijų. Konkrečiai Sun Ray sistema užtikrina iki 100 ms vėlinimą tiek vietinio tinklo, tiek globalaus tinklo aplinkose.

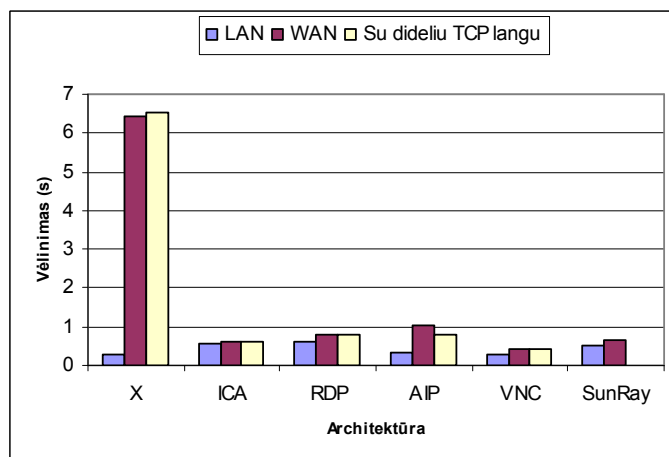


1 pav. Letter, Scroll, Fill, Red Bitmap Image ir Image operacijų apibendrinti rezultatai

2 ir 3 paveikslėliuose pateikti web aplikacijų testavimo rezultatai visoms šešioms testuojamoms „plono kliento“ sistemoms. 2 pav. pavaizduotas vidutinis persiųstų duomenų kiekis vienam web puslapiui skirtingoms tinklo architektūroms: vietiniame tinkle (LAN), globaliajame tinkle (WAN) bei naudojant didelį 1 MB TCP langą.



2 pav. Persiųsti duomenys web puslapiui



3 pav. Vėlinimas web puslapiui

3 pav. pavaizduotas vidutinis sistemos vėlinimas vienam web puslapiui atvaizduoti. Reikalaujama, kad vieno puslapio atsiuntimas turėtų užtrukti ne ilgiau nei vieną sekundę. Didžioji dalis sistemų užtikrina web puslapio persiuntimą ir atvaizdavimą iki vienos sekundės.

Atlikti matavimai rodo, kad Sun Ray architektūros sistema veikė geriausiai visuose bandymuose. Lyginant sistemas tarpusavyje pastebėta, kad skirtingų architektūrų veikimo kokybė labai skiriasi.

Testuojant video aplikacijas, didžioji dalis „plono kliento“ sistemų pasirodė neefektyvios tiek vietiniame, tiek globaliajame tinkle. Tik Sun Ray sistema užtikrino 70% norimos kokybės. Kitose sistemose kokybė nesiekė 35% ir tai buvo nepriimtina.

Gauti rezultatai remiasi atliktais matavimais, kai nebuvo atsižvelgta į paketų praradimą bei linijų persipildymą, kurie pasitaiko realiose globalaus tinklo (WAN) situacijose. Testavimai buvo atlikti emuliuojant Internet2 tinklą su atsitiktiniu 10% paketų praradimu.

3 Windows Remote Desktop Protocol (RDP) efektyvumo įvertinimas

Šiuo protokolu veikia MS Windows Terminal Services ir Remote Desktop Connection paslaugos, kurių pagalba vartotojas gali prisijungti prie nutolusio duomenų centro ir per virtualų darbstaalį dirbti su juo tinkle taip, lyg dirbtų darbo vietoje. Kadangi Windows operacijų sistema yra populiariausia kompiuteriams darbo stotims, todėl tikslinga įvertinti RDP protokolo tinkamumą „plono kliento“ sistemoms.

Šio protokolo efektyvumas tirtas atliekant įvairius elementarius veiksmus, kurie paprastai vykdomi dirbant kompiuteriu Windows aplinkoje.

Tyrimo rezultatai pateikti 1 lentelėje [2].

1 lentelė. Remote Desktop Protocol elementarių veiksmų tinklo resursų naudojimas

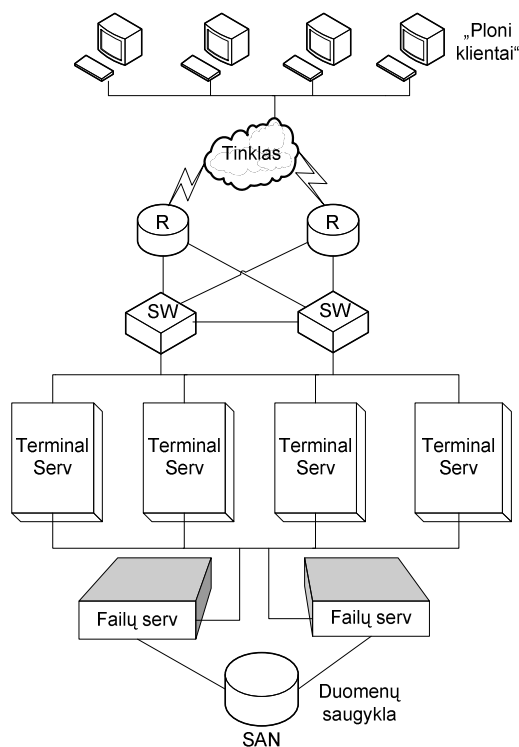
	Valdymo duomenų kanalas	Atvaizdavimo duomenų kanalas		MS Windows
		MS Word programa	WordPerfect programa	
Tekstų spausdinimas (75 žpm)	1,56 KBps	6,26 KBps	11,66 KBps	-
Pelės judesiai	1,95 KBps	-	-	-
Meniu navigacija	-	-	-	3,12 KBps
Meniu navigacija su "Alt" klavišu	-	39,82 KBps	17,13 KBps	-
Dokumento vartymas (angl. scrolling)	-	59,24 KBps	192,54 KBps	-

Kadangi RDP protokolas yra sukurtas Microsoft kompanijos ir jo veikimo specifikacija viešai neskelbiama, negalėjome pateikti detalių RDP testavimo rezultatų paaiškinimo. Testuojant buvo stengiamasi emuliuoti tipinius vartotojo veiksmus Windows sistemoje, ir buvo matuojamas tinklo apkrovimas bei atsako laikai. Iš pateiktų rezultatų galime daryti išvadą, kad pikinis tinklo resursų panaudojimas gali siekti iki 200KBps, tačiau dažniausiai neviršija 60 KBps. Taigi RDP protokolu veikiančios „plono kliento“ sistemos gali būti naudojamos tiek vietiniame, tiek ir globaliame tinkle, jeigu jis pasiekiamas per plačiąjuostes ryšio prieigas.

4 „Plono kliento“ modelio taikymas didelio pateikiamumo serverių sistemose

Plono kliento terminaliniai serveriai ir serveriai, kuriuose saugomi visi vartotojų duomenys, turi būti suprojektuoti kaip atskiri serverių klasteriai su duomenų saugyklomis ar diskų masyvais [3]. Terminaliniai serveriai tik aptarnauja klientus, t.y. juose veikia tam tikro kliento programos, tačiau visi duomenys yra pasiekiami per failų serverius, kurie turi tiesioginį ryšį su centralizuotomis didelio našumo duomenų saugyklomis. Toks terminalinių serverių klasteriavimas leidžia paskirstyti apkrovimus ir tinkamai bei kokybiškai aptarnauti didelį klientų skaičių, kai klientai nėra susieti su konkrečiu terminaliniu serveriu.

4 pav. pateikta duomenų centrų panaudojimo „plono kliento“ sistemoje schema. Duomenų centrų panaudojimo atveju tinklas, skiriantis plonus klientus ir duomenų saugyklas, paprastai yra globalus tinklas. „R“ simboliais pažymėti maršrutizatoriai, „SW“ simboliais – komutatoriai. Tai yra duomenų centro tinklo lygis, užtikrinantis duomenų centro pasiekiamumą iš „plonų klientų“. Norint įdiegti „plono kliento“ architektūrą paremtą virtualaus darbstaalio paslaugą duomenų centre, turi būti įvedamas naujas duomenų centro įrangos lygis – terminaliniai serveriai. Jų užduotis yra tiesiogiai aptarnauti „plonus klientus“: priimti prisijungimus, suteikti duomenų centrų atminties ir procesoriaus resursus bei prieigą prie duomenų. Vartotojų duomenys nesaugomi terminaliniuose serveriuose. Tam skirtos duomenų saugyklos, užtikrinančios efektyvų ir saugų centralizuotą duomenų saugojimą. Tam gali būti naudojamos duomenų saugyklos, pagrįstos SAN (angl. Storage Area Network) technologija. Duomenys iš duomenų saugyklų pasiekiami per tam skirtus failų serverius. Duomenų centruose tokie failų serveriai suteikia prieigą prie duomenų visoms duomenų centrų paslaugoms. Jie gali būti tinkamai naudojami ir „plono kliento“ architektūros atveju kaip tarpininkai tarp terminalinių serverių ir duomenų saugyklų.



4 pav. Duomenų centrų panaudojimas „plono kliento“ sistemoje

5 Išvados

Plono kliento modelis gali būti tinkamai naudojamas tiek lokaliame, tiek globaliame tinkle. Tipinis Windows taikomųjų programų naudojimas RDP protokolu nutolusiame serveryje vartoja nedaug tinklo resursų. Pikinis tinklo resursų panaudojimas gali siekti iki 200KBps, tačiau dažniausiai neviršija 60 KBps.

Norint užtikrinti didelį paslaugos pateikiamumą, paslaugos teikėjams tikslinga įdiegti serverių architektūrą su tinklo linijų bei terminalinių serverių rezervavimais. Išaugus tinklo pralaidumui, tokia sistema techniniu požiūriu gali būti ekvivalentiška įprastai personalinių kompiuterių sistemai, kai duomenys saugomi ir apdorojami atskiruose kompiuteriuose. Ekonominiu požiūriu klientui naudingiau naudotis duomenų centrų paslaugomis, nes jis turi mokėti tik už tą programinę įrangą, kurią jis naudoja savo darbe.

6 Literatūros sąrašas

- [1] Limits of Wide-Area Thin-Client Computing. *Prieiga per Internetą*: [Žiūrėta 2005-10-21]
<http://portal.acm.org/citation.cfm?id=511363>
- [2] Evaluating Windows NT Terminal Server Performance. *Prieiga per Internetą*: [Žiūrėta 2005-10-21]
<http://www.usenix.org/publications/library/proceedings/usenix-nt99/wong.html>
- [3] High-Availability System Architecture. *Prieiga per Internetą*: [Žiūrėta 2005-11-09]
<http://www.microsoft.com/technet/archive/windows2000serv/plan/hiavsys.msp>

Thin Client Model Application for Datacenters

Nowadays more often Application Service Providers provide clients the possibility to use datacenter resources (applications, data storages, data, network connections etc.) via the Internet browser or specialized software. These services are used by so called “thin clients” – hardware or/and software equipment on a client side, which is used for connecting to the remote server. In such a case a user pays the provider only for the software he/she uses for a specified time.

The efficiency of six most popular “thin client” architectures has been measured by performing three types of testing: a latency benchmark for measuring response time, a web benchmark for measuring web browsing performance, and a video benchmark for measuring video playback performance.

A service provider has to implement the server architecture with network connection and server redundancy in order to assure high service availability. A large number of clients can be distributed among the group of servers. Each user's data is stored on centralized data storages, which is accessed through dedicated second-level file servers.

11.3. Straipsnis konferencijoje „Informacinės Technologijos 2007“

WEB PASLAUGOS KOKYBĖS ĮVERTINIMAS

Rimantas Plėštys, Rokas Zakarevičius

Kauno Technologijos Universitetas, Kompiuterių tinklų katedra, Studentų g. 50

Tradicinių SLA metrikų (pateikiamumo, tinklo vėlinimo, paketų praradimo procento) garantijos neužtikrina galinio vartotojo tinklo paslaugos kokybės. Konkrečiai pateikiamumas neįvertina pačios paslaugos kokybės, kuri gali būti ir nepriimtina vartotojui. Todėl tinklo paslaugoms, kurios veikia HTTP protokolo pagrindu, buvo nuspręsta pasiūlyti į žiniatinklio paslaugą orientuotą kokybės metriką "web atsako laiką" (angl. *Web Response Time* – *WRT*). Atlikti web paslaugos kokybės vertinimo eksperimentai iš serverio ir kliento perspektyvų, siekiant įvertinti duomenų perdavimo tinklo bei serverio resursų įtaką paslaugos kokybei. Atlikti eksperimentai trejose skirtingose tinklų architektūrose, siekiant nustatyti interneto naršyklės atminties (*cache*) įtaką web paslaugos kokybei, kuri labiausiai pastebima, esant mažos spartos duomenų perdavimo kanalui tarp WWW serverio ir kliento.

1 Įžanga

Pastaruoju metu, kuriant naujas IP paslaugų SLA (angl. *Service Level Agreement* – paslaugų lygio sutarčių) užtikrinimo metodikas, pastebima, kad tradicinių SLA metrikų (pateikiamumo, tinklo vėlinimo, paketų praradimo procento) garantijos neužtikrina galinio vartotojo tinklo paslaugų kokybės. Todėl yra siekiama įvesti naujas metrikas, kurios tiksliau atspindėtų konkrečios teikiamos paslaugos kokybę [1].

Siekiant užtikrinti aukšto lygio duomenų centrų paslaugas, vien tik pateikiamumo vertinimo nebeužtenka. Pateikiamumo parametras parodo, ar paslauga yra pasiekiamą, ar nepasiekiamą vartotojui, t.y. kiek laiko per fiksuotą laikotarpį vartotojas galės naudotis paslauga. Tačiau jis neįvertina pačios paslaugos kokybės, t.y. gali būti tokių atvejų, kai paslauga bus pasiekiamą vartotojui, tačiau jos kokybė bus nepriimtina.

Tradicinė SLA daugiau skirta norint apsaugoti paslaugų teikėjo interesus, prisiimant minimalius kokybės išpareigojimus galiniam vartotojui. Plečiantis duomenų centrų paslaugoms, didesnis dėmesys yra kreipiamas į jų kokybę. Paslaugų teikėjai tampa suinteresuoti tos kokybės užtikrinimu ir išpareigojimais klientams. Todėl būtina stebėti ir vertinti konkrečios paslaugos kokybę, t.y. kaip kinta paslaugos kokybė laike, kokios yra priimtinos ir nepriimtinos kitimo ribos.

Šiame darbe buvo siekiama ištirti WWW paslaugos veikimą skirtingomis sąlygomis bei sukurti efektyvią metodiką, siekiant periodiškai stebėti duomenų centro paslaugų kokybę. Visa tai reikalinga, norint užtikrinti maksimalios kokybės paslaugą.

2 WEB atsako laikas

Nagrinėjant realius duomenų perdavimo tinklų panaudojimo pavyzdžius, galime teigti, kad didžioji dalis kritinių tinklo paslaugų yra žiniatinklio (angl. *web*) paslaugos, kurios standartiškai veikia HTTP protokolu. Todėl buvo nuspręsta pasiūlyti į žiniatinklio paslaugą orientuotą kokybės metriką, siekiant kuo tiksliau įvertinti vartotojui teikiamos paslaugos kokybę.

Šią naują metriką „web atsako laikas“ [2] (angl. *Web Response Time* – *WRT*) galima būtų apibūdinti taip: klientas periodiškai inicijuoja HTTP objekto užklausą į tam tikrą web serverį (pvz. web serveris yra pagrindiniame įmonės tinkle, o vartotojas – tam tikro įmonės padalinio tinkle). Taigi, šios užklausos įvykdymo laikas ir būtų WRT rezultatas – tai realus laikas, kurio reikėjo, kad klientas gautų reikiamą informaciją iš serverio.

Šios papildomos aplikacijų lygio metrikos naudojimas turi kelis privalumus, lyginant su tradicinėmis „ping“ pagrindu veikiančiomis metrikomis:

- WRT metrikos rezultatai įvertina tiek paketų praradimo, tiek ir tinklo vėlinimo įtaką konkrečios paslaugos (Web) kokybei.
- Yra daug lengviau interpretuoti šios metrikos rezultatus, ir vertinti tai, kaip duomenų perdavimo tinklas veikia galinių vartotojų naudojamų paslaugų kokybę.

Remsimės statistiniu percentiliu ir tarsime, kad 95% web atsako laiko (WRT) rezultatų turi būti mažiau nei 2 sekundės (skaičiuojant 2 valandų laiko periodą) [2]. Taip sukuriamo „į vartotoją orientuota“ pateikiamumo metrika, kurią pažeidus – tinklas gali būti laikomas nepasiekiamu.

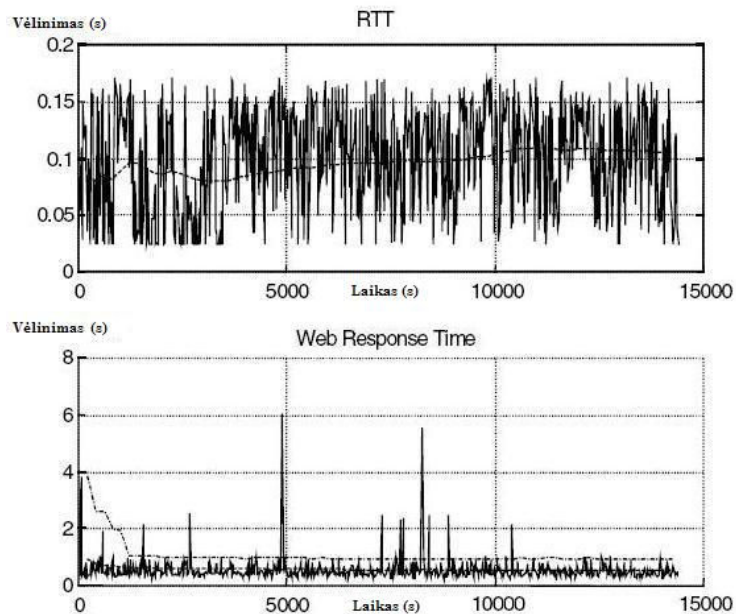
Įprastas vėlinimo išpareigojimas turi sekančius komponentus [2]:

$$RTT_{threshold} = RTT_{min} + T_{downstreamdelay} + T_{upstreamdelay} \quad (1)$$

Čia $RTT_{\min}$ - tuščios linijos abipusis vėlinimas. $T_{\text{downstreamdelay}}$ - bendras eilių vėlinimas „download“ kryptimi. Įvertinti vėlinimą, atsiradusį dėl eilių, paprastai yra gana sunku, nes jis priklauso nuo gana daug kintamųjų.

Jeigu modeliuosime atvejį, kai tinklo linijų apkrovimas yra artimas 50% ilgą laiko tarpą, gausime vėlinimo rezultatus, kad vidutinis mėnesio RTT (angl. *Round Trip Time* – abipusis vėlinimo laikas) gali būti labai artimas parametras SLA vėlinimo išpareigojimui. Esant dar blogesnėms sąlygoms, SLA išpareigojimui gali būti pažeisti, ir paslaugų teikėjas turės nuostolių.

1 pav. parodytas momentinis (išsitiesinė linija) ir vidutinis (brūkšninė linija) ping dvibusio atsako laiko (RTT) rezultatas (šiuo atveju vidutinis vėlinimas apie 100 milisekundžių). Yra sunku pagal šią statistiką spręsti, kaip tinklas veikia galinio vartotojo vartojamų paslaugų kokybę.



1 pav. Skirtingų metrikų rezultatai (tradicinio RTT ir naujo WRT) [2]

1 pav. apatinis paveikslas iliustruoja web atsako laiko metrikos skaičiavimo rezultatus. Brūkšninė kreivė atvaizduoja 95 percentilio statistinę informaciją, kuri realiu laiku buvo skaičiuojama, modeliuojant HTTP klientų darbą. Ši statistika vidurkinama ir skaičiuojama 2 valandų laikotarpiui.

Kadangi web technologija yra pakankamai elastinga ir nėra labai jautri tinklo RTT vėlinimui, tai galima sakyti, kad paslaugų teikėjui yra lengviau užtikrinti tokias SLA vėlinimo parametrų reikšmes, kurios remiasi konkrečiai WRT metrika. Taip yra todėl, kad realiaame tinkle optimalios parametrų slenkstinės reikšmės pastoviai kinta. Tuo pačiu, į web paslaugą orientuota metrika ne tik sprendžia šias problemas, bet ir suprantamiau parodo konkrečios paslaugos kokybę duomenų perdavimo tinkle.

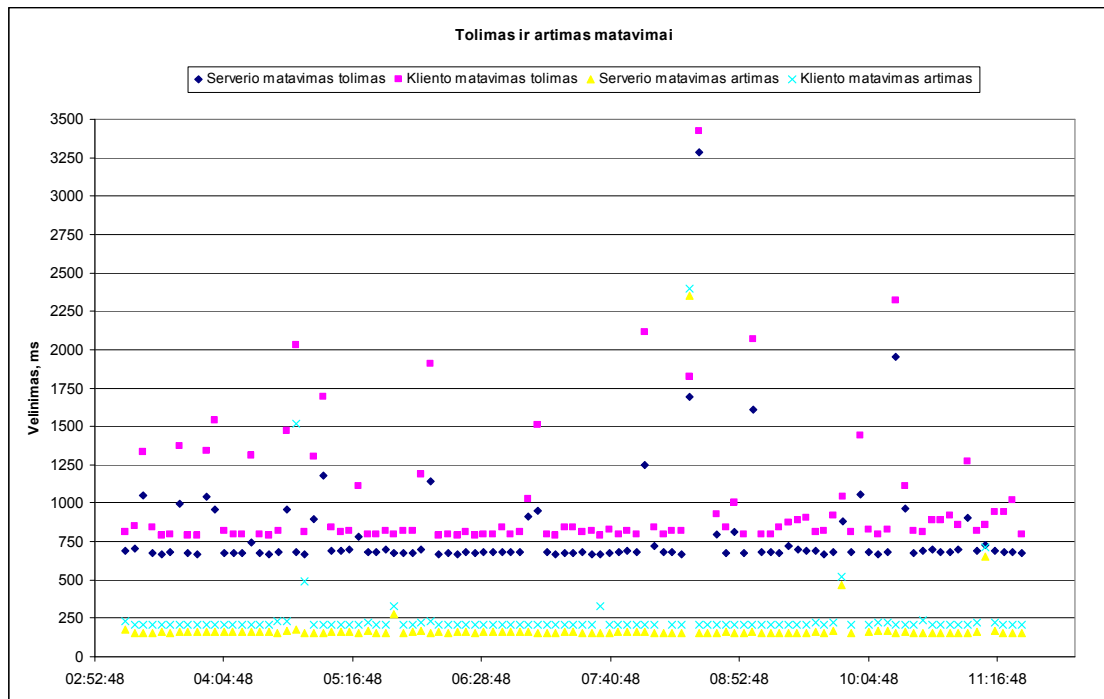
3 WEB paslaugos kokybės vertinimas iš serverio ir kliento perspektyvų

Atlikti KTU ITPI LITNET valdymo centre esančio WWW serverio dviragis.ktu.lt paslaugos vėlinimo matavimai. Serveryje veikia www.itpi.ktu.lt interneto svetainė, o taip pat visa eilė kitų interneto svetainių. Interneto www.itpi.ktu.lt svetainė periodiškai apklausinėjama iš dviejų skirtingai nutolusių klientų kompiuterių – buvo vykdomos internetinio puslapio parsisiuntimo užklauskos (HTML puslapio, bei reikiamų grafinių failų siuntimas) – kurių vykdymo metu buvo matuojamas paslaugos vykdymo laikas iš dviejų skirtingų perspektyvų (serverio ir kliento). Pasirinkti du skirtingai nutolę klientai, norint įvertinti duomenų perdavimo tinklo įtaką paslaugos kokybei. Arti esantis klientinis kompiuteris (193.219.184.93) veikė tame pačiame vietiniame tinkle (LAN), o nutolęs klientinis kompiuteris (85.206.232.6) – kito interneto paslaugų teikėjo tinkle.

Serveryje (LINUX OS) matavimai buvo atliekami abiejų skirtingų klientų užklauskoms. Buvo matuojama, naudojantis paketų sekimo/fiksavimo programine įranga „tcpdump“ [3], kuri suteikia galimybę stebėti paketus juos patogiai filtruojant, pateikiama informacija apie transakcijų trukmę (kada paketas gautas ar išsiųstas, koks yra laiko skirtumas tarp gretimų paketų ir t.t.).

Klientų pusėje (LINUX OS) automatizuotai tekstinio failų siuntimo įrankio „wget“ pagalba buvo atsisiunčiamas visas interneto svetainės puslapis (t.y. buvo vykdomas rekursyvus 1 lygio failų parsisiuntimas) ir matuojamas šios operacijos įvykdymo laikas. Matavimai kartoti kas 5 minutes. Programos „tcpdump“ pagalba gauta informacija saugota atskiruose failuose. Serveris tuo pačiu momentu užklašinėtas iš abiejų klientinių kompiuterių.

Matavimo rezultatai pateikti 2 pav. Iš nutolusio kliento užklauskos atsakymo vėlinimai visada didesni nei iš artimojo. Tačiau serverio ir kliento pusėse gauti vėlinimai koreliuoti.



2 pav. Puslapio užkrovimo vėlinimas iš serverio ir kliento perspektyvų (artimas ir tolimas matavimai kartu)

HTTP protokolas veikia viršutiniame OSI modelio lygyje. Jo veikimui įtakos turi visa eilė žemesniuose lygiuose esančių protokolų: TCP (4 OSI lygis), IP (3 OSI lygis), bei žemesniuose lygiuose veikiantys fiziniai protokolai. Tai kompiuterių tinklo įtaka web paslaugos vėlinimui. Kadangi serveris turi gaišti laiką užklauskos apdorojimui (surasdamas ir paimdamas reikiamą failą), tai bendram paslaugos vėlinimui didelę reikšmę turi ir serverio aparatinė bei programinė įranga, bei turimų resursų kiekis kiekvienu laiko momentu, nes naudotos sudėtingos HTTP užklauskos, sudarytos iš daugelio paketų.

Matavimai buvo vykdomi apie 12 val. ir buvo pastebėti matavimo rezultatų nukrypimai nuo vidurkio. Manoma, kad tai įvyko dėl serverio laisvų resursų sumažėjimo.

Iš šių matavimų galime suformuluoti išvadą, kad ne visada vien duomenų perdavimo tinklo resursai įtakoja paslaugos kokybinius parametrus, tačiau nemažą įtaką turi ir serverio turimi resursai, t.y. jo sugebėjimai aptarnauti klientines užklauskas konkrečiu laiko momentu.

4 Interneto naršyklės atminties (cache) įtaka WEB paslaugos kokybei

WWW puslapiai paprastai yra sudėtingos kompleksinės struktūros dokumentai, susidedantys iš HTML kalba rašytų tekstinių duomenų, bei grafinės, video ar garso informacijos failų. Norint vartotojui juos atvaizduoti, visa informacija turi būti parsisiunčiama į kliento kompiuterį. Kiekviena grafinė naršyklė šių puslapių saugojimui „paskiria“ tam tikrą kiekį vietos kompiuterio kietajame diske, kuri vadinama laikinąja atmintimi (angl. cache). Dažniausiai interneto puslapiai nekinta labai dažnai, todėl pakartotinis WWW puslapio parsisiuntimas turėtų užtrukti trumpiau – nes ne visa informacija bus siunčiama duomenų perdavimo tinklu, o nepakitusi informacija gali būti paimama tiesiog iš kliento kompiuterio laikinosios atminties (cache). Eksperimento tikslas ištirti interneto naršyklės laikinosios atminties (cache) poveikį bendrai (vartotojui realiai pastebimai) tam tikros WWW paslaugos kokybei.

Tyrimas atliktas trejose skirtingose tinklų architektūrose: KTU ITPI LITNET valdymo centre esančiame vietiniame tinkle – nuo klientinio kompiuterio iki WWW serverio dviragis.ktu.lt, kuriame veikia visa eilė interneto svetainių (www.litnet.lt, tinklas.ktu.lt ir pan.), iš namų kompiuterio (prieiga prie interneto per ADSL liniją) iki populiaraus WWW serverio www.delfi.lt, bei prieiga prie interneto per testinį 64kbps greičio prisijungimą iki www.delfi.lt ir www.cnn.com svetainių.

Klientiniame kompiuteryje panaudota LINUX operacinė sistema su KDE grafine sąsaja ir Mozilla Firefox interneto naršykle. Matavimų automatizavimui sudaryti komandų scenarijai (angl. *scripts*), kurių pagalba minėtos svetainės buvo periodiškai atsisiunčiamos. Laikas stebėtas naudojantis paketų sekimo/fiksavimo programine įranga „tcpdump“ [3]. Buvo atliekama po 3, 5 arba 10 matavimų atitinkamai kas 2, 1 ir 0,5 minutės kartojant kas kelias valandas. Programos „tcpdump“ pagalba gaunama informacija saugota atskiruose failuose. Tyrimo metu nustatyta kaip priklauso WWW puslapio užkrovimo trukmė nuo užkrovimo periodiškumo.

Atlikus visą eilę matavimų su skirtingais laiko periodiškumo parametrais trejose duomenų perdavimo tinklo aplinkose, buvo gauti tokie galutiniai rezultatai:

1 lentelė. Interneto naršyklės atminties (cache) įtaka WEB paslaugos kokybei – apibendrinti rezultatai.

	Testas	Pirmi puslapių atvertimai (s)	Paskutiniai puslapių atvertimai (s)	Santykinis skirtumas (%)
Artimas matavimas	1	15,83	15,65	1,14
	2	15,67	14,41	8,04
Tolimas matavimas	3	18,35	15,89	13,41
	4	15,22	14,16	6,96
Tolimas matavimas – lėtas kanalas 64kbps	5	77,42	68,66	11,32
	6	144,99	125,39	13,51

Pirmi puslapių atvertimai – tai laikas, sugaišamas parsisiųsti WWW puslapį, kai šis siuntimas vykdomas periodiškai, praėjus tam tikram ilgesniam laikui po paskutinių puslapių atvertimų. Visa eilė matavimų buvo vykdoma po 12 – 24 valandas, ir visuose jų rezultatuose matėsi, jog po ilgesnės pertraukos puslapiai buvo parsisiunčiami su didesniu vėlinimu. Santykinis skirtumas pateiktas 1 lentelėje procentų pavidalu.

5 Išvados

Apie WEB paslaugos kokybę negalima spręsti vien tik pagal įprastas tinklo metrikas (vidutinis tinklo apkrovimas, vidutinis *ping* vėlinimas, paketų praradimo procentas ar pan). Kelių dimensijų vertinimas (pagrįstas aukštesnio lygio metrikomis – mūsų atveju WRT) įgalina gauti tikslesnius į vartotoją orientuotus paslaugų kokybės vertinimo rezultatus. Web paslaugos kokybė priklauso tiek nuo duomenų perdavimo tinklo parametrų, tiek ir nuo serverio bei kliento personalinio kompiuterio aparatinės ir programinės įrangos resursų. Vartotojo kompiuterio laikinosios atminties įtaka web paslaugų kokybei labiausiai pastebima esant mažos spartos duomenų perdavimo kanalui tarp WWW serverio ir kliento.

Literatūros sąrašas

- [1] **A. Schmietendorf, R. Dumke, D. Reitz**, SLA management - challenges in the context of Web-service-based infrastructures, *Proceedings of the IEEE International Conference on Web Services (ICWS'04)*
- [2] **J. Martin, A. Nilsson**, On Service Level Agreements for IP Networks, *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE.*
- [3] <http://www.tcpdump.org>

Web service quality evaluation

The traditional SLA (Service Level Agreement) metric (availability, latency, packet loss rate) guarantees offer little assurances to the end user of the particular network service. Concerning network availability – it does not assure the quality of the service, which can be unacceptable to the client. A web oriented performance metric Web Response Time (WRT) was introduced, that can either augment or replace traditional latency and loss metrics in the network environment, there web services are main and critical.

The web service quality evaluation experiments have been made from two perspectives – server and client – in order to evaluate the network and server utilization influence on web service quality.

Internet browser cache influence on web service quality has been evaluated by performing experiments on three different network architectures. The browser cache usage is more significant on slow data communication channels between the web server and the client.