KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
PROGRAMŲ INŽINERIJOS KATEDRA

Vytautas Krakauskas

Kompiuterių tinklo srautų anomalijų aptikimo metodai

Magistro darbas

Darbo vadovas
dr. R. Kavaliūnas

Kaunas, 2006

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
PROGRAMŲ INŽINERIJOS KATEDRA

Vytautas Krakauskas

Kompiuterių tinklo srautų anomalijų aptikimo metodai

Kalbos konsultantė
Lietuvių k. katedros lekt.
dr. J. Mikelionienė
2006-05-24

Recenzentas
doc. V. Rėklaitis
2006-05-29

Vadovas
dr. R. Kavaliūnas
2006-05-29

Atliko
IFM-0/2 gr. stud.
Vytautas Krakauskas
2006-05-29

Kaunas, 2006

SUMMARY

Detection of network traffic anomalies

This paper describes various network monitoring technologies and anomaly detection methods. NetFlow were chosen for anomaly detection system being developed. Anomalies are detected using a deviation value.

After evaluating quality of developed system, new enhancements were suggested and implemented. Flow data distribution was suggested, to achieve more precise NetFlow data representation, enabling a more precise network monitoring information usage for anomaly detection. Arithmetic average calculations were replaced with more flexible Exponential Weighted Moving Average algorithm. Deviation weight was introduced to reduce false alarms. Results from experiment with real life data showed that proposed changes increased precision of NetFlow based anomaly detection system.

TURINYS

1. ĮVADAS

Magistro studijų metu, buvo iškelta užduotis sukurti programinę įrangą, kuri aptiktų anomalijas kompiuterių tinkluose. Trijų semestrų metu, buvo suprojektuota ir sukurta anomalijų aptikimo sistema. Sistemoje, kompiuterių tinklo srautų stebėjimui buvo naudojama NetFlow technologiją. Anomalijų aptikimas vykdomas skaičiuojant momentinį nuokrypį nuo vidutinio srauto kiekio. Esminiai programinės įrangos aspektai apibūdinti 3 šio dokumento skyriuje „Projektinė dalis“.

Po sistemos įdiegimo, buvo pastebėti programinės įrangos trūkumai. Šio darbo tikslas – ištirti ir realizuoti programinės įrangos kokybės pagerinimo metodus. 4 dokumento skyriuje ištirtos pagrindinės trūkumų priežastys ir pasiūlyti trys sistemos tikslumo padidinimo metodai. Pirmiausia, išnagrinėtas srauto duomenų išskaidymas per visą srauto gyvavimo laikotarpį. Vėliau, vidutinio srauto kiekio skaičiavimui pasiūlytas eksponentinio svorinio slenkančio vidurkio metodas. Galiausiai, patarta santykio skaičiavimą papildyti nuo duomenų spartos priklausančiu svoriniu įverčiu. Šie patobulinimai bus patikrinti naudojant realius Kauno Technologijos Universiteto tinklo srautus.

2. ANALITINĖ DALIS

Vartotojai naudoja kompiuterių tinklus duomenims tarp sistemų perduoti. Duomenų perdavimas dėl įvairių priežasčių gali sutrikti. Nenumatytus nukrypimus nuo įprastų duomenų srautų vadinsime anomalija. Anomalijų trukmė ir poveikis priklauso nuo problemos priežasties. Barford [1] ir Miluocheva [2] anomalijų priežastys klasifikuoja į tris apibendrintas kategorijas:

- **Tinklo veikimo problemos** – anomalijos susijusios su fizinėmis ar programinėmis tinklo infrastruktūros problemomis, pvz.: sugedę įrenginiai, neteisinga konfigūracija, resursų trūkumas.
- **Antplūdis** – tinklo resursų išeikvojimas dėl nenumatytai didelio duomenų srauto, pavyzdžiui, labai daug vartotojų siunčiasi failus.
- **Piktnaudžiavimas** – problemos susijusios su piktavališka veikla, pavyzdžiui, atkirtimo nuo paslaugos (DOS) atakos, virusai, skenavimai.

Anomalijų tipai detaliau apibūdinti 2.1 skyriuje.

Šiame dokumente nebus nagrinėjami metodai, kaip pagal aptiktų anomalijų charakteristikas identifikuoti sutrikimų atsiradimo priežastis. Pagrindinis dėmesys skirtas anomalijų aptikimui. Kompiuterių tinklų srautų anomalijoms aptikti naudojami įvairūs metodai. Thottan ir Ji [3] juos klasifikuoja į tokias grupes:

- **Taisyklėmis pagrįsti metodai** – anomalijos aptinkamos ieškant iš anksto žinomas charakteristikas atitinkančių srautų. Tokie metodai nagrinėjami literatūros šaltiniuose [4], [5] ir [6].
- **Baigtiniais automatais pagrįsti metodai** – iš anksto turima informacija apie tinklo būsenas prieš ir įvykio metu. Anomalijos aptinkamos stebint tinklo būsenų kitimą ir ieškant atitikmenų. Baigtiniais automatais pagrįsti metodai leidžia ne tik aptikti anomaliją, bet ir identifikuoti problemą. Baigtinių automatų taikymas kompiuterių tinklų problemoms aptikimui nagrinėjama Katzella [7] ir Rouvelou [8].
- **Tinklo veikimo modeliais pagrįsti metodai** – normalaus darbo metu stebimos ir fiksuojamos įvairios tinklo charakteristikos. Taip sudaromas tinklo veikimo modelis. Jei stebimi srautai neatitinka sudaryto modelio – fiksuojama anomalija [9]. Tokio tipo sistemą nagrinėja Feather et al. [10].
- **Statistine analize pagrįsti metodai** – darbo metu statistiškai vertinamas tinklo srauto kitimas. Anomalija fiksuojama, kai statistinis įvertis viršija nustatytą ribą. Dalį tokių metodų palygina Soule et al. [11]. Populiariausi šio tipo anomalijų aptikimo metodai apibūdinti 2.3 skyriuje.

Kuo tikslesnis stebėjimo metodas, tuo tiksliau įmanoma aptikti anomalijas. Galimi du

stebėjimo metodai [12]: aktyvus ir pasyvus.

- **Aktyvus metodas** – tinklo būklė stebima siunčiant duomenis, pvz.: ping, netparf. Tokio tipo stebėjimai dažniausiai naudojami tinklo kokybės charakteristikų įvertinimui, tačiau rezultatų analizė gali būti panaudota anomalijų aptikimui.
- **Pasyvus metodas** – stebimi tinklu perduodami duomenys, priklausomai nuo sistemos, informacija apie srautus iš vis neperduodama, arba perduodama atskiru tinklu, pvz.: SNMP, NetFlow, LIBPCAP.

Tinklo stebėjimo metodai apžvelgiami 2.2 skyriuje.

2.1. ANOMALIJŲ TIPAI

Tinklo anomalijos dažniausiai būna susijusios su situacijomis, kai kompiuterių tinklo darbas nukrypsta nuo įprasto. Anomalijos gali kilti dėl įvairių priežasčių, pavyzdžiui, netvarkingi įrenginiai, tinklo perkrova, žalingos atkirtimo nuo paslaugos atakos ar įsibrovimai, kurie sutrukdo normalų paslaugų teikimą. Tokie įvykiai įtakos ir stebimų srautų charakteristikas. Poveikio stiprumas ir trukmė priklauso nuo anomalijos pobūdžio [3]. Lakhina et al. [13] anomalijas pagal jų charakteristikas ir kilmę suskirstė į aštuonias grupes. Anomalijų grupės ir trumpas jų apibūdinimas pateiktas 1 lentelėje.

1 lentelė. Anomalijų tipai

Anomalija	Apibrėžimas	Charakteristikos
DOS	Atkirtimo nuo paslaugos ataka prieš konkretų adresą.	Į vieną adresą nukreiptas staigus paketų ir/arba srautų kiekio padidėjimas. Šaltinio adresai bendrų savybių neturi. Dažniausiai trunka mažiau nei 20 minučių.
Antplūdis	Neįprastai intensyvus serverio ar paslaugos naudojimas.	Staigiai padidėja srautų ir paketų kiekis į konkretų adresą su vyraujančiu pastoviu paslaugos prievadu. Bendro šaltinio nėra. Dažniausiai trunka neilgai.
Skenavimas	Skenuojami kompiuterio arba viso tinklo prievadai.	Staigiai padidėja, iš vieno šaltinio siunčiamų srautų ir paketų kiekis. Tikslo adresai ir portai ne visuomet turi bendrų savybių. Trunka neilgai, dažniausiai iki 10 minučių.
Kirminai	Save platinančios programos, kurios tinkle plinta išnaudodamos neapsaugotų sistemų spragas.	Staigiai padidėja srautų kiekis. Bendrų adresų nėra, tačiau vyrauja vienas ar keli standartiniai prievadai.
Vienas-vienas	Neįprastai didelis duomenų srautas tarp dviejų taškų.	Staigus duomenų ir/arba paketų kiekio padidėjimas. Vyrauja srautai tarp dviejų adresų. Šuolis dažniausiai trumpalaikis, ne ilgesnis nei 10 minučių.

Vienas-daug	Turinio platinimas iš vienos tarnybinės stoties daugeliui vartotojų.	Staigus duomenų ir/arba paketų kiekio padidėjimas. Vyrauja srautai iš fiksuoto šaltinio ir prievado. Prievadas dažniausiai būna standartinis ir gerai žinomas.
Tinklo gedimas	Įvykiai, dėl kurių sumažėja perduodamų srautų kiekis.	Sumažėja duomenų, paketų ir srautų kiekis. Gali trukti ilgai.
Kelio pakeitimas	Pakeičiamas duomenų perdavimo kelias.	Skirtinguose kanaluose pasikeičia srautų kiekis. Pirminiame kanale jis sumažėja. Antriniame, prie kurio buvo pereita, staigiai padidėja. Veikia daug srautų, kurie neturi tarpusavyje bendrų savybių.

Kituose šaltiniuose P. Barford ir D. Plonka [1] bei I. Miluocheva ir E. Muller [2] anomalijas grupuoja į tris apibendrintas kategorijas:

- **Tinklo veikimo** – pagal Lakhina et al. [13], tai tinklo gedimo ir kelio pakeitimo anomalijos.
- **Antplūdis** – sutampa su 1 lentelės klasifikacija. Vienas-daug tipo anomalijos taip pat turi antplūdžio savybių.
- **Piktnaudžiavimas** – atitinka DOS, skenavimų ir kirminų anomalijų tipus.

2.1.1. TINKLO VEIKIMO PROBLEMOS

Anomalijos gali atsirasti dėl tokių veiksnių, kaip tinklo įrenginių gedimas, žymus srauto skirtumas dėl konfigūracijos pakeitimų ir pan.

Dėl įrenginio gedimo atsiradusioms srautų anomalijoms būdingas staigus pasikeitimas užfiksuotame perduotų baitų, paketų ir srautų kiekyje per laiko vienetą. Sugedus įrenginiui, tikėtinas staugus perduotų duomenų kiekio sumažėjimas. Įrenginį sutaisius, dėl sąlyginai negreito nutrūkusių sesijų atstatymo, perduotų duomenų kiekis ims didėti, bet ne taip staigiai, kaip sumažėjo po gedimo, todėl anomalijos pabaiga gali būti užfiksuota netiksliai. Padidėjimas priklausys nuo gedimo trukmės. Kuo trumpiau įrenginys buvo sugedęs, tuo mažiau sesijų nutrūks ir tuo staigesnis bus duomenų srauto atsistatymas.

Jei anomalija atsirado dėl konfigūracijos pakeitimo, poveikis gali būti labai įvairus. Pakeitus ribojimus ar dali srautų nukreipus kitu duomenų kanalu, bus matomas staigus srautų sumažėjimas arba padidėjimas. Tokio tipo anomalijos pasikeitimų trukmė gali būti labai ilga, todėl bus užfiksuota tik pasikeitimo pradžia.

A. Markopoulou et al. [14] savo darbe išsamiai nagrinėja tinklo anomalijų priežastis. Jų tyrimų rezultatai parodė, kad 20% anomalijų gali būti dėl suplanuotų tinklo tvarkymo darbų. 30% iš neplanuotų gedimų atsiranda dėl maršrutizatorių ar perdavimo terpės gedimų. Svarbu atskirti planuotą tinklo priežiūrą nuo neplanuotų gedimų ir įvertinti, kiek ryšio kanalų buvo

paveikta. Jei buvo imtasi priemonių, pasikeitimai dėl tinklo priežiūros darbų bus ne tokie stiprus, kaip neplanuoto gedimo atveju. Jei gedimas įtakoja daug ryšio kanalų, poveikis bus dar stipresnis.

Anomalijos taip pat gali atsirasti dėl neteisingų duomenų apie srautus. Dažniausiai matomas perdavimo duomenų sumažėjimas, kartais visiškas nutrūkimas. Jei dėl konfigūracijos pakeitimo stebima tik dalis srautų ar neteisingas ryšio kanalas, anomalijos priežastį nustatyti gali būti sudėtinga.

2.1.2. ANTPLŪDIS

Antplūdis – tai neįprastai didelis duomenų srautų kiekis atsiradęs dėl to, kad daug vartotojų staiga bando pasiekti tą patį resursą. Dažniausiai tai kreipiniai į tinklapių ar failų serverius. Jis gali atsirasti dėl populiariame tinklapyje paskelbtos nuorodos arba išleidus naują programinės įrangos versiją. Jei nėra pasiruošta, dėl antplūdžio gali būti išnaudoti visi serverio resursai ir jis nebespės aptarnauti visų vartotojų užklausų. Jei duomenų srautas viršija duomenų perdavimo kanalo ribas, gali būti nepasiekiamas ne tik vienas serveris, bet visas tinklas.

Antplūdžio prigimtis nėra piktybinė, tačiau tinklo srautų charakteristikos sutampa su DOS tipo atakos charakteristikomis. Praktiškai šias dvi anomalijas atskirti yra labai sudėtinga. Antplūdžio ir DOS tipo anomalijų charakteristikas detaliai analizuoja J. Jung et al. [15].

2.1.3. PIKTNAUDŽIAVIMAS

Piktnaudžiavimai – tai DOS atakos, kompiuterių ar tinklų skenavimai, virusai ir kita piktybinė veikla.

Pagal CERT [16], DOS ataka – tai bandymas sutrukdyti vartotojams naudotis sistemos teikiamomis paslaugomis. Streinlein et al. [17] DOS tipo atakas skirsto į du tipus: logines ir srautines. Loginės DOS atakos pasinaudoja pažeidžiamumais sistemose, sutrikdo jų darbą ir padaro nepasiekiamas vartotojui. Srautinės DOS atakos stengiasi sistemas užgožti labai dideliu srauto kiekiu ir taip išeikvoti visą procesoriaus darbo laiką, vietą laikmenose ar tinklo pralaidumą. Sėkmingos atakos atveju, dėl resursų trūkumo, sistema nebegali aptarnauti vartotojų. Kartais, DOS atakos neturi didelės įtakos bendram tinklui, tačiau stipriai paveikia pavienius vartotojus. Kitais atvejais intensyviai naudodamos resursus, anomalijos gali sukelti tinklo perpildymą, taip paveikdamos daugelio vartotojų galimybę naudotis tinklo resursais [18]. Loginės atakos įgyvendinimui nereikalingi dideli srautai, žymiai svarbesnis perduodamos informacijos turinys. Tokio tipo atakos lengviausiai aptinkamos naudojant taisyklėmis pagrįstus aptikimo metodus. Srautinės DOS tikslas – išnaudoti resursus, todėl tokių atakų atveju, tinkle pastebimas labai didelis duomenų, paketų ar srautų kiekio padidėjimas. DOS atakos yra labai aktuali tema. Be Streinlein et al. [17], tokio tipo anomalijų aptikimus tiria Feinstein et al. [19],

Cheng et al. [20], Hussain et al. [21] ir kiti.

Skenuodamas kompiuterius ar tinklus, pažeidėjas įvairiais metodais jungiasi prie vieno kompiuterio ar kelių kompiuterių prievadų ir taip bando nustatyti, kokios paslaugos jam pasiekiamos. Jei nėra stengiamasi išvengti aptikimo sistemų, skenavimo metu siunčiama labai daug mažo dydžio paketų. Žymaus pokyčio perduodamos informacijos kiekyje nebus, tačiau bus matomas staigus paketu ir naujų srautų padidėjimas per laiko vienetą.

Kirminai – tai tokia virusų rūšis, kurios plitimui nereikalingas apsikeitimas užkrėstais failais. Kirminas bando plisti pats pats susirasdamas ir išnaudodamas kitų sistemų spragas. Poveikis kompiuterių tinklui priklausys nuo pažeidžiamumo tipo ir viruso agresyvumo. Pavyzdžiui, „MS Slammer“ viruso plitimas buvo labai staigus, ir vienas iš jo požymiu – labai didelis srautų, siunčiamų į 1433 prievadą, kiekis, sutrikdantis normalų tinklo darbą.

2.2. TINKLO STEBĖJIMO METODAI

Tinklo stebėjimas – tai gebėjimas kaupti ir analizuot duomenis apie tinklo srautus. Egzistuoja du pagrindiniai stebėjimo principai: aktyvus ir pasyvus.

Aktyvūs metodai dažniausiai naudojami kokybinių tinklo charakteristikų nustatymui, pvz.: vėlinimui ir pralaidumui. Aktyviu metodu gautų rezultatų užtenka, kad būtų galima nustatyti anomalijas. Tačiau matavimo metu siunčiami šalutiniai duomenys patys trikdo normalų tinklo darbą ir gali būti anomalijų priežastimi. Šio metodo privalumas yra tas, kad nereikalinga specializuota techninė įranga, o programinė įranga yra nesudėtinga. Dažniausiai naudojama testavimo programinė įranga pateikta 2 lentelėje.

2 lentelė. Aktyvaus testavimo įrankiai.

Programa	**Paskirtis**	**Veikimo principas**
ping	vėlinimo iki galinio įrenginio nustatymas	Nustatoma kiek laiko praėjo nuo ICMP užklausos išsiuntimo iki atsakymo gavimo. Jei tinkle yra problemų, vėlinimas bus neįprastai didelis. Įvertinus vidutinę vėlinimo trukmę, galima nustatyti ribą, kurią viršijus, butų registruojama anomalija.
traceroute	vėlinimas iki tarpinių maršrutizatorių	Siunčiami paketai su nuolat didinama TTL reikšme. Nustatoma per kiek laiko grąžinamas klaidos pranešimas. Naudojant šį metodą, galima nustatyti ne tik kada, bet ir ties kuriuo tinklo mazgu įvyko tinklo anomalija.
netperf	tinklo pralaidumo matavimai	Iš vienos tarnybinės stoties į kitą, siunčiamas labai didelis informacijos kiekis. Matuojama kiek informacijos pavyko perduoti per laiko vienetą. Nuolatinis metodo naudojimas anomalijų aptikimui gali sutrikdyti tinklo darbą. Metodą geriausia naudoti, kai reikia patikrinti kokybines

Programa	Paskirtis	Veikimo principas
		tinklo charakteristikas, kai įtariamas gedimas.

Pasyvūs metodai tinklo srautų stebėjimui naudoja specializuotą techninę įrangą. Tokią funkciją gali atlikti srautus persiunčiantis maršrutizatorius arba prie tinklo prijungtas pasiklausymo įrenginys, perduodamus srautus nukreipiantis į stebėjimo stotį. Lyginant su aktyviu, pasyvus metodas turi tokius privalumus:

- Tinklo srautų stebėjimo metu nėra siunčiama šalutinė informacija, kuri sutrikdytų normalų stebimo tinklo darbą. Į nutolusias stotis tarnybinę informaciją galima perduoti kitais tinklais. Landfeldt et al. [22] pademonstruoja, kad pasyvus stebėjimas išties turi mažesnį šalutinį poveikį.
- Apie srautus gaunama išsami informacija tiek apie tinklo protokolo parametrus, tiek perduodamos informacijos turinį. Tokius rezultatus galima efektyviau panaudoti detaliam įvairių anomalijų aptikimui.

Populiariausi pasyvaus stebėjimo įrankiai: LIBPCAP, SNMP, RMON, NetFlow.

2.2.1. LIBPCAP

LIBPCAP [23] - tai populiari paketų stebėjimui naudojama biblioteka. Ji suteikia aukšto lygio sąsają su paketų stebėjimo mechanizmais ir leidžia pasiekti visus paketus priimamus iš tinklo, net ir tuos, kurie neskirti stebinčiam įrenginiui. LIBPCAP biblioteką naudoja tcpdump, ethereal, snort ir kitos populiarios tinklo srautų analizės programos. Tai tiksliausias iš esamų metodų, nes gaunama pilna informacija apie paketus: tinklo adresai, IP protokolo adresai, prievadai, informacijos turinys ir kt. Naudojant tokią informaciją galima aptikti sutrikimus ne tik pagal duomenų ar paketų kiekius, tačiau ir pagal jų turinį. Tačiau didelis tikslumas turi ir keletą neigiamų aspektų:

- Dideliuose tinkluose, kuriais perduodami labai dideli informacijos srautai, pilnas informacijos stebėjimas reikalauja labai daug resursų. Deri [24] tyrimas parodė, kad prie didelių srautų, sistemos nespėja susidoroti su pilnu tinklo srautu, todėl dalis informacijos yra prarandama.
- Dažnai naudinga saugoti tinklo srautų istoriją ilgą laiko periodą. Pilnos srautų informacijos saugojimas nors ir būtų naudingas, tačiau dėl informacijos mastų praktiškai yra neįmanomas.

Grossglauser ir Rexford [25] siūlo tris šių problemų sprendimo galimybes:

- Sumažinti stebimos informacijos kiekį, vietoj viso paketo stebinti tik jo antraštes.
- Stebėti tik tam tikrus tinklus, naudojant filtrus pagal šaltinio ar paskirties adresus.

- Į stebėjimo stotį perduoti tik dalį paketų, pavyzdžiui, kas dešimtą.

2.2.2. SNMP/RMON

Tinklo valdymo protokolai teikia statistinę tinklo srautų informaciją. Šie protokolai palaiko kintamuosius, kurie atitinka įrenginio srautų skaitiklius. Tokius skaitiklius galima nuolatos stebėti ir pagal juos spręsti apie tinklo būseną. Vienas iš tokių tinklo valdymo protokolų - SNMP.

SNMP – tai standartas, kuris specifikuoja įrenginių valdymo informacijos bazę MIB ir prieigos prie bazėje kaupiamų duomenų protokolą [26]. MIB bazė – tai hierarchinė medžio struktūra, kurios lapai – tai įvairaus tipo kintamieji. Tinklų stebėjimui papildomai naudojama standartizuota RMON valdymo informacijos bazė, kurios tikslas – palengvinti nuotolinį tinklų stebėjimą [27]. Pagrindiniai RMON privalumai:

- galimybė stebėti tinklo srautus,
- suskaičiuoti statistinę informaciją ir ją išsaugoti gedimo atveju,
- nustatyti pavojaus signalo sąlygas ir veiksmus, kurių būtų imtasi jei sąlygos butų patenkintos, pavyzdžiui, perspėjimo išsiuntimas.
- nustatyti paketų filtravimo sąlygas ir veiksmus, pavyzdžiui paketo turinio išsaugojimas.

Toks RMON funkcionalumas yra sunkiai įgyvendinamas. Dėl šios priežasties gamintojai RMON pilnai įdiegia tik lėtesniuose įrenginiuose, o greitų įrenginių funkcionalumas dažniausiai būna stipriai apribotas [25].

2.2.3. NETFLOW

Cisco Systems [28] NetFlow srautus apibūdina taip:

NetFlow srautas – tai vienkrypčių iš siuntėjo gavėjui siunčiamų paketų aibė. Srautą identifikuoja septyni raktiniai paketo laukai: siuntėjo IP adresas, gavėjo IP adresas, siuntėjo prievado numeris, gavėjo prievado numeris, protokolo numeris, TOS reikšmė, priimančios sąsajos SNMP indeksas.

3 lentelė. NetFlows paketo struktūra

Laukas	Ilgis	Paaiškinimas
srcaddr	**4**	**Siuntėjo IP adresas**
dstaddr	**4**	**Gavėjo IP adresas**
nexthop	4	Tolimesnio maršrutizatoriaus IP adresas
input	**2**	**Priimančios sąsajos SNMP indeksas**

output	2	Išsiunčiančios sąsajos SNMP indeksas
dPkts	4	Paketų skaičius sraute
dOctets	4	Oktetų (baitų) skaičius sraute
first	4	Sistemos laikas srauto pradžioje
last	4	Sistemos laikas srauto pabaigoje
srcport	**2**	**Siuntėjo prievado numeris**
dstport	**2**	**Gavėjo prievado numeris, arba ICMP tipas ir kodas**
pad1	1	Nenaudojama, užpildoma nuliais
tcp_flags	1	TCP vėliavėlių suma panaudojant loginį „arba“.
prot	**1**	**Protokolo numeris**
tos	**1**	**IP paketo TOS reikšmė**
src_as	2	Siuntėjo autonominės sistemos numeris
dst_as	2	Gavėjo autonominės sistemos numeris
src_mask	1	Siuntėjo tinklo kaukės ilgis
dst_mask	1	Gavėjo tinklo kaukės ilgis
pad2	2	Nenaudojama, užpildoma nuliais

NetFlow srautas sukuriamas kai maršrutizatorius pamato pirmą paketą. Tie paketai, kurių raktiniai laukai sutaps, bus priskirti tam pačiam srautui. Jei nesutampa nors vienas raktinis laukas, paketas priskiriamas naujam srautui. Srautas užbaigiamas kai tenkinama viena iš sąlygu:

- TCP protokolo jungtis nutraukiama atsiunčiant paketą su RST arba FIN vėliavėle,
- paketai srautui nebuvo priskirti ilgiau nei 15 sekundžių,
- srauto trukmė viršija 30 minučių,
- jei persipildo maršrutizatoriaus srautų lenelė.

Kiekviename NetFlow sraute saugoma informacija apie paketus pateikta 3 lentelėje. Paprastos TCP sesijos metu bus sukurti du srautai: vienas, atitinkantis paketus siunčiamus iš kliento serveriui, kitas atitinkantis paketus siunčiamus iš serverio klientui. Kartais vieną TCP sesiją atitinka daugiau nei du NetFlow srautų įrašai. Tai atsitinka tokiais atvejais, kai sesija trunka ilgiau negu 30 minučių. Norit sužinoti visos sesijos paketų skaičių, reikės susumuoti informaciją iš skirtingų NetFlow srautų įrašų [29].

Ilgalaikiai NetFlow srautų įrašai taip pat įtakoja anomalijų aptikimo metodus. Analizuojant srautų informacija, būna aišku, tik kiek duomenų persiųsta per laiko intervalą, tačiau neaišku koks duomenų pasiskirstymas. Jei informacija nebus išskaidyta, ilgai trukusių srautų analizės rezultatai gali būti neteisingi.

2.3. ANOMALIJŲ APTIKIMO METODAI

Valdant tinklus, anomalijų aptikimas talkina įvairiais atvejais, pradedant DOS atakų aptikimu, baigiant tinklo konfigūracijos klaidomis. Kuo greičiau aptinkama anomalija, tuo greičiau galima pradėti spręsti problemą ir tuo mažesni nuostoliai bus patirti.

Išnagrinėjus įvairius, su anomalijų aptikimu susijusius tyrimus, buvo išskirti trys populiariausi metodai, anomalijoms kompiuterių tinkluose aptikti:

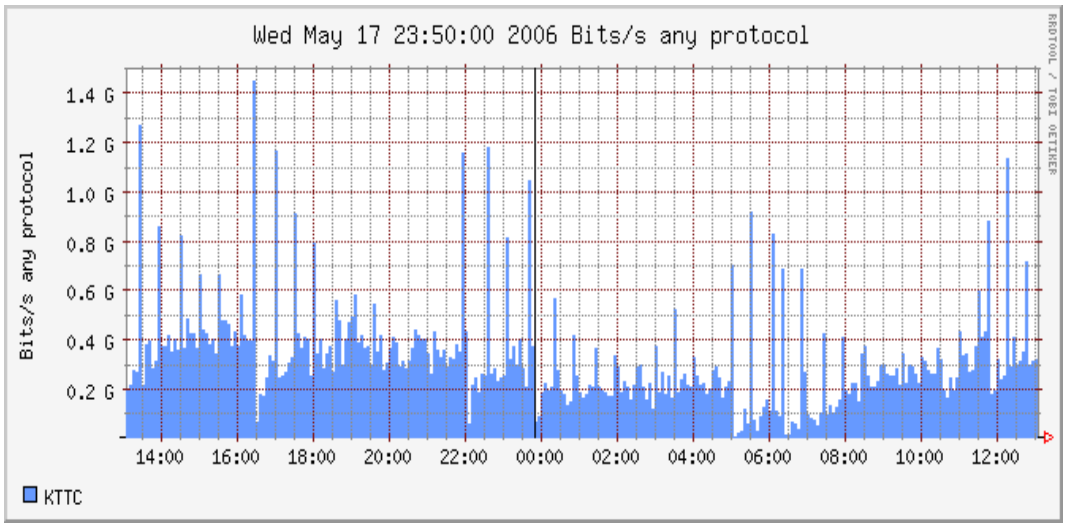
- **Bajeso metodas** – sistema išmokoma apie įvairias tinklo būsenas, vėliau nuolatos stebint tinklą, skaičiuojama tikimybė, kad esama būsena atitinka vieną iš anksčiau žinomų. Toks metodas reikalauja išankstinių žinių apie įvykusius incidentus, bei nuolatinio sistemos mokymo, atsiradus naujo tipo atakoms. Metodas geriausiai veikia tuomet, kai turima išsami informacija apie tinklo srautus. Tokios informacijos analizė dideliuose tinkluose gali pareikalauti didelių resursų. Ishiguro et al. [30] ir Hood [31] sukūrė veikiančias anomalijų aptikimo sistemas pagrįstas Bajeso metodu.
- **Bangelės metodas** – tai signalo apdorojimų metodas informaciją apie srautus išskaidantis į skirtingo dažnio dedamąsias. Metodas pagrįstas tuo, kad aukšto dažnio dedamojoje, geriau išryškėja trumpalaikės anomalijos. Atitinkamai vidutinio ir žemo dažnio dedamosiose gerai matomos vidutinės ir ilgos trukmės anomalijos. Tokiu metodu pagrįstas anomalijas nagrinėja Abry et al. [32] ir Barford et al. [33].
- **Nuokrypiu pagrįsti metodai** – tai įvairūs metodai, pagrįsti vidutinės reikšmės ir nuokrypio nuo jos skaičiavimu. Metodai tarpusavyje skiriasi vidutinės reikšmės skaičiavimo algoritmu. Vienas populiariausių algoritmų - tai eksponentinis svorinis slenkantis vidurkis [34]. Šis metodas atitinka žemo dažnio filtrą signalo apdorojimo sistemose. Jis panaikina aukšto dažnio triukšmus palikdamas tik bazinę žemo dažnio dedamąją. Holto metodas – tai populiarus eksponentinio svorinio slenkančio vidurkio variantas, papildomai įvertinantis srautų periodiškumą [35].

Visi metodai turi savų savybių, todėl negalima išskirti vieno. Šio dokumento 4 skyriuje „Tyrimas“, bus detaliau tiriamos eksponentinio svorinio slenkančio vidurkio metodo taikymas sukurtoje sistemoje.

2.4. SISTEMOS ANALOGAI

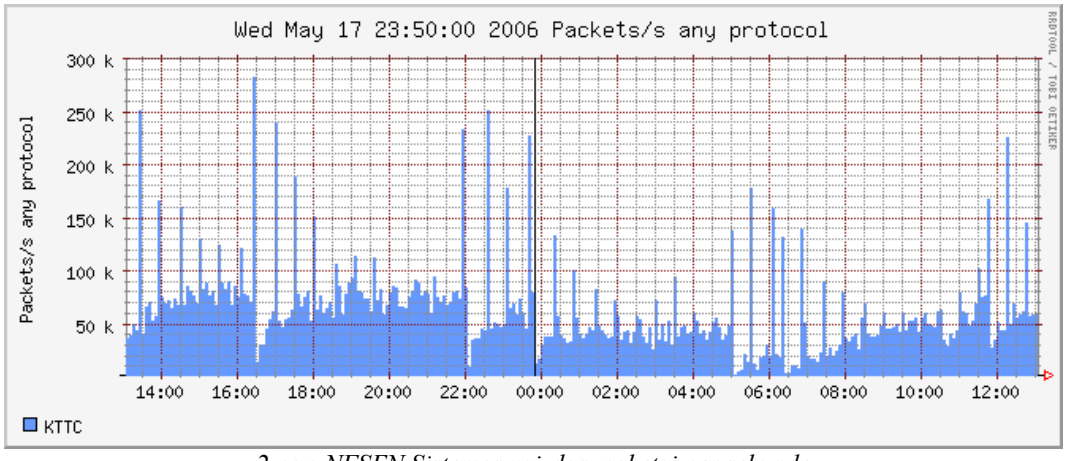
NFSEN (http://nfsen.sourceforge.net/) - nemokama Šveicarų akademinio tinklo kuriama sistema srautų informacijai peržiūrėti. Sistema naudoja Netflow technologiją srautams fiksuoti. Praktiškai išbandžius sistemą, pastebėta, kad dalis jos funkcijų veikia nekorektiškai. Sistema nuolatos registruojamus srautus vaizduoja 5 minučių intervaluose. Tačiau NetFlow įraše gali būti sukaupta informacija apie srautus, trukusius net iki 30 minučių. Gavus įrašą apie tokį

srautą, NFSEN sistema jį priskirs įprastam 5 minučių intervalui. Toliau pateiktuose 1, 2 ir 3 paveiksluose matyti, kad grafiškai vaizduojant duomenų, paketų ir srautų kiekį, atsiranda realybės neatitinkantys srautų kiekio šuoliai. Dideli šuoliai fiksuoja duomenų perdavimo greičius, viršijančius tinklo galimybes, todėl juos pastebėti lengva. Kur kas sunkiau aptinkami maži šuoliai, kurie susilieja su bendru srauto lygiu ir iškreipia realių srautų vaizdą.



1 pav. NFSEN Sistemos vaizdas, baitai per sekundę

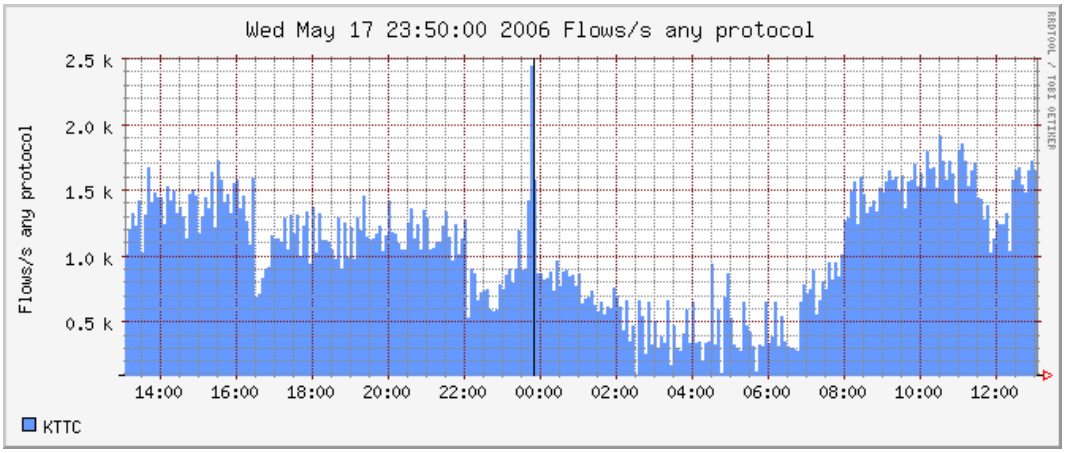
Realiai neįmanoma nustatyti užregistruotų srautų pasiskirstymo laiko tarpe, tačiau galima įvertinti srauto laiko trukmę ir jį paskirstyti pasirinktame intervale. 1 paveiksle matome, kad neišskaičius ilgų srautų, duomenų perdavimo greitis kartais siekia daugiau nei 1gigabitą per sekundę. Toks greitis viršija fizines tinklo įrangos galimybes.



2 pav. NFSEN Sistemos vaizdas, paketai per sekundę

Tokį pati efektą galima pastebėti ir perduotų paketų skaičiuje per sekundę, pavaizduotame 2 paveiksle. Nors sunku vertinti tinklo įrangos galimybes, tačiau akivaizdu, kad šuoliai sutampa

su greičio šuoliais 1 paveiksle.



3 pav. NFSEN Sistemos vaizdas, srautai per sekundę

Mažiausiai įtakos šuoliai turi srautų skaičiaus grafike, pavaizduotame 3 paveiksle. Šiame grafike išsiskiria šuolis maždaug 12 valandą nakties. Toks pats šuolis 1 ar 2 paveiksluose susilietų su kitais, ir nebūtų pastebėtas.

Galima daryti išvadą, kad NFSEN sistemos srautų skirstymo klaidos anomalijų aptikimą daro sudėtingu uždaviniu, nes pagal pateikiamus duomenis, sunku atskirti srautų pokyčius nuo duomenų vaizdavimo klaidų. Nepaisant neigiamų savybių, sistemą labai patogu naudoti detaliai srautų peržiūrai.

Be NFSEN egzistuoja ir kiti anomalijų aptikimo produktai, tokie kaip **Arbor Peakflow** (http://arbornetworks.com/), **Riverhead Guard** (http://www.riverhead.com/), **Cisco Network Analysis Module**, tačiau jie yra mokami ir tyrimo metu jų išbandyti nepavyko.

3. PROJEKTINĖ DALIS

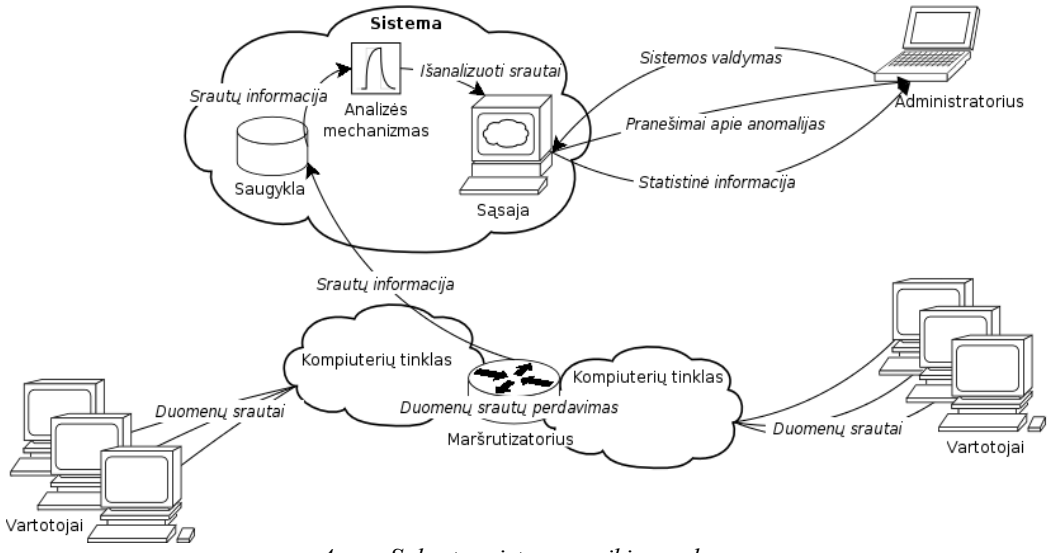
Sukurtos sistemos tikslas – palengvinti kompiuterių tinklų administratorių darbą identifikuojant nepageidaujamus srautus tinkluose. Tinklų anomalijų stebėjimo stoka apsunkina tinklo valdymą. Administratoriai dažniausiai pasiekia tik ribotą informaciją apie tinklo būseną. Atakų atveju, neretai sutrikdomas tinklo darbas, tačiau atakų aptikimas tinkluose, kur nuolatos perduodami dideli srautai, yra labai sudėtingas ir kartais neįmanomas, jei nenaudojamos specializuotos priemonės. Anomalijų aptikimo sistema, administratoriams pateikia informaciją apie neįprastą tinklo segmento veikimą, kurio priežastimi gali būti ataka, virusai, neteisinga tinklo konfigūracija ir kt.

Naudojant sistemą, administratoriai gali identifikuoti įvykusias problemas, todėl galima skirti daugiau laiko jų sprendimui ir taip sumažinti netinkamo tinklo naudojimo atvejų kiekį. Sistema suteikia galimybę identifikuoti tokias tinklo srautų anomalijas, kaip neįprastai didelis ar mažas perduodamų duomenų, paketų ar sesijų kiekis. Tai leidžią identifikuoti daugiausiai įtakos tinklo darbui turinčias problemas.

Šiame skyriuje bus apžvelgiami tik pagrindiniai patobulintos programinės įrangos techninės-projektinės dokumentacijos aspektai. Detali pradinės architektūros specifikacija, vartotojo dokumentacija ir kiti su programinės įranga susiję dokumentai pateikiami prieduose.

3.1. SISTEMOS STRUKTŪRA

Sukurtos sistemos veikimo schema pateikta 4 paveiksle.

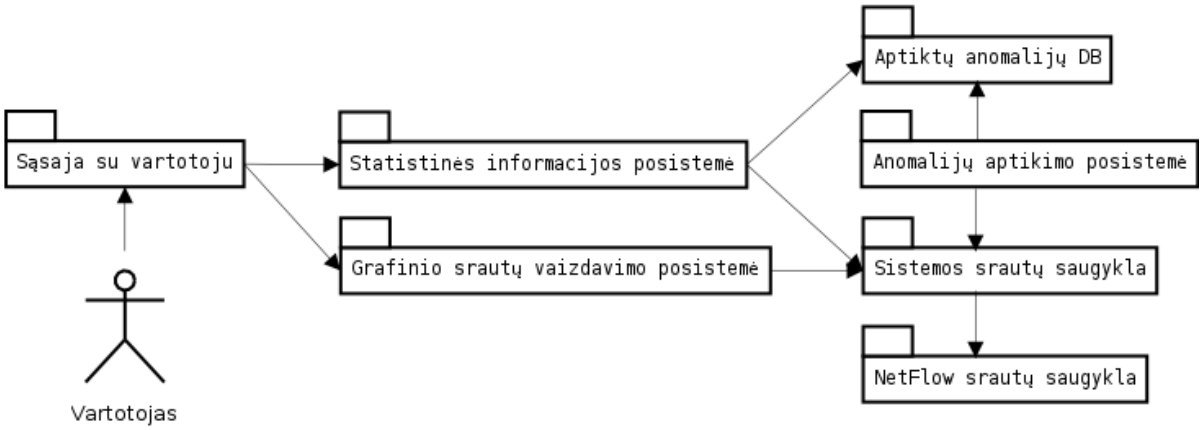


4 pav. Sukurtos sistemos veikimo schema

Vartotojų duomenų srautai, kuriuos maršrutizatorius persiunčia tarp skirtingų tinklų, yra registruojami. Užregistruotų srautų informacija yra siunčiama į sistemos saugyklą NetFlow formatu. NetFlow formatu saugomų duomenų yra labai daug, apdorojimas užtrunka ilgai, todėl

jų analizavimas realiu laiku reikalautų labai daug resursų ir praktiškai būtų sunkiai įgyvendinamas. Dėl šios priežasties, sistema laisvu metu reguliariai tikrina saugykla ir iš jos nuskaito apibendrintą informaciją apie srautus: baitų, paketų ir srautų skaičių. Tokios informacijos pakanka anomalijų aptikimo algoritmams. Reguliariais laiko intervalais sistema tikrina, ar naujai gautuose srautų įrašuose nėra didelių šuolių ar nukritimų. Tokią informaciją aptikus, anomalijos registruojamos atskiroje duomenų bazėje. Vėliau, tinklo administratorius naudodamas vartotojo sąsają gali peržiūrėti anomalijų istoriją, patyrinėti srautus ir jei reikia imtis problemos šalinimo.

Detali sistemos struktūra pavaizduota 5 pav.



5 pav. Sistemos komponentai

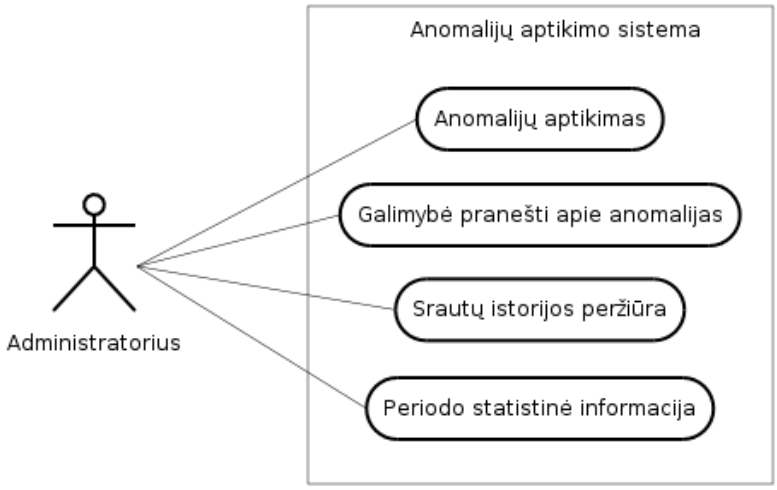
5 pav. pavaizduotų komponentų paaiškinimas:

- **Vartotojas** – žmogus, kuris naudojasi sistema
- **Sąsaja su vartotoju** – tinklapis, prie kurio vartotojas jungiasi su interneto naršyklę.
- **Statistinės informacijos posistemė** – pateikia statistinę informaciją apie pasirinktą laiko periodą ir aptiktų anomalijų sąrašą.
- **Grafinio srautų vaizdavimo posistemė** – grafiškai pavaizduoja pasirinkto tipo grafiką. Galimi tipai: baitai, paketai, aktyvus srautai, nauji srautai per sekundę.
- **Aptiktų anomalijų DB** – žurnalas, kuriame saugoma informacija apie aptiktas anomalijas.
- **Anomalijų aptikimo posistemė** – analizuoja pasirinkto periodo srautų informaciją ir aptiktas anomalijas fiksuoja anomalijų žurnale. Ši posistemė vykdoma reguliariais laiko intervalais, pavyzdžiui, kas 30 min.
- **Sistemos srautų saugykla** – procesas, kuris nuolatos tikrina, ar srautų saugykloje nėra naujų įrašų. Nauji įrašai yra perskaitomi ir atmintyje išsaugoma apibendrintą informaciją apie juos.

- **NetFlow srautų saugykla** – laikmena, kurioje NetFlow formatu saugoma iš maršrutizatoriaus gauta srautų informacija.

3.2. PANAUDOJIMO ATVEJAI

Sistemos panaudojimo atvejai pavaizduoti 6 pav.



6 pav. Panaudojimo atvejų diagrama

3.2.1. ANOMALIJŲ APTIKIMAS

Anomalijų aptikimas vyksta pagal tokią veiksmų seką:

1. Anomalijų aptikimo posistemė iškviečiama reguliariais laiko intervalais.
2. Iškviesta posistemė prisijungia prie sistemos srautų saugyklos ir nuskaito informaciją apie srautus pradedant nuo 12 valandų senumo srautų, baigiant pusvalandžio senumo srautais nuo iškvietimo momento.
3. Srautai filtruojami panaikinant aukšto dažnio svyravimus, pagal 4.2 skyriuje apibūdintą algoritmą.
4. Minutės tikslumu lyginamas santykis tarp žemo dažnio dedamosios ir srautų informacijos.
5. Santykis sumažinamas atsižvelgiant į spartą. pagal 4.3 skyriuje apibūdintą algoritmą.
6. Jei santykis neviršija pasirinktos ribos, pereinama prie kito laiko momento.
7. Jei santykis viršija nustatytą ribą, žymima anomalijos pradžia.
8. Jei santykis neviršija ribos, o ankstesniu laiko momentu buvo nustatyta anomalija, fiksuojama anomalijos pabaiga ir informacija apie ją įrašoma i duomenų bazę.
9. Jei skirtumas tarp paskutinės ir naujos anomalijos mažesnis nei 10 minučių, jos sujungiamos.

10. Baigus periodo analizę, informacija apie anomalijas įrašoma į žurnalą.

3.2.2. GALIMYBĖ GREITAI PRANEŠTI APIE ANOMALIJAS

Vartotojo sąsajoje teikiama galimybė išsiųsti pranešimą apie anomaliją tinklo administratoriams. Peržiūrint užfiksuotas anomalijas, galima pasirinkti apie kurias anomalijas pranešti administratoriui. Pasirinkus, automatiškai sugeneruojama žinutė su informacija su reikiama informacija. Žinutę galima papildyti savo komentaru ir nusiųsti vienam iš sąraše pateiktų administratorių.

3.2.3. SRAUTŲ ISTORIJOS PERŽIŪRA

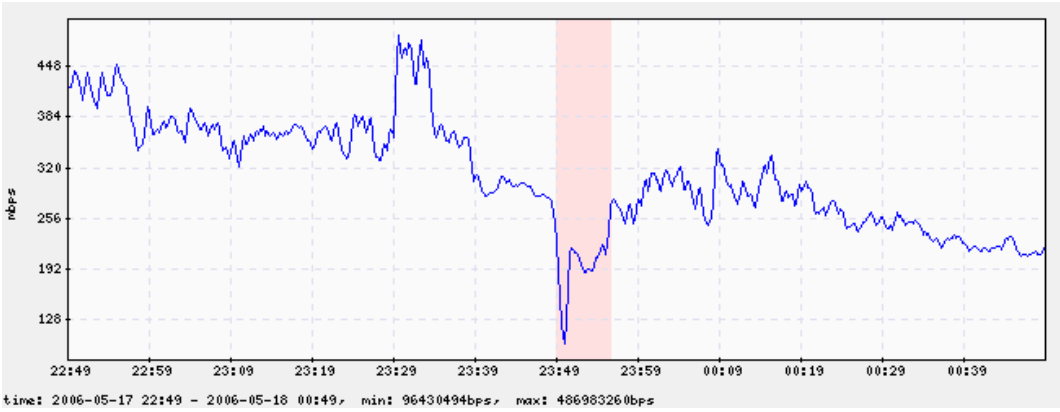
Vartotojui, sistema srautus pateikia grafiniame pavidale. Sistemos srautų saugyklos procesas iš anksto analizuoja tinklo srautų informaciją ir ją išsaugo atmintyje, todėl vartotojui paprašius, rezultatai pateikiami labai dideliu greičiu. Mažiausias periodas, kurio srautus galima peržiūrėti – 10min. Čia vienas taškas grafike atitinka vieną sekundę. Yra galimybė vaizduoti srautus, nufiltruojant aukšto dažnio svyravimus. Tokiu būdu pavaizduoti srautai aiškiau suprantami.

3.2.4. PASIRINKTO PERIODO STATISTINĖ INFORMACIJA

Pateikiamas mažiausias ir didžiausias baitų, paketų ir srautų kiekis per sekundę bei bendras perduotos informacijos kiekis per pasirinktą laiko periodą.

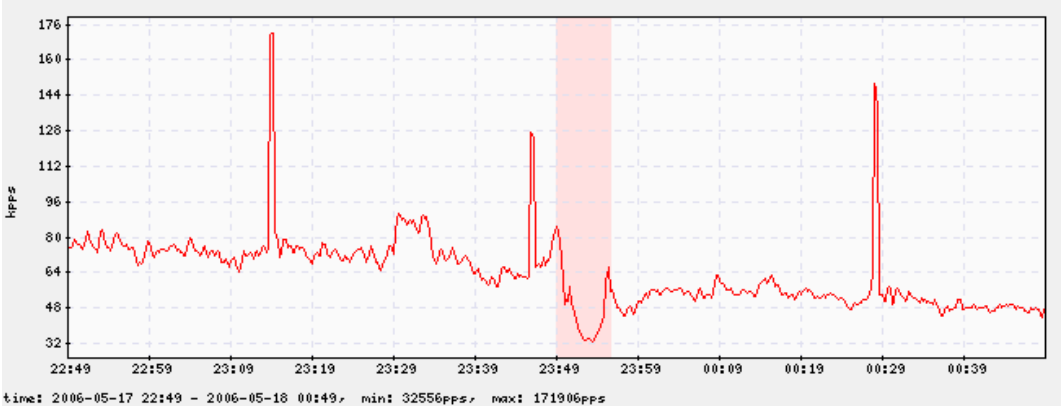
3.3. APTIKTOS ANOMALIJOS PAVYZDYS

Pavyzdyje pateiktas laiko periodas sutampa su 2.4 skyriuje naudotu periodu. 7, 8, 9 ir 10 pav. paryškintas laikotarpis, kurio metu sistema nustatė anomaliją.



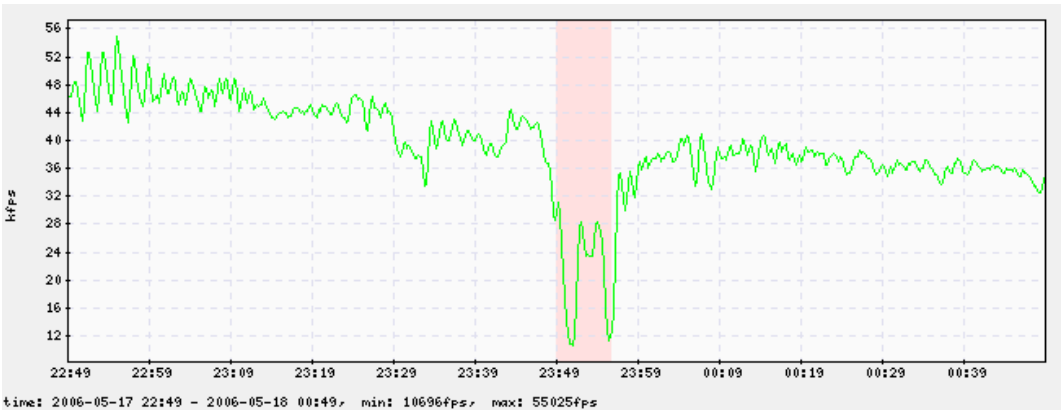
7 pav. Aptikta anomalija, baitai per sekundę

7 paveiksle matomas staigus duomenų srautų sumažėjimas. Skirtingai nei NFSEN sistemoje, ilgų duomenų srautų informacija čia yra išskaidoma per visą srauto laikotarpį, todėl nematyti didelių duomenų srautų šuolių.



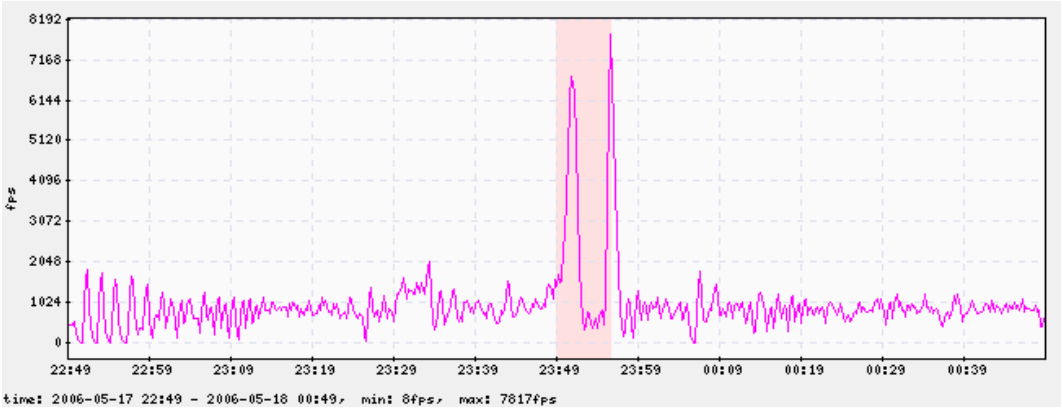
8 pav. Aptikta anomalija, paketai per sekundę

Paketų per sekundę grafike taip pat matomas sumažėjimas patenkantis į tos pačios anomalijos laikotarpį. Be minėtos anomalijos, matomi ir kitais laiko momentais įvykę paketų skaičiaus šuoliai. Kadangi informacija apie ilgų srautų paketus buvo išskaidyta pagal 4.1 pasiūlytą metodą, galima daryti išvadas, kad šie šuoliai išties įvyko duotais laiko momentais ir tai nėra duomenų vaizdavimo klaida. Juos nagrinėjant atskirai taip pat būtų galima daryti išvadas apie galimas anomalijas.



9 pav. Aptikta anomalija, aktyvūs srautai per sekundę

Aktyvių srautų grafike, situacija analogiška. NetFlow įrašuose aktyvių srautų skaičius duotu laiko momentu nėra pateikiamas. Šį parametrą sistema suskaičiuoja pagal naujų naujų srautų informaciją įvertinant srauto pradžios ir pabaigos laikus. Aktyvių srautų skaičius – tai papildoma informacija apie tinklo būseną, todėl ji gali būti naudinga tiriant tinklo anomalijų priežastis.



10 pav. Aptikta anomalija, naujai sukurti srautai per sekundę

10 paveiksle pateikiamame naujų srautų per sekundę grafike matomas padidėjimas, sutampantis su informacijos kiekio sumažėjimu 7, 8 ir 9 paveiksluose.

Sistema nedaro išvadų apie anomalijos tipą ar jos priežastis, tačiau iš patirties galima spręsti, kad tokie srautai būdingi tinklo gedimui. Gedimo atveju dauguma užmegztų sesijų nutraukta, todėl sumažėja perduodamų duomenų, paketų ir aktyviu srautų skaičius. Po nutraukimo vartotojai bando sesijas užmegzti iš naujo, dėl to atsirada staigus naujų srautų skaičiaus padidėjimas. 7, 8, 9 ir 10 pav. matomi du gretimi nuokrypiai, todėl galima spręsti, kad įvyko du gedimai. Kadangi nuokrypiai įvyko per sąlyginiai trumpą laiko tarpą, jie buvo užregistruoti kaip viena anomalija.

4. TYRIMAS

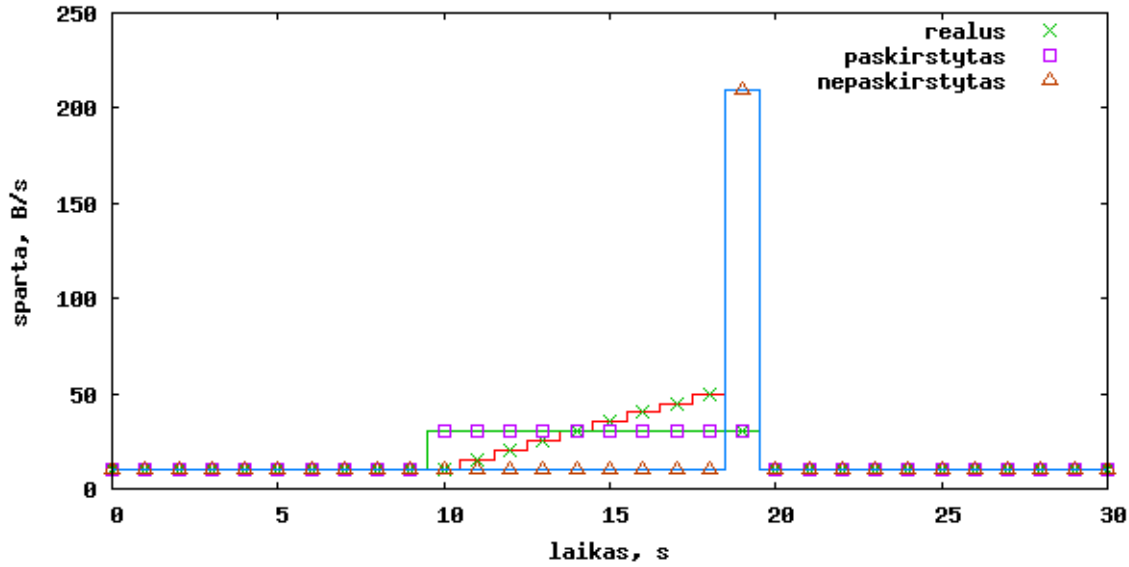
Sukūrus ir įdiegus programinę įrangą, kai kurie jos veikimo aspektai neatitiko lūkesčių.

- **Srautų paskirstymo problema** – vaizduojant srautus, matomi labai dažni duomenų srautų šuoliai, kartais viršijantys fizines duomenų kanalo galimybes. Šios problemos sprendimo metodas pateiktas 4.1 skyrelyje.
- **Netinkamos aritmetinio vidurkio charakteristikos** – esant dideliam šuoliui ar nukritimui, papildomai anomalijos užfiksuojamos prieš ar po pagrindinio nukrypimo. Problemos sprendimas pateiktas 4.2 skyrelyje.
- **Santykio įvertinimo problema** – mažo tinklo aktyvumo metu, sistema labai jautriai reaguoja į nedidelio masto pasikeitimus. Problemos sprendimas pateiktas 4.3 skyrelyje.

4.1. SRAUTŲ PASKIRSTYMO PROBLEMA

Vaizduojant srautus grafiškai matomi dažni šuoliai, kartais viršijantys fizinius duomenų kanalo pajėgumus. Problema ypač aktuali, peržiūrint trumpo laiko periodo srautus. Tokios problemos pavyzdys pateiktas 1, 2 ir 20 pav.

Problemos priežasčių pradėta ieškoti NetFlow srautų įrašuose. Buvo surasti NetFlow įrašai apie srautus su labai dideliu perduotų duomenų kiekiu. Pastebėta, kad tokie srautai pasižymi itin ilga laiko trukme, kartais siekiančia 30 minučių. Peržiūrėjus programinės įrangos srautų kaupimo algoritmą, buvo nustatyta, kad visi srautai rašomi į vieną laiko vienetą. Toks veikimo principas nėra teisingas, nes persiųsta srautų kiekis turėtų būti išskaidyta visame laiko intervale. Kaupiant srautus, buvo prarasta informacija, apie duomenų kiekio pasiskirstymą srauto gyvavimo laikotarpyje. Kadangi informacijos praradimo išvengti negalima, nuspręsta perduodamų srautų informaciją paskirstyti visame laiko intervale vienodai. Tai ne visiškai atitiks realių srautų, tačiau bus tiksliau, nei srauto priskyrimas vienam laiko vienetui. 11 pav. pavaizduoti skirtingi paskirstymo tipai.

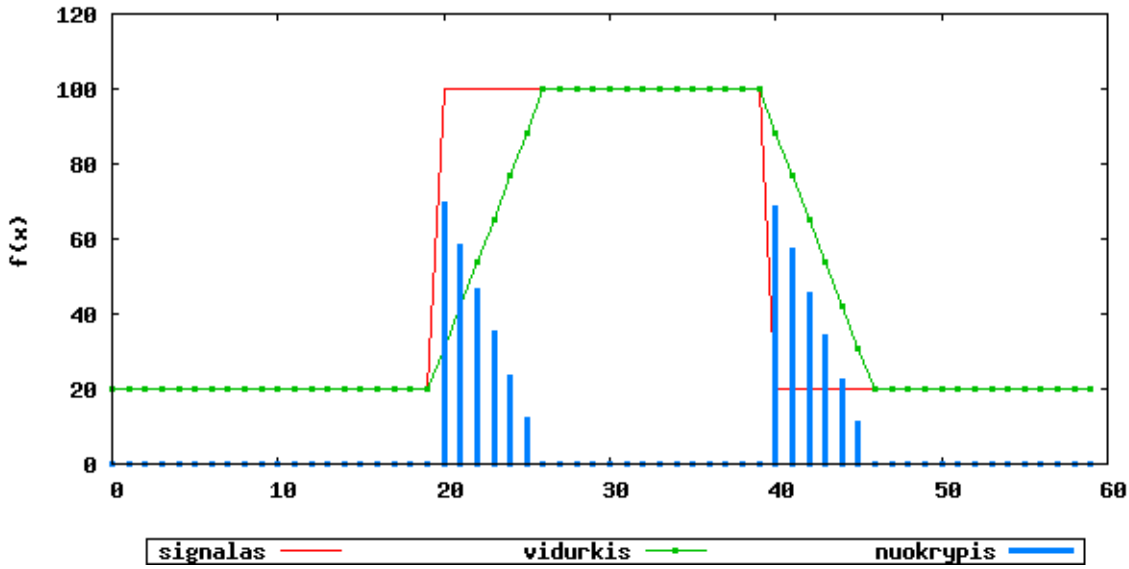


11 pav. Srautų paskirstymo metodų charakteristikos

Realaus paskirstymo atveju, pavaizduota, situacija, kai duomenų siuntimo sesija prasidėjo 10-uoju laiko momentu ir baigėsi 20-uoju. Realios sesijos metu siunčiant duomenis sparta palaipsniui didėja, kol pasiekiama maksimali sparta arba siuntimas baigiamas. Tokią situaciją atitinka kryžiukais pažymėta kreivė. Programinė įranga užfiksuotus srautus priskiria tuo laiko momentu, kada baigėsi susijungimas arba srautas viršijo maksimalų gyvavimo laikotarpį – 30 min. Persiustos informacijos kiekį priskyrus viename taške susidaro didelis šuolis, atitinkantis 11 pav. trikampiais pavaizduotą kreivę. Norint sušvelninti tokį efektą, srautus nuspręsta paskirstyti visame laiko intervale vienodai. 11 pav. toks paskirstymas pažymėtas kvadratukais. Lyginant su pirminiu variantu, nebėra didelio šuolio ir grafiškai vaizduojami srautai turėtų būti glotnesni. Toks paskirstymas tiksliausiai atitinka tas situacijas, kai maksimali siuntimo sparta pasiekiama sesijos pradžioje ir nekinta srauto gyvavimo periodo metu.

4.2. SVORINIS EKSPONENTINIS SLENKANTIS VIDURKIS

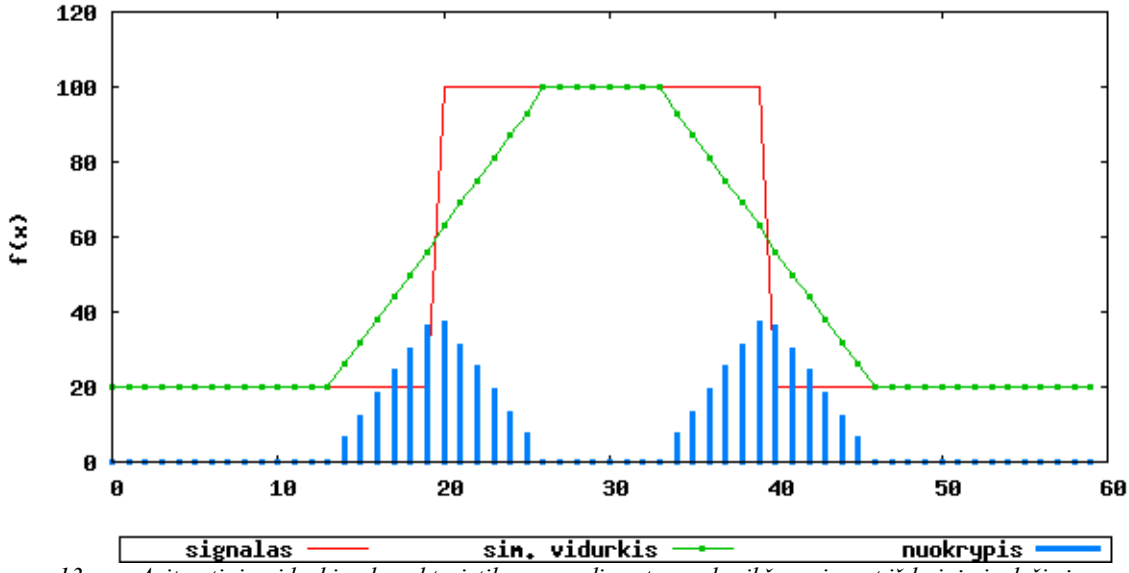
Didelio šuolio ar nukritimo, atveju anomalijos papildomai užfiksuojamos prieš ar po pagrindinio nukrypimo. Aptinkant anomalijas, skaičiuojamas santykis, tarp srauto suapvalintos reikšmės ir nesuapvalintos duotame laiko momente. Apvalinant reikšmę naudojamas paprastas aritmetinis vidurkis.



12 pav. Aritmetinio vidurkio charakteristikos, apvalinant pagal reikšmę iš kairės

Pagal aritmetikos vidurkio charakteristikas 12 pav. matyti, kad įvykus lūžiui, ties kritiniais taškais nuokrypis tarp signalo ir vidutinės reikšmės didžiausias ties lūžio taškais. Taip pat matoma, kad ties antru lūžio tašku, kur kreivė grįžta į pradinę poziciją, atsiranda labai didelis nuokrypis ir pagal sistemos algoritmą būtų fiksuojama anomalija. Tai iš dalies atitinka probleminės situacijos apibūdinimą.

Pabandykime problemą išspręsti panaudodami simetrinį vidurkį, t.y. apvaldinami naudojant reikšmes ir iš kairės ir iš dešinės.



13 pav. Aritmetinio vidurkio charakteristikos, apvalinant pagal reikšmes imant iš kairės ir dešinės

Pagal charakteristikas pateiktas 13 pav., matyti, kad nuokrypis sumažėjo ir išsiplėtė į abejas kreivės lūžio puses. Anomalijos, pagal tai, kiek reikšmių būtų įvertinama iš abiejų pusių,

šiuo atveju būtų fiksuojamos arba visame šuolio ruože ir už jo ribų, arba ties lūžio taškais.

Skaičiuojant aritmetinį vidurkį, visos reikšmės rezultatą įtakoja vienodai. Tačiau kartais naudinga, kad tolimesnės reikšmės turėtų ne tokią pačia įtaka, kaip artimesnės. Tokia savybe pasižymi eksponentinis svorinis slenkantis vidurkis (ESSV). ESSV skaičiavimo formulę išvesime iš aritmetinio vidurkio skaičiavimo formulės (1):

$$\bar{x}_k = \frac{1}{n} \cdot \sum_{i=k-n+1}^{k} x_i \tag{1}$$

čia *k* – k-tojo elemento reikšmė, *n* – elementų skaičius.

k+1 elemento reikšmę, galime apskaičiuoti pagal formulę (2):

$$\bar{x}_{k+1} = \frac{1}{n+1} \cdot \sum_{i=k-n+1}^{k+1} x_i = \frac{1}{n+1} \cdot \left(x_{k+1} + \sum_{i=k-n+1}^{k} x_i \right) \tag{2}$$

pasižymėkime, kad $\sum_{i=k-n+1}^{k} x_i = \bar{x}_k \cdot n$, tuomet (2) formulę galime užrašyti taip (3):

$$\bar{x}_{k+1} = \frac{1}{n+1} \cdot \left(x_{k+1} + \bar{x}_k \cdot n \right) = \left(\frac{1}{n+1} \right) \cdot x_{k+1} + \left(\frac{n}{n+1} \right) \cdot \bar{x}_k \tag{3}$$

Jei grįšime prie k-tojo elemento, (3) formulę pertvarkykime taip:

$$\bar{x}_k = \left(\frac{1}{n+1} \right) \cdot x_k + \left(\frac{n}{n+1} \right) \cdot \bar{x}_{k-1} \tag{4}$$

Pasižymėkime, kad $a = \left(\frac{n}{n+1} \right)$. a – proporcingumo koeficientas, $0 \leqslant a \leqslant 1$, nulemiantis, kokį svorį turės esama ir buvusi reikšmė. Įvedus a, (4) formulę galima užrašyti taip:

$$\bar{x}_k = a \cdot \bar{x}_{k-1} + (1 - a) \cdot x_k \tag{5}$$

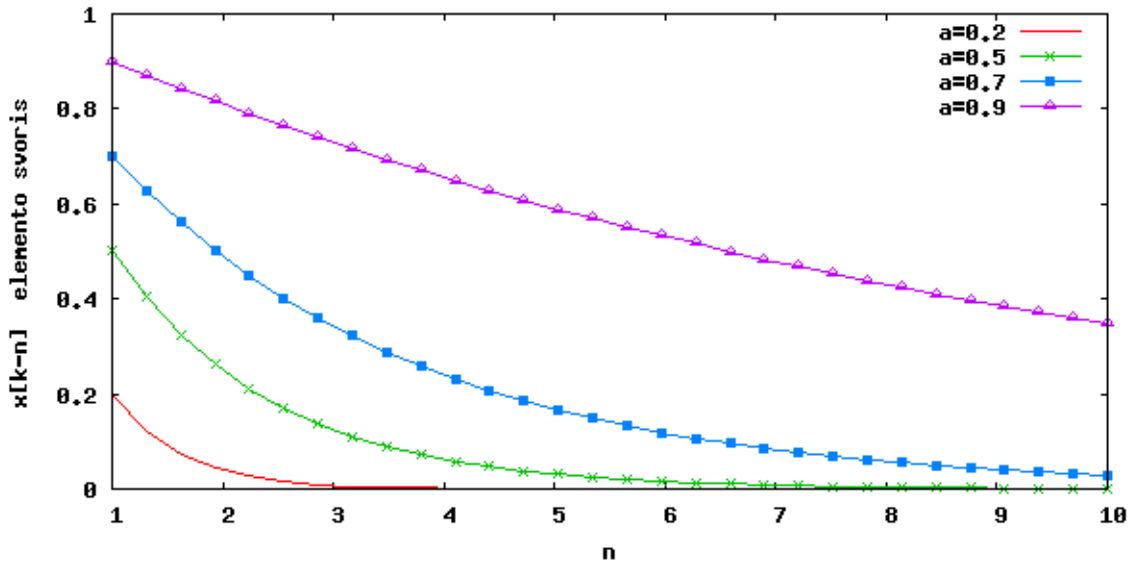
Tai eksponentinio svorinio slenkančio vidurkio apskaičiavimo formulė. Galime nesunkiai įsitikinti, kad (5) formulė iš tiesų skaičiuoja eksponentinį vidurkį:

$$\bar{x}_{k-1} = a \cdot \bar{x}_{k-2} + (1 - a) \cdot x_{k-1} \tag{6}$$

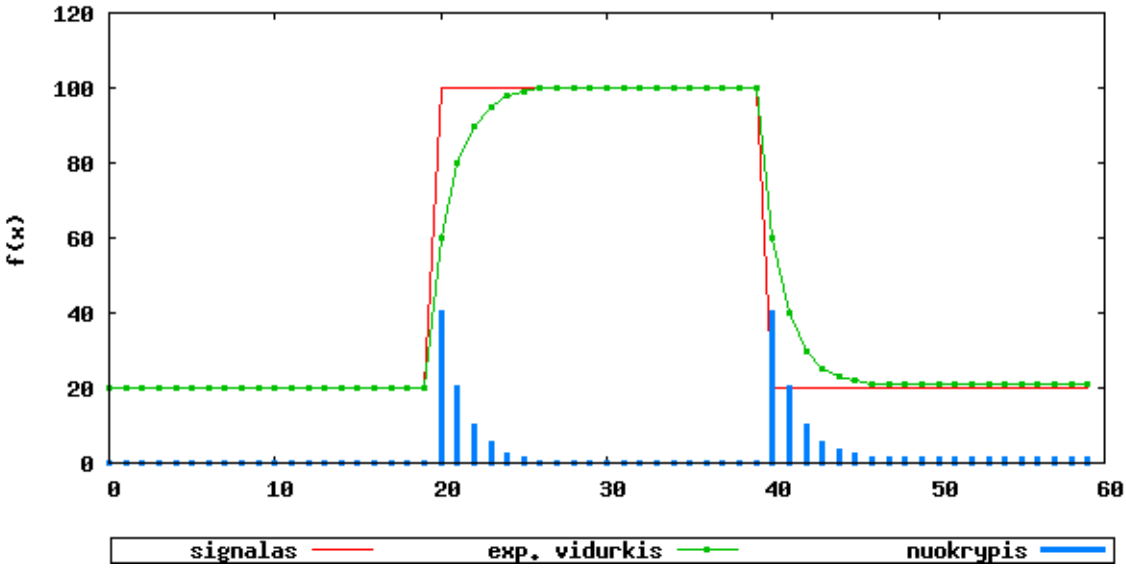
$$\bar{x}_k = a \cdot (a \cdot \bar{x}_{k-2} + (1 - a) \cdot x_{k-1}) + (1 - a) \cdot x_k \tag{7}$$

$$\bar{x}_k = a^2 \cdot \bar{x}_{k-2} + a \cdot (1 - a) \cdot x_{k-1} + (1 - a) \cdot x_k \tag{8}$$

matome, kad (8) formulėje, proporcingumo koeficientas a kinta eksponentiškai: ..., a^2, a, 1. Pagal proporcingumo koeficiento charakteristikas vaizduotas 14 pav. matyti, kad kuo didesnė *a* reikšmė, tuo daugiau įtakos turi buvę elementai ir priešingai.



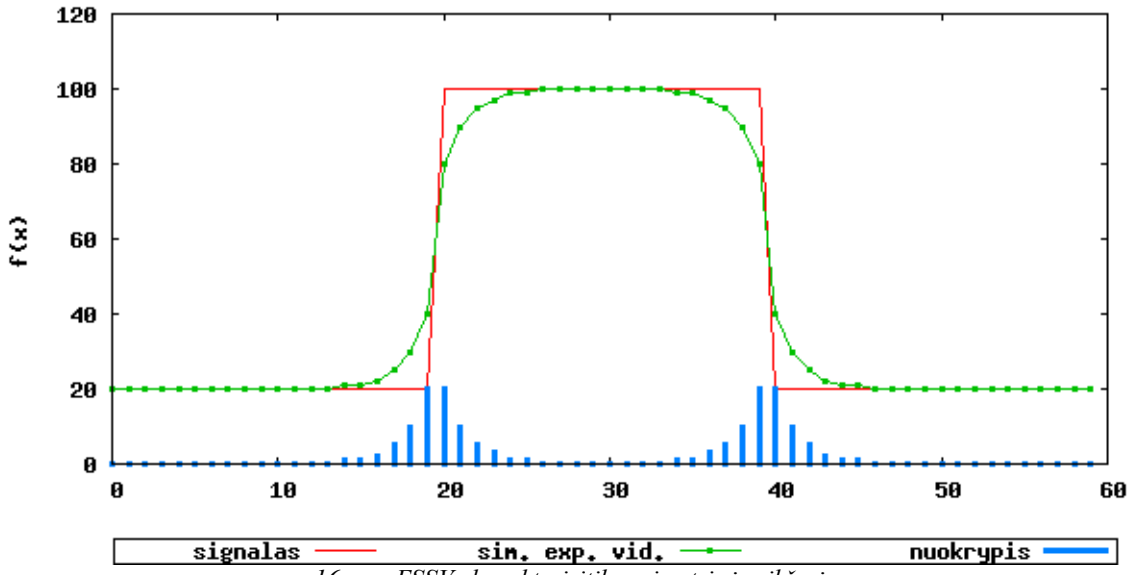
14 pav. Proporcingumo koeficiento a charakteristikos



15 pav. ESSV charakteristikos, apvalinant pagal reikšmę iš kairės

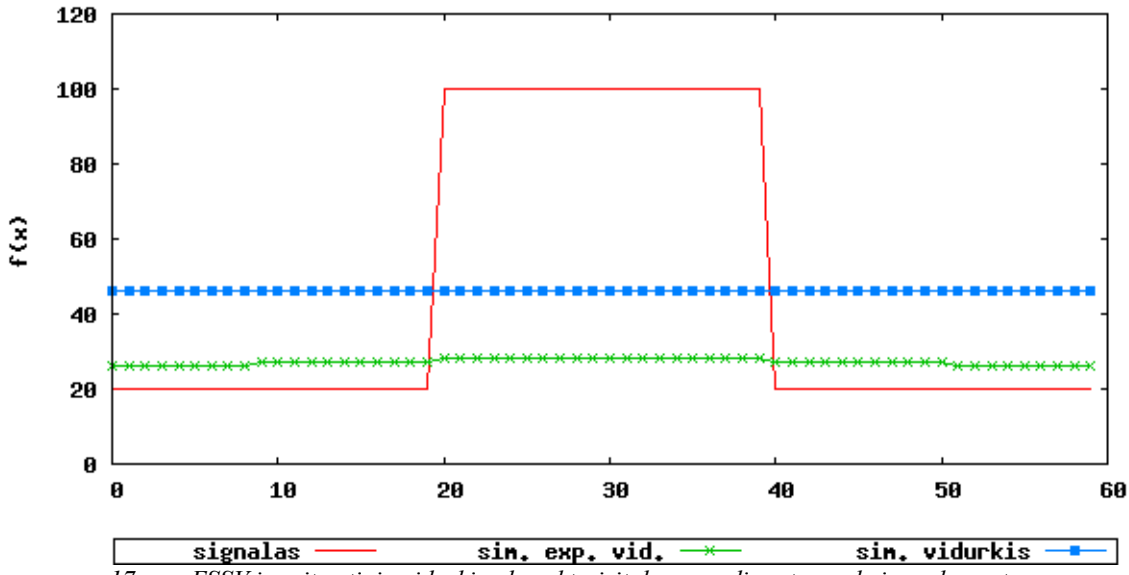
ESSV charakteristikos apskaičiuotos tokiu pačiu principu, kaip ir aritmetiniam vidurkiui, pateiktos 15 pav. Skaičiuojant naudotas proporcingumo koeficientas a=0.5. Lyginant su aritmetinio vidurkio charakteristikomis 12 pav, matyti, kad nuokrypis ties lūžio taškais yra žymiai mažesnis nei aritmetinio vidurkio atveju, tačiau vis dar pakankamai didelis.

Skirtingai nei aritmetinio vidurkio atveju, ESSV simetrinė reikšmė skaičiuojama ne imant reikšmę iš abiejų pusių, o įvykdant du ciklus: iš kairės ir iš dešinės. Simetrinė reikšmė apskaičiuojama pagal iš kairės ir iš dešinės pusės apskaičiuotų reikšmių aritmetinį vidurkį. Skaičiuojant simetrinę ESSV reikšmę, proporcingumo koeficientas a=0.6. Parinkta tokia reikšmė, kad ESSV pradėtų kisti tame pačiame taške, kaip ir aritmetinis vidurkis.



16 pav. ESSV charakterisitikos simetrinė reikšmė

16 pav. matyti, kad simetrinio ESSV atveju, nuokrypis žymiai mažesnis, nei aritmetinio vidurkio atveju 13 pav. Keičiant ESSV proporcingumo koeficientą ir aritmetinio vidurkio apvalinamų elementų kiekį, galima pasiekti skirtingų suapvalinimo laipsnių. Aptinkant anomalijas, naudojamas didelis apvalinimo laipsnis, kad trumpos anomalijos kuo mažiau įtakotų žemo dažnio dedamąją.



17 pav. ESSV ir aritmetinio vidurkio charakterisitokos apvalinant pagal visus elementus

17 pav. matyti, kad ESSV su a=0.9 proporcingumo koeficiento reikšme nuo bazinės linijos nukrypsta maždaug 20% mažiau, negu aritmetinio vidurkio reikšmė, apvalinant pagal visus periodo elementus, todėl signalo šuolis bus aptiktas geriau ir įtaka aplinkinių signalų vidutiniai reikšmei bus mažesnė.

Įvertinus ESSV ir aritmetinio vidurkio savybes, galima daryti išvadą, jog ESSV

priklausomybė nuo pakitusio signalo yra mažesnė, nei aritmetinio vidurkio atveju. Dėl šios priežasties todėl šuoliai ir nukritimai mažiau įtakos aplinkinių signalų vidutines reikšmes ir sumažės neteisingų pranešimų, atsiradusiu dėl stipraus signalo pokyčio, skaičius. Vienetinių šuolių atveju, ESSV yra arčiau bazinio signalo, todėl šuolio ir bazinio signalo santykis bus didesnis ir dėl to anomalijos bus aptinkamos tiksliau.

4.3. SANTYKIO ĮVERTINIMO PROBLEMA

Kai perduodamų duomenų sparta nėra didelė, pakankamai mažas nuokrypis turės sąlyginai didelį santykį su vidutine reikšme. Pavyzdžiui spartai pakitus nuo 1mbps iki 2mbps, santykis bus 2. Tokį patį santyki pasieksime jei reikšmė pakis nuo 100mbps iki 200mbps, nors tokį pokyti pasiekti sunkiau. Dėl tokios priežasties, kai perduodamų duomenų sparta yra maža, žymiai lengviau pasiekiamos situacijos, kada registruojama anomalija.

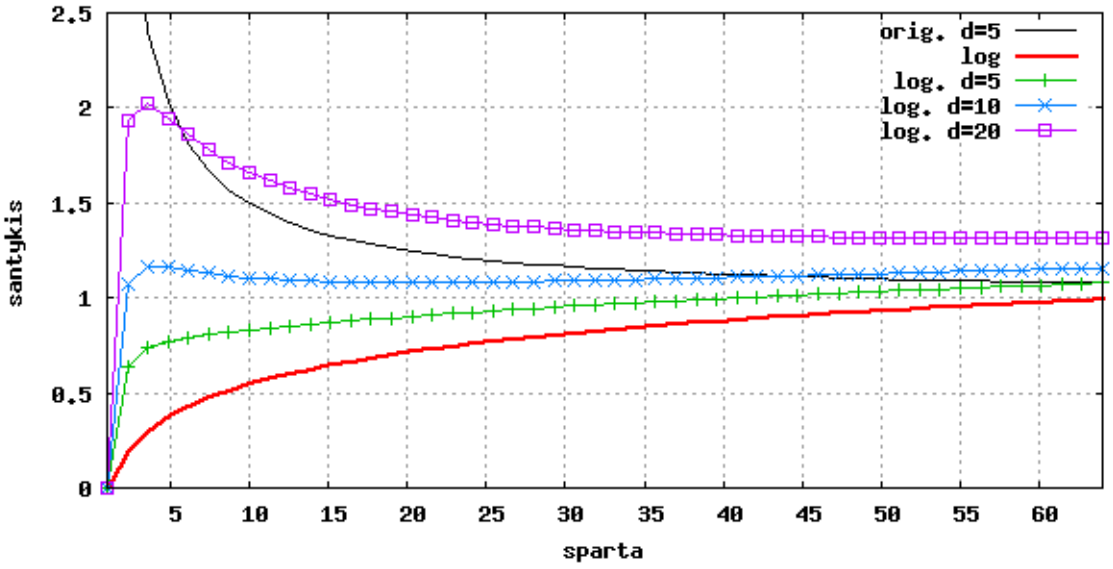
Tokios problemos sprendimui nuspręsta naudoti svorį, pagal kurį būtų pakeičiama santykio reikšmė. Svoris turėtų priklausyti nuo spartos, prie kurios pasiektas tam tikras santykis. Prie žemų spartų, svoris turėtų būti artimas 0, prie aukštų – artimas 1.

Norimas charakteristikas atitinka dvi funkcijos – logaritminė ir eksponentinė.

Logaritminės funkcijos atveju svoris skaičiuojamas pagal (9) formulę:

$$f(x)=\frac{x+d}{x}\cdot\log_{max}(x) \qquad (9)$$

čia d – nuokrypis, x – sparta, max – maksimali pasiekiama sparta. Maksimali pasiekiama sparta gali būti arba konstanta, tačiau patogiau naudoti pačią didžiausią spartą pagal turimą srautų informaciją.



18 pav. Logartiminės svorinės funkcijos charakteristikos

Logaritminės funkcijos charakteristikos pateiktos 18 pav. Vientisa plona linija,

pavaizduotas santykis tarp x+d ir x. Stora linija žymi logaritminės funkcijos svorį. Kitos linijos žymi naują svorio įvertį prie skirtingų nuokrypių. Gauti rezultatai tenkina poreikius, tačiau norėtųsi lankstesnės funkcijos, kad būtų galima reguliuoti funkcijos įtaką.

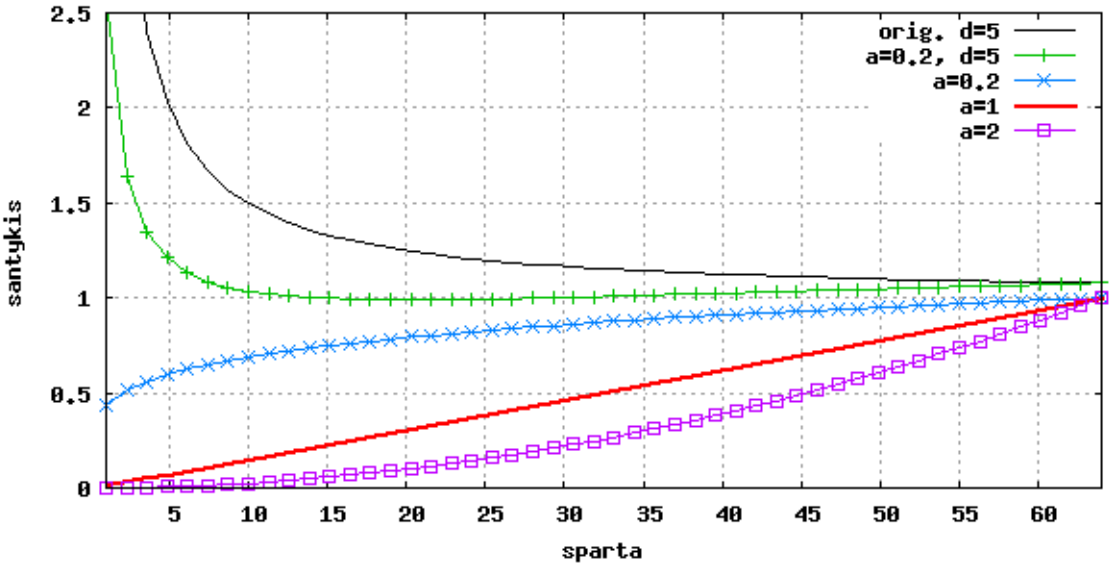
Tokius pačius veiksmus atlikime su eksponentine funkcija:

$$f(x)=\frac{x+d}{x}\cdot\left(\frac{x}{max}\right)^{a} \qquad (10)$$

čia a – koeficientas, reguliuojantis funkcijos poveikį:

- jei a=0 – funkcija visiškai neįtakoja santykio,
- jei a=1, funkcijos reikšmė kinta tiesiškai tiesinė,
- jei 0 < a < 1, funkcijos reikšmė kinta eksponentiškai.

Funkcijos charakteristikos pateiktos 19 pav.



19 pav. Eksponentinės svorinės funkcijos charakteristikos

Naudojant eksponentinį svorinį santykio įvertinimą, galima sumažinti neteisingai užregistruotų anomalijų kiekį. Funkcijos poveikio koeficientą a, reikėtų pasirinkti priklausomai nuo to, kokiame intervale svyruoja srautų sparta. Jei sparta svyruoja nuo žemos iki labai aukštos, vertėtų pasirinkti didesni koeficientą, jei sparta svyruoja nedaug, koeficientas gali būti artimas 0.

5. EKSPERIMENTAI

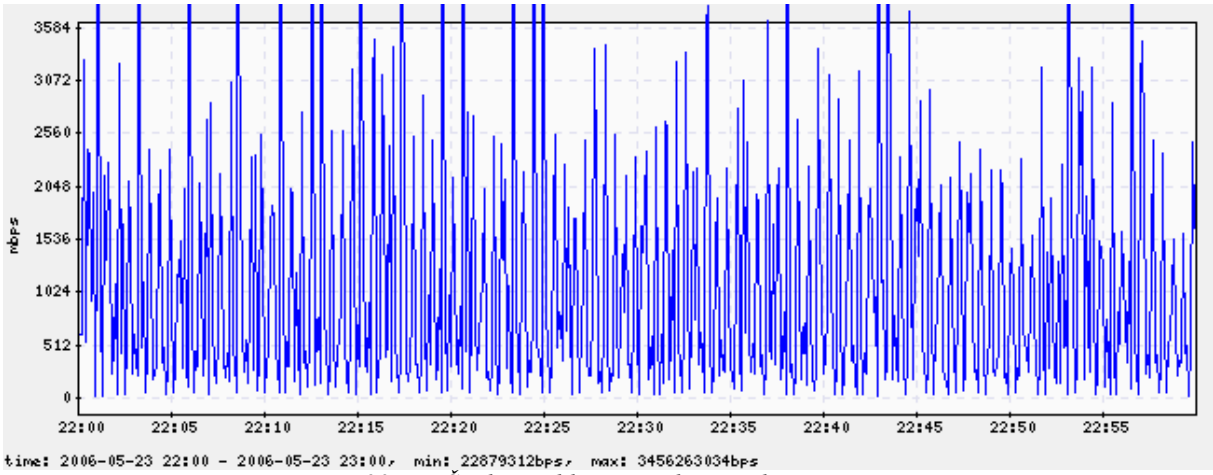
4 skyriuje „Tyrimas“, buvo pasiūlyti teoriniai programinės įrangos tobulinimo metodai. Šiame skyriuje pateikti eksperimentų rezultatai, gauti pritaikius pasiūlytus problemų sprendimo metodus. Visi eksperimentai atlikti su realiais tinklo srautų duomenimis. Eksperimentiniams duomenims buvo pasirinktas vienos valandos srautų periodas.

5.1. SRAUTŲ PASKIRSTYMO PROBLEMA

Pasirinkto periodo NetFlow duomenys buvo apdoroti naudojant du skirtingus srautų paskirstymo metodus:

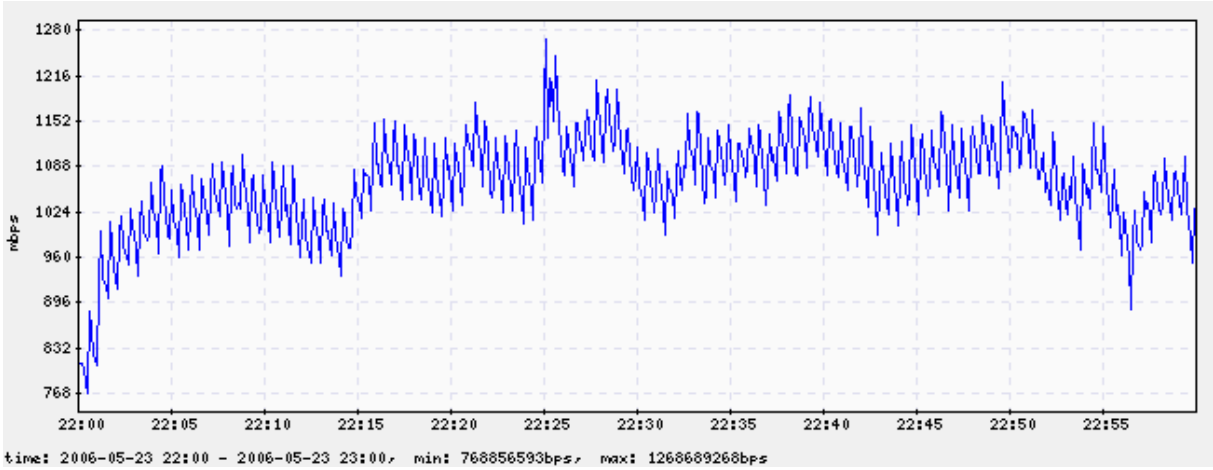
- persiųstų duomenų kiekio priskyrimą srauto gyvavimo pabaigos momentui
- tolydų duomenų paskirstymą per visą srauto gyvavimo periodą.

Duomenų paskirstymas, pagal persiųstų duomenų kiekio priskyrimą srauto gyvavimo pabaigos momentui, pateiktas 20 pav. Jei pagal tokią informaciją būtų nustatinėjamos tinklo anomalijos, duomenys būtų itin netikslūs ir praktiškai nenaudingi.



20 pav. Šuolių problema vaizduojant duomenis

Šios programinės įrangos patobulinimui buvo pasiūlytas tolydus duomenų paskirstymo metodas. 21 pav. pateiktas to paties laikotarpio srautų informacijos vaizdas. Pritaikius pakeitimus, reikšmių svyravimas sumažėjo maždaug 6 kartus. Pagal ši metodą pateiktame grafike galima išskirti atskirus momentus, kur matomas duomenų perdavimo spartos padidėjimas ar sumažėjimas. Likusius šuolius galima išlyginti prieš vaizduojant duomenis pritaikant filtrą, kuris suvienodintų aplinkinių taškų reikšmes.



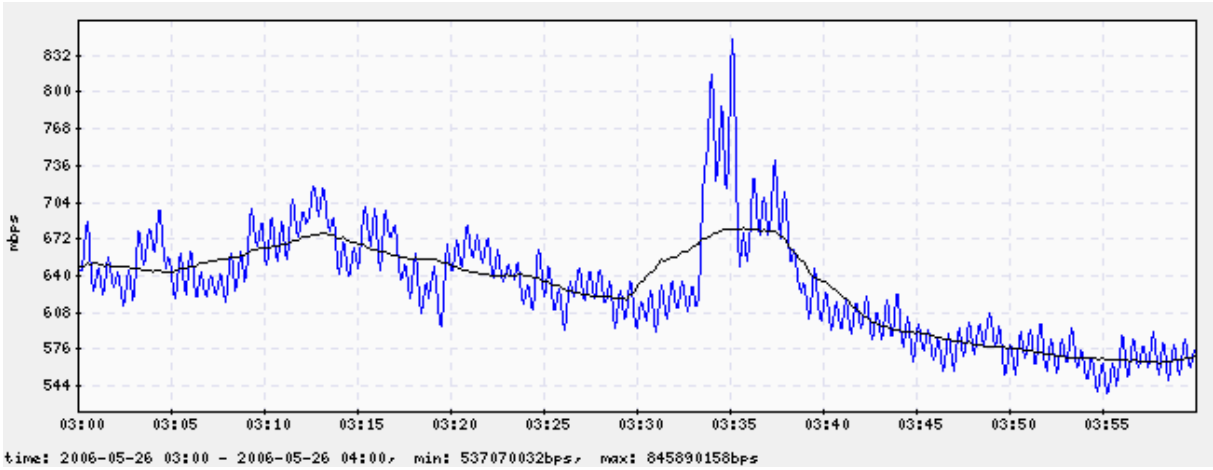
21 pav. Duomenų vaizdas po problemos išsprendimo

5.2. SVORINIS EKSPONENTINIS SLENKANTIS VIDURKIS

Pasirinkto periodo duomenys buvo filtruojami naudojant dvejus metodus:

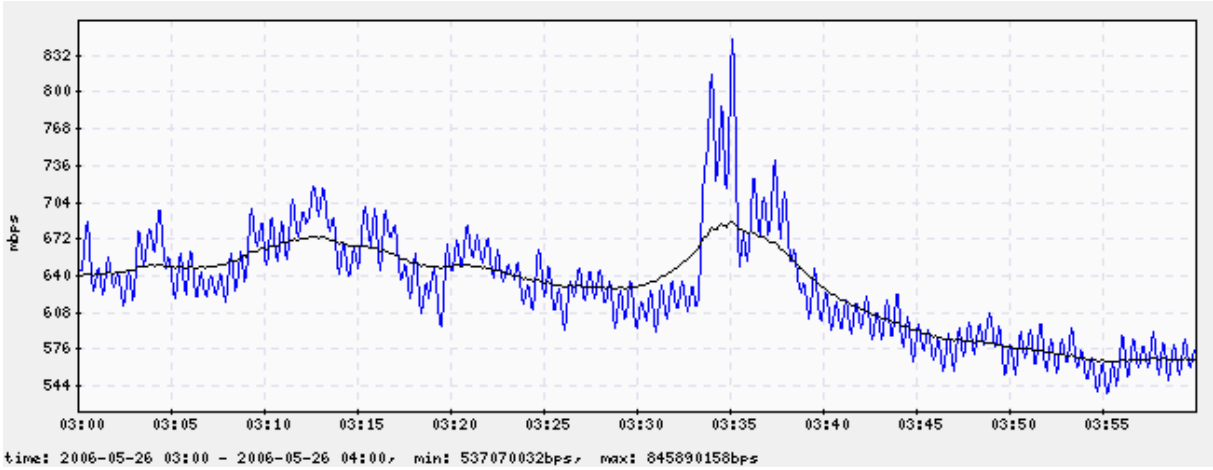
- simetrinį aritmetinį vidurkį
- simetrinį eksponentinį svorinį vidurkį

Eksperimento metu, duomenys buvo filtruojami naudojant simetrinį aritmetinį vidurkį pagal 40 iš kairės ir dešinės esančių reikšmių. Realūs duomenys ir filtruota reikšmė pavaizduota 22 pav.



22 pav. Realių srautų filtravimas naudojant simetrinį aritmetinį vidurkį

Gautus rezultatus arčiausiai atitiko ESSV filtras, su a=0.85 svoriniu įverčiu. Gauta filtruota reikšmė pateikta 23 pav.



23 pav. Realių srautų filtravimas naudojant simetrinį ESSV

Palyginus gautus rezultatus, metodų skirtumas geriausiai matomas laiko momentais 03:30 – 03:35. Srautus filtruojant aritmetinio vidurkio metodu, prieš duomenų šuolį bazinės dedamosios reikšmė pakyla žymiai ankščiau, nei filtruojant ESSV metodu. Toks rezultatas atitinka teorines prognozes, pateiktas 4.2 skyriuje.

6. IŠVADOS

Pagrindinis darbo tikslas – pagerinti sukurtos programinės įrangos kokybę. Ištyrus sukurtos programinės įrangos trūkumus, buvo pasiūlytas ir ištirtas aritmetinio vidurkio ir svorinio santykio įvertinimo patobulinimas. Patobulinus srautų stebėjimo ir anomalijų aptikimo metodus buvo sumažintas klaidingų anomalijų skaičius. Pritaikius srautų išskaidymą, buvo padidintas NetFlow formatu kaupiamų duomenų tikslumas, tapo galima atlikti tikslesnę srautų analizę bei buvo panaikintos ilgai trunkančių srautų vaizdavimo klaidos. Remiantis teorinėmis prognozėmis ir jas patvirtinančiais eksperimentų rezultatais, galima daryti išvadą, kad padidėjo sistemos tikslumas, todėl pagrindinis darbo tikslas buvo įgyvendintas.

7. LITERATŪRA

1. Barford, P. ir Plonka, D. Characteristics of Network Traffic Flow Anomalies. *ACM SIGCOMM Internet Measurement Workshop*: tarpatutinės konferencijos pranešimų medžiaga. San Franciskas, 2001, p. 69-73.
2. Miluocheva, I. ir Muller, E. A practical approach to forecast Quality of Service parameters considering outliers. *1st Int. Workshopo Inter-Domain Performance and Simulation*: tarpatutinės konferencijos pranešimų medžiaga. Salzburgas, 2003, p. 163-172.
3. Thottan, M. ir Ji, C. Anomaly Detection in IP Networks. *IEEE Transactions on signal processing*, 2003, Nr. 51.
4. Ndousse, T.D. ir Okuda, T. Computational intelligence for distributed fault management in networks using fuzzy cognitive maps. *IEEE ICC*: tarpatutinės konferencijos pranešimų medžiaga. Dalasas, 1996, p. 1558-1562.
5. Lewis, L. A case based reasoning approach to the management of faults in communication networks. *IEEE INFOCOM*: tarpatutinės konferencijos pranešimų medžiaga. San Franciskas, 1993, p. 1422-1429.
6. Franceschi, A. S., Kormann, L. F. ir Westphall, C. B. Performance evaluation for proactive network management. *IEEE ICC*: tarpatutinės konferencijos pranešimų medžiaga. Dalasas, 1996, p. 22-26.
7. Katzela, I. ir Schwarz, M. Schemes for fault identification in communication networks. *IEEE/ACM Transactions on Networking*, 1995, Nr. 3, p. 753-764.
8. Rouvellou, I. ir Hart, G. Automatic alarm correlation for fault identification. *IEEE INFOCOM*: tarpatutinės konferencijos pranešimų medžiaga. Bostonas, 1995, p. 553-561.
9. Lazarevic, A.; et al. A comparative study of anomaly detection schemes in network intrusion detection. SIAM International Conference on Data Mining: tarpatutinės konferencijos pranešimų medžiaga. Sanfranciskas, 2003.
10. Feather, F. ir Maxion, R. Fault detection in an ethernet network using anomaly signature matching. *ACM SIGCOMM*: tarpatutinės konferencijos pranešimų medžiaga. San Franciskas, 1993, p. 279-288.
11. Soule, A.; Salamatian, K. ir Taft, N. Combining filtering and statistical methods for anomaly detection. Internet Measurement Conference: tarpatutinės konferencijos pranešimų medžiaga. Berklis, 2005.
12. ImageStram Internet Solutions, Inc.. Network Monitoring. 2003 [žiūrėta 2006-05-15]. Prieiga per internetą: <http://oem.imagestream.com/Monitoring_White_Paper.PDF>.
13. Lakhina, A.; Crovella, M.; ir Diot, C. Characterization of Network-Wide Anomalies in

Traffice Flows. *4th ACM SIGCOMM conference on Internet measurement*: tarpatutinės konferencijos pranešimų medžiaga. Taormina, 2004, p. 201-206.

14. Markopoulou, A.; et al. Characterization of Failures in an IP Backbone. *IEEE INFOCOM*: tarpatutinės konferencijos pranešimų medžiaga. Honkongas, 2004, .
15. Jung, J; Krishnamurthy, B.; ir Rabinovic, M. Flash crowds and denial of service attacks: characterization and implications for CDNs and web sites. *11th international conference on World Wide Web*: tarpatutinės konferencijos pranešimų medžiaga. Honolulu, 2002, p. 293-304.
16. CERT Coordination Center. Denial of Service Attacks. 2001 [žiūrėta 2005-05-18]. Prieiga per internetą: <http://www.cert.org/tech_tips/denial_of_service.html>.
17. Streilein, W.; Fried, D.; ir Cunningham, R. Detecting Flood-based Denial-of-Service Attacks with SNMP/RMON. Workshop on Statistical and Machine Learning Techn: tarpatutinės konferencijos pranešimų medžiaga. Fairfaksas, 2003.
18. Lakhina, A; Crovella, M; ir Diot, C. Diagnosing network-wide traffic anomalies. *2004 conference on Applications, technologies, architectures, and protocols for computer communications*: tarpatutinės konferencijos pranešimų medžiaga. Portlandas, 2004, 219-230.
19. Feinstein, L; et al Statistical Approaches to DDoS Attack Detection and Response. *DARPA Information Survivability Conference and Exp*: tarpatutinės konferencijos pranešimų medžiaga. IEEE Computer Society, 2003, p. 303.
20. Cheng, C.-M.; Kung, H. T.; ir Tan, K.-S.; Use of spectral analysis in defense against DoS attacks. IEEE GLOBECOM: tarpatutinės konferencijos pranešimų medžiaga. Taipejus, 2002.
21. Hussain, A.; Heidemann, J.; ir Papadopoulos, C.; A framework for classifying denial of service attacks. *2003 conference on Applications, technologies, architectures, and protocols for computer communications*: tarpatutinės konferencijos pranešimų medžiaga. Niujorkas, 2003, p. 99-110.
22. Landfeldt, B.; Sookavatana, P.; ir Seneviratne, A. The Case for a Hybrid Passive/Active Network Monitoring Scheme in the Wirel. *8th IEEE International Conference on Networks*: tarpatutinės konferencijos pranešimų medžiaga. Vasingtonas, 2000, p. 139.
23. Nacionalinė Lorenco Berklio laboratorija. libpcap. 2005 [žiūrėta 2006-05-20]. Prieiga per internetą: <http://www.tcpdump.org/>.
24. Deri. L. Improving Passive Packet Capture: Beyond Device Polling. SANE 2004: tarpatutinės konferencijos pranešimų medžiaga. Amsterdamas, 2005.
25. Grossglauser, M.; ir Rexford, J. Passive Traffic Measurement for IP Operations.

INFORMS Telecom Meeting: tarpatutinės konferencijos pranešimų medžiaga. Lauderdeilas, 2002.

26. Stallings, W. *SNMP, SNMPv2, SNMPv3 and RMON 1 and 2, 3rd edition*. Bostonas. 1998. 619 p. ISBN 0-20-148534-6.
27. Perkins, D. *RMON: remote monitoring of SNMP-managed LANs*. Prentice Hall. 1998. 440 p. ISBN 0-13-096163-9.
28. Cisco Systems. NetFlow Services Solutions Guide. 2004 [žiūrėta 2006-05-20]. Prieiga per internetą: <http://www.cisco.com/univercd/cc/td/doc/cisintwk/intsolns/netflsol/nfwhite.pdf>.
29. Romig, S; Fullmer, M; ir Luman, R.; The OSU Flow-tools Package and CISCO NetFlow Logs. *14th USENIX conference on System administration*: tarpatutinės konferencijos pranešimų medžiaga. Berklis, 2000, p. 291-304.
30. Ishiguro, M; Suzuki, H,; Murase, I.; ir Ohno, H. Internet threat detection system using Bayesian estimation. 16th Annual FIRST Conference on Computer Security: tarpatutinės konferencijos pranešimų medžiaga. Budapestas, 2004.
31. Hood, C.; ir Ji, C. Proactive network fault detection. *INFOCOM '97*: tarpatutinės konferencijos pranešimų medžiaga. Vasingtonas, 1997, p. 1147.
32. Abry, P; ir Veitch, D. Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory*, 1998, Nr. 1, p. 2-15.
33. Barford, P; et al. A signal analysis of network traffic anomalies. *2nd ACM SIGCOMM Workshop on Internet measurment*: tarpatutinės konferencijos pranešimų medžiaga. Marseile, 2002, .
34. Roughan, M; et al. IP forwarding anomalies and improving their detection using multiple data sources. *ACM SIGCOMM workshop on Network troubleshooting*: tarpatutinės konferencijos pranešimų medžiaga. Portlandas, 2004, p. 307-312.
35. Brutlag, J. Aberrant behavior detection in time series for network monitoring. *14th USENIX conference on System administration*: tarpatutinės konferencijos pranešimų medžiaga. Naujasis Orleanas, 2000, p. 139-146.

8. PAVEIKSLŲ SĄRAŠAS

9. LENTELIŲ SĄRAŠAS

10. SANTRUMPŲ IR TERMINŲ ŽODYNAS

1. **CERT** (angl. *Computer Emergency Responce Team*) – kompiuterinių incidentų tyrimo tarnyba.
2. **DOS** (angl. *Denial of Service*) – atkirtimo nuo paslaugos ataka.
3. **ESSV** – eksponentinis svorinis slenkantis vidurkis.
4. **HTTP** (angl. *Hyper-Text Transfer Protocol*) – Tinklapių informacijos perdavimo protokolas.
5. **ICMP** (angl. *Internet Control Message Protocol*) – Interneto valdymo pranešimų protokolas. Naudojamas pranešimų apie klaidas bei maršrutizavimo sprendimus apsikeitimui tarp maršrutizatorių bei tinklinių įrenginių .
6. **IP** – (angl. *Internet Protocol*) Interneto protokolas, suteikiantis adresavimo ir fragmentavimo mechanizmus datagramų perdavimui paketais komutuojamuose tinkluose
7. **NetFlow** – tinklo srautų registravimo metodas.
8. **MIB** (angl. *Management Information Base*) – SNMP protokolo valdymo informacijos bazė.
9. **RMON** (angl. *Remote Monitoring*) – telekomunikacijų įrangos standartas leidžiantis nuotolinį tinklo įrangos stebėjimą naudojant SNMP protokolą.
10. **SNMP** (angl. *Simple Network Management Protocol*) – tinklo valdymo protokolas naudojamas nuotoliniam įrenginių stebėjimui ir valdymui.
11. **TCP** (angl. *Transmission Control Protocol*) – protokolas, skirtas patikimam duomenų perdavimui internete tarp dviejų įrenginių.
12. **TOS** (angl. Type Of Service) – IP protokolo parametras, nurodantis paslaugos kokybės charakteristikas.
13. **TTL** (angl. *Time To Live*) – IP protokolo parametras, nurodantis kiek kartų paketas gali būti nusiųstas tolesniam maršrutizatoriui.

11. PRIEDAI

Prie dokumento pateikiamas kompaktinis diskas. Disko turinys pateikiamas 4 lentelėje.

4 lentelė. Kompaktinio disko turinys

Failo vardas	Dydis	Turinys
baigiamasis_darbas.pdf	427K	Šis dokumentas
programos_dokumentacija.pdf	3.7M	Sukurtos programinės įrangos dokumentacija
nfbad.tgz	176K	Darbiniai programos failai
sasaja.tgz	11K	Programos sąsaja su vartotoju.