KAUNO TECHNOLOGIJOS UNIVERSITETAS

INFORMATIKOS FAKULTETAS

Dalius Lapėnas

EL. PAŠTO SAUGOS PRIEMONIŲ EFEKTYVUMO TYRIMAS

Baigiamasis magistro darbas

Vadovas
Doc. dr. Nerijus Morkevičius

KAUNAS, 2016

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
KOMPIUTERIŲ KATEDRA

EL. PAŠTO SAUGOS PRIEMONIŲ EFEKTYVUMO TYRIMAS

Baigiamasis magistro darbas
Informacijos ir informacinių technologijų sauga (kodas 621E10003)

Vadovas
(parašas) Doc. dr. Nerijus Morkevičius
(data)

Recenzentas
(parašas) Doc. dr. Tomas Adomkus
(data)

Projektą atliko
(parašas) Dalius Lapėnas
(data)

KAUNAS, 2016

KAUNO TECHNOLOGIJOS UNIVERSITETAS

Informatikos fakultetas

(Fakultetas)

Dalius Lapėnas

(Studento vardas, pavardė)

Informacijos ir informacinių technologijų sauga (kodas 621E10003)

(Studijų programos pavadinimas, kodas)

„EL. PAŠTO SAUGOS PRIEMONIŲ EFEKTYVUMO TYRIMAS“

AKADEMINIO SĄŽININGUMO DEKLARACIJA

20 16 m. Gegužės 16 d.

Kaunas

Patvirtinu, kad mano **Daliaus Lapėno** baigiamasis projektas tema „EL. PAŠTO SAUGOS PRIEMONIŲ EFEKTYVUMO TYRIMAS“ yra parašytas visiškai savarankiškai, o visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka) *(parašas)*

Lapenas, D. „El. pašto saugos priemonių efektyvumo tyrimas“. Magistro baigiamasis projektas / vadovas doc. dr. Nerijus Morkevičius; Kauno technologijos universitetas, Informatikos fakultetas, Kompiuterių katedra.

Kaunas, 2016. 57 p.

SANTRAUKA

Šiame darbe analizuojama elektroninio pašto saugumo problema, t.y. kaip apsaugoti elektroninio pašto sistemų resursus nuo perteklinio nepageidaujamų arba žalingų laiškų apdorojimo. Išanalizavus literatūros šaltinius, įsitikinta, kad elektroninis paštas yra svarbus ir pažeidžiamas elektroninių komunikacijų kanalas, o vieningo, visais atvejais efektyvaus nepageidaujamų laiškų filtravimo metodo ar metodų sistemos, nėra.

Darbe buvo apžvelgti populiariausi el. laiškų filtravimo metodai ir eksperimentiškai ištirta jų įtaka tarnybinių stočių operaciniams resursams. Pagal surinktus statistinis duomenis – nustatytas santykinis šių metodų efektyvumas ir sukurtas kompiuterinis imitacinis modelis, leidžiantis imituoti filtravimo sistemos veikimą, keisti jos komponentų nustatymus.

Imitacinio modelio validavimo metu nustatyta, kad modelis iš esmės atitinka realios elektroninių laiškų filtravimo sistemos veikimą, atkartoja pasaulinę elektroninio pašto kanalu perduodamų laiškų dydžių statistiką, ir tinka simuliuoti įvairių filtravimo metodų vėlinimą laiškų pristatymui bei santykinį efektyvumą.

Lapėnas, Dalius. *Research of Effectiveness of Email Security Methods: Master's* thesis in Information and Information Technology Security / supervisor assoc. prof. Nerijus Morkevičius. The Faculty of Informatics, Kaunas University of Technology.

Research area and field: email security, computer simulation

Key words: email, spam, simulation, effectiveness, model

Kaunas, 2016. 57 p.

SUMMARY

This study analyses email security problem with regards to an operational overhead related to unsolicited junk mail reception and processing. It was discovered that emails as such is still very important electronic communication channel susceptible to malicious threat and that there is no unified universal and effective email filter method that would perform well in every environment under any situation.

The most popular email filter methods were reviewed. Using the methods in subject, experiments were carried out to identify impact on operating system resource utilization when using each of them. Based on the data collected the relative efficiency of these methods were defined and simulation model was created. This model allows a user to simulate an email filter system's operation with ability to change parameters.

During the validation of the model it was determined that the actual model corresponds to the real system operation and replicates global email distribution based on email size and can be used to simulate various email filtration methods, their processing delay and efficiency.

TURINYS

LENTELIŲ SĄRAŠAS

PAVEIKSLŲ SĄRAŠAS

TERMINŲ IR SANTRUMPŲ ŽODYNAS

Santrumpa	Reikšmė
ASCII	Angl. *American Standard Code for Information Interchange*. Simbolių kodavimo schema.
DNS	Ang. *Domain Name Service*. Vardų srities paslauga.
SMTP	Angl. *Simple Mail transfer protocol*. Parasto pašto perdavimo protokolas.
DNSBL	Angl. *DNS Blacklist*. DNS juodasis sąrašas
URIBL	Angl. *Uniform Resource Identifier*. Suvienodintas resursu identifikatorius.
RBL	Angl. *Real-time Blackhole List*. Realaus laiko „juodosios skylės“ sąrašas.
DKIM	Angl. *DomainKeys Identified Mail*. Srities raktais identifikuojamas paštas.
DANE	Angl. *DNS-based Authentication of Named Entities*. DNS pagrindu authentifikuoti įvardinti subjektai.
UDP	Angl. *User Datagram protocol*. Varotojo datagramos protokolas
IP	Angl. *Internet protocol.* Interneto protokolas
TCP	Angl. *Transmision control protocol*. Perdavimo kontrolės protokolas
RFC	Angl. *Request for comments*. Komentarų užklausimas.
IETF	Angl. *Internet Engineering Task Force.*
MSA	Angl. *Message submission Agent*. Žinučių pateikimo agentas.
MTA	Angl. *Message transfer Agent*. Žinučių perdavimo agentas.
MDA	Angl. *Message delivery Agent*. Žinučių pristatymo agentas
MX	Angl. *Mail Exchanger*. Pašto apsiketimo paslauga.
SMIME	Angl. *Secure Multipurpose Internet Mail extensions.*
SPF	Angl. *Sender Policy framework*. Siuntėjo politkos struktūra.
DMARC	Angl. *Domain-Based Message authentication, reporting and conformance.* Žinučių autentifikavimas, ataskaitos ir atitiktis vardų srities pagrindu
SIMAN	Angl. *Simulation and Automation*. ARENA apdorojimo ir modeliavimo kalba.
SURBL	Angl. *Spam URI RBL*. Laiškuose naudojamų adresų juodasis sąrašas.
SPAM	Angl. *Electronic spamming.* Brukalas
TLS	Angl. *Transport layer security*. Perdavimo lygio apsauga.
SSL	Angl. *Secure sockets layer.* Saugių jungčių lygmuo.
SQL	Angl. *Structured Query Language.* Duomenų užklausų aprašomoji kalba.
XML	Angl. *Extensible Markup Language.* Duomenų struktūrų aprašomoji kalba.

ĮVADAS

Įvairios elektroninių žinučių sistemos buvo naudojamos jau nuo 1960 m. Elektroninis paštas toks, kokį mes naudojame šiandien buvo pradėtas naudoti maždaug nuo 1980 m. ir atsirado kartu su elektroninio pašto protokolu, kuris vėliau buvo pakeistas į SMTP (angl. *simple mail transfer protocol*). Protokolas, su keletu išplėtimų ir patobulinimų, yra naudojamas ir šiandien. Padedant šiam protokolui, elektroninis paštas perduodamas tarp kliento ir tarnybinės stoties ir tarp tarnybinių stočių.

Galima pamanyti, kad ganėtinai seno apsikeitimo žinutėmis protokolo populiarumas ir efektyvumas krenta, bet panašu, kad populiarėjant mobiliesiems prietaisams bei išmaniesiems telefonams, žmonių naudojimosi el. paštu tendencijos tik didėja. El. paštas išlieka pagrindiniu elektroninės komunikacijos būdu nustelbdamas tiek socialinius tinklus, tiek realaus laiko apsikeitimo žinutėmis sistemas. Retas žmogus ar organizacija nesinaudoja el. paštu, o ir kuriant paskyras bei primenant slaptažodžius – el. paštas yra *de facto* standartas.

Dėl ekonominių sumetimų, siekiant maksimaliai optimizuoti kaštus, tuo pačiu neprarandant el. pašto produktyvumo – patikimumo ir saugumo užtikrinimas yra labai svarbus. Grėsmės, keliančios pavojų minėtoms savybėms, yra šios: nepageidaujami elektroniniai laiškai, virusai, sukčiavimo atvejai (angl. *phishing*), taip pat el. pašto žinučių vientisumo pažeidimai bei siuntėjo autentiškumo ar neišsiginamumo problemos.

Kovojant su grėsmėmis yra naudojami įvairūs el. pašto žinučių filtravimo metodai, tačiau žalingų ar nepageidaujamų laiškų kiekis internete yra toks didelis, kad yra perkraunamos pasaulinio tinklo bei el. pašto infrastruktūros resursai. Pagal tai galima spręsti, kad nepaisant šiuo metu populiarių filtravimo būdų, kova su šiukšlėmis internete yra dar toli gražu ne laimėta.

Dažnai, testuojant sistemas, tai yra daroma su realiomis sistemomis ar tarnybinėmis stotimis, supaprastintomis, ar išskaidytomis atskiromis dalimis, pasikliaujant praktika, visuotinai priimtomis normomis, konsultantų patarimais ar intuicija. Tokių bandymų rezultatai būna ne visai tokie, kokių sulaukiama realioje situacijoje. Pagrindinės priežastys, dėl kurių tokių testavimų efektyvumas netenkina, yra sistemos komponentų tarpusavio ryšių ir įtakos numatymo sudėtingumas, priimtų sprendimų įtaką sistemai, nelankstus komponentų parametrų keitimas ir panašiai. Modeliavimas naudojant realias sistemas yra brangus, užima daug laiko ir ne visais atvejais įmanomas. Tokiais atvejais į pagalbą pasitelkiamas kompiuterinis imitacinis modeliavimas. Modeliavimo tikslas - rasti optimalius sprendimus sistemoms su iš anksto apsibrėžtais tikslais, rasti esamų sistemų silpnąsias vietas ar įvertinti efektyvumą, lanksčiai keisti sistemos komponentų parametrus, realiu laiku stebėti pakeitimų įtaką bendrai sistemai. Magistro darbe naudojama imitacinis kompiuterinis modeliavimas, kaip įrankis brukalo filtravimo sistemos simuliacijai, ir pristatomas validuotas tokios sistemos modelis.

Atsižvelgiant į prieš tai išsakytas mintis, šiame darbe buvo:

- išanalizuota grėsmė el. pašto saugai,
- išsiaiškinta esama saugumo ir filtravimo architektūra ir metodai,
- minėtų metodų įtaką tarnybinių stočių resursų panaudojimui,
- šių metodų efektyvumas,
- pasiūlytas ir validuotas modelis, leidžiantis imituoti filtravimo metodų veikimą.

Naudojant sukurtą modelį bus galima simuliuoti įvairių filtravimo metodų įtaką sisteminiams resursams, el. pašto žinučių perdavimo ir apdorojimo trukmę.

Šis darbas parašytas studijuojant Informacijos ir IT saugą Kauno technologijos universitete.

Darbo problematika ir aktualumas

Nepageidaujamų elektroninių laiškų, dažniausiai reklaminių, siuntimas (angl. *email spam*) [1], yra eksponentiškai plintantis reiškinys, prasidėjęs maždaug 1990m. ir šiuo metu, pagal konservatyvius skaičiavimus, sudaro apie 80 – 85 proc. viso perduodamo elektroninio pašto pasaulyje. Pasaulio vyriausybės, įdėjusios didžiules pastangas paskelbiant nepageidaujamo el. pašto siuntimą nelegalia veikla, interneto paslaugų tiekėjai ėmęsi filtravimo, panašu, kad sugebėjo pristabdyti šio reiškinio plitimą. Korporacija „Symantec“ savo pranešime „2014 Internet Security

Threat Report, Volume 19“ [2] skelbia, kad dabartiniai skaičiai sumažėjo iki 66 proc. (1 pav.) arba 23 milijardų, atitinkamai Lietuvoje – 87 proc., Latvijoje 63 proc., Estijoje – 75 proc. 2015 metų leidinyje skelbiama, kad pasaulinis brukalo kiekis dar mažėjo iki 60 proc.

Reikia pastebėti, kad vis dažniau nepageidaujamus laiškus siunčia kompiuteriai „zombiai“ arba virusais užkrėsti kompiuteriai. Kitais atvejais yra net siūlomos debesų kompiuterijos paslaugos siųsti naujienlaiškiams. Naujienlaiškis nuo brukalo skiriasi tik tuo ar gavėjas pageidauja gauti tam tikro turinio informaciją ar ne, todėl nubausti ar klasifikuoti tokią paslaugą siūlančias kompanijas yra sunku. Dažniausia tokios įmonės atsakomybę perkelia paslaugos užsakovui, kurį retai galima atsekti dėl silpnos registracijos verifikavimo politikos.

1 pav. Nepageidaujamo el. pašto lygis

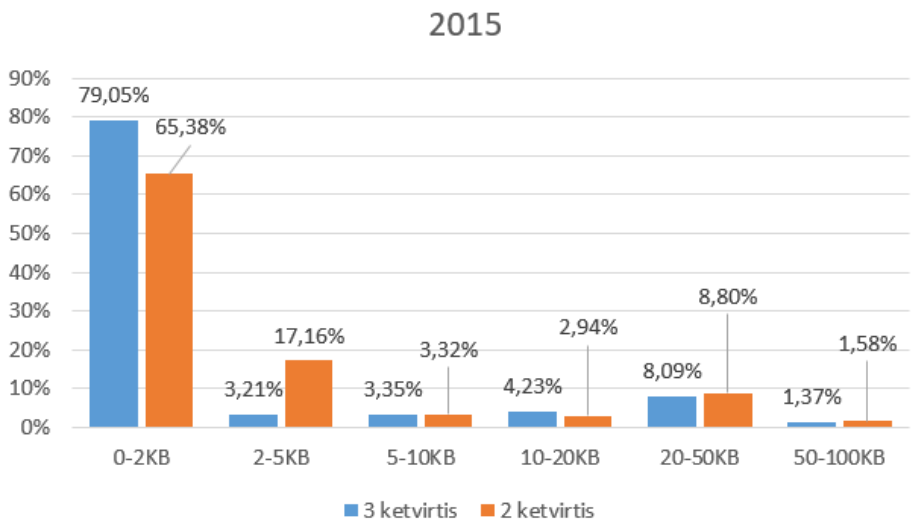
Kita pavojinga grėsmė yra žalingo kodo arba virusų platinimas elektroninio pašto kanalu. Ši problema yra daug retesnė, nes dauguma vartotojų jau turi įdiegtą antivirusinę programinę įrangą arba el. pašto srautas yra automatiškai filtruojamas interneto ar el. pašto paslaugų tiekėjų. Visgi, „Symantec“ korporacijos duomenimis [2] statistiškai tokių atvejų daugėja, 2 pav. parodo kokiame kiekyje laiškų tikėtinas vieno virusu užkrėsto laiško pasitaikymas.

2 pav. Virusas laiškų imtyje

Stebimas augantis sukčiavimo laiškų siuntimas (angl. *spear-phishing*), 2013 metais buvo užfiksuotas 91 proc. tikslingų atakų augimas, taip pat 3 kartus pailgėjusių tikslingų atakų kampanijų vykdymo laikas. Per tokio tipo atakas yra klastojamas siuntėjo adresas arba žinutės turinys taip klaidinant gavėją, kad el. pašto žinutė atėjo iš patikimo siuntėjo. SMTP protokolas pats vienas neturi tikrojo siuntėjo susiejimo su el. pašto dėžutės savininku mechanizmo, kovai prieš siuntėjo adreso klastojimą pasitelkiami kriptografiniai bei kitokie metodai, kurie turi įtakos sistemos resursų naudojimui.

Nepageidaujamų laiškų siuntimas yra brangi veikla, ypač tinklo infrastruktūros prasme, todėl 2015 metais tokie laiškai buvo siunčiami kaip įmanoma mažesnio dydžio – didžioji dauguma jų buvo iki 1 KB dydžio (3 pav.) [3]. Elektroninių laiškų dydis leidžia daryti prielaidą, kad laiško turiniui nebėra skiriama daug dėmesio, bandoma vartotoją suklaidinti arba sudominti, kad būtų paspaudžiamos nuorodos. Dalis laiškų taip pat yra koduojami „Base64“ arba „Uuencode“ algoritmais dažniausia naudojamais dvejetainiams duomenims el. pašto laiškuose koduoti, t.y. perduodant

prisegtas ne tekstines bylas, kurios, kaip žinia, gali būti kenksmingo kodo šaltinis. Taip pat taip elgiamasi bandant apeiti laiško turinio apdorojimo filtrus. Apibendrinant – nepageidaujamuose el. pašto laiškuose bandoma patraukti gavėjo dėmesį priverčiant jį paspausti ant nuorodų arba įkelianti nutolusį turinį iš tinkle esančios tarnybinės stoties, tai leidžia įsitikinti, kad vartotojas tikrina el. pašto dėžutę ir yra aktyvus, iš kokios geografinės vietovės yra ir kokią operacinės sistemos kalbą naudoja. Patvirtintų – aktyvių vartotojų el. pašto adresų duomenų bazės juodojoje rinkoje yra žymiai geriau vertinamos.



3 pav. Nepageidautino laiško dydis (2015)

Yra pastebėta, kad elektroninio brukalo akcijos yra pradedamos po įvairių stichinių nelaimių ir tragedijų, ypač didelio masto ir kai pasaulinės organizacijos pradeda rinkti aukas šiems regionams paremti. Siunčiant tokio tipo laiškus yra klastojami siuntėjų adresai, patys laiškai atrodo lyg būtų siunčiami raudonojo kryžiaus ar panašios organizacijos su nuoroda į tinklalapį, kuriame galima paaukoti pinigų. Tokiu būdu yra daroma tiesioginė finansinė žala pasinaudojant elektroninio pašto kanalais ir situacija kaip įrankiais vykdant nusikalstamą sukčiavimą.

Daugiamilijoniniai nuostoliai patiriami dėl komunikacijos kanalų perkrovos, kurią prižiūrėti reikia vis daugiau aukštos kvalifikacijos specialistų. Vartotojų ar darbuotojų sugaišto laiko, kai žmogus vietoje 10-30 elektroninių laiškų gali susidurti su 160-180 nepageidaujamų žinučių ir užtrunka 5-6 valandas per mėnesį vien tik trindamas šiuos laiškus. Sistemos vartotojų emocinis nepasitenkinimas, išblaškymas ir svarbių laiškų netyčiniai ištrynimai – tai tik finansinė žalos dalis, kurią sukelia brukalo siuntėjai.

Nepageidaujamas idėjinis turinys taip pat yra pavojingas. Yra pranešama apie atvejus kai pažeidžiamos visuomenės grupės, pvz.: vaikai nukenčia nuo elektroninių nusikaltėlių naudojančių elektroninio pašto kanalus skleisti vaizdinei ir tekstinei medžiagai, kuri paprastai neturėtų būti pasiekiama. Religijos protegavimas ar kitos religijos asmenų grupių užgauliojimas, politinių pažiūrų skleidimas – pasitaikantis reiškinys.

Darbo tikslas ir uždaviniai

Šio magistrinio projekto tikslas – sukurti, realizuoti ir praktiškai ištestuoti modelį skirtą įvairių pašto serverio apsaugos priemonių efektyvumui įvertinti. Užsibrėžtam tikslui pasiekti turėsime įgyvendinti šias užduotis:

- Išanalizuoti egzistuojančius el. pašto laiškų filtravimo metodų efektyvumą.
- Sukurti el. pašto tarnybos, filtruojančios nepageidaujamus el. pašto laiškus modelį.
- Ištirti modelio charakteristikas lyginant jį su realiame pašto serveryje vykstančiais procesais.
- Pritaikyti sukurtą modelį efektyvesnėms el. laiškų filtravimo sistemoms kurti.

1. PROBLEMINĖS SRITIES ANALIZĖ

Metodo, visiškai apsaugančio nuo nepageidaujamų elektroninių laiškų, nėra. Skirtingi metodai turi skirtingą efektyvumą ir saugo nuo skirtingų tipų atakų. Yra bandoma įvairaus tipo filtrus naudoti įvairiu eiliškumu siekiant atrasti aukso vidurį arba toleruotiną nepageidaujamų laiškų identifikavimo lygį atsižvelgiant į organizacijos užsibrėžtą ribą. Nepageidaujamų laiškų atmetimo metodo pritaikymas yra besitęsiantis konfigūracijos ir įvairių metodų žongliravimas bandant prisitaikyti prie dinamiškai kintančių internetinių šiukšlių formų, esamos situacijos ir tendencijų. Manipuliacija konfigūracija neretai priveda prie klaidingai laiško priėmimo arba klaidingo laiško atmetimo, ypač su „nulinės dienos“ (angl. *zero day*) – naujomis grėsmėmis. „Nulinės dienos“ grėsmės, tai tokios grėsmės kurios dar plačiai nežinomos ir prieš kurias nėra sukurta kovos priemonių. Galima daryti prielaidą, kad aukojant daugiau sisteminių resursų (finansinių resursų) galima pasiekti aukštesnį efektyvumą.

Šiame darbo skyriuje bus apžvelgti ir palyginti populiariausi el. pašto laiškų filtravimo metodai.

1.1. Analizės tikslas

SMTP (trumpinys nuo angl. *Simple Mail Transfer Protocol* – paprastas pašto perdavimo protokolas) yra *de facto* standartas el. pašto laiškams perduoti internete. Naudojamas elektroniniams laiškams pristatyti į gavėjo el. pašto dėžutę. SMTP yra 7 lygio pagal OSI modelį protokolas ir naudoja TCP transporto protokolo prievadus 25 arba 587. Jeigu serveris palaiko “saugaus perdavimo protokolą” (SMTP SSL) jungtis gali būti 465. 1995 metais IETF publikavo RFC 1869, kuris apibrėžė išplėstinį SMTP protokolą – ESMTP.

SMTP protokolas yra santykinai paprastas. Siunčiančioji pusė (klientas) duoda tekstines komandas (4 pav.), gaunančioji (serveris) apie komandų vykdymo rezultatus praneša grąžindama „klaidų (būsenos) kodus“. Originali SMTP protokolo versija reikalavo, kad tiek komandos, tiek laiško turinys būtų šifruojamas ASCII koduote. Dėl to buvo sudėtinga siųsti prisegtus skaitmeninius failus. Šiai problemai spręsti buvo sugalvotas MIME formatas arba 8BITMIME išplėtimas. Standartinė SMTP komandų seka susideda iš trijų komandų ir atsakymų į jas:

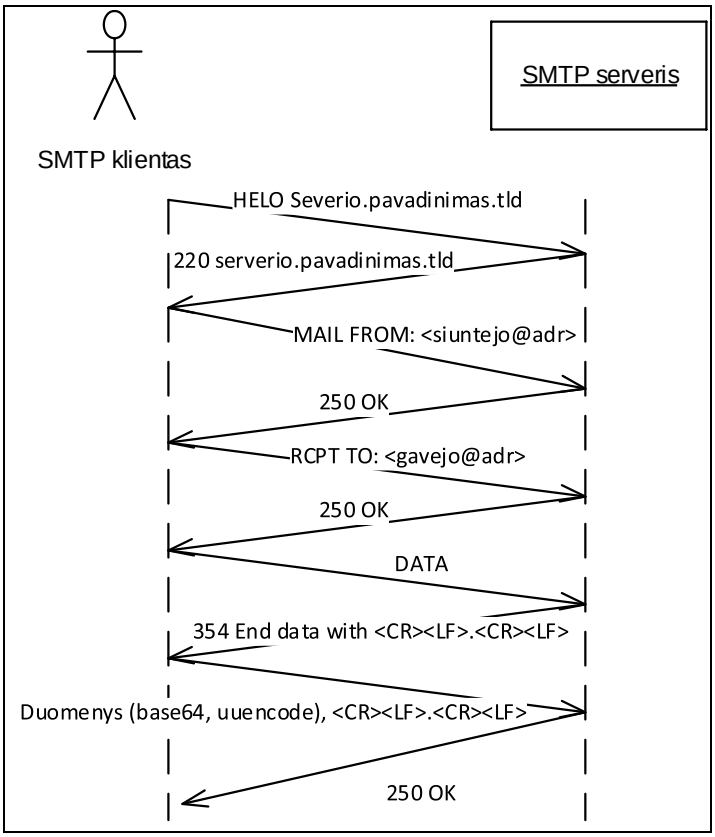
1) *MAIL* komanda skirta grįžtamajam keliui nurodyti, jame nurodomas siuntėjo adresas.
2) *RCPT* komanda skirta žinutės gavėjui nurodyti, gali būti kartojama keletą kartų norint nurodyti keletą gavėjų, šie adresai taip pat būna įdedami į žinutės „voką“.
3) *DATA* praneša pašto tarnybai, kad bus pradėtas perduoti žinutės tekstas ir antraštės.

SMTP aprašo žinutės perdavimą, bet ne žinutės turinį, todėl SMTP reikšmės yra įrašomos į žinutės „voką“, bet ne antraštes. Anksčiau minėtos komandos yra aprašytos RFC-821, tuo tarpu žinutės „vokas“ – RFC-822. Dažnai literatūroje yra minima RFC821 ir RFC822 parametrai, todėl turime suprasti, kad 821 yra komunikacijos lygmens aparatiniai nustatymai, 822 – žinutės turinyje esantys parametrai.

SMTP protokolas aprašo standartinius tarnybinės stoties atsakymus į komandas, jie gali būti arba teigiami (2xx eilės atsakymai) arba neigiami. Neigiami atsakymai gali būti laikini arba griežti atitinkamai 4xx eilės ir 5xx eilės (1 lentelė). Klientas gavęs atmetimo atsakymą, pvz.: 550, neturėtų bandyti pristatyti vėliau ir nusiųsti siuntėjui atitinkamą pranešimą apie nepavykusį pristatymą. Priešingai, gavęs 450 laikino atmetimo atsakymą, klientas turi bandyti pristatyti vėliau. Pakartotinių pristatymo bandymų intervalas priklauso nuo besijungiančiosios pusės nustatymų, bet paprastai bandoma pristatyti kartojant po 15, 30, 60, 240 minučių iki žinutės galiojimo pasibaigimo, kuris gali būti nustatomas savaitės laikotarpiui.

SMTP tarnybinės stotys pagal paskirtį gali būti skirtomis į tas kurios priima paštą organizacijos vartotojams, ir į tas, kurios skirtos transportui. Pastarosios dažnai naudojamos interneto paslaugų tiekėjų ir leidžią jų vartotojų IP adresų rėžiams pateikti el. pašto žinutes skirtas gavėjams internete. Tarnybinės stotys turi būti sukonfigūruotos taip, kad vienaip ar kitaip neautorizuoti

vartotojai negalėtų jomis naudotis, autorizacija gali būti pagrįsta sudarant prieigos kontrolės sąrašus pagal IP adresus arba panaudojant SMTP plėtinį įgalinantį autentifikavimą pagal vartotojo vardą ir slaptažodį.



4 pav. SMTP komunikacija

Elektroninį paštą priimančios tarnybinės stotys yra pažeidžiamos dėl savo pagrindinės paskirties – būti prieinamoms nenumatytiems gavėjams. Ar laiškai bus priimti, ar atmesti, priklauso nuo naudojamų filtravimo metodų, bet šis prieinamumas per internetą sukuria atakos vektorių, kurį naudojant galima paveikti paslaugos kokybę. Analizės tikslas yra išsiaiškinti populiariausius elektroninį paštą priimančių tarnybinių stočių saugos metodus.

1 lentelė. SMTP tarnybos atsakymų kodai [4]

Kodas	**Reikšmė**
211	Sistemos būklė arba atsakymas į pagalbos žinutę
214	Pagalbos žinutė
220	Tarnyba pasiruošusi (nurodo ir vardo srities adresą)
221	Tarnyba uždaro perdavimo kanalą (nurodo ir vardo srities adresą)
250	Teigiamas atsakymas į komandą
251	Nepavyko nustatyti ar vartotojas egzistuoja, bet bus bandoma pristatyti (atsakymas į VRFY komandą)
354	Leidžia pradėti perdavinėti žinutės duomenis
421	Paslauga negalima, uždaromas perdavimo kanalas
450	Nepavyko atlikti komandos, dėžutė neprieinama
451	Nepavyko apdoroti, vidinės tarnybos problemos
452	Nepavyko, trūksta vietos duomenų saugykloje
500	Sintaksės klaida, nežinoma komanda
501	Sintakses klaida, nurodant parametrus ar argumentus
502	Komanda negalioja
503	Blogas komandų eiliškumas
504	Komandos parametras negalimas

521	Vardų sritis nepriima el. pašto žinučių
530	Priėjimas uždraustas
550	Veiksmas neįvykdytas, pašto dėžutė neprieinama
551	Nurodo kur persiusti žinutę, jei vartotojas ne vietinis
552	Veiksmas neįvykdytas, perpildytas duomenų saugyklos vieta
553	Veiksmas neįvykdytas, dėžutės pavadinimas neleistinas
554	Transakcijos klaida

1.1.1. ESMTP plėtiniai

ESMTP – išplečia standartinio SMTP protokolo funkcijas, bet jų nepakeičia. Identifikacinis ESMTP tarnybinės stoties atsakymas yra pranešimas apie palaikomus plėtinius klientui prisistačius raktažodžiu EHLO, vietoje HELO (SMTP). Klientui prisistačius „EHLO kliento.adresas“ komanda, ESMTP palaikanti tarnybinė stotis turi atsakyti „250 OK“. Kiekvienas plėtinys turi atskirą registruotą aprašymą RFC. Populiariausios plėtinių komandos „MAIL“ ir „RCPT“ plėtiniai aprašyti 2 lentelėje.

2 lentelė. Populiarūs ESMTP plėtinių raktažodžiai

Raktažodis	Aprašymas	RFC
8BITMIME	8 bitų duomenų perdavimas	RFC-6152
TURN	El. pašto perdavimas pagal pareikalavimą	RFC-2645
ATRN	El. pašto perdavimas pagal pareikalavimą su autentifikacija	RFC-2645
AUTH	SMTP su kliento autentifikacija	RFC-4954
CHUNKING	Didelių žinučių perdavimas išskaidžius porcijomis	RFC-3030
DSN	Pranešimas apie pristatymo būklę	RFC-3461
ETRN	Nutolusios žinučių eilės valdymo komanda	RFC-1985
HELP	Pagalbinės informacijos pateikimas	RFC-821
PIPELINING	SMTP komandų apjungimas	RFC-2920
SIZE	SIZE žinutės dydžio paskelbimas	RFC-1870
STARTTLS	Transporto lygio sauga	RFC-3207 (2002)
SMTPUTF8	Leisti UTF-8 kodavimą el. pašto dėžučių pavadinimuose ir antraštėse	RFC-6531

SMTP protokolas pats neužtikrina konfidencialumo, todėl, kaip ir gerai žinomas ir dažnai naudojamas WWW standarto protokolas – HTTPS, naudoja TLS (angl. *transport layer security*) protokolą konfidencialumui užtikrinti. TLS gali būti iškviečiamas dviem būdais: arba jungiantis į tam skirtą saugaus ryšio prievadą 465 (SMTPS – angl. *SMTP – secure*) kuomet TLS derinimas pradedamas iš karto, arba prisijungus prie nesaugaus prievado ir (jei tarnybinė stotis palaiko saugų ryšį) inicijuojant STARTTLS komandą, kuri perjungia komunikavimo režimą į saugų.

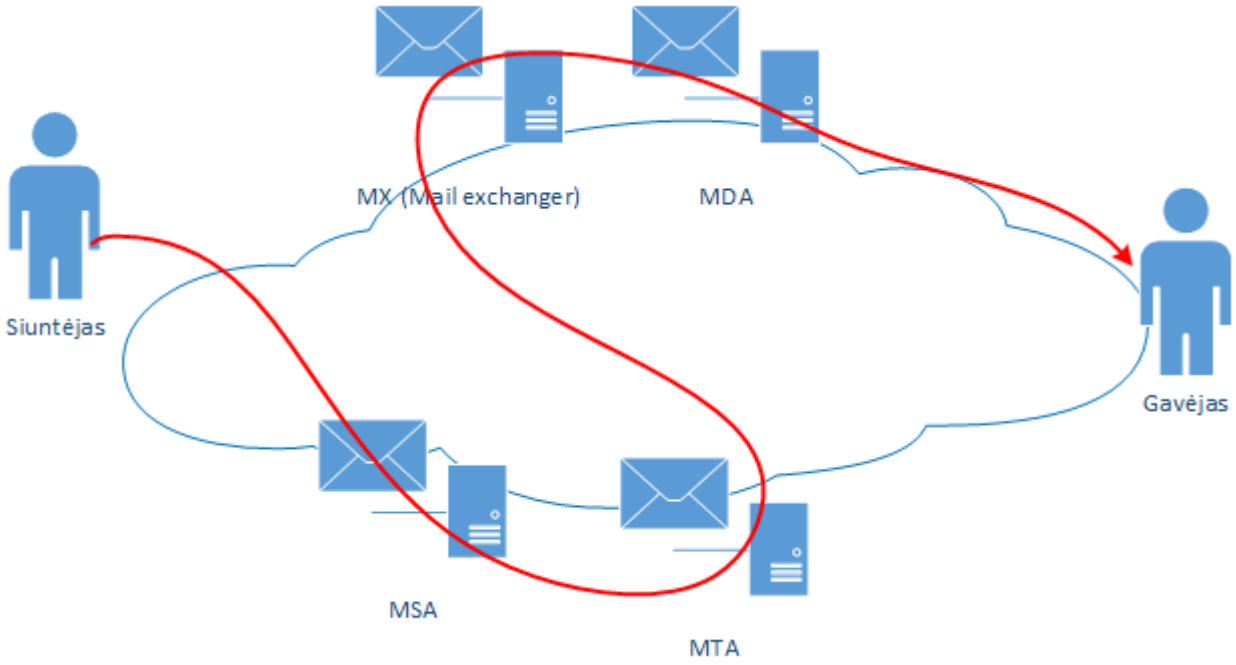
1.1.2. El. pašto infrastruktūra

Elektroninio pašto žinutės yra išsiunčiamos iš el. pašto kliento (MUA – angl. *mail user agent*) ir pateikiamos tarnybinei stočiai (MSA – angl. *mail submission agent*) naudojant SMTP protokolą ir 587 prievadą (5 pav.). Kai kurie paslaugų tiekėjai vis dar leidžia žinučių perdavimą 25 prievadu, tačiau didžioji dauguma Europos interneto paslaugų tiekėjų blokuoja 25 prievadą namų vartotojams ir tokiu būdu kovoja su nepageidaujamų laiškų srautais. Kai naudojami „zombiais“ paversti kompiuteriai ir jų tinklai tai – efektyvi priemonė. Iš MSA laiškai yra perduodami MTA (angl. *mail transfer agent*) – paslaugai, skirtai žinutės perdavimui tinklu. Dažnai MSA ir MTA veikia toje pačioje tarnybinėje stotyje, tai priklauso nuo konfigūracijos ypatumų.

Kraštinis MTA, taip vadinamas, nes dažnu atveju turi galimybę naudotis DNS (angl. *domain name service*) ir gali surasti el. pašto gavėjo tarnybines stoties serverio adresą, kuris pagal susitarimą yra skelbiamas DNS MX tipo įrašuose. Jeigu MX įrašas nerastas – bandoma pristatyti tarnybinei stočiai, aprašytai pagrindiniame vardų srities (po „@”) DNS A tipo įraše.

DNS MX įraše nurodoma tarnybinė stotis atlieka pašto apsikeitimo paslaugos rolę (angl. *mail exchanger*) ir šio etapo metu vyksta gavėjo adreso verifikavimas – tikrinimas ar tokia elektroninio pašto dėžutė egzistuoja organizacijos el. pašto sistemoje. Pašto apsikeitimo tarnyba perduoda žinutę pašto pristatymo tarnybai, kuri išsaugo žinutę reikiamu formatu. Šios rolės gali būti apjungtos ir abi tarnybos veikti vienoje tarnybinėje stotyje arba išskirtos į atskiras. Išskaidyta infrastruktūra yra lankstesnė tolimesnei plėtrai, tačiau brangesnė.

Žinutės patalpinimu į dėžutę SMTP protokolo veikimo ribos baigiasi, nuo šio taško vartotojas jungiasi į tarnybas, kurios yra atsakingos už elektroninio pašto atidavimą vartotojui.



5 pav. El. pašto apdorojimo modelis

1.2. El. pašto laiškų filtravimo metodai

Atsižvelgiant į el. laiškų struktūrą ir juose esančius metaduomenis kuriami atitinkami filtrai. Vieni iš jų analizuoja laiško „antraštę“, vertinant žinutės perdavime dalyvavusių serverių IP adresus, šių adresų reputaciją, priklausomybę, tipą, jų skaičių. Analizės metu dažnai naudojamos atvirkštines DNS užklausos, vertinami jų rezultatai.

Laiško turinys yra daugiausia informacijos turinti elektroninės žinutės dalis, nes jame yra informacija kurią tiesiogiai mato vartotojas ir kuri vartotojo atžvilgiu gali būti nepageidautina ar žalinga. Turinio analizei pritaikomi filtrai gali tikrinti dažnai pasitaikančias frazes, tekste esančias nuorodas, tam tikrai organizacijai būdingus žodžius, frazes ir panašiai.

Egzistuoja technologinės galimybės užtikrinti siuntėjo arba siunčiamo laiško kilmę todėl turi būti ir mechanizmas patikrinti ar laiške yra panaudotos šios priemonės ar ne. Tai – dar vienas filtravimo būdas.

Šie ir kiti filtravimo metodai bus išsamiau apžvelgti šiame skyriuje.

1.2.1. Kliento IP adreso reputacijos patikra realaus laiko juoduosiuose sąrašuose

Interneto juodieji sąrašai (angl. *DNS Black Lists*) – nepriklausomų organizacijų administruojamos, realaus laiko duomenų bazės, kuriose saugomi probleminių serverių IP adresai. Šie sąrašai plačiai naudojami laiškų filtravimui, juose patikrinama, ar nėra siuntėjo serveris arba domenas susijęs su probleminiais IP, o jei yra – laiškas atmetamas arba patalpinamas „šlamšto“ pašto dėžutėje. Patekimo į DNSBL priežastys būna įvairios: serverių saugumo spragos, nepageidaujamų laiškų siuntimas ir kt. Jei domeną aptarnaujantis serveris patenka į DNSBL, pirmosios problemos pasireiškia tuo, kad laiškai iš to domeno nebepasiekia dalies adresatų. Dauguma el. pašto MX

tarnybų gali būti nustatytos naudoti šiuos sąrašus ir atmesti prisijungimus dar net nepriėmus, nepradėjus apdoroti laiško.

DNSBL veikia OSI 7 lygio DNS – vardų srities paslaugos protokolu, kuris, savo ruožtu naudoja UDP (angl. – *user datagram protocol*) kaip transporto protokolą. Vidutinis DNS užklausos dydis sudaro apie 200 baitų, o turint omenyje, kad tai UDP protokolas, kuris turi labai mažą pertekliškumą, nes neužtikrina perduoto paketo pristatymo – DNSBL gali būti naudojamas su labai nedideliu pertekliumi kiekvieno IP adreso reputacijai patikrinti. Greitai gavus atsakymą galima spręsti ar prisijungimą atmesti, ar priimti. Užklausa susideda iš trijų pagrindinių žingsnių:

1) Įeinančio prisijungimo IP adresas paverčiamas atbuline oktetų tvarka, pvz.: 192.168.0.1 tampa 1.0.168.192
2) Apverstas IP adreso oktetų rinkinys pridedamas prie DNSBL tiekėjo vardų zonos dalies, pvz.: 1.0.168.192.dnsbl.pazyzdys.lt.
3) Vykdoma DNS A tipo įrašo užklausa, atsakanti tarnybinė stotis gražina arba atsakymą (127.0.0.2), arba NXDOMAIN, atsakymas nurodo ar IP adresas yra juodajame sąraše ar ne.
4) Šis žingsnis yra neprivalomas, tačiau dauguma juodųjų sąrašų paslaugų atsako ir į DNS TXT tipo įrašų užklausas apie adresus aptiktus tračiajame žingsnyje, o atsakymuose pateikiama informacija apie įtraukimo į sąrašą priežastį.

Nors realaus laiko juodųjų sąrašų veikimo principo mažas pertekliškumas ir greitis gali atrodyti kaip labai geras metodas priešakinei nepageidaujamų laiškų filtravimo linijai, jis vis dažniau sulaukia kritikos iš informacinių technologijų saugumo bendruomenės. Visų pirma, naudojant paslaugą, reikia vienareikšmiškai pasitikėti organizacija, kuri ją teikia, reikia būti susipažinus su jos paslaugų teikimo politika. Didžiulė rizika, kad tokios organizacijos gali tapti kibernetinių atakų taikiniu, todėl svarbu atitinkamai konfigūruoti savo serverius situacijai, kai nebėra atsakymo iš nutolusios šalies. Šios organizacijos gali būti atakuojamos norint sukompromituoti trečiąsias, niekaip nesusijusias šalis, pvz.: imituojama engėjiška veikla „ktu.lt“ pašto dėžučių būdu, kuris pagal DNSBL paslaugos tiekėjo politiką atitinka kriterijus, pagal kuriuos ktu.lt MTA IP adresas įtraukiamas į šiuos sąrašus. Tokiu būdu pasiekiama, kad sistemos, naudojančios DNSBL atmetinėtų el. pašto žinutes iš „ktu.lt“ ir tokiu būdu kenkti šiai trečiajai šaliai. Yra dalis interneto paslaugų tiekėjų, kurie neleidžia vartotojams tiesiogiai perduoti el. pašto srauto, reikalaudami, kad jie naudotųsi tarpiniu tiekėjo serveriu, būtent šio serverio IP adresas gali greitai pakliūti į juoduosius sąrašus dėl vieno ar kelių kenkėjišku kodu apkrėstų kompiuterių. Dinaminių IP adresų rėžiai taip pat dažniai įtraukiami į sąrašą. Tokiais būdais varžoma vartotojų teisė tiesiogiai keistis el. pašto žinutėmis nekontaktuojant tarpinio serverio, kuriame laiškai gali būti archyvuojami, skaitomi ar kitaip pažeidžiamas komunikacijos konfidencialumas.

1.2.2. El. pašto žinutės antraščių apdorojimas

Elektroninio laiško byla susideda iš antraštės ir turinio, kuris gali būti išskaidytas į MIME (angl. *Multi-Purpose Internet Mail Extensions*) dalis. Atskiros laiško dalys gali būti koduotos UUENCODE arba Base64 algoritmais. Šioje darbo dalyje nagrinėjame SMTP protokole aprašomų antraščių dalį, kurioje saugoma informacija apie siuntėją, gavėją, datą, tarpinius serverius, papildomą informaciją. Laiško antraštėje metaduomenys įrašomi iš visų el. pašto apdorojimo tarpinių komponentų (5 pav.) todėl yra kritiškai svarbi. El. laiško siuntimo klientinė programinė įranga prideda duomenis apie siuntėją, gavėja, laiško išsiuntimo datą, žinutės identifikacinius duomenis, kodavimo tipą ir panašiai. Tarpiniai perdavimo serveriai pildo antraštę pridėdami gavimo į konkretų serverį laiką, siunčiančio ir gaunančio serverių IP adresus. Iš šių duomenų galima daryti reikšmingus sprendimus klasifikuojant gaunamą paštą.

1.2.3. DKIM

Naujo elektroninio pašto instrumento specifikacijos, įvestos 2007 metais, Internet Engineering Task Force (IETF) organizacijos, ir padeda pagerinti apsaugą nuo žvejybos (angl. *phishing*) ir elektroninių šiukšlių (SPAM).

Technologija DKIM (angl. *DomainKeys Identified Mail*) apjungia kelis egzistuojančius „antižvejybos“ ir „antispamo“ metodus tam, kad būtų pagerinta „gerų“ elektroninių žinučių klasifikacijos ir identifikacijos schema. Be tradicinio IP adreso, siuntėjo identifikacijai DKIM prideda prie laiško skaitmeninį parašą, susijusį su organizacijos domeno vardu. Šis parašas automatiškai patikrinamas gavėjo pusėje, po to, siuntėjo reputacijos nustatymui, taikomi baltieji ir juodieji sąrašai ar kiti metodai.

„DomainKeys“ esmė yra tai, kad siuntėjų autentifikavimas vyksta ne pagal IP adresus, bet pagal domenų vardus, kadangi, kūrėjų manymu, domeno vardai yra kur kas stabilesni ir labiau siejasi su organizacija, nei IP adresas. Kitas DKIM bruožas yra tai, kad viešųjų raktų perdavimui yra naudojama egzistuojanti DNS sistema, todėl nereikia kurti papildomo infrastruktūros lygio viešųjų raktų skelbimui. DKIM privalumas yra ir tai, kad visas apdorojimas vyksta automatiškai, t.y. vartotojui nereikia pačiam nuspręsti priimti ar atmesti atėjusią žinutę nes tai įvykdoma tarnybinėje stotyje.

„DomainKeys“ projektą pradėjo kompanija „Yahoo“, o „Identified Internet Mail“ inicijavo „Cisco Systems“. Prie šių projektų vėliau prisidėjo ir kitos stambios kompanijos ir jų neformali grupė maždaug metus kūrė naująją DKIM specifikaciją. 2005 metų liepą, ši specifikacija buvo perduota IETF grupei tam, kad būtų parengtas naujas kovos su elektroninio pašto grėsmėmis standartas.

DKIM schema susideda iš dviejų dalių: parašo, pridedamo prie el. pašto žinutės, ir DNS TXT tipo įrašo. Parašo specifikacija parodyta žemiau:

```
DKIM-Signature: v=1; a=rsa-sha256; d=example.net; s=brisbane;
c=relaxed/simple; q=dns/txt; l=1234; t=1117574938; x=1118006938;
h=from:to:subject:date:keywords:keywords;
bh=MTIzNDU2Nzg5MDEyMzQ1Njc4OTAxMjM0NTY3ODkwMTI=;
b=dzdVyOfAKCdLXdJOc9G2q8LoXSlEniSbav+yuU4zGeeruD00lszZVoG4ZHRNiYzR
```

DKIM parašas sudaromas iš rinkinio parametrų ir reikšmių, kurias apžvelgsime 3 lentelėje.

3 lentelė. DKIM parašo parametrų reikšmės [5]

Raktas	**Reikšmė**	**Komentaras**
v	versija	kolkas - 1
a	parašo algoritmas	pvz.: rsa-sha1, rsa-sha256 (numatytasis). Kolkas naudojami tik šie su galimybe ateityje pasirinkti kitus, tačiau tiek pasirašanti, tiek tikrinanti šalys turi būti įsidiegusios susitartąjį algoritmą.
c	antraštės ir turinio kanonizavimo algoritmas	*simple antraštei* – nekeičia antraštės laukų. Antraštės laukai turi būti tokie patys, kokie ir buvo pasirašyti *relaxed antraštei* – antraštės laukų pavadinimai pertvarkomi paverčiant juos mažosiomis raidėmis. Visos WSP (angl. *whitespace*) simbolių sekos paverčiamos į vieną SP (angl. *space*) simbolį. Ištrina visus WSP simbolius antraštės laukų reikšmių pabaigose. Ištrina visus tarpo simbolius (WSP) esančius tarp dvitaškio simbolio, skiriančio lauko pavadinimą ir reikšmę - „pavadinimas:reikšmė“. *simple turiniui* – nekreipia dėmesio į tuščias linijas laiško pabaigoje. Jeigu laiško turinio pabaigoje nėra naujos eilutės simbolio – jis pridedamas, tačiau jei turinio išvis nėra, tai nereiškia, kad po formatavimo bus pridėtas naujos eilutės simbolis. *relaxed turiniui* – pašalina dėmesio į tarpo simbolius eilučių pabaigoje, bet nepašalina CRLF (naujos eilutės simbolis) iš eilutės pabaigos. Simbolio „tarpas“ sekos pakeičiamos į vieną tarpą. Pašalina visas tuščias eilutes turinio pabaigoje. Pažymima, kad šis algoritmas gali atverti galimybes tam tikroms ASCII atakoms, koreguojant tarpus tarp žodžių. nurodoma atskiriant „/“ simboliu, pvz.: „antr/turi“
d	pasirašantis domenas	pvz.: lapenas.lt
s	selektorius	naudojamas kai yra reikalavimas naudoti keletą raktų tam pačiam domenui

		pasirašyti, selektorius nurodo kurį DNS TCT tipo įrašą tikrinančioji šalis turi užklausti, kad gautų viešąjį raktą parašui patikrinti
q	užklausos tipas	naudojamas dns/text (tipas/rūšis)
l	kanonizuoto turinio ilgis	išreiškiama baitais
t	parašo laikas	išreiškiama sekundėms nuo 1970 01 01 00:00:00
x	parašo galiojimo laikas	išreiškiama sekundėms nuo 1970 01 01 00:00:00
h	pasirašytos antraštės laukai	nurodoma kurie antraštės laukai buvo pasirašyti
bh	laiško turinio maiša	„l“ parametre nurodyto ilgio kanonizuoto turinio maišos reikšmė (base64 koduota)
b	skaitmeninis parašas	DKIM parašas, konvertuotas base64 funkcija, duomenų perdavimo palengvinimui

4 lentelė. DKIM DNS įrašo struktūra

laukas._domainkey.lapenas.lt. IN TXT "k=rsa; t=s; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQC8U0xr9fAwXkYKEHkEvxk76e3YECsArh8PP4fnEbwPEb5T2Fv9cmRGJVHUhWQGO7O2bVToKiEq1wvfTLUeto4RHTSG+BiU7drWhx2K5V4mcWpKAJXzIezXgi6UFKxEvfVujQWl1+G5FzO2OsUoyisy5O3LfClFSlftU4m89b/UxQIDAQAB"

5 lentelė. DKIM DNS įrašo parametrų reikšmės

Parametras	Reikšmė	Komentaras
k	viešojo rakto tipas, paprastai „rsa“	šiuo metu kiti algoritmai nenaudojami
h	maišos algoritmas	sha1 arba sha254
p	viešasis raktas	base64 koduotas DER formato viešasis raktas
s	paslaugos tipas	Pagal nutylejimą reikšmė yra „*“ – visos paslaugos, kita galima reikšmė yra „email“, kuri nebūtinai apsiriboja SMTP protokolu
t	papildomi parametrai	„y“ – galioja bandomasis režimas

DKIM DNS TXT įrašas yra skirtas viešai publikuoti pasirašančiosios vardų srities viešąjį raktą ir papildomus parametrus, jis užtikrina neišsiginamumą grindžiant tuo, kad niekas kitas, tik šalis, kuri pasirašė, gali turėti prieigą prie DNS įrašų valdymo. DNS įrašas turi prasidėti „_domainkey”, pvz.: “_domainkey.lapenas.lt”. Selektorius gali būti naudojamas pridedant jį prie DKIM įrašo pavadinimo, pvz.: „selektorius._domainkey.lapenas.lt“, kita informacija nurodoma TXT įrašo viduje, kaip parodyta pavyzdyje 4 lentelėje.
Kadangi DKIM yra pagrįstas RSA algoritmu, kuris savo ruožtu remiasi dviejų didelių pirminių skaičių sandaugos problema todėl labai svarbu naudoti pakankamo dydžio raktus, šiai dienai 2048 bitai, manoma, yra saugus rakto dydis.

1.2.4. SPF

SPF (angl. *sender policy framework*) struktūra leidžia nustatyti siunčiamų laiškų siuntimo politiką ir nurodyti kokie serveriai turi teisę siųsti laiškus tam tikro domeno vardu. SPF yra DNS TXT tipo įrašas siuntėjo vardų srityje, kuriame nurodoma siuntimo politika: kas gali siųsti ir kaip elgtis, jei priimamas laiškas atėjo iš kito serverio, nei nurodytas SPF įraše.

"v=spf1 apibrėžimasmechanizmas +mx -all"

6 pav. Įrašo sandara

SPF aprašomi mechanizmai:

- *a* - leisti siųsti laišką, jei serverio IP adresas yra domeno "A" įrašas;

- *mx* - leisti siųsti laišką, jei serverio IP adresas yra srities "MX" įrašas;
- *ptr* - leisti siųsti laišką domeno vardu iš serverio IP adreso, kuris turi atvirkštinį DNS įrašą rodantį į siuntėjo vardų zoną, šis tipas laikomas nebegaliojančiu ir IRTF neberekomenduoja jo naudoti;
- *ip4* - nurodomas IP (4 versija) adresas, iš kurį galima siųsti laiškus;
- *ipv6* - nurodomas IP (6 versija) adresas, iš kurį galima siųsti laiškus;
- *all* – visi adresai, dažnai naudojamas uždrausti visus nepaminėtus mechanizmus;
- *exists* – nurodo, kad duotoji zona egzistuoja, t.y. išverčiama į bet kokį adresą (naudojama retai, DNSBL kontekste);
- *include* – nurodo, kad siuntėjo IP adresas turi tenkinti politiką nurodytą „include“ mechanizmo parametre, pvz.: „include:spf.lapenas.lt“;

po kiekvieno iš šių mechanizmų (išskyrus „all“) galima naudoti dvitaškio simbolį ir nurodyti papildoma vardų sritį, kurios atžvilgiu naudoti duotąjį mechanizmą, pvz.: *mx:ktu.lt,* bus tikrinama ar siunčiančio serverio IP adresas yra MX įrašas *ktu.lt* vardų srityje, nors pats SPF įrašas gali būti *lapenas.lt* zonoje.

Galimi apibrėžimai:

- „+“ pažymi leidimą. Paprastai „+“ nėra rašomas, nes +mx yra tas pats kas mx;
- „?“ nurodo neutralumą, t.y. neaprašo, kaip elgtis jei mechanizmas buvo patenkintas ar nepatenkintas;
- „~“ nurodo vidutinišką leidimą, jei mechanizmas buvo netenkintas laiškas neturėtų būti atmestas, tačiau gali būti pažymėtas;
- „-“ griežtas negatyvas, laiškas turi būti atmestas;

SPF įrašo pavyzdžiai:

„*lapenas.lt. IN TXT "v=spf1 ip4:8.8.8.8 a –all*“. Jei siunčiančio serverio IP (4 versija) adresas yra 8.8.8.8 arba *lapenas.lt* vardų srities A tipo įrašas – priimti, visus kitus adresus – atmesti

„*lapenas.lt. IN TXT "v=spf1 ~ipv4:8.8.8.8 mx ?all*“ jei siunčiančio serverio IP (4 versija) adresas yra 8.8.8.8 – priimti, bet pažymėti, *lapenas.lt* vardų srities MX tipo įrašas – priimti, visus kitus adresus laikyti neutraliais [6].

Atskiri serveriai gali turėti individualia politiką, kaip rūšiuoti mechanizmus, pvz.: praleisti tik tenkinančius arba atmesti tik griežtai nurodytus. Paprastai SPF įrašo patikrinimas yra atliekamas per tiek DNS tipo užklausų kiek yra „a“, „mx“, „ptr“ mechanizmų plius pagrindinės vardų srities TXT, tačiau, jei yra naudojamos įterptinės politikos („include“), tuomet užklausų skaičius gali išaugti neribotai.

Apibendrinant šį filtravimo metodą galima būtų pažymėti, kad jis yra gana efektyvus, nes leidžia aiškiai apibrėžti siuntėją ir užkirsti kelią siuntėjo adreso klastojimui. Metodas veikia tik tada kai priimančioji pusė savo saugumo politikoje yra numačiusi SPF tikrinimą, o siunčiančioji – įgyvendinusi SPF politiką DNS įrašuose. Neveikia, kai laiškas yra persiunčiamas. Pertekliškumas – nežymus dėl DNS paslaugos specifikos.

1.2.5. DMARC

DMARC (angl. *Message authentication, reporting and Conformance*) yra dar vienas el. pašto autentifikavimo mechanizmas skirtas kovoti su neteisėtai siunčiamais laiškais, ypač klastojant siuntėjo adresą ir taip apgaunant gavėją. DMARC buvo aprašytas 2011 metais, išplečiant jau egzistuojančius SPF ir DKIM. Šiame darbe išsiaiškinta, kad DKIM yra specifikacija, nurodanti kaip reiktų pasirašyti ir patikrinti siunčiamas žinutes, tačiau ji nenurodo kaip elgtis priimančiajam serveriui, jei parašo patikrinimas nepavyko, ar atmesti, ar pažymėti, ar priimti. SPF, savo ruožtu, turi metodą nurodantį šį apsisprendimą. DMARC – sujungia SPF ir DKIM, įveda politiką, kurios atžvilgiu gaunantieji serveriai turėtų priimti sprendimus, taip pat sukuria galimybę gauti ataskaitas iš trečiųjų šalių apie tam tikros vardų srities vardu laiškus siunčiančius serverius.

Pagrindinės problemos dėl kurių DMARC buvo aprašytas – daugybė siuntėjų turinčių sudėtingas el. pašto siuntimo sistemas, dažnai už juos laiškus siunčia ir trečiosios šalys todėl

užtikrinti, kad kiekviena žinutė yra autentifikuota SPF ir DKIM, tampa sudėtinga. Jeigu vardų srities savininkas siunčia ir pasirašytas, atitinkančias SPF nurodymus žinutes, ir žinutes, kurios nėra pasirašytos – pasidaro labai sunku atskirti, kurios nepasirašytos žinutės ateina iš tikrojo siuntėjo, o kurios yra suklastotos ir būna atmetamos patikros metu. Siuntėjams yra sunku sužinoti arba „sugaudyti“ kurie laiškai, siunčiami jų arba trečiųjų šalių sistemų ir netenkina tiek SPF tiek DKIM patikrinimų, paprastas pavyzdys: iš *ktu.lt* el.pašto sistemos laiškai tiek pasirašyti, tiek atitinka SPF tikrinimą, o trečiosios šalies siunčiami, tarkime „DreamSpark“ – netenkina nei SPF, nei pasirašyti DKIM. Apie tokį atvejį *ktu.lt* el. pašto sistemos administratorius gali net nežinoti, o laiškai iš „DreamSpark“ gali būti labai svarbūs, tačiau yra atmetami. Netgi jeigu siuntėjas sugeba surinkti visas sistemas arba unifikuoti pašto siuntimą ir siųsti visus laiškus iš vienos vietos, kad visos žinutės tenkintų abiejų mechanizmų tikrinimus, gaunančioji pusė apie tai nežino ir gali paprasčiausiai bijoti atmetinėti neautentiškų žinučių, galvodama, kad gali atmesti reikalingą komunikaciją.

Būdas, spręsti aukščiau išvardintas problemas, yra siuntėjui keistis informacija su gavėju. Gavėjai pateikia informaciją apie jiems „matomą“ siuntėjų laiškų autentiškumą, o siuntėjai „pasako“ gavėjams ką daryti kai gauta žinutė netenkina mechanizmų patikros kriterijų. Pirmieji pabandę šį metodą 2007 metais buvo „PayPal“ ir „Yahoo“, vėliau ir „Google“. Rezultatai buvo stebinantys, apgaulingų laiškų nuo „Paypal“ „Google“ ir „Yahoo“ pašto sistemų vartotojai beveik nebegaudavo.

DMARC yra suprojektuotas, kad būti įdiegtas į jau veikiančias gaunamo pašto sistemas be didelių pakeitimų. Jis veikia taip, kad padeda pašto gavėjams nustatyti ar žinutė tenkina patikros algoritmus, o jei ne, DMARC naudojamas nurodyti ką daryti su žinute. 7 paveiksle pateikiamas žinutės tikrinimo pavyzdys. Pagrindiniai tikslai:

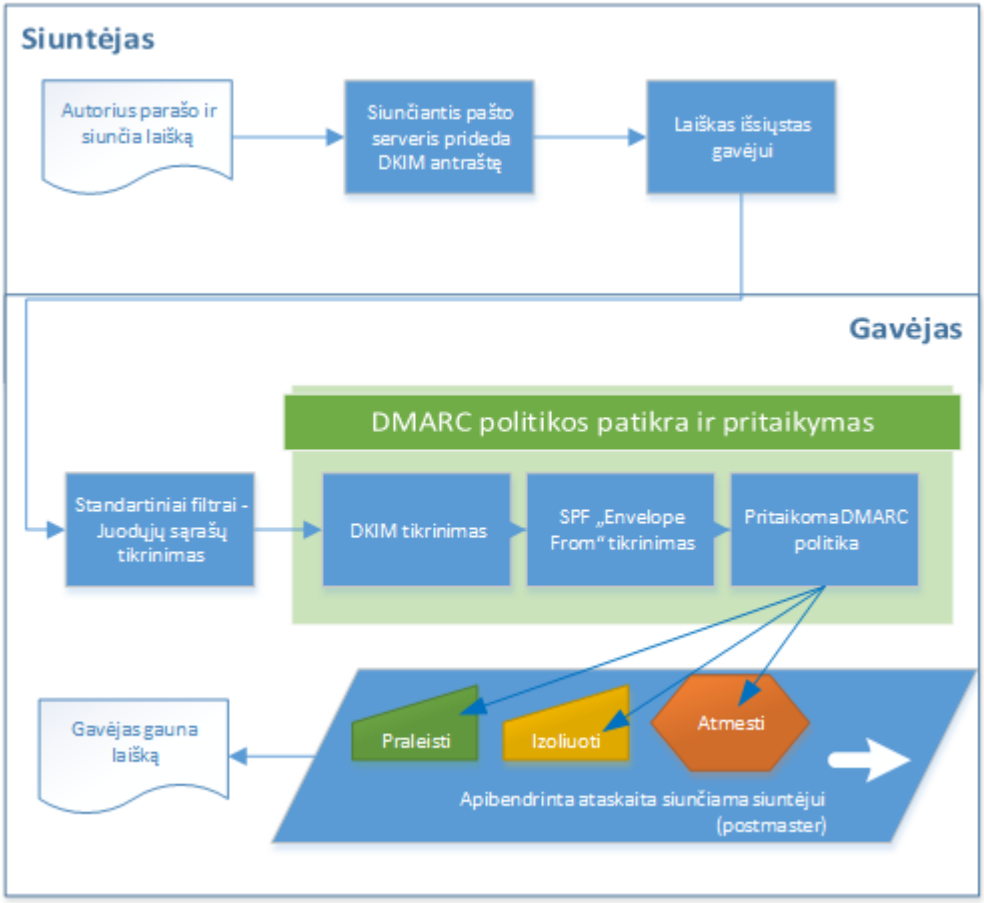
- Sumažinti klaidingai teigiamus filtravimo sprendimus
- Teikti autentiškumo ataskaitas
- Pritaikyti siuntėjo politiką gaunant laiškus
- Sumažinti žvejojimo (angl. *phishing*) siuntėjo klastojimo atakų efektyvumą
- Veikti globalaus interneto lygiu
- Minimizuoti sudėtingumą

DMARC politiką siuntėjas viešina tokiu pačiu būdu kaip ir SPF arba DKIM, manant, kad DNS įrašus zonai modifikuoti gali tik zonos šeimininkas,ir tokiu būtu apsaugant šios politikos vientisumą. DNS TXT tipo įrašas susideda iš kabliataškiu atskirtų parametrų ir reikšmių porų pvz.: „*v=DMARC1;p=reject;pct=100;rua=mailto:postmaster@dmarcdomain.com*“.

Gavėjas, savo elektroninio pašto filtravimo sistemoje, įsidiegęs DMARC tikrinimo įrankius, kaip papildomą funkcionalumą SPF ir DKIM tikrinimui, priėmus laišką įvykdo atitinkamus patikrinimus ir papildomai daro DNS užklausas bandant gauti DMARC politiką. Gavus teigiamą atsakymą, kad DMARC politika egzistuoja – vykdo siuntėjo sistemos administratorių aprašytus politikos veiksmus: atmeta arba priima laiškus, kurie nebuvai tinkamai autentifikuoti SPF arba DKIM. Pašto sistemų administratoriai diegiantys DMARC politiką, gali nurodyti kokią procentinę dalį laiškų, netenkinančių algoritmo patikros, atmesti. Tokiu būdu išvengiant galimų paslaugos sutrikimų jei pvz.: nežinoma apie trečiąja šalį siunčiančią laiškus organizacijos vardu. Bet kuriuo atveju, politikoje nurodytu el. pašto adresu bus siunčiama ataskaita apie atmestas žinutes, pagal tai administratoriai galės nustatyti „klaidingai teisingus“ atvejus.

Priklausomai nuo DMARC politiką įsidiegusių siuntėjų skaičiaus ir laiškus siunčiančių serverių kiekio, gavėjo sistemoms gali tekti perteklinias krūvis kaupti duomenis apie kiekvienos DMARC politiką įsidiegusios organizacijos legalius ir nelegalius SMTP serverius, periodiškai generuoti, archyvuoti bei išsiųsti ataskaitas. Gaunamos ataskaitos yra XML aprašytos XML kalba, kuri nėra patogi sistemų administratoriui perskaityti, todėl šis procesas turi būti automatizuotas ir pranešti sistemą prižiūrinčiam personalui suprantamu formatu.

Dėl šių priežasčių tik didžiausi el. pašto paslaugų tiekėjai aktyviai naudoja ir pilnai įdiegia šios papildomos el. pašto autentifikavimo priemonės saugumo komponentus.



7 pav. DMARC veikimo schema

DMARC politiką galima naudoti ir „egoistiškai“, DNS įraše nurodžius politiką, bet pačiam neįgyvendinus protokolo plačiąja prasme, tokiu atveju, yra gaunamos apibendrintos ataskaitos apie vardų sritį ir situaciją apie siuntėjus, siunčiančius nurodytos srities vardu. Protokolo pertekliškumas yra mažas, o perėjimas prie griežtos politikos gali būti atliekamas etapais, nes protokolas leidžia nurodyti žinučių, kurioms bus taikoma politika kiekį. Pradžioje rekomenduojama naudoti „none“ politiką, 100 proc. žinučių, tokiu būdu siunčiant instrukciją gavėjams pradėti rinkti duomenis ir siųsti ataskaitas. Gaunama informacija leidžia daryti išvadas ir planuoti.

6 lentelė. DMARC DNS įrašo parametrai ir jų reikšmės [7]

Parametras	**Reikšmė**	**Komentaras**
v	Protokolo versija	Kolkas - 1
pct	Žinučių procentas	Kokiai žinučių daliai bus taikoma politika
ruf	Ataskaitų apie atmestas žinutes siuntimo adresas	El. pašto adresas, kur siusti ataskaitą apie IP adresus serverių, kurie siuntė žinutes, nepraėjusias autentiškumo testų. Pvz.: authfail@lapenas.lt
rua	Apibendrintų ataskaitų siuntimo adresas	El. pašto adresas, kur siusti apibendrintas ataskaitas kas ir kiek žinučių siuntė iš viso ir kiek ir kieno siųstos žinutes perėjo/neperėjo patikrinimus. Pvz.: dmarcreport@lapenas.lt
p	Vardų srities politika	Galimos reikšmės: „none“ – nepaisyti, „quarantine“ – izoliuoti, „reject“ - atmesti

sp	Aukštesnio lygio vardų sričių politika	Galimos reikšmės: „none“ – nepaisyti, „quarantine“ – izoliuoti, „reject“ - atmesti
adkim	DKIM paisymo lygis	„r“ – nebūtinas „s“ - griežtas
aspf	SPF paisymo lygis	„r“ – nebūtinas „s“ – griežtas
ri	Ataskaitų siuntimo intervalas	Nurodoma kaip dažnai siųsti apibendrintą ataskaitą (laikas, sekundėmis)
fo	Ataskaitos apie nepavykusias autentifikacijas	0 – generuoti ataskaitą apie nepavykusias autentifikacijas jei visi autentifikacijos mechanizmai nepavyko 1 - generuoti ataskaitą apie nepavykusias autentifikacijas jei visi bent vienas iš mechanizmų nepavyko d – tik DKIM s – tik SPF
rf	Individualių žinučių ataskaitos formatas, jei nepavyko	neprivalomas

1.2.6. El. pašto žinutės turinio apdorojimas

Interneto adresų juodieji sąrašai

URIBL (angl. *uniform resource identifier blacklist*) ir SURBL (angl. *spam url blacklist*), kitaip sakant – internetinių adresų su protokolų, vardų srities ir kelio iki resursų nuorodų juodieji sąrašai. Tokie sąrašai paprastai yra vystomi nepelno siekiančių, su internetiniais brukalais kovojančių organizacijų ir yra teikiami per anksčiau jau minėtą DNS paslaugą. Užklausos struktūra yra lygiai tokia pati kaip ir RBL užklausos, aprašytos anksčiau, tik vietoje IP adreso yra pridedamas vardų srities adresas bei kelias iki resursų.

Vienas iš svarbiausių internetinių brukalo siuntėjų tikslų yra reklama arba tikslas į el. pašto laiško turinį įterpti nuorodą į puslapį ar objektą internete ir įtikinti, sudominti gavėją ja pasekti. Kadangi brukalas yra siunčiamas daugybei gavėjų vienu metu, atsiranda galimybė, kad šį nepageidaujamą laišką gavę gavėjai bus pranešę apie įvyki juodiesiems sąrašams. Nepageidaujamo el. pašto laiško išeitinį kodą galimą pateikti URIBL sąrašus prižiūrinčioms organizacijoms, kur jie yra apdorojami ir, atsiradus kritiniam pranešimų skaičiui, patalpinami juoduosiuose sąrašuose, kurie yra viešai prieinami interneto vartotojams. Paprastai tokiuose sąrašuose yra nuorodos, kurios nėra randamos „geruose“ laiškuose. Populiarėja ir jungtinės paslaugos kaip „multi.surbl.org“, kurios už vartotoją įvykdo užklausas populiariausiems juodųjų sąrašų tiekėjams ir gražina atitinkamą atsakymą klientui su žyma, pagal kurią galima spręsti kuriame iš juodųjų sąrašų nuoroda buvo rasta.

Šiam filtravimo tipui galioja tokios pačios rizikos kaip ir IP adresų juostiesiems sąrašams, t.y. pasitikėjimas paslaugos tiekėju, jo reputacija, ir taikoma saugumo bei įtraukimo į sąrašus politika. Galima pažymėti, kad šis būdas yra efektyvesnis ir turėtų turėti mažiau teigiamai neigiamų atmetimų, nei filtruojant pagal siuntėjo IP adreso reputaciją, kuri gali būti kintanti.

Juodieji raktažodžių sąrašai

Žodžiai, dažniausiai pasitaikantys nepageidaujamame sraute, yra įvedami į duomenų bazę ir kiekvienas priimtas laiškas yra apdorojamas ieškant jame tam tikrų raktažodžių, kurie gali padidinti tikimybę, kad laiškas yra nepageidaujamas. Tokios duomenų bazės yra pasiekiamos per internetą, jas galima nemokamai parsisiųsti ir naudoti. Papildomas dėmesys turi būti skiriamas duomenų bazės atnaujinimui, nes brukalo metodai ir pobūdis kinta. Svorio koeficientas filtravimo sistemoje šiam metodui turėtų būti žemas.

Adaptyviniai Bejeso filtrai

Populiariausias adaptyvinio filtravimo metodas remiasi Tomo Bajeso (1701 – 1761) 1763 m. pristatyta teorema. Bajesas įrodė, kad įvykio tikimybę galima daug tiksliau nustatyti naudojant iš anksto žinomą informaciją ir naujų stebėjimų duomenis. Ši teorema yra tikimybių teorijos statistinė teorema, kuri nustato įvykio tikimybę, kai žinoma tik dalis informacijos apie įvykius [8].

Bajeso klasifikatorius yra statistinis metodas el. pašto filtravimui siejantis tekste randamus žodžius ir (kartais) kitus objektus „geruose“ ir nepageidaujamuose laiškuose. Tuomet apskaičiuojama tikimybė ar tai yra pageidaujamas laiškas ar ne. Nepageidaujamų laiškų filtravimui šis matematinis modelis pradėtas taikyti dar 1990 metais dėl jo žemo neigiamai teigiamo atmetimų dažnio ir laikomas vienu seniausių kovos su brukalu metodų.

7 lentelė. Tikimybės skaičiavimas Bajeso metodu **[9]**

$p(c\|\vec{x}) = \frac{p(c_s)\cdot p(\vec{x}\|c_s)}{p(c_s)\cdot p(\vec{x}\|c_s) + p(c_h)\cdot p(\vec{x}\|c_h)}$	
$p(c\|\vec{x})$	Tikimybė, kad žinutė yra brukalas, žinant, kad joje yra žodis „velnias“
$p(c_s)$	Bendra tikimybė, kad žinutė yra brukalas
$p(\vec{x}\|c_s)$	Tikimybė, kad žodis „velnias“ būna brukalo žinutėse
$p(c_h)$	Bendra tikimybė, kad bet kuri duotoji žinutė nėra brukalas
$p(\vec{x}\|c_h)$	Tikimybė, kad žodis „velnias“ pasitaiko gerose žinutėse

Bajeso brukalo filtravimo metodas pagrįstas principu, kad dauguma įvykių yra susiję ir, kad tikimybė įvykiui įvykti ateityje gali būti numatoma iš istorinių atsitikimų, taip pat yra daroma ir su brukalo klasifikavimu. Jei teksto ištrauka ar žodis dažnai pasitaiko nepageidaujamoje korespondencijoje, bet nepasitaiko pageidaujamoje, tuomet galima manyti, kad žinutė yra brukalas (7 lentelė).

Klasifikatoriaus veikimui reikalinga turėti žodžių ir jų žymų duomenų bazę, žyma gali būti sudaryta iš paties žodžio, siuntėjo IP adreso, vardų srities ir t.t. surinktų iš pavyzdinių brukalų ir „gerų“ laiškų. Pageidautina leisti vartotojams patiems pažymėti laiškus, taip kaupiant vis didesnę duomenų bazę. Techninis tokios galimybes realizavimas galėtų būti papildomas funkcionalumas el. pašto vartotojo sąsajoje, prieinamoje per naršyklę, arba visiems žinomas el. pašto adresas, kur būtų persiunčiamas brukalas, kuris periodiškai būtų apdorojamas ir tokiu būdu būtų papildoma duomenų bazė. Duomenų bazės papildymas „gerais“ laiškais gali būti vykdomas analizuojant vartotojų išeinančius laiškus, atsižvelgiant, kad jie patys nesiunčia brukalo. Minėti duomenų bazės arba filtro „apmokymo“ metodai turi du pagrindinius pažeidžiamumus: 1. Žinant kuo užsiima įmonė, kuriai yra pritaikyta „gerų“ žodžių duomenų bazė, galima daryti gana tikslius spėjimus kokius žodžius naudoti, kad apeiti klasifikatorių, 2. Jei „gerų“ žodžių duomenų bazė nėra pakankamai išsami ir pritaikyta tam tikrai įmonei, iškyla didelė tikimybė neigiamai teigiamiems atsitikimams.

Priežastys kodėl Bajeso klasifikavimo metodas yra plačiai naudojamas:

1) Bajeso metodas apdoroja visą el. laiško tekstą, išanalizuoja visus žodžius, kurie identifikuoja brukalą, bet tuo pat metu atpažįsta žodžius, kurie būna tik „geruose“ laiškuose ir pateikia tikimybę. Šis metodas yra daug tikslesnis nei ieškoti tam tikrų raktinių žodžių tekste (juodasis sąrašas) ir tik pagal juos spręsti apie žinutę.
2) Bajeso filtras yra dinamiškas ir realiu laiku „mokosi“ iš įeinančio ar išeinančio pašto srauto, tuo didindamas algoritmo tikslumą. Dažnai brukalo siuntėjai bando apeiti juodųjų sąrašu (statinius) filtrus naudodami skaičius vietoje raidžių pvz.: „nu0la1da“ arba deda tarpus „n-u-o-l-a-i-d-a“, Bajeso filtras, skaičiuodamas tikimybę, tokius žodžius „pastebi“, nes tikimybė, kad jie pasitaikė „gerame“ sraute yra mažesnė, nei brukale.
3) Bajeso metodas „prisiderina“ prie vartotojų įpročių. Žodžiai „kreditas“ turės skirtingas tikimybės bakinio sektoriaus organizacijoje, nei bibliotekoje.

4) Filtras yra daugiakalbis, priešingai, nei raktažodžių bibliotekos ar duomenų bazės.
5) Bajeso klasifikatorių sunkiau apeiti nei statinę raktažodžių duomenų bazę, nes ji yra kiekvienam vartotojui atskira prisitaikiusi prie būtent jo siunčiamų ir gaunamų laiškų.

1.2.7. Nepatikimų geografinių vietovių blokavimas

Šiam tikslui pritaikytos el. pašto filtravimo nuo nepageidaujamų laiškų sistemos kartais filtruoja siuntėjus pagal IP adresų geografinę padėtį. Toks būdas yra retai naudojamas dėl universalumo ir prieinamumo nuostolių, tačiau gali būti pateisinamas kai reikia komunikuoti tik dalykiniais klausimais tarp įmonės departamentų ar šalies vyriausybinių organizacijų.

1.3. Kriptografiniai metodai, naudojami el. pašto pristatyme bei vientisumo užtikrinime

1.3.1. DANE

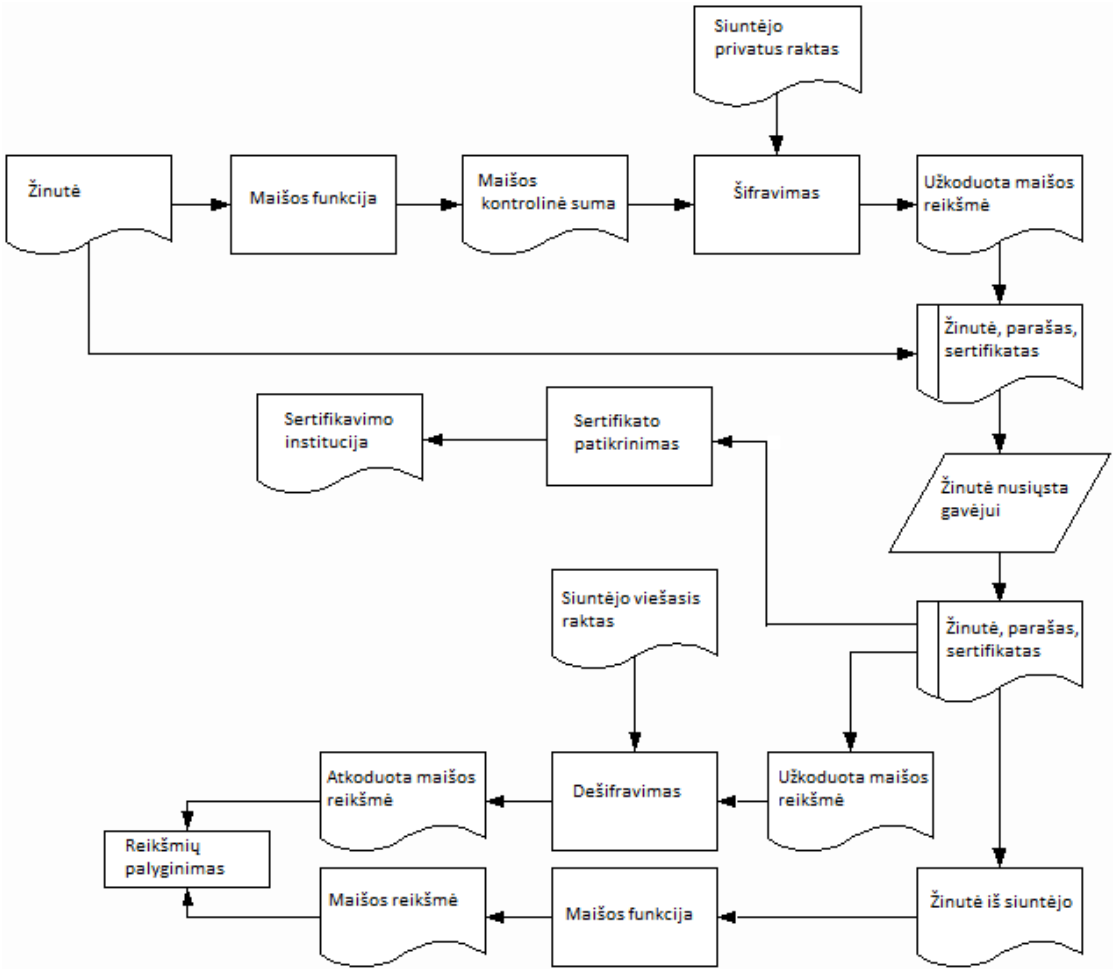
DANE (angl. *DNS based Authentication of named Entities*) – DNS paremta paslaugų autentifikacija TLS kontekste. Kadangi SMTP protokolas naudoja TLS kaip transporto lygio kodavimo protokolą, o TLS – X.509 sertifikatus nutolusios šalies autentiškumui nustatyti. Besijungiančioji šalis turi pasitikėti sertifikatų išrašymo centru kaip vieninteliu autoritetu. Sertifikavimo institucijų skaičius vis didėja, ir vis daugiau jų įtraukiama į operacinių sistemų ar programinių paketų pasitikėjimo sąrašus, todėl atsiranda rizika, kad vienos iš jų privatusis raktas gali būti panaudotas pasirašyti fiktyviems sertifikatams ir sulaužyti visą pasitikėjimo grandinę. DANE protokolas aprašo galimybę pasitikėti viešuoju raktu be sertifikavimo institucijos įsikišimo. Vardų srityje nurodomas specialaus TLSA – specialaus tipo DNS įrašas vardų srityje. Įraše nurodomas prievadas ir protokolas, o reikšmės dalyje yra nurodoma naudojimo ribos, įrašo tipas, maišos funkcija ir viešojo rakto maišos reikšmė. Pavyzdys žemiau:
„_25._tcp.mail.laukas.lt IN TLSA 3 1 1
EA518948D39E59403F65751F632F7032447A37E80C8CA841E49AF04A7564E83F“.

Klientas, besijungiantis prie paslaugos, daro DNS užklausą, kurios pagalba gauna viešojo rakto maišos reikšmę, tuomet, prisijungęs prie paslaugos prievado ir gavęs viešąjį raktą, gali patikrinti ar tai tas pats viešasis raktas, kuris buvo nurodytas vardų srities įraše. DNS paslauga yra išskaidyta paslauga, turinti daugybę tarpinių stočių iki autoritatyvaus tam tikrai zonai serverio, todėl atsiranda galimybė, kad bet kurioje vietoje gali įvykti klaida arba ataka, kuri leistų suklastoti TLSA įrašą. Minėtą problemą išsprendžia DNSSEC (viešo rakto kriptografijos pagrindu pasirašyti DNS įrašai) plėtinys, kuris yra būtinas naudojant DANE, tiek TLSA įrašų publikavimui, tiek užklausimui.

1.3.2. SMIME

„S/MIME“ (angl. *Secure/Multipurpose Internet Mail Extensions*) – saugūs daugiatikslio elektroninio pašto plėtiniai. SMIME yra viešojo rakto kriptografijos standartas elektroninių laiškų pasirašymui ir šifravimu. Standartas yra paremtas X.509 sertifikatais ir sertifikavimo institucijų pasitikėjimo modeliu, t.y. siuntėjas pasirašo išsiunčiamą laišką savo privačiuoju raktu ir prisega sertifikavimo institucijos pasirašytą savo sertifikatą, kurio subjektas yra siuntėjo elektroninio pašto adresas. Tik siuntėjas gali turėti sertifikato, kuriame yra jo el. pašto adresas ir sertifikavimo tarnybos parašas privatųjį raktą. Gavėjas patikrina sertifikato galiojimą, tuomet patikrina parašą su siuntėjo viešuoju raktu įsitikinant, kad žinutės turinys nebuvo keičiamas nuo jo išsiuntimo momento. SMIME pasirašymo operacija pavaizduota 8 paveiksle.

SMIME taip pat naudojamas žinutės konfidencialumo tarp gavėjo ir siuntėjo užtikrinimui, tai praktiškai pašalina šifravimo reikalingumą tinklo transporto lygmenyje. Žinutės šifravimas yra vykdomas gavėjo viešu raktu, todėl prieš pirmą karta išsiunčiant SMIME šifruotą žinutę, reikia turėti gavėjo viešąjį raktą, arba jau būti gavus gavėjo pasirašytą laišką. Šis viešojo rakto apsikeitimo veiksmas sukelia papildomų nepatogumų vartotojams nes nėra viešos sertifikatų „telefonų knygos“, kurioje galima būtų rasti siuntėjo viešus duomenis pirmą kartą gavus šifruotą laišką. Todėl SMIME šifravimo funkcionalumas yra dažniau naudojamas padidinto saugumo reikalaujančiose organizacijose, o bendrai paėmus rečiau naudojamas nei pasirašymas.



8 pav. SMIME parašo veikimas

1.4. Analizės apibendrinimas

Šiame skyrelyje pateikiamos el. pašto saugumo priemonių efektyvumo analizės išvados. Apžvelgus elektroninio pašto perdavimo infrastruktūrą bei jau naudojamus el. laiškų filtravimo metodus, galime padaryti šias išvadas:

1. Išanalizuotas SMTP protokolo veikimo principas, apdorojimo infrastruktūra ir nustatyta, kad jos pažeidžiamumas yra aukštas, kad yra nesudėtinga klastoti siuntėjo adresą ir be didelių išteklių siųsti didelius kiekius nepageidaujamų laiškų.
2. Nustatyta, kad tiek el. pašto saugumas, tiek tarnybinių stočių resursų saugumas apdorojant didelius kiekius nepageidaujamo el. pašto žinučių yra aktuali problema.
3. Išsiaiškinti el. pašto filtravimui ir klasifikavimui naudojami metodai ir nustatyta, kad vieni iš jų veikia tiksliau, kiti prasčiau, vieni yra skirti užtikrinti siuntėjo autentiškumą, kiti apdoroti el. laiško struktūrą, turinį, ar priimti klasifikavimo sprendimus. Nustatyta:
    1) Nėra standartinio metodo, kaip ir kokia tvarka, skirtingų tipų filtravimo metodai turėtų būti taikomi įeinančiam paštui, kad sumažinti sistemos resursų naudojimą juos apdorojant.
    2) Siuntėjo autentiškumo įtaka siunčiamų laiškų klasifikacijai (ar autentifikuotas siuntėjas būtinai reiškia, kad laiškas yra pageidaujamas)
4. Apžvelgta populiariausi kriptografiniai metodai, naudojamus el. pašto pristatyme bei turinio vientisumo užtikrinime.
5. Literatūros šaltinių analizėje susipažinta su įrankiais, kuriais naudojantis bus galima sukurti, realizuoti ir praktiškai ištestuoti modelį, skirtą įvairių pašto serverio apsaugos priemonių efektyvumo įvertinimui ir jį ištirti kiekybiškai bei kokybiškai.

2. SIŪLOMAS EL. PAŠTO FILTRAVIMO BŪDŲ BEI JŲ ĮTAKOS TARNYBINĖS STOTIES RESURSŲ NAUDOJIMUI IMITAVIMO METODAS

Atlikus nepageidaujamų laiškų filtravimo metodų analizę, nustatyta, kad vienintelio ir visiškai teisingo būdo, kuris tiktų visais atvejais ir visiems – nėra. Norint kuo efektyviau sukonfigūruoti sistemą, reikia naudoti filtravimo metodų kombinaciją, kuri geriausiai tinka tam tikroje aplinkoje ir yra pakankamai efektyvi. Filtravimo metodai pagal jų funkcionalumą skiriasi ir, galbūt, kai kuriuos reikia naudoti laiško priėmimo pradžioje, atmetant prisijungimą prieš pradedant detaliau analizuoti laiško turinį ir tokiu būdu išsaugoti skaičiavimo ir atminties išteklių „geruųjų laiškų“ apdorojimui.

Siūlomas imitacinis modelis, paremtas realiais duomenimis iš tarnybinės stoties leis įvertinti įvairių filtravimo metodų bendro naudojimo architektūros efektyvumą ir jų įtaką tarnybinės stoties resursų naudojimui. Toks modelis padės įvertinti įrangos galimybes, patikrinti ar numatoma įsigyti įranga gali aptarnauti užsibrėžtą ar padidėjusį elektroninių laiškų srautą.

Įmonei ar organizacijai planuojant naujos infrastruktūros plėtrą ar atnaujinimą siūloma pasinaudoti sukurta metodika: susikurti sistemos modelį (pavyzdinis modelis pateiktas šiame darbe), modelio sudedamųjų komponentų parametrus nustatant pagal šiame darbe ištirtas filtravimo metodų charakteristikas, paremtas praktiniais tyrimais. Tokiu būdu bus galima įvertinti kokių išteklių reikės diegiant tokią sistemą.

2.1. Metodo koncepcija

Siūlomas tyrimo metodas leis ištirti ir imituoti el. pašto filtravimo būdus ir technikas. Galutinis sprendimas bus pagrįstas modeliu, kurio pagalba bus vertinama skirtingų filtrų įtaka sisteminiams resursams, laiško apdorojimo vėlinimas ir efektyvumas. Naudojant modelį bus galima priimti sprendimus dėl realios paslaugos kūrimo ir parametrų nustatymo.

Kaip pavaizduota 9 pav. modelis bus paremtas realioje tarnybinėje stotyje, kurioje bus įjungtas filtravimas, surinktais duomenimis ir charakteristikomis. Šie duomenys bus nustatomi kaip modelio parametrai skirtingiems filtravimo mechanizmams.

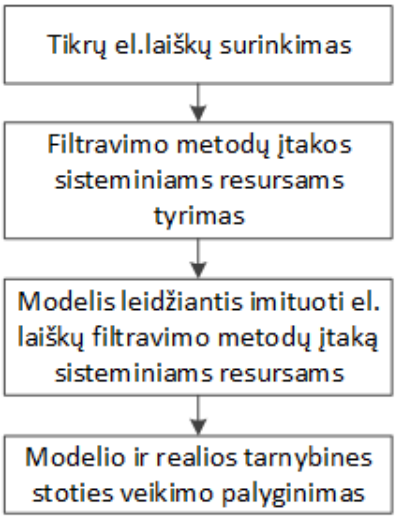
Pirmojoje ir antrojoje eksperimentinių tyrimų stadijoje bus renkami laiškai realiame pašto serveryje, kurie vėliau bus naudojami atlikti tyrimams: kiek laiko skirtingo dydžio laiškai yra apdorojami, kaip tai įtakoja sisteminių resursų apkrovimą, kaip skirtingų filtravimo metodų išjungimas ar įjungimas įtakoja laiškų apdorojimą, kaip keletas vienu metu apdorojamų laiškų įtakoja apdorojimo laiką. Yra daugybė el. pašto filtravimo metodų, tačiau dėl darbo apimties, tyrime bus naudojami šie populiarūs filtravimo metodai:

- „ClamAV“ antivirusinis filtras
- „Spamassassin“ nepageidaujamų laiškų filtravimo paketo komponentai:
    - Juodųjų sąrašų, atvirkštinių DNS (angl. *domain name system*) vardų, SPF įrašų tikrinimas. Apibendrinus – DNS naudojantys tikrinimai.
    - „Razor2“ ir „Pyzor“ patikra. Apibendrinus – laiško santraukos tikrinimas centralizuotose nepageidaujamų laiškų duomenų bazėse.
    - Raktažodžių, frazių ir antraštės analizė. Apibendrinant – laiško turinio analizė naudojant reguliarias išraiškas.
    - Bajeso tikimybės tikrinimas. Apibendrinant – žodžių tikrinimas Bajeso duomenų bazėje, kuri bus apmokyta apie 100 „pageidaujamų laiškų“.

Atliekant tyrimą bus stebimas apdorojimo laikas ir šios charakteristikos:

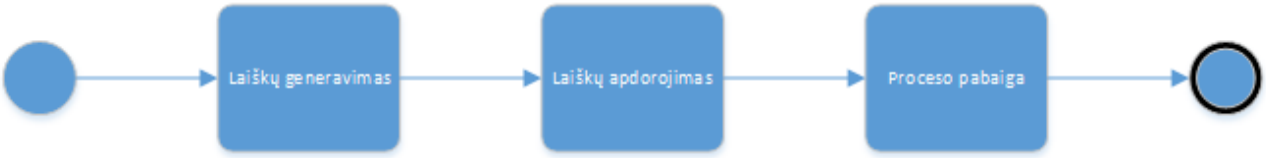
- Procesoriaus apkrovimas
- Procentinis disko kreipties posistemė apkrovimas
- Absoliutus tinklo greitaveikos apkrovimas

Sisteminių resursų charakteristikų surinkimui bus naudojami „Munin“ įskiepiai ir atvaizdavimo paslauga, kuri leis grafiškai atvaizduoti procesoriaus, atminties ar tinklo resursų naudojimą 5 minučių vidurkio intervalais.



9 pav. Metodo schema

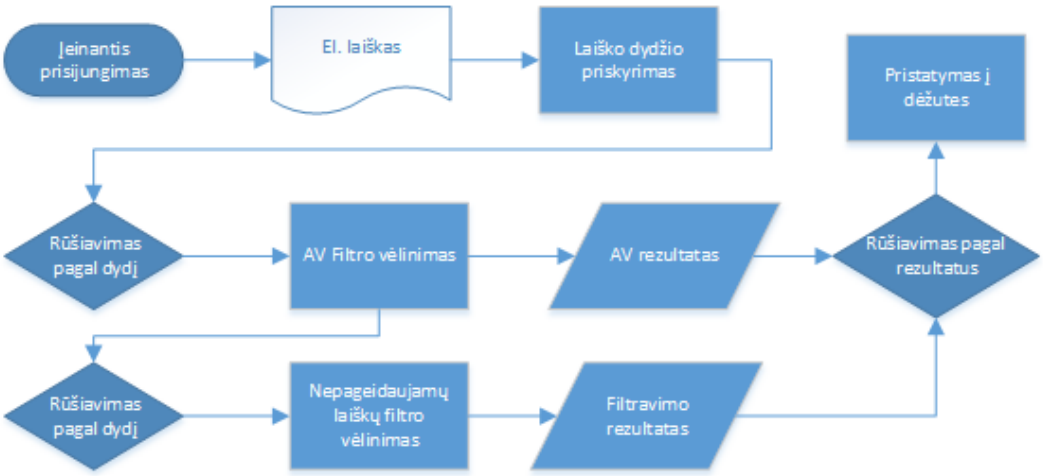
Trečiame etape bus kuriamas kompiuterinis imitacinis modelis „Rockwell Software ARENA“ aplinkoje. Šios programinės įrangos gamintojas suteikia galimybę studentams naudotis nemokama versija, kuri neriboja funkcionalumo, tačiau komercinis jos taikymas yra draudžiamas. Modelis bus kuriamas pagal 10 paveiksle pavaizduotą schemą, kur laiškai bus generuojami, apdorojami ir pristatomi į el. pašto dėžutes.



10 pav. Supaprastinta modelio schema

Modelyje bus nustatomi parametrai iš su realiais duomenimis surinktų eksperimentų, o modelio logika bus bandoma atkartoti realios tarnybinės stoties filtravimo sistemos architektūrą ir nustatymus.

2.2. El. laiškų filtravimo sistemos imitacinis modelis



11 pav. Nepageidaujamų laiškų filtravimo sistemos imitacinio modelio schema

Imitacinis kompiuterinis modeliavimas apima skaičiavimo ir matematikos metodus, imituojant ir analizuojant įvairių sistemų funkcionalumą. Modeliavimo tikslai: įžvalga, geresnis sistemos ar jos dalies pažinimas, jos funkcionalumo supratimas bei reagavimas į pokyčius [10].

Sukurtas imitacinis modelis leis keisti el. laiškų priėmimo sistemos architektūrą modeliuojant ir nekonfigūruojant realios tarnybinės stoties. Simuliacijos rezultatai leis spręsti apie resursų panaudojimą, veikiant vienokiems ar kitokiems filtrams ir daryti išvadas ar modeliuojama sistemos koncepcija tenkina poreikius: toleruojamas nenufiltruotų laiškų procentas prieš pastatant laiško apdorojimo laiką naudojant užduotus sisteminius parametrus.

Modelio logika bus kuriama iš struktūrinių modulių, kurie atitiks filtravimo arba laiškų apdorojimo procesus realioje tarnybinėje stotyje. Duomenys modulyje bus laiškų dydžiai ateinantys ir apdorojami sistemoje. Laiškų atėjimo greičio, atsitiktinės paskirstymo funkcijos ir tikimybės reikšmes bus nustatomos iš eksperimentų, atliktų tikroje tarnybinėje stotyje duomenų. ir Filtravimo sistemos imitacinio modelio logine schema pavaizduota 11 paveiksle.

Priklausomai nuo imitacinės aplinkos programinės įrangos galimybių bus leidžiama keisti pagrindinius modelio vėlinimo parametrus ir filtrų efektyvumo rodiklius.

2.3. Realios tarnybinės stoties architektūra

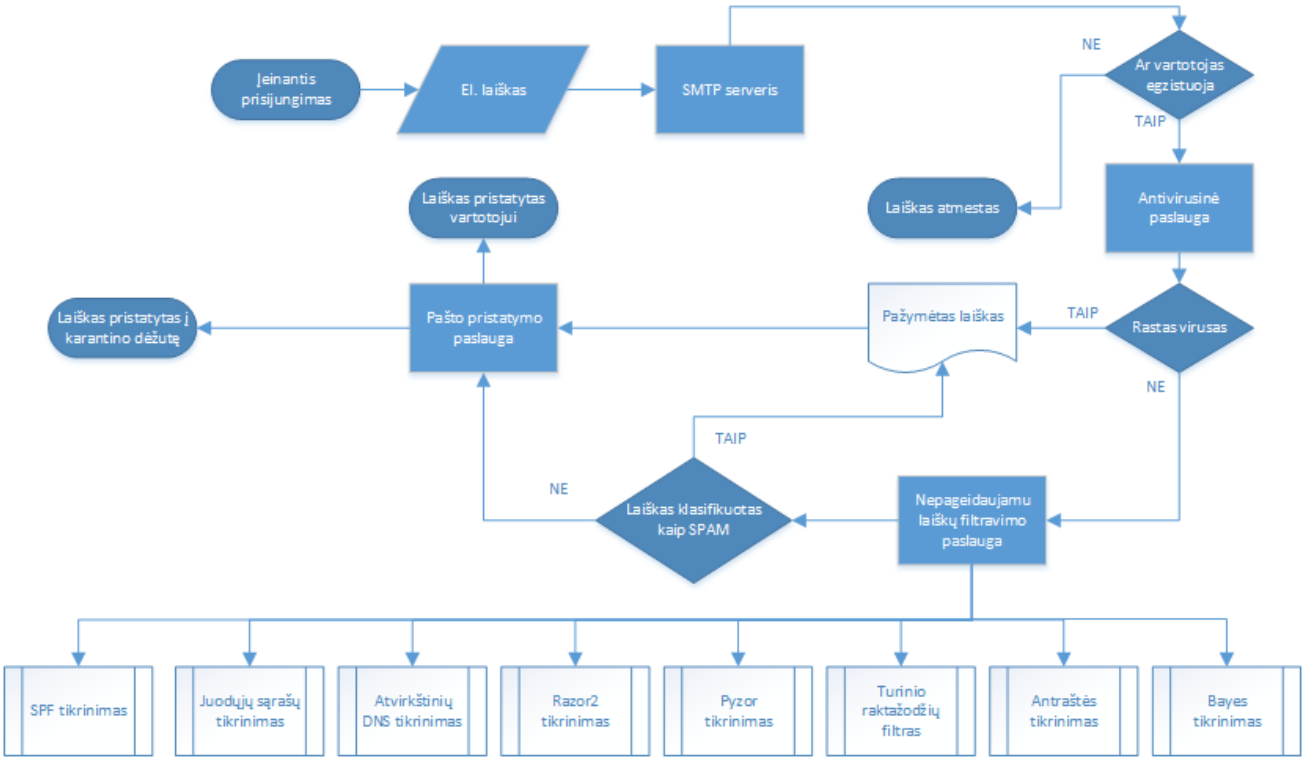
Elektroninių laiškų priėmimo ir filtravimo sistemai kurti bus naudojama architektūra pavaizduota 12 pav. Būtent tokia loginė sistemos konfigūraciją buvo pasirinkta todėl, kad tai yra dažnai pasitaikanti tiek finansinėse institucijose, tiek ne pelno siekiančiose organizacijose, tiek privačiame versle. Paprastai norima atsikratyti dviejų pagrindinių blogybių: laiškų su bylomis, kuriose yra žalingas kodas arba virusai, ir nepageidaujami, dažniausiai, reklaminiai laiškai. Pradinėje konfigūracijoje bus naudojami du filtravimo lygiai: 1) antivirusinis lygis, 2) nepageidaujamų laiškų filtravimo lygis. Pirmame lygyje bus tikrinama ar laiškuose nėra pavojingo kodo apraiškų ir, esant teigiamam rezultatui, bus vykdomas numatytas veiksmas. Daugeliu atvejų nebėra svarbu ar laiškas gali būti klasifikuojamas kaip nepageidaujamas (angl. "*SPAM*"), jei jame buvo rastas virusas, tačiau eksperimento tikslu, būsimoje sistemoje apdorojimas bus tęsiamas statistikai surinkti.

Laiškas, patekęs į paskutinį filtravimo lygmenį, bus apdorojamas specifinės, tik nepageidaujamiems laiškams apdoroti skirtos, programinės įrangos, kuri naudos aibę skirtingų filtravimo metodų. Kiekvienas iš jų prisidės prie bendros tikimybės skaičiavimo. Kiekvienas iš filtravimo mechanizmų skirtingai apkrauna tarnybinės stoties procesorių, atmintį, disko bei tinklo kreipties posistemes ir dėl to ilgėja laiško apdorojimo laikas.

Antrame ir trečiame lygiuose teigiamus analizės rezultatus „gavę" laiškai nebus trinami – bus pažymimi „X" tipo žyma ir perduodami pristatymo paslaugai (LDA - angl. *local delivery agent*), kuri pristatys laišką į atitinkamas dėžutes. Paprastai tokios dėžutės kuriamos siekiant išvengti neteisingų atmetimų ir leidžia vartotojui peržiūrėti visus gautus laiškus, net jei jie ir nepageidaujami..

Virusų filtravimo variklis filtravimo serveryje veiks nuosekliai nepageidajamų laiškų filtrui nepriklausomai nuo rezultato. Apdorojamas laiškas filtravimo serveryje bus tik pažymimas ir grąžinamas pašto serveriui, kuris pagal žymas, iš abiejų filtravimo variklių, priims sprendimą apie laiško pristatymą į dėžutes.

Tikrai tarnybinei stočiai panaudotos virtualios operacines sistemos, veikiančias „CentOS" Linux distribucijos pagrindu. Sprendimą sudarys SMTP serveris naudojantis „maildrop" LDA, kuris turi filtravimo funkcionalumą. „Maildrop" bus sukonfigūruotas perduoti duomenis į „ClamAV" antivirusinės paslaugos serverį, vėliau į „Spamassassin" – nepageidaujamų laiškų filtrą.
Skirtingų šaltinių duomenimis vienas populiaresnių SMTP serverių programinės įrangos paketų yra „Postfix", ši programinė įranga ir bus naudojama tyrimo metu. SMTP serverio dalis ir filtravimo funkcijas atliekančio serverio dalis bus atskirtos ir bendraus per tinklo sąsają kaip pavaizduota 13 paveiksle, toks sprendimas leis lanksčiai keisti nustatymus ir simuliuoti laiškų apdorojimą įjungiant ar išjungiant operacinius komponentus. Galima numanyti, kad filtravimo serveris naudos daugiau skaičiavimo t.y. procesoriaus išteklių, todėl, turint, atskiras sistemas, viena neįtakos kitos ir taip leis objektyviau įvertinti skirtingų komponentų įtaką.



12 pav. Nepageidaujamų laiškų filtravimo sistemos loginė schema

Nebus atsižvelgiama į tinklo vėlinimą tarp SMTP serverio ir filtravimo komponentų, nes tinklo sąsajos tarp virtualių mašinų taip pat yra virtualios ir valdomos hipervizoriaus.

Pasaulinio tinklo vėlinimas iki nutolusių resursų, tokių kaip juodųjų sąrašų serveriai, maišos reikšmių duomenų bazės ar atvirkštinės DNS užklausos taip pat bus stebimas tik kaip santykinis dydis, nes jo įtakoti ar objektyviai įvertinti nėra reikalo.

2.4. Antivirusinis filtras „ClamAV“

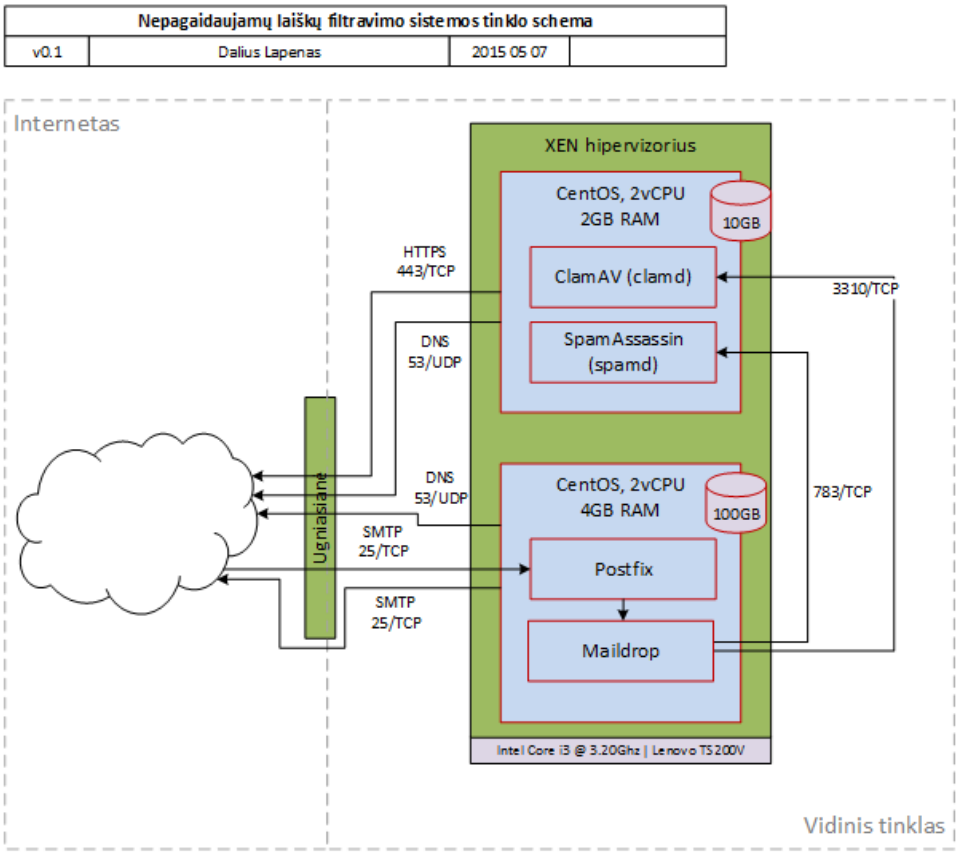
Clam Antivirus (ClamAV) yra nemokamas atviro kodo įrankis skirtas aptikti įvairiam kenkėjiškam kodui ir virusams. Vienas pagrindinių šio programinio paketo panaudojimo atvejų – pašto serveriuose kaip el. pašto filtras, skirtas duomenų sraute aptikti žinomų virusų struktūras.

ClamAV susideda iš komandinės eilutės taikomosios programos, virusų žymų duomenų bazės bei daugiaprocesinio serverio. Šiuo atveju naudosiu serverio komponentą, išvengiant papildomo pertekliškumo kiekvieno laiško tikrinimo atveju iš naujo paleidžiant šio filtro komponentą. Serverinė dalis nuolat veiks atmintyje ir gaus duomenis per tinklo sąsają iš maildrop vietinės laiškų pristatymo paslaugos.

Programinė įranga palaiko daugumą elektroniniu paštu siunčiamų bylų formatų: *Zip, RAR, Tar, Gzip, Bzip2, OLE2, Cabinet, CHM, BinHex, SIS, PE, UPX, FSG, Petite, NsPack, wwpack32, MEW, Upack* įvairius klaidinimo formatus kaip SUE, Y0da. Dažnai pasitaikančius *Microsoft Office, HTML, RTF, PDF* formatus. Dauguma išvardintų bylų yra suspaustos arba koduotos, todėl jų apdorojimas greičiausiai santykinai smarkiai apkraus procesoriaus darbą ir atsispindės galutinėse išvadose.

2.5. Nepageidaujamų laiškų filtravimo sistema „SpamAssassin“

„SpamAssassin“ yra el. pašto filtravimo sistema skirta identifikuoti nepageidaujamus el. laiškus. Sistema naudoja įvairias filtravimo metodus pritaikytus tiek laiško antraštei, tiek turiniui analizuoti. Modulinė struktūra leidžia pagal pageidavimą išjungti ar įjungti vieną ar kitą metodą.



13 pav. Nepageidaujamų laiškų filtravimo sistemos tinklo schema

Panašiai kaip ir ClamAV atveju, naudosiu serverio programinę įrangą, išvengiant programinės įrangos paleidimo vėlinimo kiekvieno apdorojamo laiško metu. „Maildrop“ LDA naudos specialų klientą kreipimuisi į filtrą per tinklo sąsają.

Metodo koncepcijoje minėtiems filtravimo būdams bus įjungti arba išjungti šie „SpamAssassin“ moduliai:

- Mail::SpamAssassin::Plugin::SPF
- Mail::SpamAssassin::Plugin::DNSEval
- Mail::SpamAssassin::Plugin::Razor2
- Mail::SpamAssassin::Plugin::Pyzor
- Mail::SpamAssassin::Plugin::Bayes.

Programinė įranga periodiškai atsinaujina filtravimo metodų nustatymų bylas, tokiu būdu prisitaikydama prie kintančių nepageidaujamų laiškų siuntimo sąlygų bei tendencijų. Priklausomai nuo operacinės sistemos distribucijos, jų pavadinimai gali būti skirtingi, tačiau Centos OS planuoju redaguoti šias bylas: */etc/mail/spamassassin/init.pre*, */etc/mail/spamassassin/v310.pre*, */etc/mail/spamassassin/v312.pre*, */etc/mail/spamassassin/v320.pre*, */etc/mail/spamassassin/v330.pre*.

2.6. Apibendrinimas

1. Pasiūlytas metodas surinkti statistinei informacijai realioje tarnybinėje stotyje, kuri apdoros ir filtruos elektroninius laiškus.
2. Numatyta kokie filtravimo metodai bus naudojami ir kokios tarnybinės stoties charakteristikos bus vertinamos.
3. Sukurta pavyzdinė kompiuterinio imitacinio modelio schema, kuri bus realizuojama pasirinktoje modeliavimo aplinkoje.

3. EL. PAŠTO FILTRAVIMO BŪDŲ IMITAVIMO METODO EKSPERIMENTINIS TYRIMAS

Prieš pradedant kurti sistemos modelį buvo eksperimentiškai ištirtos el. pašto siuntimo, gavimo ir filtravimo situacijos. Kiekvieno eksperimento metu keičiant filtrų ir sisteminių resursų nustatymus. Iš viso buvo atlikta 18 eksperimentų ir surinkti duomenys, kurie vėliau panaudoti modelio kūrime.

Eksperimentinę infrastruktūrą sudarė du nepriklausomi el. pašto siuntimo šaltiniai tuo pačiu metu siunčiantys el. laiškus, kurie buvo surinkti 4 mėnesių laikotarpyje nuo 2015 liepos iki spalio mėnesių, priimant visus laiškus iš interneto. Laiškų surinkimui buvo naudojama Internete registruota "m5.lt" vardų sritis. El. laiškų priėmimo ir filtravimo infrastruktūra pavaizduota buvo įdiegta kaip pavaizduota 13 paveiksle, o pagrindinių programinės įrangos komponentų inventorius – 8 lentelėje.

8 lentelė. Programinės įrangos sąrašas

Programinės įrangos paskirtis	**Programinės įrangos pavadinimas, versija**
Operacinė sistema	Centos Linux 6.6
El. pašto (SMTP serveris)	Postfix 2.6.6-6.el6_5
El. pašto filtro klientas	SpamC 3.3.1
El. pašto antivirusinis klientas	ClamdScan 0.98.7/21512
Antivirusinė programinė įranga	ClamAV 0.98.7-1.el6
Nepageidaujamų laiškų filtravimo programinė įranga	SpamAssassin 3.3.1-3.el6
El. pašto filtrų valdiklis / MDA	Maildrop 2.8.3

Pagrindiniai tyrime naudojamos programinės įrangos elementai yra el. pašto serveriai, filtravimo serverio ir filtravimo kliento programinė įranga. Ši programinė įranga buvo nustatyta, kad tenkintų eksperimentinių situacijų reikmes:

- CentOS operacinės sistema buvo įdiegta „XEN 4.4" virtualioje aplinkoje, o pačios sistemos parametrai eksperimentinės infrastruktūros kūrimo metu keičiami nebuvo.
- Postfix SMTP serveris buvo sukonfigūruotas priimti laiškus „m5.lt" domenui ir nepriklausomai nuo el. pašto adresato pristatyti visus laiškus adresu „dalius@m5.lt". Vartotojas "dalius" – vienintelis sistemos vartotojas. Už el. pašto pristatymą atsakinga programa buvo pakeista į „Maildrop" – virtualų žinučių pristatymo agentą.
- Maildrop MDA buvo nustatytas patikrinti laiško dydį, ir jei jis mažesnis nei 15 MB, nuosekliai iškviesti filtravimo tarnybų klientus, kurių pagalba laiškai buvo perduodami į filtravimo serverį. Filtravimo tarnyboms baigus laiško apdorojimą, yra pažymimas „x" tipo žyma pagal kurią „maildrop" buvo nustatytas pristatyti laišką į atitinkamai virusų ar nepageidaujamų laiškų el. pašto dėžutę „karantinas" arba „spam".
- „ClamAV" antivirusinis filtravimo paketas, pagal nutylėjimą, veikia kaip programa, kuriai reikia nuroti skenuojamą bylą. Eksperimentinės infrastruktūros kūrimo metu „ClamAV" buvo sukonfigūruotas veikti serviniame režime, kurio vykimo principas yra veikti tinklo režimu ir filtruoti gaunamą tinklo srautą. Šiam srautui perduoti į serverį naudojamas „ClamdScan" klientas.
- „SpamAssassin" nepageidaujamų laiškų filtravimo programinis paketas pagal nutylėjimą taip pat veikia programos režime, todėl jis buvo nustatytas veikti tinklo serverio konfigūracija. Tokie nustatymai reikalingi todėl, kad eksperimentinės sistemos filtravimo ir laiškų gavimo serveriai yra atskiros, tinklu komunikuojančios sistemos. SMTP serveryje įdiegtas programinis klientas „spamc".

Sukurta sistema buvo išbandyta ir paleista laiškų surinkimui. Vėliau ši sistema buvo naudojama eksperimentuose statistinei informacijai surinkti. Eksperimentų metu buvo keičiami hipervizoriaus nustatymai, kurie leido keisti operacinių sistemų aparatinių resursų, tokių kaip procesorių skaičius, operatyvines atminties vertes.

3.1. El. pašto filtrų efektyvumo ir įtakos sisteminiams resursams tyrimo eiga

Naudojant skirtingo skaičiavimo pajėgumo virtualias mašinas atlikti eksperimentai, siekiant išsiaiškinti skirtingų filtrų naudojimo įtaką sisteminiams resursams. Taip pat buvo renkama statistika kiek laiko skirtingų dydžių laiškai yra apdorojami prie skirtingų filtravimo mechanizmų ir virtalių mašinų skaičiavimo pajėgumo. Kaip atskaitos tašką, kuris nenaudotų jokių nepageidaujamų laiškų filtravimo metodų, buvo pasirinktas SMTP serveris, kuris tik priima laiškus ir pristato į vartotojo dėžutę. Tokios konfigūracijos sistema yra 0 proc. arba neefektyvi nepageidaujamų laiškų filtravimo atžvilgiu, tačiau padeda suprasti kokie yra pradiniai sisteminių parametrų rodikliai. Eksperimentų rezultatai atvaizduoti 10 lentelėje (11 lentelėje nurodyti sutrumpinimų paaiškinimai). Įjungus visus užsibrėžtus filtravimo metodus 62 proc. visų surinktų laiškų yra klasifikuojami kaip nepageidautini, o iš jų tik apie 0,43 proc. turi kenkėjiško kodo požymių (14 pav.). Rezultatai yra panašios eilės kaip surinkti pasauliniai duomenys darbo analizės metu (9 lentelė).

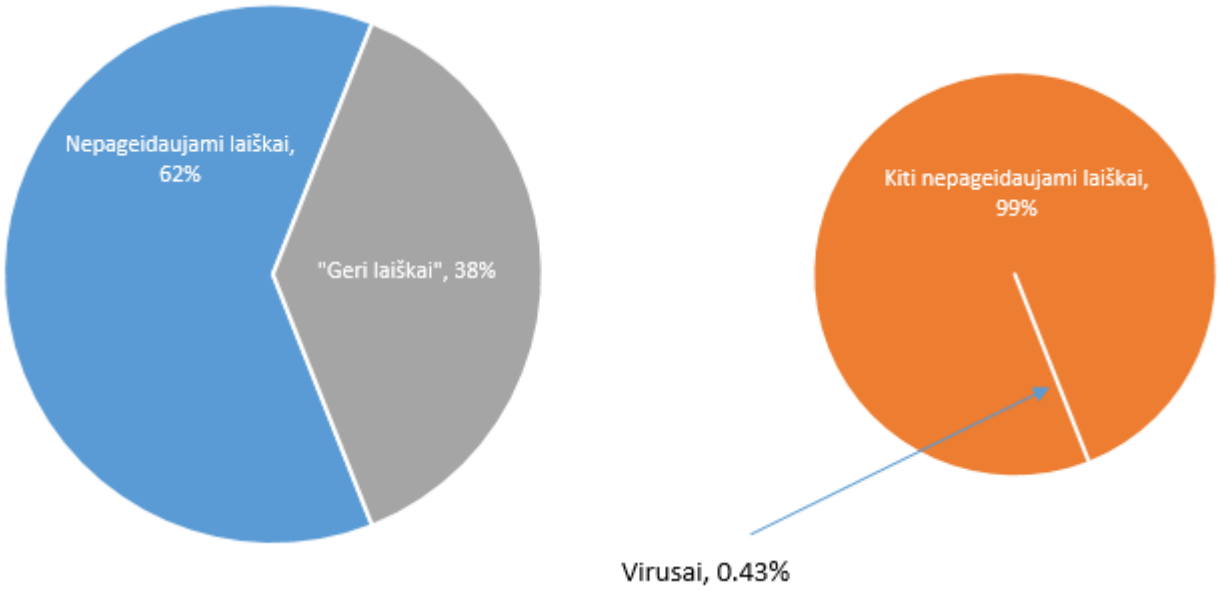
9 lentelė. Sisteminių resursų būklė eksperimentų metu

Tipas	**Pasaulinis lygis 2012 m.**	**Pasaulinis lygis 2013m.**	**Pasaulinis lygis 2014m.**	**Eksperimentais nustatytas 2015m.**
„Spam“	69%	66%	60%	62%
Virusų turintys laiškai	1 iš 291	1 iš 196	1 iš 244	1 iš 365

Visų įjungtų filtrų rezultatai bus laikomi maksimalia sistemos galimybių rasti nepageidaujamus laiškus riba, todėl nusistačius pradinius ir maksimalius taškus – kitų filtrų rezultatai buvo apskaičiuoti santykine procentine išraiška pateikiama efektyvumo stulpelyje.

Elektroniniams laiškams siųsti į tiriamą sistemą buvo naudojamas „bash“ komandinės aplinkos scenarijus, kurio tikslas buvo skaityti direktoriją, kurioje yra surinkti laiškai ir siųsti juos į SMPT serverį. Padidinti išsiunčiamų laiškų kiekį šis scenarijus buvo paleistas dviejuose nepriklausomuose serveriuose. Tuo pat metu iš kiekvieno el. laiško antraštės buvo ištrintos „x-original-to“ bei „delivered-to“ skiltys, tokiu būdu apeinant gaunančiosios sistemos apsaugą nuo ciklinio laiškų priėmimo.

Eksperimentų metu nustatyta, kad sistemos vidutinis laiškų priėmimo greitis yra nuo 10 iki 17 laiškų per sekundę pastebimai priklausomai nuo laiško dydžio. Laiškų siuntimo greitis nepriklausė nuo tinklo greičio, nes tuo pat metu stebint duomenų kiekį perduodamą tinklo sąsaja – išmatuotas vidutinis 3 Mb/s eilės greitis, o serveriai tarpusavyje buvo sujungti 1 Gb/s greičio tinklu.



14 pav. Nepageidaujamų laiškų pasiskirstymas

10 lentelė. Sisteminių resursų būklė eksperimentų metu

Eksperimento Nr.	**Laiškų kiekis**	**Įjungti filtrai***	**vCPU kiekis**	**Aptikta Nepageidaujamų laiškų**	**Efektyvumas**	**vCPU parkrovimas (SMTP|AV)****	**Vidutinis tinklo apkrovimas*** (SMTP|AV)****	**Disko apkrovimas (SMTP|AV)****	**Atminties sunaudojimas (SMTP|AV)****
1.	30960	-	2	0	0%	65% | 0%	2mbs/180kbps	2% | 0%	65MB | 0MB
2.	30960	-	4	0	0%	90% | 0%	2.25mbps/192kbps	3% | 0%	67MB | 0MB
3.	30960	-	8	0	0%	93% | 0%	2.6mbps/196kbps	2% | 0 %	69MB | 0MB
4.	30960	AV	2	84	0.43%	142% | 10%	2.78mbps/2.88mbps | 2.88mbps/73kbps	2% | 0%	69MB |432MB
5.	30960	AV	4	84	0.43%	171% | 14.6%	2.72mbps/2.72mbps | 2.7mbps/73kbps	2% | 0%	71MB | 432MB
6.	30960	AV	8	84	0.43%	203% | 12%	2.78mbps/2.88mbps | 2.88mbps/73kbps	2% | 0%	72MB|435MB
7.	30960	AV+DNS	2	16962	87%	70% | 195%	2.45mbps/2.45mbps | 2.5mbps/1.37mbps	2% | 0 %	126MB|702MB
8.	30960	AV+DNS	4	16962	87%	160% | 288%	3.72mbps/3.75mbps | 3.75mbps/2.05mbps	2% | 0%	115MB|700MB
9.	30960	AV+DNS	8	16962	87%	200% | 415%	4.18mbps/4.16mbps | 4.19mbps/2.23mbps	2% | 0%	116MB|733MB
10.	30960	AV+DNS+RazrPyzr	2	17690	91%	70% | 190%	1.8mbps/1.9mbps | 2mbps/1.1mbps	2% | 0%	132MB | 813MB
11.	30960	AV+DNS+RazrPyzr	4	17690	91%	107% | 262%	2.47.mbsp/2.47mbps | 2.6mbps/1.41mbps	2% | 0%	133MB | 808MB
12.	30960	AV+DNS+RazrPyzr	8	17690	91%	110 %|300%	2.5mbps/2.5mbps | 2.7mbps/1.46mbps	2% | 0%	132MB | 821MB
13.	30960	AV+DNS+RazrPyzr+Content	2	18547	95%	40% | 198%	1.15mbps/625kbps | 723kbps/1.11mbps	2% | 0%	132MB | 956MB
14.	30960	AV+DNS+RazrPyzr+Content	4	18547	95%	91% | 370%	1.97mbps/2.06mbps | 2.19mbps/1.19mbps	2% | 0%	133MB | 977MB
15.	30960	AV+DNS+RazrPyzr+Content	8	18547	95%	96% | 470%	2.39mbps/2.43mbps | 2.28mbps/1.4mbps	2% | 0%	134MB | 967MB
16.	30960	AV+DNS+RazrPyzr+Content+Bayes	2	19490	100%	30% | 190%	1.07mbps/996kbbs | 1.07mpbs/532kbps	1% | 0%	132MB | 971 MB
17.	30960	AV+DNS+RazrPyzr+Content+Bayes	4	19490	100%	83% | 360%	1.9mbps/1.87mbps | 1.98mbps/1.08mbps	1% | 0%	134MB | 982MB
18.	30960	AV+DNS+RazrPyzr+Content+Bayes	8	19490	100%	76% | 450%	2.22mbps/2.18mpbs | 2.31mbps/1.25mbps	1% | 0%	136MB | 992MB

11 lentelė. Eksperimentų rezultatų lentelės sutrumpinimų paaiškinimai

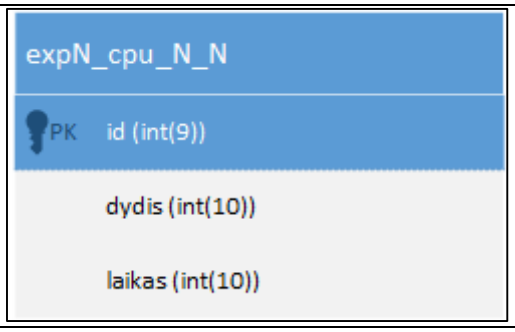
Nuoroda	Reikšmė
* Filtrų tipai	„-“ - Nėra įjungtų filtrų, laiškai tik priimami
	„AV“ - Įjungiamas Antivirus (ClamAV) filtravimas
	„AV+DNS“ - Įjungiamas SpamAssassin ir DNS tipo filtrai
	„AV+DNS+RazrPyzr“ - Įjungiami Razor2 ir Pyzor filtrai
	„AV+DNS+RazrPyzr+Content“ - Įjungiama turinio tikrinimo sistema
	„AV+DNS+RazrPyzr+Content+Bayes“ - Įjungiamas Bayes tikimybių filtras
** (SMTP|AV)	Pateikiama SMTP ir filtravimo serverio (AV) sisteminio resursų apkrovos reikšmės atskirai
*** Vidutinis tinklo apkrovimas	„/“ simboliu atskirta vidutinė įeinančio ir išeinančių tinklo srautų greičių reikšmės

3.1.1. El. pašto žinučių dydžiai

Pradedant eksperimentus nebuvo žinoma kokio dydžio elektroninių laiškų galima tikėtis, todėl priimamų laiškų dydžio limitas buvo padidintas iki 400 MB. Paprastai elektroninio pašto paslaugų tiekėjai šį limitą nustato daug mažesnį. Toks limitas buvo pasirinktas norint priimti visus laiškus. Iš analizės metu surinktos informacijos buvo aišku, kad pagal „Kaspersky Labs“ duomenis, didžioji dauguma nepageidaujamų laiškų yra santykinai mažo dydžio, tačiau tikslas buvo priimti absoliučiai viską ir įsitikinti ar realioje tarnybinėje stotyje galios tos pačios tendencijos.

Po kiekvieno eksperimento, kurio metu laiškai buvo siunčiami į el. pašto priėmimo bei filtravimo sistemą sukurta žurnalo byla „maillog“, buvo apdorojama „bash“ komandinės aplinkos scenarijumi, kuris išrenka unikalius priimtų žinučių identifikatorius, apskaičiuoja apdorojimo laiką ir žinutės dydį, tuomet duomenis patalpina „MySQL“ duomenų bazės lentelėse. Lentelės struktūra ir SQL užklausos statistikai surinkti pavaizduotos 12 lentelėje.

12 lentelė. Duomenų bazės lentelė ir SQL užklausos statistikai apskaičiuoti



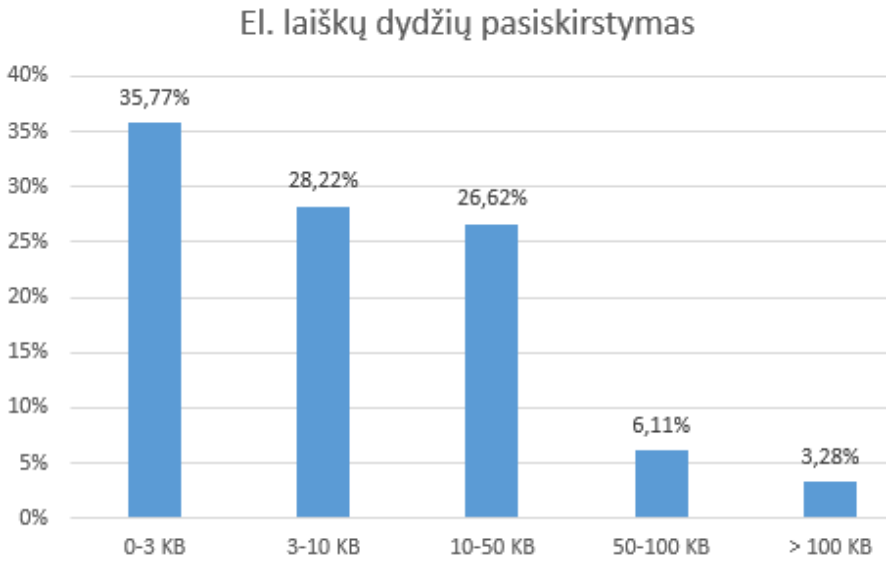

```
SELECT count(*), avg(laikas) FROM `exp1_cpu1_0`
WHERE dydis < 102400;
SELECT count(*), avg(laikas) FROM `exp1_cpu1_0`
WHERE dydis between 102400 and 512000;
SELECT count(*), avg(laikas) FROM `exp1_cpu1_0`
WHERE dydis between 512000 and 1024000;
SELECT count(*), avg(laikas) FROM `exp1_cpu1_0`
WHERE dydis >  1024000;
```

Atlikus statistinius skaičiavimus paaiškėjo, kad iš 30960 laiškų, didžioji dauguma (96 proc.) yra mažesni nei 100 KB ir tik 0,66 proc. surinktų laiškų yra didesni nei 500KB. Detalesni rezultatai pateikiami 15 paveiksle. Pastebėta, kad 2015 metų tendencija neprieštarauja Kaspersky Labs 2013 metais atliktam tyrimui ir galima daryti išvadą, kad nepageidaujamų laiškų siuntėjams vis dar neapsimoka siųsti didesnes žinutes su, galbūt, kenksmingu kodu.

Siunčiami laiškai yra maži, todėl, jiems siųsti naudojama mažiau tinklo perdavimo išteklių, o įvedus tam tikrą maišą žinutės tekste, jos tampa unikalios ir tokiu būdu lengviau apeiti nepageidaujamų laiškų filtrus. Tokiu būdu apeinami santraukos ar kontrolinių sumų principu paremti filtrai, kurie ieško jau žinomų nepageidaujamų laiškų centrinėse duomenų bazėse.

Žinant pasaulines tendencijas dėl žinutės dydžio dauguma antivirusinių, tiek nepageidaujamų laiškų filtravimo programinės įrangos paketų, pagal nutylėjimą, neapdoroja didesnių el. pašto bylų taip taupant sisteminius išteklius. Tokie sprendimai gali būti priimti pasvėrus riziką ir atmetus tikimybę, kad tokios bylose gali būti žalingo kodo ar turinio.

Eksperimentų metu „Postfix“ aktyvios eilės valdiklis pagal nutylėjimą nesiuntė didesnių nei 15 MB bylų į filtravimo serverį. Didžiausias „Spamassassin“ nepageidaujamų laiškų filtravimo sistemos aptiktas brukalas buvo 1 MB dydžio, todėl šis limitas nebuvo pasiektas.

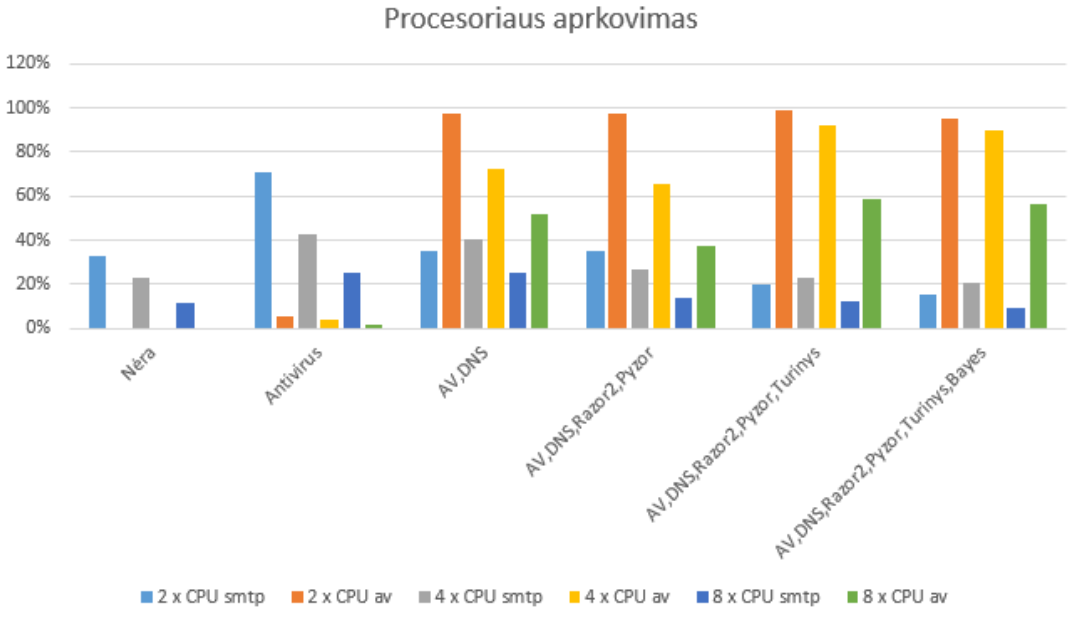


15 pav. El. pašto žinučių pasiskirstymas pagal dydį

Palyginimui, didžiausia žinutė, turinti žalingo kodo požymių buvo 371 KB dydžio ir turėjo prisegtą teksto redagavimo programos bylą.

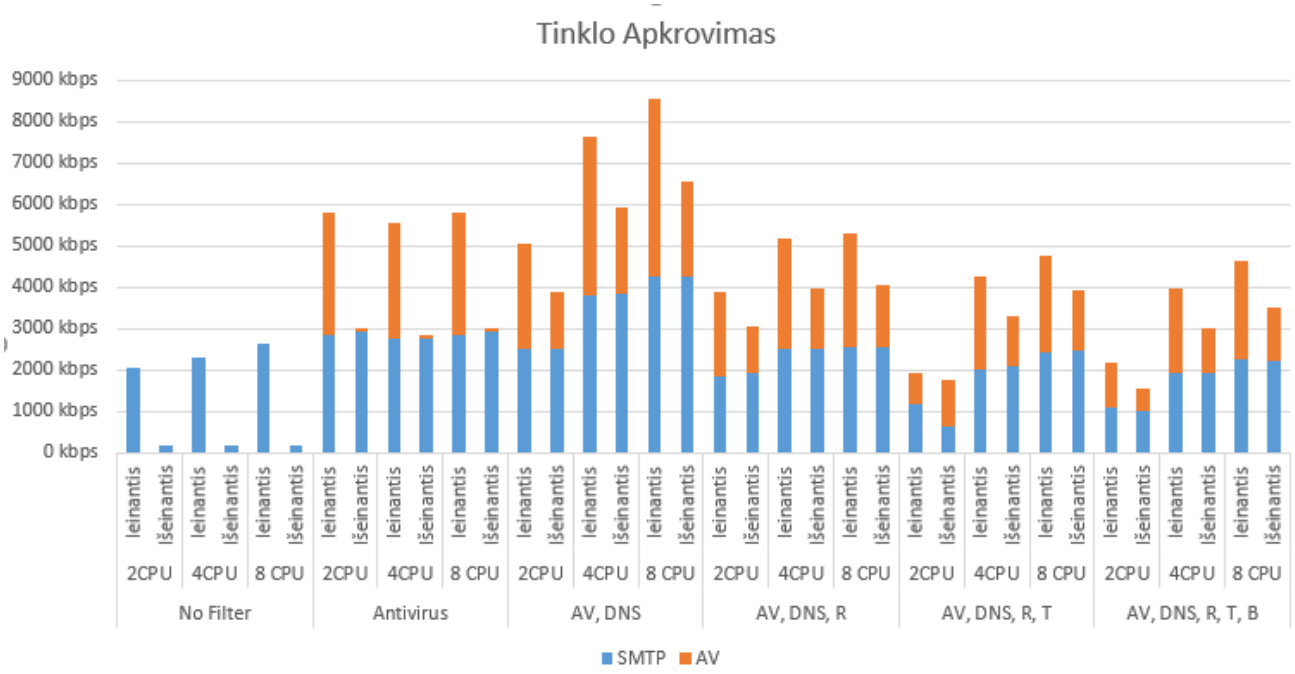
3.2. El. pašto filtrų įtakos sisteminiams resursams rezultatai

Iš surinktų duomenų sudarius virtualių mašinų resursų naudojimo grafikus, matyti, kad filtravimo funkciją atliekantis serveris naudojo daugiau skaičiavimo resursų nei už el. pašto priėmimą ir pristatymą atsakinga tarnybinė stotis. Taip pat akivaizdu, kad 2 procesorių operacinė sistema, įjungus „SpamAssassin“ filtravimo sistemą, pilnai išnaudoja procesorių išteklius. Panagrinėkime atskirus filtravimo metodus ir jų įtaką sisteminiams resursams (grafikuose „smtp“ yra el. pašto, „av“ yra filtravimo serveris):



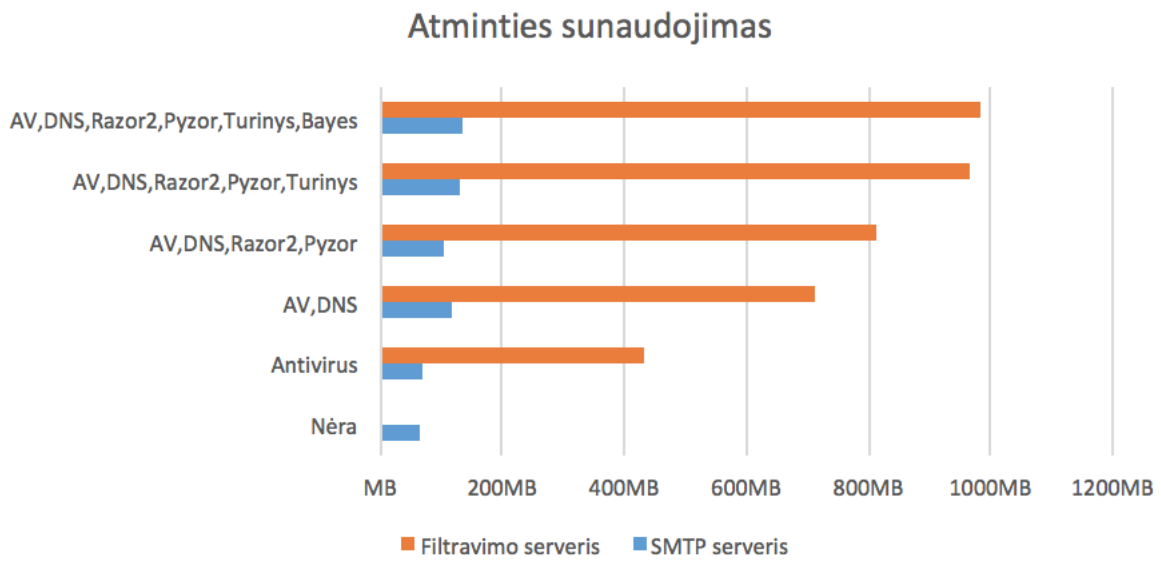
16 pav. Procesoriaus apkrovimas eksperimentų metu

- „ClamAV“ antivirusinis filtras veikia labai greitai ir didelės įtakos šia programine įranga laiškus apdorojančio serverio procesoriui neturi (16 pav.). Tuo tarpu SMTP serverio procesoriaus naudojimas išauga, tai galima būtų paaiškinti tuo, kad perduoti laišką skenavimo serveriui, gauti atsakymą, pristatyti laišką ir vesti veiksmų žurnalą, reikia daugiau skaičiavimo resursų nei ieškoti kenkėjiško kodo požymių. Ta pati tendencija išlieka su visomis procesorių konfigūracijomis. „ClamAV“ antivirusinis paketas veikia operatyvinėje atmintyje, serveriniame režime, todėl 18 paveiksle matosi, kad skenavimo serveryje, atliekant eksperimentus su šiuo filtru, stebimas aktyvus jos naudojimo padidėjimas. Tinklo sąsajos apkrova nuspėjama: SMTP serverio įeinantis tinklo srauto greitis yra lygus išeinančiam, tai reiškia, kad be vėlinimo gaunami laiškai yra perduodi skenavimo serveriui, kurio tinklo sąsajos rodikliai adekvačiai tai patvirtina (17 pav.).



17 pav. Tinklo apkrovimas eksperimentų metu

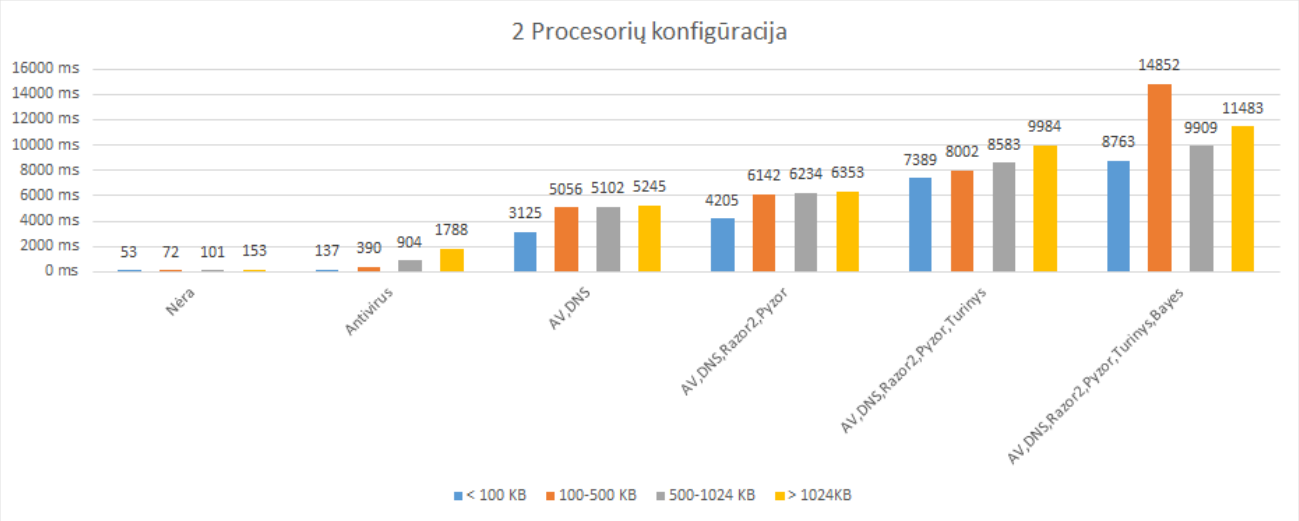
- DNS tipo filtrai, naudojami eksperimentuose yra dalis „SpamAssassin“ programinio paketo, todėl jo įgalinimas turi įtakos filtravimo serverio operatyvinei atminčiai, nes reikia paleisti dar vieną serverinį procesą ir užkrauti taisykles į atmintį. Norint analizuoti laiško perdavimo procese naudojamus tarpinius serverius skenavimo procesui reikia atidaryti ir analizuoti el. laiško turinį, šiuo atveju antraštę, ir siųsti užklausas į internete talpinamus juoduosius sąrašus. DKIM parašų, SPF įrašų tikrinimas reikalauja nemažai skaičiavimo ir dėl to padidina procesorių apkrovą. Tuo tarpu SMTP serverio procesorių apkrova sąlyginai sumažėja, palyginus su eksperimentu, kai buvo naudojamas tik antivirusinis filtras. Tai galima paaiškinti atsiradusiu vėlinimu skenavimo serveriui apdorojant žinutę laukiant atsakymo iš pasaulinių juodųjų sąrašų. Ryšio kanalo apkrovimą pasidaro kaip didžiausias iš visų eksperimentų dėl to, kad DNS užklausos ir atsakymai yra pakankamai trumpi ir greitai apdorojami, taigi kuo daugiau skaičiavimo galios naudojame, tuo greičiau apdorojame laiškus, tuo daugiau užklausų siunčiame trečiosioms šalims. 19 paveiksle matosi, kad serveris, apdorodamas laiškus DNS tipo filtrais sugaišta neproporcingai daugiau laiko, palyginus su sudėtingesniais ir, teoriškai, daugiau resursų reikalaujančiais metodais. Tokius rezultatus galima paaiškinti nutolusių serverių atsakymo laiko priklausomybe nuo pasaulinio tinklo būklės bei trečiųjų šalių paslaugų kokybės.

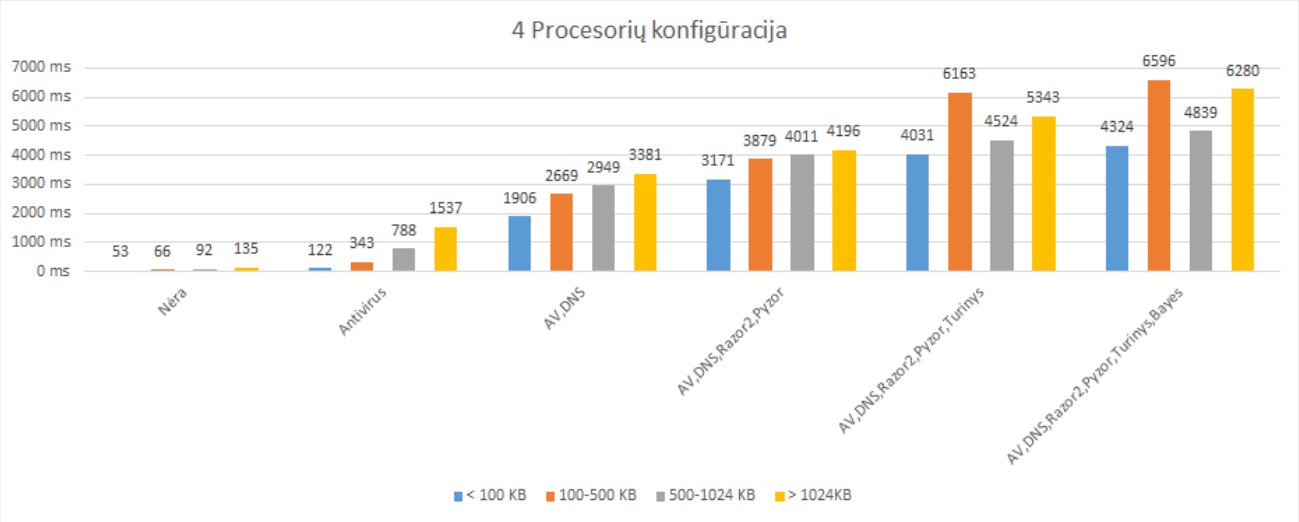


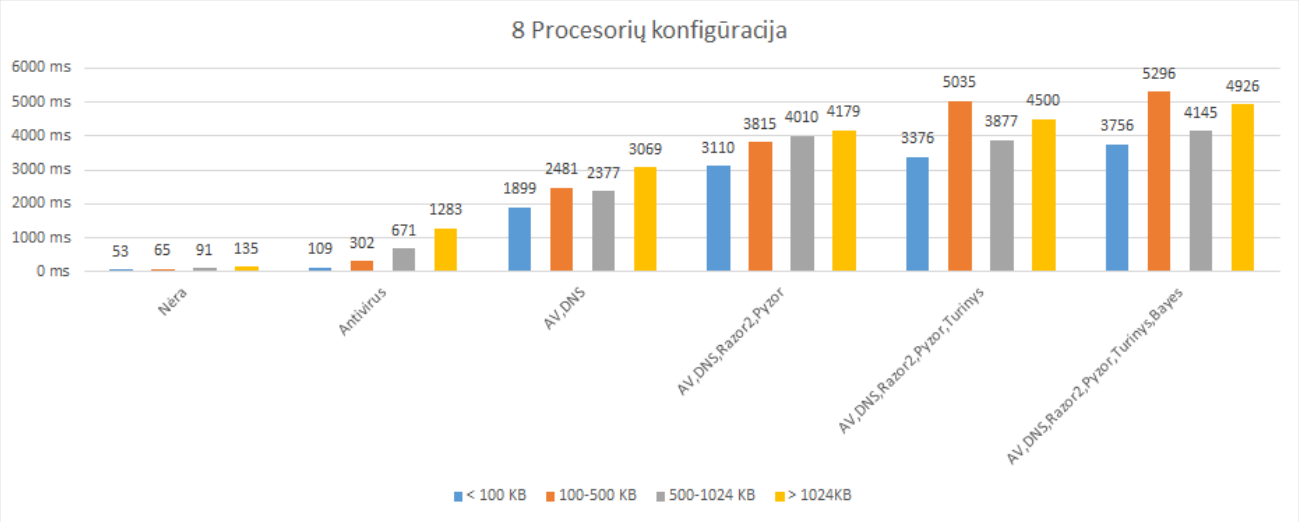
18 pav. Operatyvinės atminties naudojimas eksperimentų metu

- Razor2 ir Pyzor. Šių filtravimo mechanizmo įtaka skaičiavimo resursams yra minimali, turint omenyje, kad laiškas jau atidarytas ir perskaitytas apdorojimui DNS tipo filtrams. Kadangi kiekvieno laiško turinio maišos reikšmės užklausos apdorojimas nutolusioje duomenų bazėje užtrunka todėl ir ryšio kanalo santykinis momentinis pralaidumas yra mažesnis dėl padidėjusio vėlinimo iš trečiosios šalies. Reikia paminėti, kad šie metodai naudoja TCP protokolą maišos reikšmių tikrinimui, o šis protokolas yra, iš principo, lėtesnis už DNS užklausoms naudojamą UDP.
- Turinio filtrai. Laiško turinio filtravimui reikalinga daugiau atminties, nes reikia užkrauti reguliariųjų išraiškų reikšmes, HTML kodo pavydžius, paveiksliukų analizės variklio duomenis, taip pat nemokamo el. pašto paslaugas teikiančių tiekėjų vardų sričių sąrašus. Priešingai nei Razor2/Pyzor filtrai, šis filtravimo metodas reikalauja ir daugiau procesoriaus resursų dėl sudėtingų turinio apdorojimo taisyklių, o kadangi įjungiant vis daugiau apsaugos priemonių laiško apdorojimo laikas didėja – tinklo aprovimas santykinai mažėja. Matome, kad turint mažiausią procesorių konfigūraciją skenavimo serveryje tinklo apkrovimas ženkliai nukrenta, nes skaičiavimo procedūros sukelia žymų vėlinimą ir atitinkamai „paliekant“ daugiau laiko tinklo prieigos reikalaujančioms užduotims.
- Bajeso tikimybių ir statistikos pagrindu grįstas filtras iš esmės nedaro įtakos operatyvinės atminties naudojimui tačiau procesorių apkrovimas – padidėja. 19 paveiksle pastebėtas ir šiek tiek padidėjęs apdorojimo vėlinimas, kuris, savo ruožtu, sumažina perduodamų duomenų greitį. Eksperimentų metu buvo naudojama 100 „gerais“ ir „blogais“ el. pašto laiškais apmokytas Bajeso variklis. Šis variklis duomenis saugo „Berkeley“ duomenų bazėje, kuri yra užkraunama į atmintį. Ištrynus šia duomenų bazę – pastebėtas operatyvinės atminties naudojimo sumažėjimas, taigi turint didelę duomenų bazė tikėtinas reikšmingesnis operatyvinės atminties užimtumas.

Apibendrinant filtravimo metodų įtaką kompiuterinės sistemos resursams galima būtų rekomenduoti minimalius reikalavimus sistemoms. Iš rezultatų galima spręsti, kad el. laiškų sistemai priimanti 10-17 laiškų per sekundę, kurių dydžių pasiskirstymas pavaizduotas 15 paveiksle, reikalingi bent jau 2 procesoriai ir apie 200 MB operatyvinės atminties. Filtravimo serveriui rekomenduotina apie 1GB operatyvinės atminties ir mažiausiai 4 procesoriai.





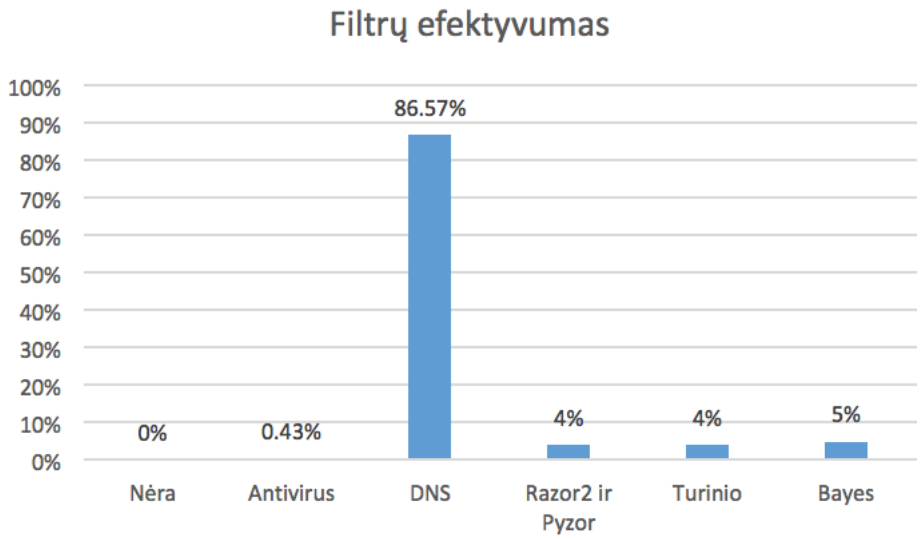


19 pav. Laiškų apdorojimo laiko priklausomybė nuo įjungtų filtrų tipo iš laiško dydžio

3.3. El. pašto filtrų efektyvumo tyrimo rezultatai

Apdorojus eksperimentų duomenis filtrų efektyvumo atžvilgiu t.y. kuris filtras aptiko daugiausiai nepageidaujamų laiškų nustatyta, kad DNS tipo filtrais buvo identifikuota apie 87 proc. brukalo (20 paveikslas). Iš šių rezultatų galima daryti prielaidą, kad didžioji dauguma internete perduodamo šlamšto yra siunčiama iš žinomų šaltinių, kurie gana greitai yra patalpinami į

pasaulinius juoduosius sąrašus, neturi SPF arba DKIM/DMARC politikos arba ji yra neteisinga. Kadangi DNS užklausos į juoduosius sąrašus, iš anksčiau aprašytų pastebėjimų, yra vykdomos pakankamai greitai ir laiško apdorojimas užtrunka trumpą laiko tarpą, peršasi išvada, kad DNS tipo filtras „SpamAssassin“ programiniame pakete yra efektyviausias. Atliekant eksperimentus šis filtras buvo naudojamas kaip pirmasis ir būtent su juo buvo paleidžiamas „SpamAssassin“ serveris, todėl atminties sunaudojimas atrodo tik santykinai didžiausias.



20 pav. Santykinis el. pašto filtrų efektyvumas

3.4. El. pašto filtrų įtakos sisteminiams ir jų efektyvumo apibendrinimas

Atlikus tyrimą paaiškėjo, kad 62 proc. laiškų yra identifikuojami kaip nepageidaujami. Palyginus su analizės metu surinktais duomenimis kur 2012 metais pasaulinis nepageidaujamų laiškų kiekis yra 69 proc., 2013 – 66 proc. galima daryti išvadą, kad šiuo atžvilgiu tendencija yra mažėjanti. Didžioji dauguma laiškų yra mažesni nei 100KB, tai taip pat išlaiko ryšį su globaliais rodikliais.

Iš visų nepageidaujamų laiškų tik 0.43 proc. buvo aptikti su žalingo kodo požymiais. Dauguma populiariausių modernių operacinių sistemų arba jau turi integruotą apsaugą nuo virusų, o ir žmonių įprotis atidarinėti prisegtas bylas iš nežinomų šaltinių, sąlygoja, mažą tokio brukalo populiarumą. Dėl šių priežasčių atrodytų, kad galima būtų mąstyti apie tokio tipo filtrų atsisakymą, tačiau reikia prisiminti, kad virusai gali pridaryti daug daugiau žalos nei nepageidaujami laiškai.

Eksperimentais nustatyta, kad efektyviausias filtras yra DNS tipo juodųjų sąrašų tikrinimas, o tai parodo, kad dauguma nepageidautino turinio atkeliauja iš žinomų ir dažniai pasikartojančių šaltinių, kurie jau būna užregistruoti ir klasifikuoti.

Žinant santykinį laiškų dydžių pasiskirstymą, apdorojimo trukmę, procesorių, operatyvinės atminties ir tinklo resursų įtaką filtravimo greičiui, taip pat filtrų tipų efektyvumo duomenų turėtų pakakti el. pašto filtravimo sistemos imitacinio modelio parametrų nustatymui ir palyginimui su realia tarnybine stotimi.

4. EL. PAŠTO FILTRAVIMO SISTEMOS IMITACINIO MODELIO KŪRIMAS

„Rockwell Software ARENA“ yra nesusijusiems įvykiams modeliuoti skirta programinė įranga. Šios programinės įrangos gamintojas suteikia galimybę studentams naudotis nemokama versija, kuri neriboja funkcionalumo, tačiau komercinis jos taikymas yra draudžiamas.

4.1. ARENA imitacinio kompiuterinio modeliavimo komponentų pritaikymas

Modeliavimo aplinka leidžia kurti imitacinius modelius naudojant grafinę vartotojo sąsają nerašant programinio kodo. Grafiškai sukurtas modelis automatiškai yra paverčiamas į „SIMAN“ programinį kodą ir įvykdomas. Pagrindiniai komponentai ARENA kompiuterinio modeliavimo aplinkoje, kurie bus panaudoti modeliui:

- Esybės (angl. *entities*) – modelio objektas, judantis per modulius ir keičiantis savo būseną. Šie objektai gali sužadinti kitus komponentus, juos pakeisti arba pasikeisti patys. Paprastai jos yra sugeneruojamos ir sunaikinamos. El. pašto filtravimo sistemos modelyje elektriniai laiškai bus esybėmis.
- Atributai (angl. *attributes*) – kintamieji priskiriami esybėms. Kiekviena esybė keliaujanti sistema gali turėti savo unikalius atributus, kurie gali keistis imitacijos eigoje. Modelio kontekste elektroninių laiškų dydis, viruso ar nepageidaujamo laiško antraštės, bus atributai priskirti laiško esybei.
- Eilės (angl. *queues*) – esybės modelyje gali sustoti ir negalėti judėti dėl įvairių priežasčių, tokiu atveju esybės „stoja“ į eilę ir laukia kol galės pajudėti. Projekto modelio kontekste SMTP serverio laiškų eilė bus modeliuojama šio komponento pagalba.
- Resursai (angl. *resources*) – apdorojamos esybės gali naudoti resursus, pavyzdžiui aptarnaujamas laiškas gali naudoti filtravimo proceso „darbuotoją“ ir kol yra apdorojamas, resursas yra užimtas ir yra atlaisvinamas kai apdorojimas yra baigtas.
- Imitacinis laikrodis (angl. *simulation clock*) – imitacinėje aplinkoje laikas gali būti pagreitintas, palėtintas, atsukamas atgal ir panašiai.
- Moduliai (angl. *modules*) – turbūt svarbiausi imitacinės aplinkos komponentai. Įvairūs moduliai atlieka skirtingas užduotis, jie gali generuoti esybes, jas naikinti, apdoroti, sulaikyti, apjungti ir panašiai. Praktiškai dauguma esybės veiksmų, užlaikymų, sprendimų bus vykdoma šių struktūrinių modelių pagalba.
- Statistikos kaupikliai (angl. *statistics collection*) – ARENA aplinka pagal nutylėjimą kaupia ir atiduoda daug statistinės informacijos apie asybes, jų apdorojimą, eiles ir panašiai. Kartais prireikia apskaičiuoti nenumatytas statistikas kaip pvz.: vidutinis generuojamos esybės atributo reikšmės vidurkis (elektroninio laiško bylos vidutis dydis).
- Gedimai (angl. *failures*) – simuliuojant įrengimą, netgi kompiuterines sistemas kiekvienas iš sistemos komponentų, turi vidutinį veikimo laiką, po kurio jis sugenda ir turi būti pakeistas. Šis programinės įrangos aplinkos komponentas leidžia nustatyti gedimo laiką, tikimybę ir uždelsimo laiką, kuris reikalingas komponentui pakeisti.

ARENA aplinka turi ir papildomų komponentų, kurie gali būti dinamiškai užkrauti iš išorinių bibliotekų, tokiu būdu praplečiant modeliavimo logikos galimybes, tačiau projekto modeliui bus naudojami standartiniai ir integruoti aukštesnio lygio komponentai.

4.2. Loginio modelio įgyvendinimas kompiuteryje, imitacinio modelio kūrimas.

Remiantis laiškų filtravimo sistemos imitacinio modelio metodu ir 11 paveikslo logine schema buvo sukurtas imitacinis modelis, leidžiantis imituoti laiškų kelią elektroninio pašto priėmimo, filtravimo ir pristatymo sistemose.

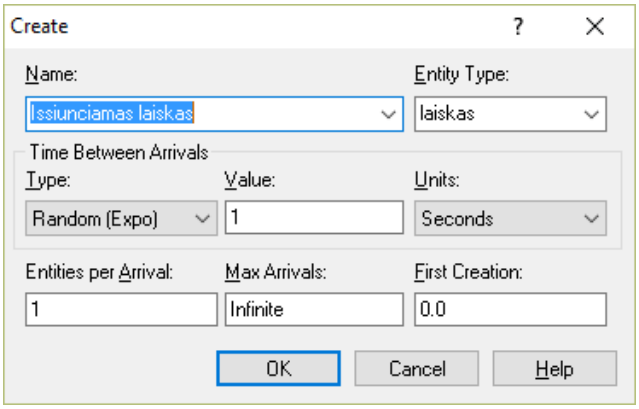
Pagrindinė esybė – elektroninis laiškas, aprašomas modeliavimo aplinkoje, jam priskiriamas voko paveiksliukas, nustatoma statistikos rinkimo apie esybę žyma (21 pav.), kiti nustatymai paliekami nepakeisti, nes jie nebus naudojami modelyje.

Entity - Basic Process

	Entity Type	Initial Picture	Holding Cost / Hour	Initial VA Cost	Initial NVA Cost	Initial Waiting Cost	Initial Tran Cost	Initial Other Cost	Report Statistics
1	laiskas	Picture.Envelope	0.0	0.0	0.0	0.0	0.0	0.0	☑

21 pav. Esybės nustatymai

Esybė sukuriama įtraukiant kūrimo (angl. *create*) modulį į modelį. Kūrimo modulis leidžia pasirinkti atvykimo laikų intervalus, šiuo atveju, po kiek ir kokiu intervalu bus generuojami laiškai. Kaip pradinį pasirinkimą nustatoma, kad vidutiniškai po vieną laišką bus gaunama kas sekundę. Vėliau ši reikšmė bus keičiama, pritaikant prie realios situacijos arba norimos imitacijos. 22 paveiksle parodyti



22 pav. Laiško (esybės) kūrimas

esybės kūrimo nustatymai. *Random(Expo)* yra eksponentinė pasiskirstymo funkcija. Didžiausias generuojamų esybių kiekis nustatomas kaip begalinis, tai reiškia, kad simuliaciją, esant reikalui bus nutraukiama rankiniu būdu. Pirmosios esybės kūrimas pradedamas iš karto, be užlaikymo (0.0).

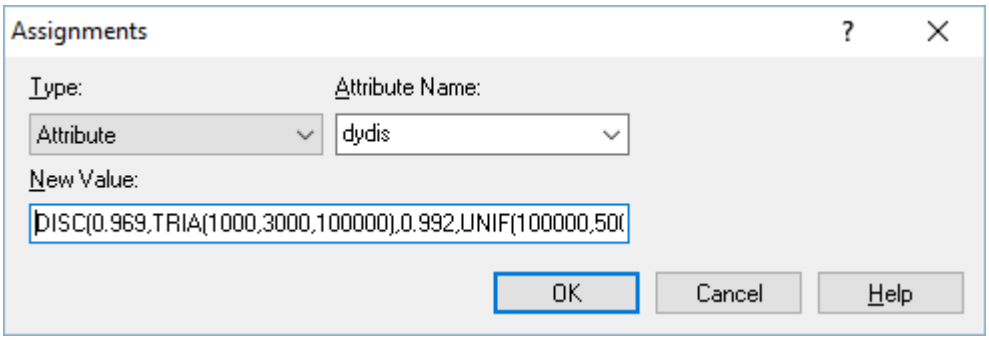
Eksperimentų realybėje metu buvo nustatytas tam tikras pasiskirstymas tarp laikų dydžių (3.1.1 skyrius). Elektroniniam laiškui, arba esybei, reikia priskirti atributą „dydis“, kuris aprašomas atributų komponente (23 pav.), taip pat parenkamas realiųjų skaičių aibės tipas.

Attribute - Basic Process

	Name	Rows	Columns	Data Type	Initial Values
1	dydis			Real	0 rows

23 pav. Laiško dydžio atributas

Atributo ir jo reikšmės priskyrimas vykdomas priskyrimo modulio (angl. *assign*) pagalba.



24 pav. Atributo priskyrimas esybei

Daugeliu atvejų ARENA modeliavimo aplinkoje reikšmę galima aprašyti funkcija arba išraiška. Elektroninio laiško esybei priskiriamas dydžio atributas kurio išraiška yra: *DISC(0.969,TRIA(1000,3000,100000),0.992,UNIF(100000,500000),0.998,UNIF(500000,1024000),1.0,UNIF(1024000,1465130))*, ir įrašomas kaip reikšmė (pavaizduota 24 paveiksle). Šioje išraiškoje panaudotos trys atsitiktinio pasiskirstymo funkcijos, kurių pagalba aprašomas laiško dyždio atributas. Išraiška apibrėžia tikimybinį skirstinį kuris apie 97 proc. atvejų pasiskirstys pagal trikampio skirstinį: dažniausia pasitaikanti reikšmė bus 3000, mažiau 1000 ir rečiausia – 100000. Ppie 2 proc. tolygiuju skirstiniu reikšme nuo 100000 iki 50000, likusi, mažai tikėtina dalis, nuo 500000 iki 1024000 ir nuo 1024000 iki 1465130. Ši išraiška buvo išvesta naudojant realybėje atliktų eksperimentų metu surinktų laiškų statistiką, tačiau validavimo metu pastebėta, kad modelio generuojami laiškai yra maždaug 50 KB dyžio ir neatitinka realybės, todėl išraiška buvo patobulinta į: *DISC(0.3577,UNIF(500, 2999), 0.6399,UNIF(3000, 9999), 0.9061,UNIF(10000,49999), 0.9672,UNIF(50000,99999), 1.0,UNIF(100000,1465130))* ir tiksliai atitiko statistiką tikrovėje (validavimo rezultatai pateikti šiame darbe 4.2.3.1. skyriuje).

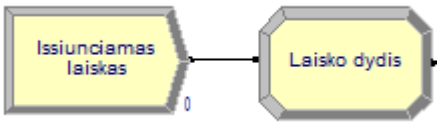
Panaudotų pasiskirstymo funkcijų aprašymai pateikiami žemiau:

DISC (angl. *discrete*). Nesusijusių įvykių tikimybės funkcija. SIMAN modeliavimo kalba generuoja skaičių nuo 0 iki 1 ir ieško tikimybės reikšmės. Pavyzdžiui: funkcija *disc(0.3,1, 0.8,2, 1.0,3)* grąžins skaičių „1“ su 30 proc. tikimybe, kad skaičius „2“ bus grąžinamas maždaug 50 proc., o „3“ – 20 proc.

TRIA (angl. *triangle*). Atsitiktinis trikampio skirstinys. Trikampis skirstinys gali būti naudojamas, jei yra žinomi pasiskirstymo minimali ir maksimali reikšmės, ir bent apytikriai – moda, t.y. dažniausiai pasikartojanti reikšmė.

UNIF (angl. *uniform*). Tolygusis skirstinys modeliuojant naudojamas, jeigu yra žinoma, kad imties reikšmės yra pasiskirsčiusios tarp tam tikrų ribų, tačiau nėra žinoma, ar vienos reikšmės įvyksta dažniau nei kitos.

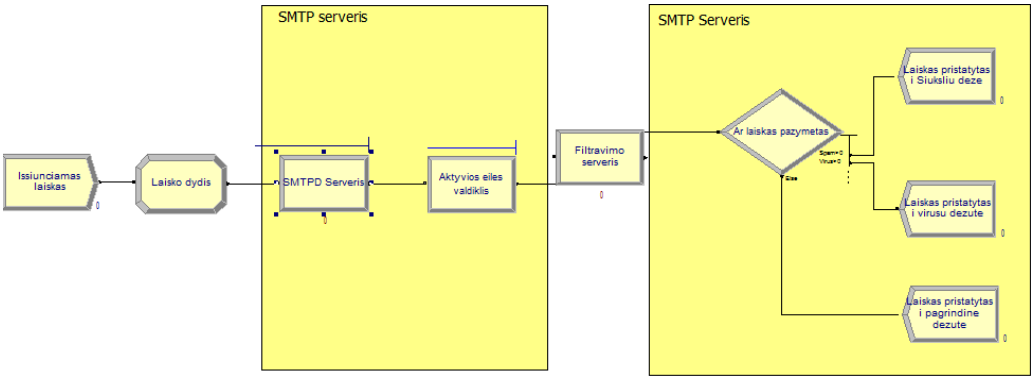
Apibendrinant atliktą darbą šioje stadijoje jau turima procesą, kuris pasirinktu dažnumu generuoja esybes, joms priskiria unikalius, pagal realius eksperimentus nustatytus atributus (25 pav.).



25 pav. Laiško generavimo ir dydžio priskyrimo moduliai

4.2.1. SMTP Serveris

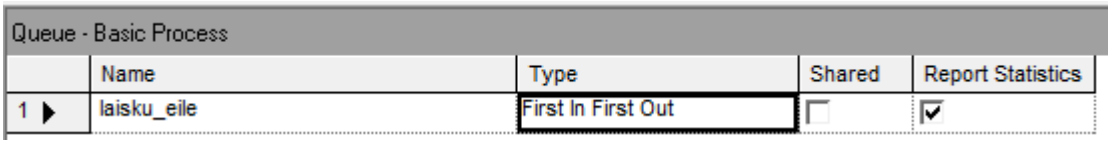
Modeliuojamo SMTP serverio paskirtis sistemoje yra laiškų priėmimas, eilės valdymas, laiškų perdavimas filtravimo sistemai, laiškų rūšiavimas po filtravimo ir pristatymas į vieną iš trijų el. pašto dėžučių: virusų karantino, nepageidaujamų laiškų ir numatytąja – normaliems laiškams. Loginė SMTP serverio veiksmų schema, kaip sumodeliuota ARENA aplinkoje, parodyta 26 paveiksle.



26 pav. SMTP serverio modelio schema

Eksperimentų metu nustatyta, kad nepriklausomai nuo procesorių skaičiaus, laiškų priėmimas vyksta labai greitai. Įtakos gali turėti „Postfix“ priėmimo posistemė arba tai, kad nebuvo pasiektas kritinis apkrovimas, po kurio vėlinimas išaugtų. Buvo išvestas vidurkis, kad laiško priėmimas užtrunka maždaug tiek laiko, kiek laiško dydį baitais padalinus iš 80000000. Vėlinimui modeliuoti buvo panaudotas proceso modulis. Numatytoji proceso loginė veiksmų seka yra užimti išteklius, vėlinti, ir atlaisvinti išteklius. Šiuo atveju procesoriaus vėlinimo eksperimentiniu būdu nustatyti nepavyko, o tinklo vėlinimą apibrėžia laiško dydžio daliklis iš nustatytos konstantos. Kadangi modelyje nenaudojami išteklių komponentai – proceso modulio veiksmas pakeičiamas į uždelsimą (angl. *delay*).

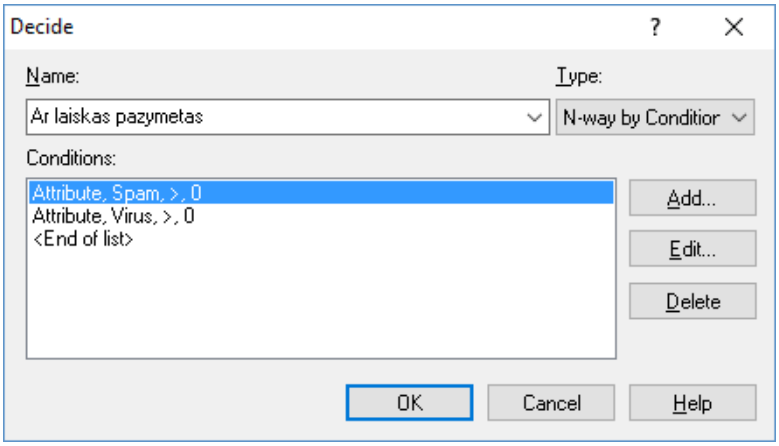
„Postfix“ el. pašto serverio aktyvios eiles valdikliui modeliuoti buvo pasirinktas užlaikymo (angl. *hold*) modulis, šis modulis leidžia nustatyti eilės kaupimosi priežastį be privalomo išteklių naudojimo. Užlaikymo modulis nustatomas pradėti kaupti eilę tik tada, kai visų filtravimo serverio procesų apdorojamų esybių suma yra daugiau nei apibrėžti ištekliai, kurie apibrėžiami kaip atskiras duomenų masyvas (angl. *resource*). Užlaikymo modulis nustatomas būtinai laukti priežasties prieš pradedant kaupti eilę. Pilna vėlinimo modulio priežasties išraiška apibrėžiama taip: *„(ClamAV Filtras.WIP + ClamAV Filtras 2.WIP + ClamAV Filtras 3.WIP + ClamAV Filtras 4.WIP + Filtro procesas 1.WIP + Filtro procesas 2.WIP + Filtro procesas 3.WIP + Filtro procesas 4.WIP) < MR(Filtro procesas)“*. Naudojami kintamieji yra apibrėžti filtravimo serverio modelyje, kuris bus aprašytas atskirame skyrelyje, o laiškų eilė apibrėžiama kaip „fifo“ (angl. *first in, first out*) pirmas į vidų – pirmas į išorę ir reiškia principą, kad tas, kas pirmiau patenka į eilę, pirmiau iš jos ir išeina. Eilės nustatymai parodyti 27 paveiksle.



27 pav. Laiškų eilės nustatymai

Daugumoje modulių nustatymų galima pastebėti priskyrimo (angl. *allocation*) skiltis, kuriose galima pažymėti ar esybė keliaudama per modulį ir vėlinama sukuria vertę. Ši skiltis skirta ataskaitos formavimui ir tinka verslo procesų modeliavimui. Statistikos nustatymai, reikalingi modelio validavimui ir bus sukurti atskirai, prisitaikant prie modelio ir reikalingų ištirti parametrų.

Filtravimo serverio sistema laiškams priskiria filtravimo rezultato atributus, pagal juos SMTP serveris turi nuspręsti į kurią dėžutę laiškas turi būti pristatomas. Šią funkciją modeliavimo aplinkoje atlieka sprendimo modulis. Sprendimo modulis nustatomas tikrinti „Spam“ ir „Virus“ atributų reikšmes (28 pav.). Jeigu kažkuri nepageidaujamo laiško atributo reikšmė didesnė už „0“, sprendimo modulis atitinkamai parenka laiško kelią iki dėžutės.



28 pav. Sprendimo modulio nustatymai

Susiklosčius situacijai kai laiškas yra klasifikuojamas kaip virusas ir kaip nepageidaujamas t.y. abu atributai pažymėti – laiškas bus pristatomas į nepageidaujamų laiškų dėžutę, tokia pat rūšiavimo tvarka buvo nustatyta ir realioje tarnybinėje stotyje.

Esybių sunaikinimui, arba laiško pristatymui į dėžutę modeliuoti, naudojamas sunaikinimo (angl. *dispose*) modulis. Moduliui priskiriamas pavadinimas ir nustatoma, kad jis rinktų statistiką apie iš sistemos išeinančius laiškus.

Iš sistemos išėję, sunaikintos esybės, baigia imitacinio modelio kelią palikdamos statistinius duomenis ataskaitų generavimo posistemėje. Šie duomenys bus analizuojami modelio validavimo metu. Užbėgant už akių, sukuriama statistikos modulio elementų, kurie gali būti naudingi analizuojant modelio veikimą. Vidutinio laiškų dydžio, apdorojimo laiko statistinių duomenų nustatymai pavaizduoti 29 paveiksle. Tiek dydžio, tiek apdorojimo laiko statistikos tipas parenkamas kaip priklausantis nuo laiko. Vidutinio apdorojimo laiko reikšmė nustatoma „TAVG“ funkcija, kuri skaičiuoja vidutinį laiką iš visų esybių vidutinės „TotalTime“ (angl. *visas laikas*) reikšmės. „Total Time“ yra esybės užtruktas laikas sistemoje.

Statistic - Advanced Process

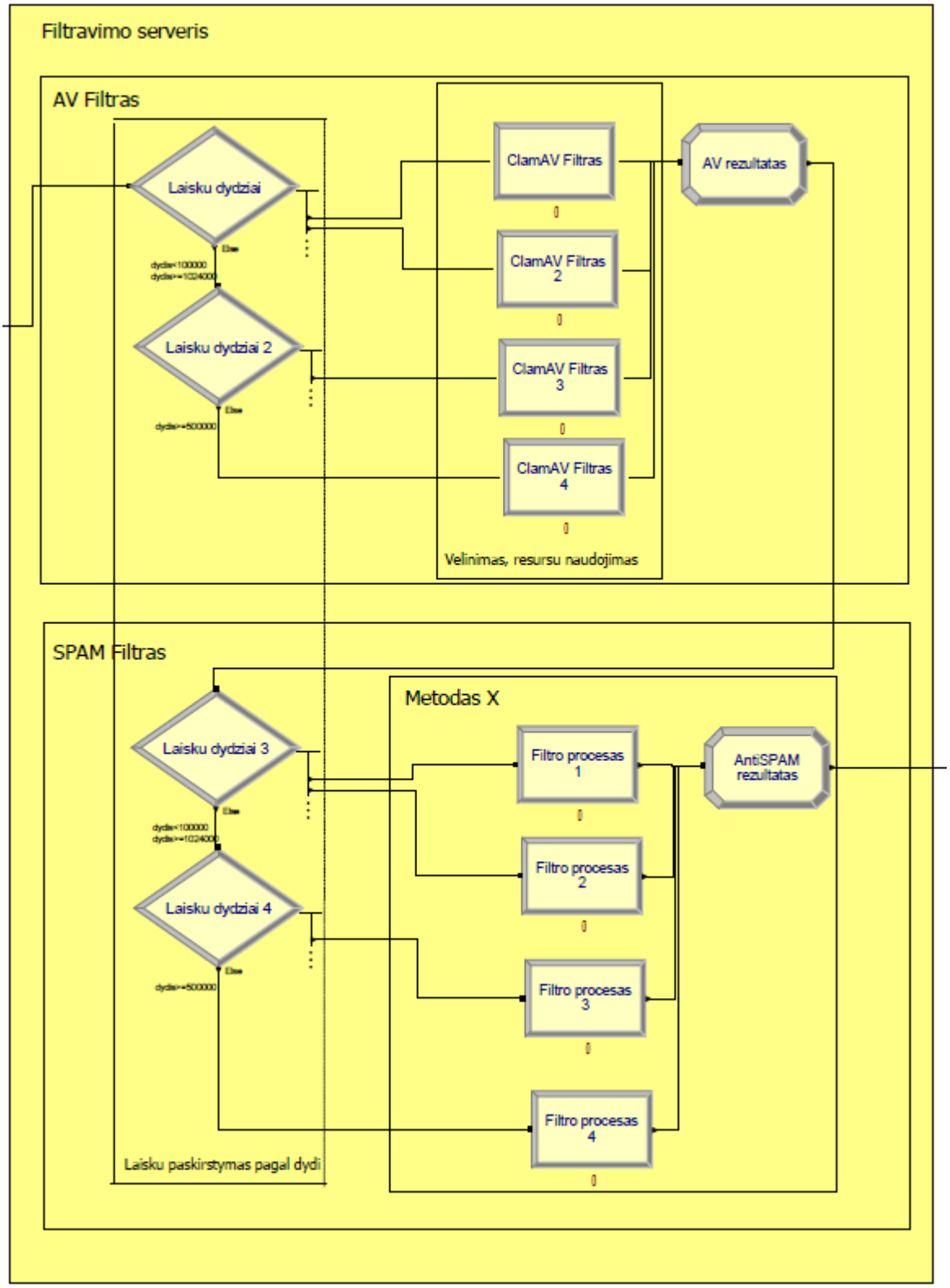
	Name	Type	Counter Name	Limit	Expression	Collection Period	Initialization Option	Counter Output File	Report Label
1 ▶	Laisku dydis	Time-Persistent	Counter 2		dydis	Entire Replication	Replicate		Laisku dydis
2	Apdorojimo laikas	Time-Persistent	Counter 3		TAVG(laiskas.TotalTime)	Entire Replication	Replicate		Apdorojimo laikas

29 pav. Papildomos statistikos nustatymai

4.2.2. Filtravimo serveris

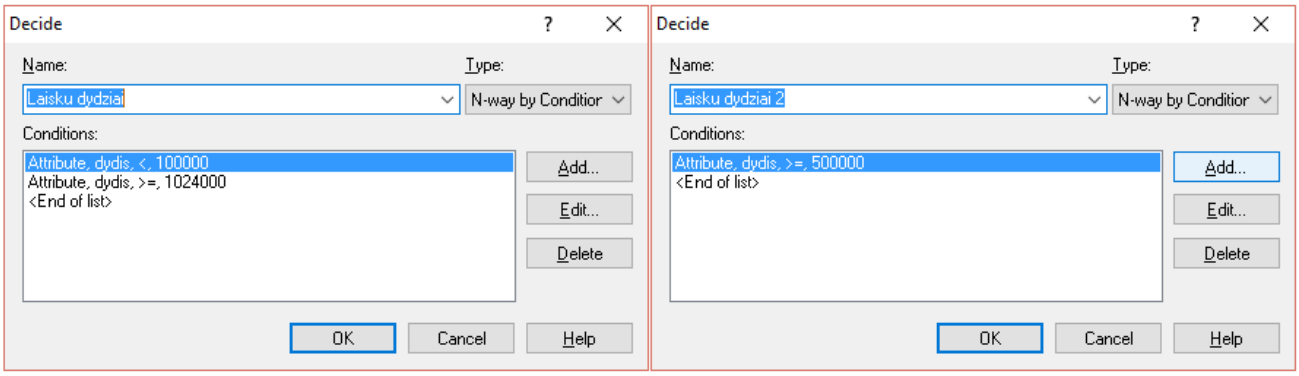
Filtravimo serveris yra sudėtingiausia modeliuojamos sistemos dalis, šios sistemos dalies modelis buvo sukurtas iš dviejų komponentų: antivirusinio filtro ir nepageidaujamų laiškų filtro (30 pav.). Dėl tam tikrų modeliavimo aplinkos logikos apribojimų, laiškų paskirstymui pagal dydį panaudota „sprendimo“ modulių sistema, paskirstanti laiškus į skirtingo vėlinimo filtravimo procesus. Pagal 6.1 priede eksperimentų metu surinktus duomenis laiškai buvo skirstomi į 4 dydžių grupes: mažiau nei 100 KB, 100 – 500 KB, 500 – 1024 KB ir virš 1024 KB. Kiekvienai dydžių grupei buvo priskiriamas vėlinimo procesas, o vėlinimo reikšmė priklauso nuo laiško dydžio, procesorių kiekio filtravimo serveryje, bei filtravimo metodo. Modeliavimo metu, esybės arba laiškai yra generuojami dinamiškai pagal nekintančią atsitiktinio paskirstymo išraišką, antivirusinio filtravimo vėlinimas turi būti keičiamas priklausomai nuo modeliuojamoje situacijoje įsivaizduojamų procesorių skaičiaus. Nepageidaujamų laiškų filtro vėlinimas priklauso ne tik nuo modeliuojamų procesorių skaičiaus, tačiau ir nuo modeliuojamo filtro tipo, todėl keičiant modelio filtrų tipą, turi būti atitinkamai nustatomi ir filtrų vėlinimai.

Atkartojant realios tarnybinės stoties veikimą sukurtame modelyje, filtravimo metodai yra sujungti nuosekliai ir filtravimo rezultatai yra įrašomi į atitinkamą esybės atributą. Nepageidaujamų laiškų filtras yra iškviečiamas nepriklausomai nuo antivirusinio filtro rezultato. Sprendimo modulių veikimo logika pritaikoma pirmajai sąlygai, kitos sąlygos nebėra vertinamos, todėl viename modulyje galima naudoti tik dvi sąlygas laiškų pagal dydį paskirstymui, pvz.: pirmasis modulis paskirsto laiškus „mažesnius“ už 100KB ir didesnius nei 1024000, jei nei viena, nei kita sąlyga netenkinamos – siunčiama į antrąjį sprendimo modulį. Antrasis modulis paskirsto didesnius arba lygius 500 KB ir kitus likusius, tokiu būdu padengiamos visos keturios užsibrėžtos laiškų dydžių grupės. Sprendimo modulių nustatymai pavaizduoti 31 paveiksle.



30 pav. Filtravimo serverio modelio schema

Laiškams patekus į atitinkamą filtro proceso modulį yra vykdoma uždelsimo logika. Užlaikymas šiame modelyje apibrėžiamas kaip Gauso, arba kaip „normalusis“ skirstinys. Pagal normalųjį skirstinį laiško dydžių reikšmes dažniausiai pasiskirsto dydžius, kuriuos veikia labai daug nepriklausomų veiksnių, kurių kiekvienas turi nežymią įtaką ir sukelia nedidelį vertės pokytį.
„Normalusis“ skirstinys taikomas, nes jis gerai apibūdina matavimo paklaidą, kuri nustatyta eksperimentų metu: *0,15ms* antivirusiniam filtravimo metodui ir *100ms* nepageidaujamų laiškų filtravimo varikliui.



31 pav. Laiškų paskirstymas pagal dydį

Eksperimentais nustatyto užlaikymo prie skirtingų nustatymų reikšmių lentelė pateikiama 6.1 priede, o pavyzdys kaip užlaikymas nustatomas filtrų procesų moduliuose – 32 paveiksle.

Process - Basic Process

	Name	Type	Action	Delay Type	Units	Allocation	Expression	Report Statistics
1	SMTPD Serveris	Standard	Delay	Expression	Seconds	Value Added	dydis/80000000	☑
2	ClamAV Filtras 3	Standard	Delay	Expression	Seconds	Value Added	NORM(904,0.15)/1000	☑
3	ClamAV Filtras 2	Standard	Delay	Expression	Seconds	Value Added	NORM(1788,0.15)/1000	☑
4	ClamAV Filtras	Standard	Delay	Expression	Seconds	Value Added	NORM(137,0.15)/1000	☑
5	ClamAV Filtras 4	Standard	Delay	Expression	Seconds	Value Added	NORM(390,0.15)/1000	☑
6	Filtro procesas 1	Standard	Delay	Expression	Seconds	Value Added	NORM(2988,100)/1000	☑
7	Filtro procesas 2	Standard	Delay	Expression	Seconds	Value Added	NORM(3457,100)/1000	☑
8	Filtro procesas 3	Standard	Delay	Expression	Seconds	Value Added	NORM(4198,100)/1000	☑
9 ▶	Filtro procesas 4	Standard	Delay	Expression	Seconds	Value Added	NORM(4666,100)/1000	☑

32 pav. Filtravimo procesų užlaikymo reikšmės

Filtravimo rezultatų priskyrimas vykdomas naudojant priskyrimo modulius, priskiriant „virus“ ir „spam“ atributų reikšmes iš filtravimo procesų atėjusiems laiškams. Kadangi iš visų eksperimentais tirtų laiškų tik 0,27 proc. turėjo žalingo kodo požymių priskyrimui, naudojama nesusijusių įvykių tikimybės funkcijos išraiška *DISC(0.00271, 1, 1.0, 0)*. Tokiu būdu nustatytas priskyrimo modulis priskirs „virus“ atributo reikšmę „1“ 0,27 proc. atvejų, visais kitais atvejais bus priskiriama „0“. Modelis gali būti nustatytas imituoti skirtingus nepageidaujamų laiškų metodus keičiant nepageidaujamu laiškų filtro procesų vėlinimo reikšmes pagal eksperimentų metu surinktus duomenis. 32 paveiksle yra modeliuojamas DNS filtravimo metodas virtualioje operacinėje sistemoje, kurioje veikia 2 procesoriai, o pats metodas aptinka 16878 laiškus iš 30960, kas yra 54,5 procentai visų laiškų. Rezultato priskyrimo išraiška tokiam filtrui nustatoma kaip: *DISC(0.545, 1, 1.0, 0)*.

Laiško priėmimo, apdorojimo ir pristatymo komponentai sudaro pilną nepageidaujamų laiškų filtravimo sistemos modelį, leidžiantį imituoti filtravimo metodų veikimą. Tokio modelio pagalba galima vykdyti simuliaciją prieš priimant sprendimą realios tarnybinių stoties įsigijimui ar projektavimo metu. Šis modelis buvo verifikuotas – realizuotas skaičiuojamasis modelis buvo patikrintas ar atitinka loginį modelį ir ar jame nėra programavimo sintaksės ar parametrų nustatymo klaidų. Modeliavimo aplinkoje „Arena“ buvo paleista simuliacija, kurios imitacinis laikrodis pagreitintas iki 24 valandų nenutrūkstamo veikimo. Papildomi kompiuterinės grafikos elementai vaizduojantys laiškų dydžius ir atributus buvo įvesti į modelį siekiant realiu laiku stebėti pagrindinių parametrų kitimą. Akivaizdžių simuliacijos sutrikimų ar klaidų pranešimų nepastebėta.

4.2.3. Modelio patvirtinimas (validavimas)

Validavimas yra procesas kurio metu sukurtas modelis yra patikrinamas ar jis yra teisingas ir atitinka tiriamą realią sistemą. Patvirtinimas reiškia, kad svarbių sistemos funkcionavimo įverčių aibė, gauta iš imitacinio modelio, atitinka ar pagrįstai sutampa su analogiškais stebėjimais, gautais iš realios sistemos. [10]

Eksperimentų realioje sistemoje metu laiškų siuntimas ir apdorojimas, priklausomai nuo naudojamų procesorių kiekio, užtrukdavo 30-50 minučių. Modelio eksperimentinio patvirtinimo metu bus naudojama atsitiktinio paskirstymo funkciją laiškų generavimui, laiškai bus gaunami 10 – 17 esybių per sekundę greičiu. Kūrimo modulio nustatymuose pasirenkamas esybės tipas „laiškas“, atvykimo intervalo nustatymuose pasirenkama, kad laiškai bus generuojami kas sekundę, o laiškų kiekis bus paskirstomas tolygiojo skirstinio (angl. *uniform*) principu. Simuliacijos statistika bus skaičiuojama po 40 minučių sistemos veikimo, tokiu būdu nenukrypstant nuo realios sistemos vidutinio apdorojimo laiko. Taip pat, tyrimo praplėtimui, statistika bus užfiksuota ir po 5 valandų simuliacijos.

4.2.3.1. Generuojamų laiškų dydžio validavimas

Pagal nutylėjimą, ARENA modeliavimo aplinka, baigus simuliaciją, generuoja ataskaitą apie esybes, jų apdorojimo laiką, procesus, eiles, išteklius, tačiau atributai yra vartotojo priskiriama reikšmė, todėl statistinius duomenis modeliuotojas turi surinkti pats. Laiškų dydžių vidurkiui nustatyti buvo sukurti specifiniai statistikos nustatymai pavaizduoti 29 paveiksle. Šiuo nustatymu surinktų duomenų nepakanka sulyginti su laiškų pasiskirstymu realioje tarnybinėje stotyje, kur buvo užfiksuotos 4 laiškų dydžių grupės. Užduočiai išpildyti esamas modelis buvo papildytas įrašymo komponentais (angl. *record*), kurių paskirtis yra įrašyti papildomus duomenis į statistikos kaupiklius, kurie, šiuo atveju, buvo pasirinkti „apskaitos“ tipo. Įrašymo moduliai buvo prijungti prie nepageidaujamų laiškų filtro išėjimo, tokiu būdu suskaičiuojant išeinančias skirtingo dydžio bylas (7.3 priedas).

Atlikus pirmąsias simuliacijas paaiškėjo, kad didžioji dauguma laiškų didesni nei 50 KB, o tai neatitinka tikrovės. Toks rezultatas buvo tikėtinas, o kompiuterinis modeliavimas yra iteratyvinis procesas, kurio metu nustatymai gali būti keičiami einant bandymų ir klaidų keliu. Laiško dydžio priskyrimo išraiška pakeičiama iš teorinškai išvestos išraiškos: *DISC(0.969,TRIA(1000,3000,100000),0.992,UNIF(100000,500000),0.998,UNIF(500000,1024000),1.0,UNIF(1024000,1465130))*, į praktiškai sukurtą:
DISC(0.3577,UNIF(500, 2999), 0.6399,UNIF(3000, 9999), 0.9061,UNIF(10000,49999), 0.9672,UNIF(50000,99999), 1.0,UNIF(100000,1465130)), kad tenkintų 15 paveiksle realios tarnybinės stoties duomenis. Pakoregavus dydžio priskyrimo modulio nustatymus gauti statistiniai rezultatai pateikiami 13 lentelėje.

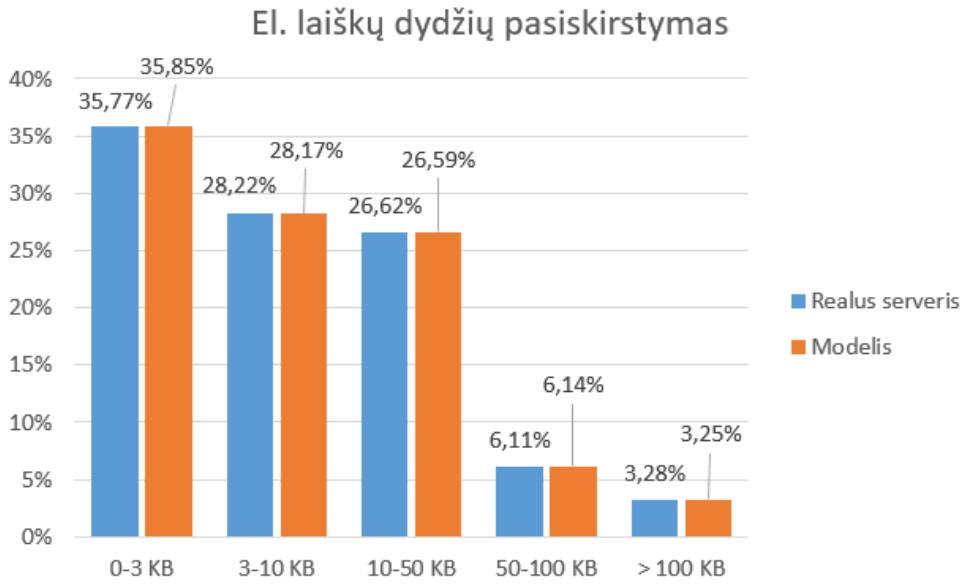
13 lentelė. Laiško dydžio modeliavimo palyginimas su tikrove

Pavadinimas	**Reikšmė modelyje (40 min simuliacija)**	**Reikšmė modelyje (5 h simuliacija)**
Gauta laiškų, vnt.	31233	561960
Pristatyta laiškų, vnt.	31181	561905
Mažiausio laiško dydis, B	500 B	500 B
Didžiausio laiško dydis, B	1465046 B	1465096 B
Matematinis laiškų dydžių vidurkis, B	88103 B	86882 B
Laiškų dydžio pusplotis, B	7306 B	1817 B

Eksperimentų su realia tarnybine stotimi metu apdorojimo laikas buvo neribojamas, tačiau buvo naudojamas baigtinis apdorojamų laiškų skaičius todėl rezultatai gali būti palyginti:

- Realioje tarnybinėje stotyje buvo apdorojama 30960 laiškų ir tai užtruko apie 40 minučių, modelis per 40 minučių apdoroja 31181 laiškų. Rezultatai modelyje ir realybėje yra panašūs. Reikia pastebėti, kad simuliacija yra stabdoma praėjus nustatytam laikui, tuo metu iš sistemos lieka neišėję 40-50 laiškų.
- Mažiausio ir didžiausio laiško dydžiai modelyje ir realybėje – sutampa.
- Matematinis laiškų dydžio vidurkis, gautas iš ARENA aplinkos ataskaitos didelės vertės nesuteikia, tačiau pasiskirstymo aibės pusplotis yra 7,3 KB. Atlikus simuliacija ilgesnį laiką (5 valandas), puspločio reikšmė artėja prie 2 KB, o tai yra labai arti realios situacijos.

Panaudojus įrašymo moduliais surinktus duomenis buvo suskaičiuota skirtingų laiškų dydžių pasiskirstymas, kuris atvaizduotas 33 paveiksle ir palygintas su realios tarnybinės stoties duomenimis. Atvaizdavimui naudojami 5 valandų simuliacijos statistiniai duomenys.



33 pav. El. laiškų dydžio pasiskirstymo palyginimas

Iš el. laiškų dydžių pasiskirstymo grafiko akivaizdžiai matosi, kad modelyje generuojamų laiškų dydžių kiekybinis pasiskirstymas sutampa su realybėje surinktais duomenimis.

4.2.3.2. Aktyvios eilės valdiklio veikimo validavimas

El. pašto serverio aktyvios eilės valdiklio paskirtis yra perduoti laiškus į filtravimo serverį, nes filtravimo serveris neturi laiškų eilės komponento. Eksperimentų su „Postfix“ serveriu metu, pagal nutylėjimą, laiškai buvo siunčiami į eilę laukti, kol bus perduoti filtravimo serveriui, jei tuo metu būdavo daugiau nei 43 filtruojamų laiškų.

Vykdant 40 minučių ir 5 valandų simuliacijas, nustatyt, kad laiškų eilė pradeda kauptis, o vienu metu antivirusinio ir nepageidaujamų laiškų filtruose esančių laiškų suma niekada nepasidaro didesnė nei nustatytas dydis – 43. Iš to galima daryti išvadą, kad valdiklis veikia ir atitinka realią situaciją.

ARENA ataskaitų sistema pagal nutylėjimą parodo eilių statistiką. Patikrinti ar filtravimo serveryje nėra daugiau laiškų nei nustatyta, buvo sukurtas papildomas statistikos kaupimo modulis, kurio funkcija skaičiavo visų filtravimo serverio procesuose esančių laiškų skaičių. Statistikos duomenys pateikiami 14 lentelėje. Įdomumo dėlei, bendras simuliacijos metu sistemoje esančių esybių vidurkis yra 45. Šis skaičius neprieštarauja surinktiems statistiniams duomenims ir gaunamas sudėjus vidutinį eilėje esančių laiškų skaičių su esančiais filtravimo serveryje.

14 lentelė. Laiškų eilės ir apdorojamų laiškų statytiniai duomenys (40min. simuliacija)

Pavadinimas	**Reikšmė**
Vidutinis laukimas laiškų eilėje, sek.	0,3046
Maksimalus laukimo eilėje laikas, sek.	1,187
Vidutinis laiškų kiekis eilėje, vnt.	3,9620
Maksimalus laiškų skaičius eilėje, vnt.	26
Vidutinis laiškų skaičius filtravimo serveryje, vnt.	41,4
Maksimalus laiškų skaičius filtravimo serveryje, vnt.	**43**

4.2.3.3. Nepageidaujamų laiškų aptikimo validavimas

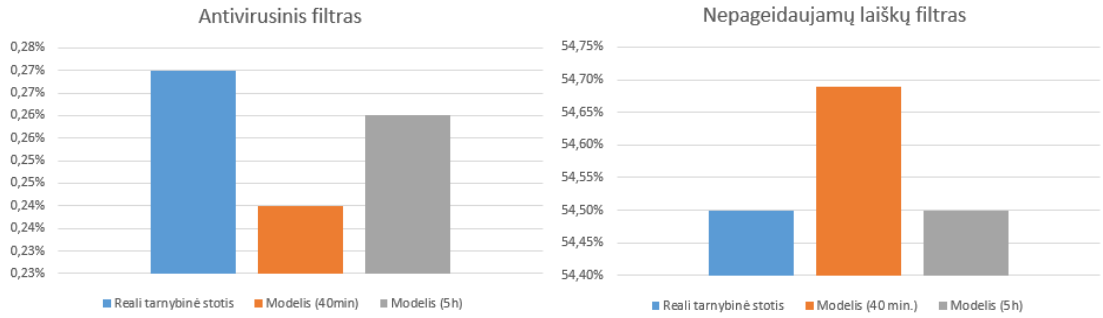
Iš eksperimentais prie 2 procesorių sistemos nustatytų duomenų, pateikiamų 8 lentelėje (4 ir 7 eilutės) žinoma, kad realioje sistemoje buvo apdorojama 30960 laiškų. Antivirusinis filtravimo

metodas aptiko 89 laiškus, tai sudaro 0,27 proc. visų laiškų. Nepageidaujamų laiškų filtras aptiko 16878 – 54,5 proc. visų laiškų. Duomenų palyginimui, validavimo duomenys pateikiami 15 lentelėje ir 34 paveiksle.

15 lentelė. Nepageidaujamų laiškų aptikimo duomenys

Filtravimo metodas	Aptiktų laiškų dalis realybėje	Aptiktų laiškų dalis modelyje (40min. simuliacija)	Aptiktų laiškų dalis modelyje (5h simuliacija)
Antivirusinis filtras	0,27%	0.09%	0.118%
Nepageidaujamų laiškų filtras – DNS metodai	54,5%	54.69%	54,5%

Validavimo rezultatai parodo, kad nepageidaujamų laiškų filtro rezultatai modelyje ir realybėje – sutampa. Antivirusinio filtravimo metodo aptiktų laiškų dalis modelyje atitinka maždaug pusę realioje tarnybinėje stotyje aptinkamų laiškų kiekio. Toks rezultatas gali būti įtakotas ypatingai mažos žalingą kodą turinčių laiškų imties ir dėl to prastai veikiančios skirstinio funkcijos. Eksperimentiniu būdu modulyje naudotą *DISC(0.00271, 1, 1.0, 0)* išraišką pakeitus į *DISC(0.000542, 1, 1.0, 0),* tikimybę padauginus iš 2, gaunami stebėtinai artimi realiai tarnybinei stočiai rezultatai: 0,24 proc., 0,26 proc., atitinkamai po 40 minučių ir 5 valandų ir 12 valandų simuliacijos.



34 pav. Nepageidaujamų laiškų aptikimas

Aptikus šį įvykių funkcijos netobulumą rekomenduotina naudoti pakoreguotą antivirusinio filtravimo metodo skirstinio išraišką.

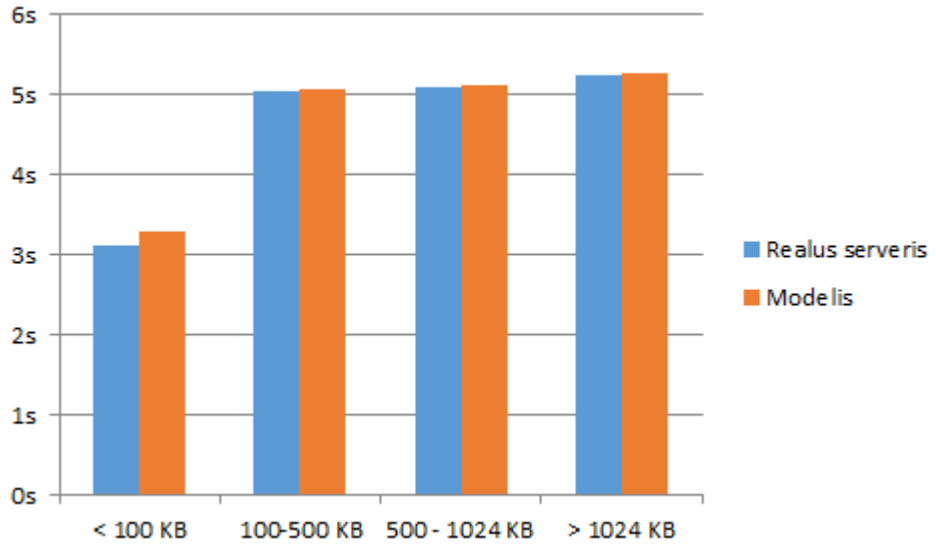
4.2.3.4. Apdorojimo laiko validavimas

Apdorojimo laiko validavimo užduočiai buvo pasirinktas 7-asis eksperimentas, nustatymai kai įjungtas antivirusinis ir DNS tipo filtrai realioje sistemoje. Imituojant kitą procesorių ar filtravimo metodų variaciją, reikėtų parinkti kitas vėlinimo reikšmes iš 7.1 priede pateiktos lentelės, o pati modelio logika nesikeistų, todėl tiktų bet kuris variantas, kuriame veikia antivirusinis ir bent vienas papildomas filtravimo metodas.

Įvykdžius 40min. ir 5 valandų simuliacijas, matematinis laiškų apdorojimo vidurkis yra 3,52 sekundės, tuo tarpu, tokių duomenų realiame serveryje nebuvo surinkta. Realybėje vėlinimas buvo skaičiuojamas skirtingoms laiškų dydžių grupėms. Norint gauti atitinkamus modelio sistemos duomenis palyginimui laiškų dydžio priskyrimo modulio nustatymai buvo pakeisti, kad atitiktų kiekvieną iš dydžių grupių ir 40 minučių simuliacija buvo paleista visus penkis kartus. Rezultatai pateikti 16 lentelėje, o grafiškai atvaizduoti 35 paveiksle.

16 lentelė. Nepageidaujamų laiškų pardorojimo vėlinimo validavimo rezultatai

Sistema	Vėlinimas kai laiškas < 100 KB	Vėlinimas kai laiškas 100-500 KB	Vėlinimas kai laiškas 500 - 1024 KB	Vėlinimas kai laiškas > 1024KB
Reali sistema, sek.	3,125	5,056	5,102	5,245
Modelis, sek.	3,3	5,058	5,109	5,258

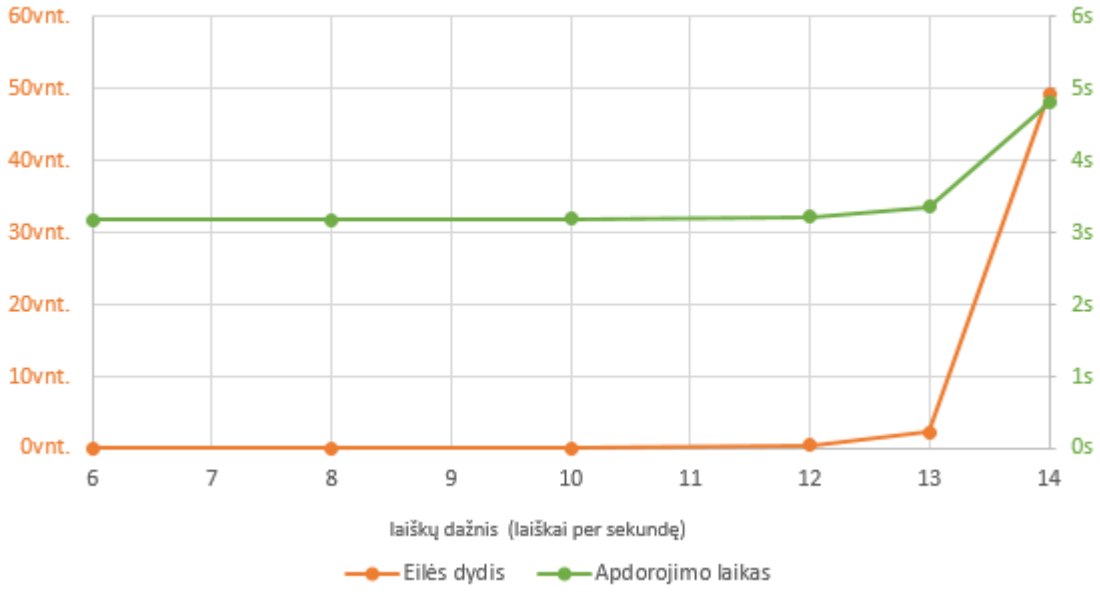


35 pav. El. laiškų apdorojimo vėlinimas pagal dydį

4.2.3.5. Imitacinio modelio panaudojimo pavyzdys

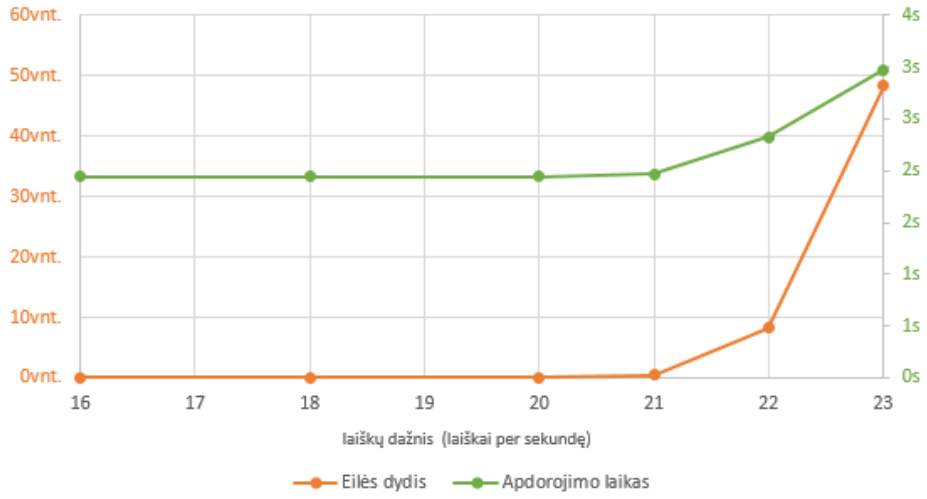
Validavimo metu buvo įsitikinta, kad sukurtas imitacinis modelis yra tinkamas nepageidaujamų elektroninių laiškų filtravimo sistemoms modeliuoti.

Praktiniam modelio išbandymui buvo sumodeliuota situacija kai filtravimo sistema veikia 87 proc. efektyvumu, t.y. „randa“ tokią dalį nepageidaujamų laiškų įjungus tik DNS ir antivirusinius filtravimo metodus (šis nustatymas gali būti keičiamas). Laiškai į sistemą siunčiami 1 sekundės intervalu, atsiunčiant vis didėjantį laiškų, vienu metu siunčiamų į sistemą, kiekį. Modelis yra tinkamas ir bet kokiam kitokiam laiškų siuntimo dažniui, pvz.: panaudojus tikimybinį skirstinį intervalui nustatyti, tačiau, supaprastinus, kad būtų lengviau suprasti rezultatus, buvo pasirinktas statinis intervalas ir tiesiškai didėjantis laiškų skaičius. Tikslas – nustatyti kada pradeda kauptis aktyvios eilės valiklio eilė ir pradeda augti vėlinimas, kaip tai priklauso nuo naudojamų procesorių skaičiaus, kokią operacinės sistemos resursų konfigūracija yra geriausia nustatytam filtravimo efektyvumui. Kiekviena simuliacija buvo pagreitinta ir leidžiama 1 valandą, o grafikuose žemiau, *x* ašyje pavaizduotas laiškų kiekis siunčiamas kas sekundę, kairėje *y* ašyje – laiškų kiekis eilėje, dešinėje *y* ašyje – vidutinis laiško apdorojimo laikas.



36 pav. Modeliuojamas gaunamų laiškų dažnis, 2 procesorių sistema

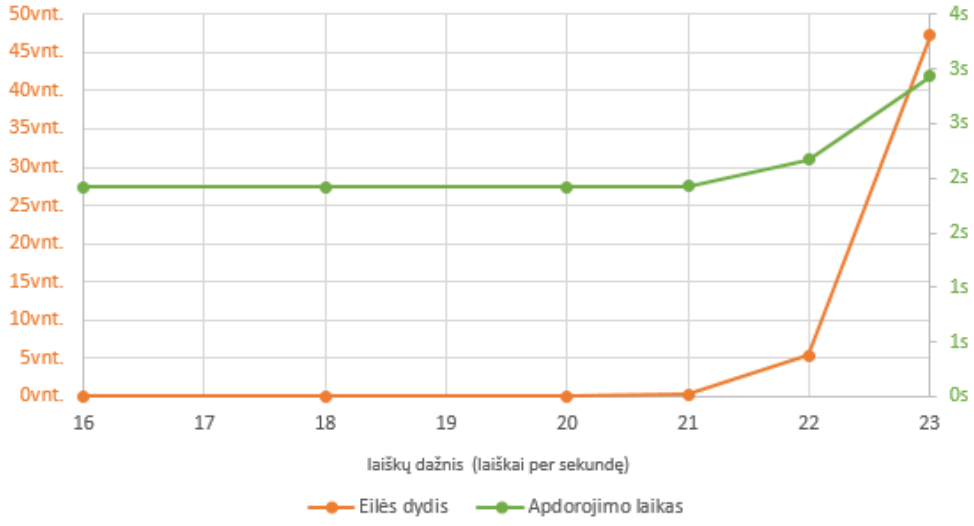
Laiškų kūrimo modulis buvo nustatytas 6 laiškams per sekundę dažniui ir kiekvienos simuliacijos metu ši reikšmė didinama kol pastebėtas eilėje esančių laiškų atsiradimas. 36 paveiksle pavaizduoti 2 procesorių sistemos simuliacijos rezultatai. Iš grafiko matyti, kad 2 procesorių sistema, kurioje veikia antivirusinis ir DNS tipo nepageidaujamų laiškų filtrai, pradeda kaupti eilę nuo 12 laiškų per sekundę. Dažniui didėjant – atsiranda vėlinimas.



37 pav. Modeliuojamas gaunamų laiškų dažnis, 4 procesorių sistema

Nustačius 4 procesorių operacinės sistemos modeliavimo parametrus, paaiškėjo, kad tokios filtravimo sistemos „atsparumas“ gaunamų laiškų dažniui padidėtų beveik 2 kartus ir sistema pradėtų kaupti eilę esant 22 laiškų per sekundę priėmimui (37 paveikslas). Studentams skirta nemokama modeliavimo aplinkos ARENA licencija apriboja modelio veikimą iki 150 esybių, tai reiškia, kad vienu metu modelyje negali būti apdorojamos ar laukti eilėje. Šis apribojimas neleidžia simuliuoti sistemos veikimo kai eilėje yra daug laiškų ir kaip tai įtakotų vėlinimą, tačiau galima numatyti, kad vėlinimas augtų eksponentiškai.

38 paveiksle pateikiami simuliacijos su 8 procesorių laiškų priėmimo ir filtravimo sistema rezultatai. Remiantis naujai pasiūlyto modelio simuliacijos duomenimis, galima daryti išvadą, kad 4 procesorių sistema yra geriausia antivirusinio ir DNS tipo filtravimo sistema, galinti atpažinti 87 procentus nepageidaujamų ar žalingo kodo turinčius laiškų. 8 procesorių sistema tik nežymiai pagreitina apdorojimo laiką, o laiškų eilė pradedama kaupti praktiškai tuo pat metu kaip ir 4 procesorių sistemoje.



38 pav. Modeliuojamas gaunamų laiškų dažnis, 8 procesorių sistema

4.2.3.6. Modelio validavimo išvados

Imitacinio modelio išvestis – statistiniai duomenys, kurie buvo lyginami su praktiškai realioje tarnybinėje stotyje surinktais statistiniais duomenimis. Šie palyginimai leidžia daryti išvadas ar modelio veikimas yra sėkmingas ir ar reikia papildomų modelio, ar jo reikšmingų faktorių korekcijų.

Atlikus kompiuterinio imitacinio modelio realizacijos validavimą kai modelis nustatytas veikti imituojant antivirusinį ir DNS tipo filtravimo metodus, buvo nustatyta kad:

1) Atlikus elektroninių laiškų dydžių pasiskirstymo modelyje tyrimus, buvo patobulinta laiškų kūrimo tikimybių, kuri labai gerai atitinka realioje tarnybinėje stotyje stebėtas tendencijas.
2) Aktyvios SMTP eilės valdiklio veikimas buvo patikrintas ir atitinka jam užduotus reikalavimus pradėti kaupti eilę filtravimo serveryje susikaupus kritiniam užduotam apdorojamų laiškų kiekiui.
3) Filtravimo metodų efektyvumo validavimo metu buvo nustatyta, kad „SpamAssassin“ nepageidaujamų laiškų aptikimo tikimybė atitinka realios tarnybinės stoties duomenis. „ClamAV“ filtro tikimybės išraiška buvo patobulinta. Šio filtravimo metodo efektyvumo rodiklis – atitinka realioje tarnybinėje stotyje užfiksuotos reikšmes.
4) Apdorojimo laikas modelyje iš esmės atitinka apdorojimo laiką realioje sistemoje. Paskirsčius laikus pagal tikrovėje darytus eksperimentus – kiekvienos laiškų dydžių grupės apdorojimas buvo simuliuojamas ir lyginimas su atitinkamais duomenimis serveryje.

Realizuotas kompiuterinis imitacinis modelis atitinka realioje serverių sistemoje atliktų eksperimentų rezultatus ir leidžia imituoti laiškų gavimą, apdorojimą, pristatymą į el. pašto dėžutes. Imitacinis modelis leidžia nustatyti skirtingus filtravimo metodus prie skirtingų procesorių nustatymų ir atitinkamai susijusių vėlinimo verčių. Sukurtas modelis leidžia priimti sprendimus realios tarnybinės stoties projektavimui.

5. IŠVADOS

1. Atlikus elektroninio pašto paslaugos pažeidžiamumo analizę išsiaiškinta, kad elektroninio pašto tarnybinės stotys yra pažeidžiamos dėl didžiulio perduodamo nepageidaujamų ar žalingu kodu apkrėstų laiškų kiekio. Nors pasaulinės tendencijos rodo, kad nepageidaujamų laiškų dalis santykinai mažėja, mobiliųjų įrenginių prieinamumas paskatino platesnį paslaugos naudojimą, tuo pačiu ir perduodamų, bei apdorojamų duomenų kiekio padidėjimą. Apie 60 proc. perduodamų laiškų pasaulyje yra priskiriami nepageidaujamiems, o Lietuvoje, 2015 metais, net apie 90 proc. [11] [12]. Atlikti tyrimai rodo, kad nepageidaujamų laiškų kiekis 2015 metais buvo 62 proc.
2. Išsiaiškinus elektroninio pašto saugumo ir filtravimo metodus, paaiškėjo, kad bendro, universalaus ir visoms situacijoms ir organizacijoms tinkamo apsaugos metodo – nėra. Įmonės ir organizacijos pritaiko savo sistemas prie toleruotino filtravimo lygio režimo ir keičia nustatymus prisitaikydamos prie naujų tendencijų. Analizės metu paaiškėjo, kad pagrindiniai filtravimo metodai yra: 1) siuntėjo autentiškumo užtikrinimo politikos, dažniausiai panaudojant DNS įrašus kaip patikrinimo šaltinį, 2) laiško teksto, jame esančių adresų ar raktažodžių vertinimas, 3) laiško antraštės metaduomenų analizė, 4) žalingo kodo aptikimo metodai.
3. Atlikus eksperimentinį tyrimą realiose tarnybinėse stotyse nustatyta, kad „ClamAV“ antivirusinio filtro įtaka operacinės sistemos procesoriams yra labai maža. „SpamAssassin“ nepageidaujamų laiškų filtravimo sistemos beveik pilnai apkrauna 2 virtualios operacinės sistemos procesorius, net ir su išjungtais filtrais, galima manyti, kad taip įvyksta, dėl laiško išskaidymo dalimis. Ši filtravimo sistema taip pat reikalauja trečdaliu daugiau operatyvinės atminties nei „ClamAV“. Rekomenduojama turėti bent jau 1Gb operatyvinės atminties ir bent 4 procesorius elektroninio pašto filtravimo sistemos veikimui.
4. Darbo metu atlikti eksperimentiniai tyrimai rodo, kad efektyviausi filtravimo metodai „SpamAssassin“ filtravimo sistemoje yra DNS tipo metodai: a) siuntėjo autentiškumo nustatymo politikos SPF ir DKIM, b) DNS protokolu užklausiamų centrinių juodųjų ir baltųjų sąrašų patikra. Kadangi didelė dalis (apie 86 proc.) laiškų buvo pažymimi kaip nepageidaujami po juodųjų sąrašų patikros. Galima daryti išvadą, kad dauguma brukalo šaltinių atkeliauja iš jau žinomų ir blogą reputaciją turinčių šaltinių arba dinaminių IP adresų rėžių, kurie yra automatiškai įtraukiami į juoduosius sąrašus.
5. Eksperimentų metu surinktų statistinių duomenų pagrindu buvo sukurtas kompiuterinis imitacinis modelis ARENA aplinkoje. Modelis leidžia imituoti laiškų gavimo ir filtravimo sistemas, naudojami tikimybės paskirstymo funkcijos laiškų generavimui, laiškų dydžių priskyrimui, filtravimo vėlinimui ir filtravimo metodų koeficientui nustatyti.
6. Atliktas kompiuterinio imitacinio modelio validavimas – patikrinimas su realios tarnybinės stoties duomenimis. Validavimo metu nustatyta, kad modelis iš esmės atkartoja: a) elektroninių laiškų dydžių pasiskirstymą realybėje, b) „postfix“ aktyvios (angl. *active queue*) aktyvios eilės valdiklio veikimą, c) filtravimo metodų efektyvumą, d) apdorojimo laiką realioje sistemoje.
7. Sukurtas imitacinis kompiuterinis modelis leidžia eksperimentuoti su elektroninio pašto gavimo ir filtravimo sistemomis. Eksperimentai realioje sistemoje buvo vykdyti ir duomenys renkami nuosekliai nustatant 2, 4 ir 8 procesorius. Šie procesorių nustatymai yra ir imitacinio modelio ribos, norint atlikti sistemos modeliavimą su kitokiomis operacinių sistemų variacijomis – reikia atlikti papildomus eksperimentinius tyrimus.
8. Tęsiant pradėtus tyrimus reikėtų ištirti naudojamo procesoriaus branduolio ir atitinkamu filtravimo metodu apdorojamo laiško proceso trukmę ir išsiaiškinti trukmės priklausomybę nuo laiško dydžių. Tokiu atveju būtų galima patobulinti modelį ir

naudoti procesorių branduolius kaip modelio išteklius, tokiu būdu modelio veikimas detaliau atkartotų operacinėje sistemoje vykstančius procesus ir nebūtų priklausomas nuo tyrimo metu naudotų procesorių skaičiaus. Tolimesni darbai, išplečiant modelio universalumą, galėtų apimti ir detalesnius filtrų grupių tyrimus, nustatant kuris algoritmas iš grupės yra efektyviausias resursų naudojimo ir laiškų aptikimo aspektais.

6. LITERATŪRA

[1] The Spamhaus Project Ltd, "The Definition of Spam," The Spamhaus Project Ltd, 12 12 2014. [Tinkle]. Available: https://www.spamhaus.org/consumer/definition/. [Kreiptasi 16 12 2014].

[2] B. N. K. C. P. Wood, "INTERNET SECURITY THREAT REPORT 2014," Symantec Corporation, 04 2014. [Tinkle]. Available: http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf. [Kreiptasi 16 12 2014].

[3] T. Shcherbakova, "SPAM in Q3 2015," [Tinkle]. Available: https://securelist.com/analysis/quarterly-spam-reports/72724/spam-and-phishing-in-q3-2015/. [Kreiptasi 26 04 2016].

[4] R. Kettlewell, "SMTP reply codes," [Tinkle]. Available: http://www.greenend.org.uk/rjk/tech/smtpreplies.html. [Kreiptasi 18 12 2014].

[5] T. Hansen, "DomainKeys Identified Mail (DKIM) Service Overview," IETF, 07 2009. [Tinkle]. Available: http://tools.ietf.org/html/rfc5585. [Kreiptasi 26 01 2015].

[6] M. W. Wong, "SPF Overview," , 01 04 2004. [Tinkle]. Available: http://www.linuxjournal.com/article/7327?page=0,0. [Kreiptasi 27 01 2015].

[7] P. Resnick, "Domain-based Message Authentication, Reporting and Conformance (DMARC)," 2015. [Tinkle]. Available: https://datatracker.ietf.org/doc/draft-kucherawy-dmarc-base/?include_text=1. [Kreiptasi 29 01 2015].

[8] The Editors of Encyclopedia Britannica, "Thomas Bayes," Encyclopedia Britannica, [Tinkle]. Available: http://www.britannica.com/biography/Thomas-Bayes. [Kreiptasi 29 01 2015].

[9] V. Metsis, I. Androutsopoulos and G. Paliouras, "Spam Filtering with Naive Bayes – Which Naive Bayes?," in *Third Conference on Email and Anti-Spam (CEAS 2006)*, Mountai View, 2006.

[10] Paulauskaitė-Taraševičienė, Agnė, Šutienė, Kristina, Sistemų imitacinis modeliavimas ir analizė Arena aplinkoje [elektroninis išteklius] : mokomoji knyga / Kauno technologijos universitetas. Taikomosios informatikos katedra. Kaunas : Technologija, 2014. 99 p. ISBN 9786090211076.

[11] Trend Micro, Inc, "Email Reputation Services," Trend Micro, Inc, 2014. [Tinkle]. Available: https://ers.trendmicro.com/pages/global. [Kreiptasi 16 12 2014].

[12] P. Wood and B. Nahorney, "INTERNET SECURITY THREAT REPORT 2015," Symantec Corporation, 04 2015. [Tinkle]. Available: https://www4.symantec.com/mktginfo/whitepaper/ISTR/21347932_GA-internet-security-threat-report-volume-20-2015-social_v2.pdf. [Kreiptasi 06 04 2016].

[13] Mediabistro Holdings, "www.adweek.com," 27 10 2014. [Tinkle]. Available: http://www.adweek.com/sa-article/5-trends-defining-email-today-161013. [Kreiptasi 16 12 2014].

[14] Interneto Vizija, UAB, "SPF įrašas," Interneto Vizija, UAB, [Tinkle]. Available: http://pagalba.iv.lt/SPF_%C4%AFra%C5%A1as. [Kreiptasi 27 01 2015].

[15] M. Delany, "DomainKeys Identified Mail (DKIM) Signatures," [Tinkle]. Available: http://dkim.org/specs/rfc4871-dkimbase.html#rfc.section.7.2. [Kreiptasi 28 01 2015].

[16] GFI, "Why Bayesian filtering is the most effective anti-spam technology," [Tinkle]. Available: http://www.gfi.com/whitepapers/why-bayesian-filtering.pdf. [Kreiptasi 29 01 2015].

[17] Arena Simulation Software, "Video Library," Arena Simulation Software, 2016. [Tinkle]. Available: https://www.arenasimulation.com/video-library . [Kreiptasi 22 03 2016].

[18] R. L. Barnes, "DANE: Taking TLS Authentication to the Next Level Using DNSSEC," *the IETF Journal*, vol. 7, no. 2, pp. 1-7, 2011.

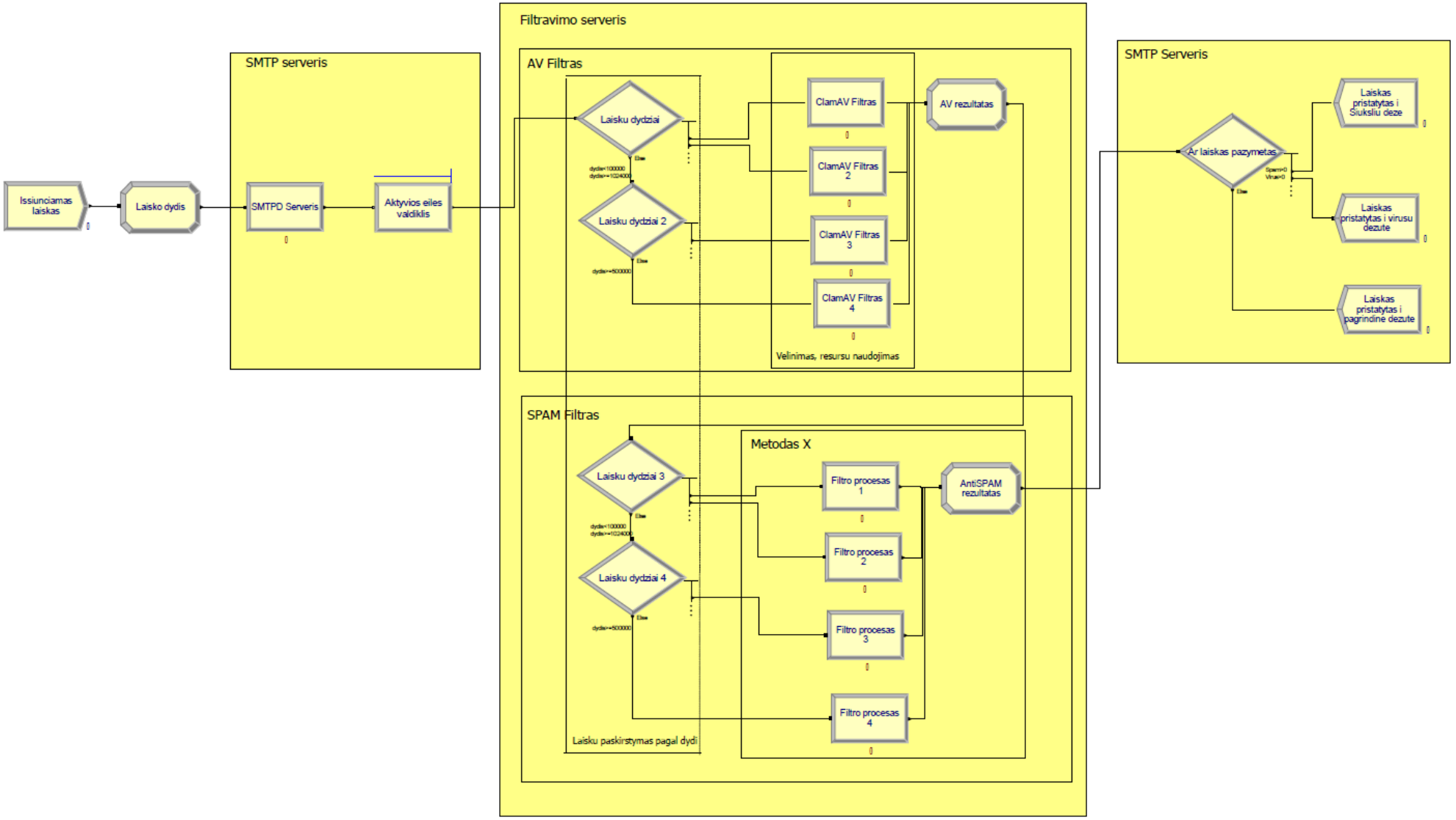
[19] J. Klensin, "Simple Mail Transfer Protocol RFC," IETF, 10 2008. [Tinkle]. Available: https://tools.ietf.org/html/rfc5321. [Kreiptasi 17 12 2014].

7. PRIEDAI

7.1. Priedas. Apdorojimo laiko duomenų lentelė

Eksperimento Nr.	**Dydis < 100KB**	**Dydis 100-500KB**	**Dydis 500-1024KB**	**Dydis > 1024KB**
1	53	72	101	153
2	53	66	92	135
3	53	65	91	135
4	137	390	904	1788
5	122	343	788	1537
6	109	302	671	1283
7	3125	5056	5102	5245
8	1906	2669	2949	3381
9	1899	2481	2377	3069
10	4205	6142	6234	6353
11	3171	3879	4011	4196
12	3110	3815	4010	4179
13	7389	8002	8583	9984
14	4031	6163	4524	5343
15	3376	5035	3877	4500
16	8763	14852	9909	11483
17	4324	6596	4839	6280
18	3756	5296	4145	4926

7.2. Priedas. Elektroninio pašto priėmimo ir filtravimo sistemos modelis ARENA aplinkoje



7.3. Priedas. Elektroninio pašto priėmimo ir filtravimo sistemos modelis ARENA aplinkoje validavimo metu