**RĪGAS TEHNISKĀ UNIVERSITĀTE**

Datorzinātnes un informācijas tehnoloģijas fakultāte

Informācijas tehnoloģijas institūts

**Pāvels OSIPOVS**

Doktora studiju programmas „Informācijas Tehnoloģija" doktorants

# ELEKTRONISKĀS INFORMĀCIJAS SISTĒMASLIETOTĀJU PERSONĪGĀ ADAPTĪVĀ PROFILA PIELIETOŠANA ANOMĀLAS UZVEDĪBAS ATKLAŠANAI

**Promocijas darba kopsavilkums**

Zinātniskais vadītājs

Dr. habil. sc. comp., profesors

A. BORISOVS

**Rīga 2015**

Osipovs P. Elektroniskās informācijas sistēmas lietotāju personīgā adaptīvā profila pielietošana anomālas uzvedības atklašanai. Promocijas darba kopsavilkums.— R.: RTU, 2015.— 38 lpp.

Iespiests saskaņā ar 2014. gada 3. jūlija Informācijas Tehnoloģijas institūta padomes sēdes lēmumu, protokola Nr. 12100–4.1/4.


Šis darbs izstrādāts ar Eiropas sociālā fonda atbalstu projektā "Atbalsts RTU doktora studiju īstenošanai".

ISBN 978-9934-10-679-8

**PROMOCIJAS DARBS**
**IZVIRZĪTS INŽENIERZINĀTŅU DOKTORA GRĀDA**
**IEGŪŠANAI RĪGAS TEHNISKAJĀ UNIVERSITĀTĒ**

Promocijas darbs inženierzinātņu doktora grāda iegūšanai tiek publiski aizstāvēts 2015. g. 1. jūnijā plkst. 14:30 Rīgas Tehniskās universitātes Datorzinātnes un informācijas tehnoloģijas fakultātē, Meža ielā 1, 3. korpusā, 202. auditorijā.

OFICIĀLIE RECENZENTI

Profesors, Dr. habil. sc. comp. Valērijs Zagurskis
Rīgas Tehniskā universitāte, Latvija

Dr. sc. ing. Mihails Savrasovs
Transporta un sakaru institūts, Latvija

Profesors, Dr. tech. sc. Aleksandrs Koļesņikovs
Imanuela Kanta Baltijas federālā universitāte, Krievija

APSTIPRINĀJUMS

Apstiprinu, ka esmu izstrādājis doto promocijas darbu, kas iesniegts izskatīšanai Rīgas Tehniskajā universitātē inženierzinātņu doktora grāda iegūšanai. Promocijas darbs nav iesniegts nevienā citā universitātē zinātniskā grāda iegūšanai.

Pāvels Osipovs …… (*paraksts*)

Datums: …….

Promocijas darbs ir uzrakstīts angļu valodā, satur ievadu, 5 nodaļas, rezultātu analīzi un secinājumus, 21 tabulas un 91 attēlu, kopā 144 lappuses. Bibliogrāfijas sarakstā ir 108 nosaukumi.

**CONTENTS**

# IEVADS

Darbā tiek pētīta elektroniskas informācijas sistēmas lietotāja uzvedības anomalitātes līmeņa aprēķināšanas problēma. Šāda tipa uzdevums ir aktuāls dažādām elektroniskām sistēmām, kuras sniedz iespēju piekļūt finanšu, medicīniskiem, militāriem un tamlīdzīgiem sensitīviem datiem. Šādas sistēmas no ārējiem draudiem ir labi aizsargātas. Lielu bīstamību šobrīd rada iekšējie draudi, kuri rodas jau autorizēto lietotāju uzraudzības neesamības dēļ. Taču finanšu, reputācijas un informācijas zaudējumi, kas rodas sensitīvo datu noplūdes dēļ, lieliem uzņēmumiem var sasniegt milzīgus apmērus [18].

## Tēmas aktualitāte

Anomālas uzvedības atklāšanas uzdevums sarežģītu elektronisku informācijas sistēmu ietvaros šobrīd nav pilnībā atrisināts [71]. Pastāvošajos risinājumos uzsvars tiek likts uz mērķa sistēmas atsevišķu sastāvdaļu anomālijas novērošanu. Kompleksas pieejas izstrāde, nodrošinot aizsardzību visos sarežģītas informācijas sistēmas līmeņos, ir darbietilpīgs teorētisks un tehnisks uzdevums. Šajā pētījumā piedāvātā **bāzes pieeja** balstās uz Markova ķēžu (MĶ) izmantošanu [35] un ļauj iegūt efektīvu un ātru (pateicoties izmantotajai Markova atmiņas neesamības īpašībai [50, 62]) lietotāja pieprasījuma anomalitātes līmeņa klasifikatoru. Arī turpmāka šīs pieejas darbības ātruma paaugstināšana paliek aktuāla, jo prasības pret lietotāja pieprasījuma apstrādes laiku arvien turpina kļūt stingrākas. Tāpat ir izstrādātas bāzes pieejas efektivitātes paaugstināšanas metodes.

## Darba mērķis

**Pētījuma galvenais mērķis** ir anomālas lietotāja uzvedības atklāšanas efektivitātes uzlabošanas iespēju izstrāde un izpēte, izmantojot uzvedības profilu, kas adaptīvi pielāgojas konkrēta cilvēka uzvedības specifikai, un ņemot vērā sākotnēji uzstādītas prasības.

Lai sasniegtu pētījumā izvirzīto mērķi, ņemot vērā prasības, ir jāizpilda šādi pētījuma uzdevumi:

- Izpētīt eksistējošas lietotāja uzvedības formalizācijas pieejas un to izmantošanu anomālas aktivitātes atklāšanā.
- Izpētīt Lietotāju klases Uzvedības Profila (LUP) un Lietotāja Uzvedības Personīgā Adaptīvā Profila realizēšanas iespējas (LUPAP).
- Izstrādāt LUP un LUPAP efektivitātes salīdzināšanas metodiku un veikt salīdzinošo analīzi, to izmantojot.
- Izstrādāt eksperimentālu programmvidi un veikt eksperimentus, lai salīdzinātu anomālas uzvedības atklāšanas efektivitāti, izmantojot abas pieejas. Pēc iespējas

parādīt, ka personīgais adaptīvais uzvedības profils kopumā būs labāks anomālas uzvedības klasifikators.

- Izstrādāt metodiku piedāvātās pieejas ieviešanai mūsdienīgās informācijas sistēmās ar atšķirīgu struktūru.

## Pētījuma objekts

**Pētījuma objekts** ir lietotāja uzvedības profils, kas balstīts uz Markova ķēdēm. Šī pieeja tiek izmantota uzvedības anomalitātes līmeņa noteikšanai laikā, kad lietotājs strādā ar sarežģītām elektroniskām informācijas sistēmām.

## Pētījuma hipotēzes

**Galvenā šajā darbā izvirzītā hipotēze** ir tā, ka lietotāja uzvedības profils, kas balstās tikai uz tiem datiem, kuri atspoguļo viņa personīgās intereses mērķa informācijas sistēmas ietvaros, spēs efektīvāk noteikt viņa (lietotāja) uzvedības anomalitātes līmeni. Galvenās hipotēzes ietvaros izvirzītas citas, tai pakārtotas, hipotēzes:

- Markova ķēdes grafa izmantošana ir ātrs un efektīvs lietotāja uzvedības īpatnību formalizācijas veids.
- Pastāv iespēja salīdzināt lietotāja uzvedības profilu efektivitāti.
- Izmantojot atšķirīgu mērķa sistēmas lietotāju uzvedību, tiks izveidoti atšķirīgi šo lietotāju uzvedības profili.

## Pētījuma metodes

Veidojot lietotāja uzvedības profilu, tika izmantotas grafu teorijas un Markova ķēžu metodes. Izveidoto uzvedības profilu efektivitātes novērtēšana veikta, izmantojot informācijas teorijas, matemātiskās statistikas, varbūtību teorijas un frekvenču analīzes metodes.

Lai ģenerētu sintētiskos datus par lietotāju uzvedību, izmantotas datu ieguves un varbūtību teorijas metodes, bet uzvedības profila apmācības un tā validēšanas laikā izmantotas mašīnapmācības pieejas.

Galvenās pētījuma metodes ziņā ir izveidota metodika eksperimentu veikšanai, kas ietver eksperimentālas sistēmas izstrādi, kas nodrošina vidi, kurā veikt dažādu izstrādājamo profilu raksturlielumu testēšanu.

## Zinātniskā novitāte

Galvenie darbā aprakstītie sasniegumi un rezultāti, kuri atbilst zinātniskās novitātes kritērijam, ir šādi:

- Kā uzvedības anomalitātes klasifikators ir piedāvāta metode, kas ņem vērā katra mērķa informācijas sistēmas lietotāja uzvedības īpatnības.

- Piedāvāta anomalitātes līmeņa sliekšņa dinamiskās izmaiņas metode, kas nodrošina augstāku uzvedības tipa gala klasifikatora darba precizitāti.
- Piedāvāta dažādas interešu prioritātes saturošu sintētisko datu ģenerēšanas metode, izmantojot eksponenciālo sadalījumu ar atšķirīgiem tā parametriem.
- Piedāvāta metodika uzvedības profilu efektivitātes salīdzināšanai, ja profili realizēti ar atšķirīgām pieejām un apmācīti ar datiem, kurus raksturo atšķirīgi raksturlielumi un kuri satur atšķirīgas iekšējo mainīgo vērtības.
- Iegūti uzvedības profilu iekšējo parametru ietekmes novērtējumi attiecībā pret uzvedības tipa klasifikācijas precizitāti.

## Praktiskais nozīmīgums

Apskatītā uzbrukumu tipa galvenā atšķirība no citām iebrukuma tehnikām ir tā, ka pēc veiksmīga visu autorizācijas un autentifikācijas procedūru iziešanas lietotājam ir pieeja plašam dažādu servisu klāstam, bet viņa profesionālās darbības specifika palielina viņa darbošanās biežumu ar tiem pakalpojumiem, ar kuriem jāstrādā visbiežāk. Tāpēc izvirzīts uzdevums atklāt tieši anomālu sistēmas pakalpojumu lietošanu tajos gadījumos, kad visi pakalpojumi lietotājam ir vienlīdz pieejami.

Šobrīd iekšējo uzbrukumu draudi ir aktuāli lielam skaitam dažādu elektronisko informācijas sistēmu. Iegūtie rezultāti var tikt izmantoti šādās jomās:

- Dažādās valsts un privātās organizācijās tiek glabāta un izmantota sensitīva informācija, taču pieeja tai ir jāierobežo.
- Mobilajās tehnoloģijās: kļūst iespējams izveidot tipiska pārnēsājamo ierīču īpašnieka profilu, piemēram, mobilajos telefonos. Un šajā gadījumā ierīces lietošanas īpatnības būs analogs līdzeklis cilvēka digitālajam parakstam.
- Uzdevumos, kuros ir dati, kuru struktūru un īpašības nosaka to autora specifika, piemēram, rakstnieka autora stila profila izveide, balstoties uz viņa darbiem.
- Izstrādātā pieeja sintētisko sistēmas lietošanas datu ģenerēšanai, kas ievēr dažādas interešu prioritātes, var tikt izmantota dažāda pielietojuma programmsistēmu izstrādē.

## Darba aprobācija

Pētījuma pamata etapi un to rezultāti prezentēti un apspriesti 11 starptautiskās zinātniskās konferencēs:

1. DMC 2013 — prudsys User Days 2013, Jule 2013, Berlin. www.data-mining-cup.de/en/ *(Osipov P. Anomaly Detection Using User Behavior Profile and Its Implementation into a Distributed Information System).*

2. IADIS Multi Conference on Computer Science and Information Systems 2013 (MCCSIS 2013), July 2013, Prague. www.iadisportal.org/mccsis2013 *(Osipov P. User Behavior Profile Implementation for Distributed Information System).*
3. XVI International Youth Forum „Radio Electronics and Youth in XXI century", 17–19 April 2012, Kharkov, Ukraine *(Osipov P. A. Identification of Transaction Types using standard Clinical Document Architecture).*
4. 1st International Symposium "Hybrid and Synergies Intelligent Systems: Theory and Practice", (GISIS 2012), 29 June – 2 July 2012, Kaliningrad, Russia *(Osipovs P., Borisovs A. Approaches to the analysis of information systems user behavior modeling).*
5. RTU 53rd International Scientific Conference dedicated to the 150th anniversary and the 1st Congress of World Engineers and Riga Polytechnical Institute / RTU Alumni, 11/12 October 2012, Riga *(Osipovs P., Borisovs A. Modern Approaches to Creating User Behavior Models).*
6. 8th International Scientific School "Modeling and Analysis of Safety and Risk in Complex Systems" (MA SR — 2011), 28 June – 02 July 2011, Saint-Peterburg, Russia. *(Osipov P. A., Borisov A. N. eHealth System Anomaly Activity Detection, Based on User Behavior Model).*
7. RTU 52rd International Scientific Conference. 13–16 October 2011, Riga, Latvia *(Osipovs P., Borisovs A. Deferred — A New Approach to Time-Critical Task Realisation).*
8. Baltic Congress on Future Internet and Communications (BCFIC Riga 2011), 15–18 February 2011, Riga, Latvia. *(Osipov P., Borisov A. Simulation of Typical Behavior User Using Markov Models).*
9. International Scientific Conference of WEB-designers (WebConf 2010), May 2010, Riga, Latvija. *(Osipovs P. Detection of Authorised Users Anomaly Behavior).*
10. RTU 51st International Scientific Conference. Subsection "Information Technology and Management Science". Riga, Latvia. 2010. *(Osipovs P., Borisovs A. Improvement of Markov models for Anomaly Detection Systems).*
11. RTU 50th International Scientific Conference. Riga, Latvija. 2009 *(Osipovs P., Borisovs A. Usage of Ontologies in Systems of Data Exchange).*

Pētījuma rezultāti, kas iegūti promocijas pētījuma laikā, aprobēti lietiskā projekta ietvaros: "e-StepControl, identifying suspicious activities", SIA "ABC SOFTWARE", 10.2013–04.2015.

## Publikācijas

Par darba gaitā veikto pētījumu rezultātiem ir rakstīts 15 zinātniskajās publikācijās. Lielākoties raksti ir citēti dažādās starptautiskajās elektroniskajās bibliotēkās. Pilnais autora publikāciju saraksts ir sniegts kopējā literatūras sarakstā, kas iekļauts šī kopsavilkuma beigās.

## Galvenie rezultāti

Galvenie darba rezultāti, kas sasniegti šī pētījuma gaitā un promocijas darba izstrādes laikā, ir šādi:

1. Veikta eksistējošu mūsdienīgu informācijas sistēmu lietotāju anomālas uzvedības atklāšanas pieeju analīze. Pētījumu gaitā secināts, ka neviena no izskatītajām metodēm nenodrošina pilnīgu visu izvirzīto prasību realizāciju.
2. Izpētītas un izanalizētas pastāvošās lietotāju uzvedības formalizācijas izveides metodes, kas balstītas uz dažādu Baijesa tīklu variantu, ontoloģiju un mobilo aģentu izmantošanu.
3. Izpētītas uz demonstrētas Markova ķēžu pielietošanas iespējas lietotāju uzvedības profilu formalizācijas uzdevumos, kas apstiprina izvirzītās zemākā līmeņa hipotēzes patiesumu.
4. Izstrādāta bāzes pieeja lietotāja uzvedības vispārīga profila izveidei un izmantošanai.
5. Izstrādāta metode, kas izmanto Lietotāja Uzvedības Personīgo Adaptīvo Profilu ar mērķi uzlabot bāzes pieejas efektivitāti, atklājot iebrukumus un uzvedības netipiskumu.
6. Izstrādāta metodika uzvedības profilu efektivitātes novērtēšanai, kas ļauj veikt personīgā adaptīvā un kopējā lietotāju uzvedības profilu salīdzinošo analīzi.
7. Izstrādāta dinamiska anomalitātes līmeņa sliekšņa metode, kas ļauj uzlabot efektivitāti lietotāju klasifikācijai „*normālos*” un „*anomālos*”.
8. Izmantojot izstrādāto metodiku anomalitātes noteikšanai lietotāju uzvedībā, veikta LUP un LUPAP uzvedības profilu efektivitātes salīdzinoša eksperimentāla analīze.
9. Izstrādātas un programmatūras sistēmas veidā realizētas dažādas eksperimentu sērijas, kas ļauj novērtēt dažādus uzvedības profilu raksturlielumus.

## Darba struktūra un saturs

Darbs sastāv no ievada, piecām nodaļām, secinājumiem un izmantotās literatūras saraksta. Darba apjoms — 144 lappuses, 91 attēls un 21 tabulas. Bibliogrāfijas saraksts satur 108 avotus.

**Ievadā** aprakstīta pētījuma nozares sfēra, definēti pamatjēdzieni, noformulēti pētījuma mērķi un izvirzītās hipotēzes, kā arī sniegts darba uzdevumu saraksts.

**Pirmā nodaļa** veltīta šā brīža situācijai anomālas uzvedības atklāšanas sfērā. Apskatītas dažādas pieejas gan anomāliju noteikšanai datos, gan lietotāju uzvedības profilu uzbūvei.

**Otrā nodaļa** veltīta bāzes algoritma formālam aprakstam, klasifikācijā izmantojamo metriku noteikšanai, vispārējās algoritma sarežģītības novērtēšanai un tā priekšrocību un trūkumu noteikšanai. Lai uzlabotu bāzes algoritma efektivitāti, tiek piedāvāta pāreja uz Lietotāja Uzvedības Personīgo Adaptīvo Profilu un dinamisku anomalitātes novērtēšanas slieksni.

**Trešajā nodaļā** apskatītas teorētiskās pieejas lietotāju uzvedības profilu efektivitātes novērtēšanai, kas balstās uz varbūtīgiem raksturlielumiem un informācijas teorijas metodēm.

**Ceturtā nodaļa** veltīta izstrādātās eksperimentālās sistēmas aprakstam. Tajā formulēti galvenie ierobežojumi, kas attiecas uz katra lietotāja pieprasījuma anomalitātes līmeņa aprēķināšanas procesu. Ir definēta specifikācija prasībām, kas izvirzītas pret eksperimentālo sistēmu. Kā arī aprakstīta izstrādātās sistēmas kopējā struktūra un tās realizācijā izmantotās galvenās pieejas un tehnoloģijas.

**Piektā nodaļa** veltīta visu veikto eksperimentu aprakstam. Saistībā ar galvenā izvirzītā mērķa sasniegšanu katra eksperimentu sērija apraksta uzvedības profilu darbību no atšķirīga skatupunkta, analizējot to īpašības.

Darbu noslēdz galveno rezultātu apraksts un paveiktajā darba iegūto secinājumu formulējums.

## ĪSS DARBA NODAĻU APRAKSTS

### Pirmā nodaļa. Iebrukumu un anomāliju atklāšana elektroniskās informācijas sistēmās

Šajā pētījumā termins *„anomāla uzvedība”* apzīmē tādu uzvedību, kura atšķiras no tipiskas konkrēta lietotāja uzvedības. Kopumā jēdziens *„anomālija”* apzīmē novirzi no normas, kādu nepareizumu, atšķirību no tipiskas likumsakarības (sistēmas mērķa likumsakarības). Lai atklātu anomālijas, pirmām kārtām ir jāizvēlas mērķa objekta parametru kopums, kas tiks izmantots tā raksturlielumu vērtību aprēķinam. Tad ar šo izmērāmo raksturlielumu kopumu ir iespējams novērtēt mērķa objektu, balstoties uz tā vērtībām. Rezultātā vienu un to pašu objektu var pasludināt par *„tipisku”* vai *„netipisku”*, izmantojot vienu un to pašu pieeju, bet izvēloties atšķirīgus pamata raksturlielumus.

Jēdziens *„iebrukums”* (saistībā ar informācijas drošības garantēšanu) apzīmē darbību kopumu, kas tiek veikts ar datoru vai datoru tīklu ar mērķi iegūt neleģitīmu pieeju tajā glabātajai informācijai [9]. Iebrukums var tikt veikts gan no mērķa sistēmas iekšienes, gan no tās ārpuses. Pastāv šādi datu apdraudējumu pamata tipi:

- **Atklāšanas (konfidencialitātes pārkāpuma) draudi** rodas, ja konfidenciāla informācija kļūst pieejama nepiederošām personām.
- **Integritātes draudi** rodas, ja informācija tiek apzināti modificēta.
- **Pieejamības atteikuma draudi** rodas, ja leģitīms lietotājs nespēj iegūt pieprasīto informāciju. Atkarībā no mērķa dokumentu nepieejamības, tie var gan saglabāt, gan zaudēt savu nozīmību.

Sistēma tiek uzskatīta par kompromitētu (tajā pastāv drošības politikas pārkāpums), ja pastāv nozīmīga varbūtība iestāties vismaz vienam no uzskaitītajiem apdraudējumiem.

Iebrukumu atklāšanas tehnikas pēc to darbības veidiem, atklājot iebrukumu, var iedalīt trīs pamata tipos [9]:

- Uz **ļaunprātīgas izmantošanas atklāšanas (ĻIA)** pieejas balstītās sistēmas izmanto tipveida iespējamo uzbrukumu veidnes.
- **Anomāliju atklāšanas (AA)** tehnikas ļauj atklāt jaunus iebrukumu tipus, balstoties uz to izmantotajiem modeļiem.
- **Hibrīdās tehnikas** iebrukumu atklāšanas sistēmu realizācijai var apvienot gan ĻIA un AA priekšrocības, gan to trūkumus. Visbiežāk sistēmas, kas realizē šādu ideoloģiju, spēj atklāt gan tipveida uzbrukumus, gan jaunus, līdz tam nezināmus, uzbrukumus.

Šajā pētījumā izmantota pieeja, kas balstīta uz anomāliju atklāšanas idejām, jo uzbrukumu veidņu izmantošana nesniedz iespēju operatīvi reaģēt uz pēkšņu ļaunprātīga lietotāja anomālu uzvedību, kas nav aprakstīta pastāvošās veidnēs.

Pētījuma sākumā tika izvirzītas prasības pret iebrukumu atklāšanas sistēmu, kas tika formulētas šādi:

- Jāpastāv sistēmas pašapmācības spējām, neiesaistot ekspertus.
- Ir jābūt minimālam informācijas apjomam, kas tiek pieprasīts no analizējamās mērķa sistēmas.
- Izstrādājamās sistēmas realizācija ir maksimāli jānošķir no mērķa sistēmas.
- Pieprasījuma anomalitātes līmeņa aprēķina ātrumam jābūt tādam, lai to varētu veikt reāllaika režīmā.

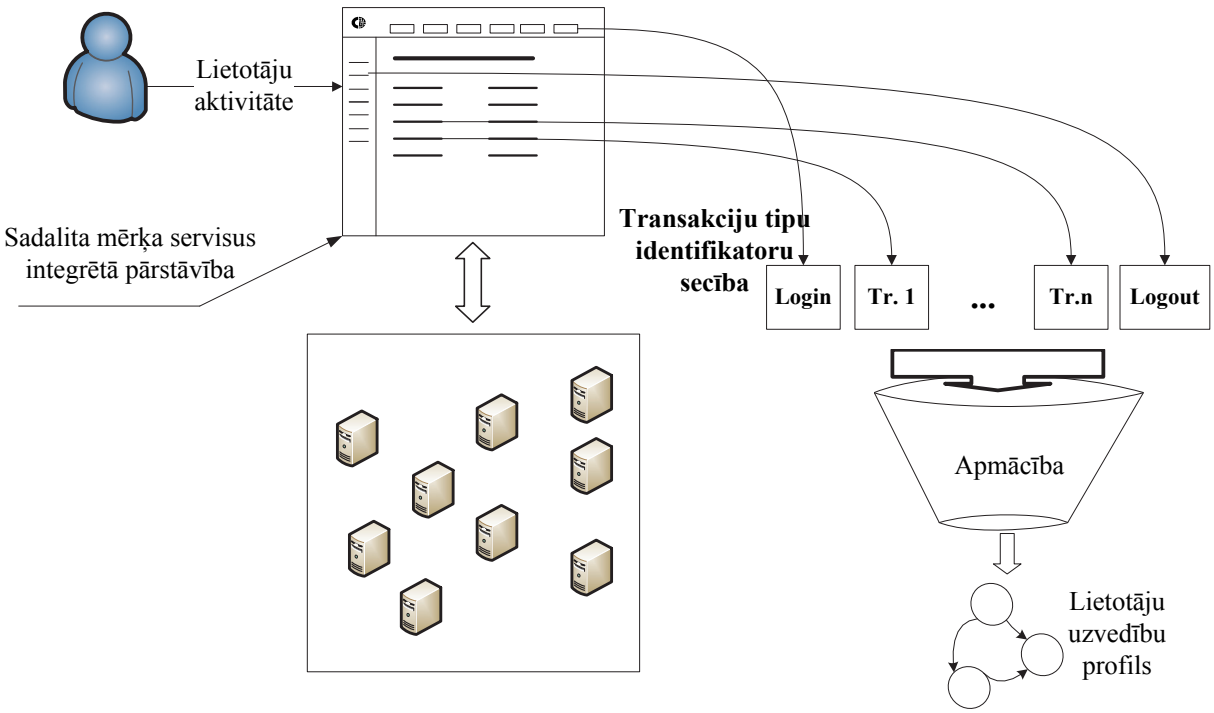
Veiktā pastāvošo pieeju funkcionalitātes analīze parādīja, ka, ņemot vērā izvirzītās prasības, neviena no izskatītajām pieejām nespēj nodrošināt nepieciešamo funkcionalitāti.

Izpētot metodes [80, 82], kas tiek izmantotas lietotāja uzvedības formālu profilu izveidei, tika atklāta nepieciešamība izstrādāt savu uzvedības profila tipu, ņemot vērā visu pieprasīto funkcionalitāti. Anomāliju atklāšanas uzdevuma risināšanas pieeju analīze [79] sniedza

pamatojumu izvēlēties vispiemērotāko jaunizveidojamās sistēmas realizācijas variantu šajā sfērā.

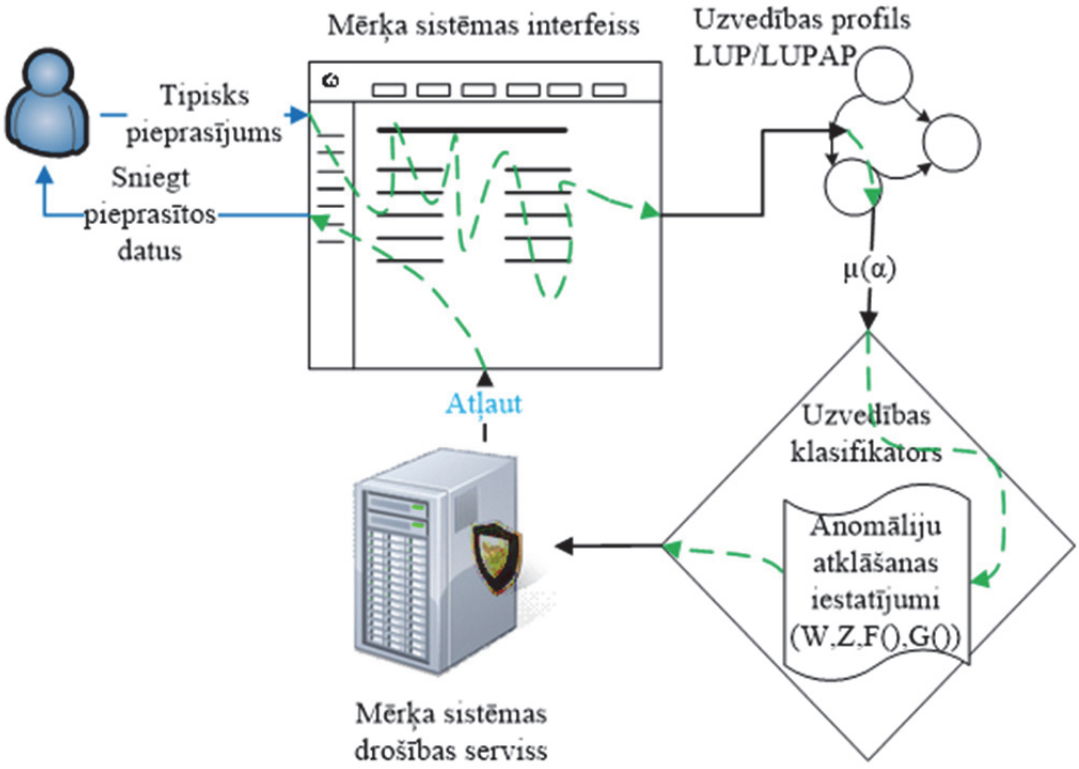
## Otrā nodaļa. Markova ķēdes anomālas aktivitātes atklāšanas uzdevumā

Izmantotās pieejas pamatā tika izmantots Markova ķēdes grafs, kas attēlo datus par lietotāja uzvedības īpatnības [76, 74, 71]. Kopējais uzvedības profila izveides process, kad, balstoties uz lietotāja uzvedības datiem, tiek būvēts profilu aprakstošs Markova ķēdes grafs, ir parādīts 1. attēlā.




1. att. Lietotāja uzvedības profila izveides vispārīga pieeja

Tālāk jau apmācītais profils tiek izmantots jaunas lietotāja uzvedības analizēšanai. Šis process ir detalizētāk parādīts 2. attēlā. Lietotājs mērķa sistēmai nosūta noteiktas transakcijas pieprasījumu. Iekšējais sistēmas drošības modulis pieprasa šī lietotāja pieprasījuma anomalitātes līmeni no anomālas uzvedības atklāšanas moduļa. Anomalitātes izskaitļošanas modulis pieprasa tā brīža lietotājam atbilstošo Lietotāju klases Uzvedības Profilu. To izmantojot kopā ar anomalitātes līmeņa metrikas iegūšanas funkcijām, modulis iegūst nepieciešamo vērtību. Ja anomalitātes līmenis pārsniedz noteikto slieksni, anomāla uzvedība tiek uzskatīta par atklātu, bet, ja anomalitātes līmenis nepārsniedz šo slieksni, tad pieprasījums tiek uzskatīts par tipisku šim lietotājam. Rezultējošais līmenis tiek nosūtīts kā atbilde sistēmas drošības modulim, kurš „pieņem lēmumu", cik lielā mērā ņemt vērā saņemto moduļa atbildi.




2. att. Izmantotās pieejas darbības kopējā shēma tipiskas lietotāja uzvedības gadījumā

**Bāzes algoritma formals apraksts.** Anomālas lietotāja aktivitātes atklāšanas bāzes algoritms savā loģikā iekļauj zemāk minētu soļu izpildi:

- lietotāja uzvedības profila izveide,
- katra lietotāja pieprasījuma anomalitātes līmeņa izskaitļošana,
- metrikas vērtības aprēķins un anomalitātes līmeņa klasifikatora vērtības noteikšana,
- uzvedības profila atjaunināšana, balstoties uz jauniem datiem par lietotāja uzvedību.

***Profila apmācība.*** Tālākajā procesa un izskaitļošanas aprakstā tiek pieņemti apzīmējumi:

- alfabēts Σ — pilna visu iespējamo sistēmas transakciju tipu kopa,
- loga lielums *w* — alfabēta Σ elementu daudzums, kuru kopums veido vienu Markova ķēdes grafa mezglu,
- *Z* — maksimālais piešķiramā soda līmenis, ja neeksistē pieprasītajam stāvoklim nepieciešams mērķa mezgls,
- *r* — metrikas vērtības līmenis, kas nepieciešams, lai atzītu pieprasījumu par anomālu.

Alfabētam Σ pievienots īpašs tukšs simbols (nulles darbība) Ø. Tiek noteikta sākotnējā loga lieluma vērtība (*w*). MĶ sākotnējais stāvoklis tiek noteikts tāds, lai tas atbilstu *w* garuma trasei un sastāvētu no nulles simboliem.

Ir noteiktas divas operācijas, kas darbojas ar ceļiem: 1) operācija next (σ), kas atgriež pirmo trases σ simbolu un pārbīda *σ* par vienu operāciju pa kreisi, tas ir, *next («abcd»)* atgriež

vērtību *„a”* un atjaunina trasi par *„bcd”*; 2) operācija *shift(σ, x)*, kas nobīda trasi σ pa kreisi un pievieno simbolu *x* trases beigās, tas ir, *shift («aba», c) = «bac»*.

Cikliskais Markova ķēdes izveides process katrā iterācijā sastāv no zemāk parādītajiem soļiem un operācijām:

1. piešķirt $c = next(\sigma)$;
2. noteikt <nākamais stāvoklis> = shift(<šā brīža stāvoklis>,c);
3. palielināt skaitītājus stāvoklim <*nākamais stāvoklis*> un pārejai (<*šā brīža stāvoklis*>, <*nākamais stāvoklis*>);
4. atjaunināt <šā brīža stāvoklis> par vērtību <nākamais stāvoklis>.

Tāpat tiek norādīta vai atjaunināta pārejas varbūtības vērtība lokam starp mezgliem, kas apraksta iepriekšējo un konkrētā brīža stāvokļus.

***Lietotāja pieprasījuma anomalitātes līmeņa izskaitļošana.*** Šīs procedūras izpildīšana katrai lietotāja pieprasītajai transakcijai. Ir ieviesti divi iekšējie mainīgie *X* un *Y*, kuru vērtības tiek izsekotas visas sesijas, kuras laikā lietotājs strādā ar sistēmu, analīzes laikā. Sākotnēji (pirmajai transakcijai — *„ieeja sistēmā”*) tām piešķir fiksētas vērtības, vai tās tiek dinamiski izskaitļotas, balstoties uz dažādu iekšējo parametru vērtībām.

Balstoties uz iepriekš apmācītu *„normālas”* uzvedības veidni, katram analizējamās sesijas solim tiek izskaitļotas jaunas mainīgo *X* un *Y* vērtības, kuras tiek izmantotas galīgās anomalitātes līmeņa vērtības (metrika μ(α)) aprēķinā.

Katrā solī ir iespējami divi varianti mainīgo *X* un *Y* jauno vērtību aprēķinam, kas ir atkarīgi no pārejas no iepriekšējā stāvokļa uz pastāvošo stāvokli $\beta_i \rightarrow \beta_{i+1}$ klātbūtnes apskatāmajā grafā.

Pāreja pastāv:
$$Y = Y + F(s, (s, s'))$$
$$X = X + G(s, (s, s')).$$

Pāreja nepastāv:
$$Y = Y + Z$$
$$X = X + 1.$$

**Metrikas μ(α) vērtība** tiek izskaitļota kā attiecība *Y/X*. Metrika μ(α) norāda, cik labi pastāvošais profils prognozē trasi α, tas ir, jo mazāka ir tās vērtība, jo precīzāk MĶ prognozē trasi α. Tā kā metriku μ parametrizē funkcijas *F*, *G* un vērtība *Z*, arī parametru *F* un *G* atšķirīga izvēle mainīs klasifikatora darbību, kas sniedz iespēju klasifikatoru niansēti uzstādīt atbilstoši izmantojamās problēmsfēras specifiku.

Klasifikators *f*, kas ļauj noteikt, vai lietotāja uzvedība ir normāla vai anomāla, darbojas, balstoties uz metrikas μ vērtību un izmantojot aprēķinus pēc formulas ( 1 ), kur vērtība „1” atbilst anomālai, bet „0” — normālai lietotāja uzvedībai.

$$f(a)=\begin{cases}1, & \mu(a)>r\\ 0, & \text{citādi}\end{cases}. \qquad (1)$$

Pastāv vairākas metodes funkciju *X* un *Y* izskaitļošanai [71]. Pētījumā izmantotas šādas pieejas: *frekvenču un varbūtību metrikas* un *uz lokālās entropijas vērtības balstīta metrika*. Vispiemērotākā metode jāpiemeklē atbilstoši pielietojuma sfēras īpatnībām. Iespējama arī izskaitļošana ar vairākām metodēm un anomalitātes klātbūtnes noteikšana vismaz vienā no rezultātiem.

***Algoritma parametri un to ietekme uz rezultāta precizitāti.*** Tā kā klasifikatora izveidē tiek izmantoti daudzi parametri, to vērtību noteikšanai ir tieša ietekme uz rezultējošā klasifikatora kvalitāti.

Izveidojamā klasifikatora precizitāte ir atkarīga no sākotnējās apmācības kopas $T_{traning}$ un šādiem papildus parametriem: loga lieluma *w*, funkciju *F*(*s*,(*s*,*s*`)) un *G*(*s*,(*s*,*s*`)) veida, parametra *Z* vērtība un slieksnis *r*. Iespējamie *F*, *G* un *Z* izskaitļošanas veidi jau tika apskatīti iepriekš, tāpēc tālāk apskatīti parametri *w* un *r*.

Izvēloties loga lielumu *w*, pastāv dilemma — mazas tā vērtības satur maz informācijas, bet lielas vērtības pārāk precīzi pielāgojas apmācības kopai (pārapmācība [41] — klasiska datu ieguves problēma). Ja metriku μ(σ) pieņem par atšķirības lielumu starp Markova ķēdi un analizējamo soli, tad, tai samazinoties, uzlabojas sadalīšana. Tad nesakritību starp visu ceļu un visu apmācības kopu $D(T_{tt})$ var noteikt, izmantojot formulu $\frac{\sum \sigma \in T_{tt} \mu(\sigma)}{\sum \sigma \in T_{tt} |\sigma|}$.

Tas ir, tā ir vienāda visu soļu metriku vērtību summas un vērtību, kas piešķirtas katram Markova ķēdes mezglam, summas attiecībai.

Tādā pašā veidā tiek izskaitļota neatbilstība starp ceļu un anomālu kopu $D(T_{an})$. Loga lielumu *w* palielina, līdz starpība $D(T_{an})-D(T_{tt})$ ir lielāka par noteiktu vērtību, t. i., līdz klasifikators spēj atšķirt šīs divas apakškopas.

Attiecībā uz parametru *r*, kas apzīmē līmeni, kurā soļa metrikas vērtība tiek jau atpazīta kā anomāla, arī šajā gadījumā ir jāpiemeklē optimāla vērtība. Šajā pētījumā tā vērtība tiek piemeklēta tā, lai brīdinājumu aktivizēšanas kopa $T_{an}$ aptuveni atbilstu apmācības kopai ar anomālās uzvedības veidnēm.

***Bāzes algoritma priekšrocības un trūkumi.*** Galveno priekšrocību un trūkumu saraksts ir parādīts 1. tabulā. Neskatoties uz to, ka trūkumu ir vairāk, kvalitatīvi pozitīvās iezīmes ir svarīgākas. It sevišķi, ja ņem vērā pastāvošos ierobežojumus un prasības pret sistēmu.

1. tabula

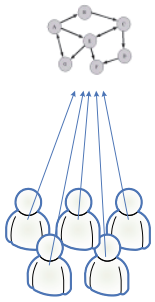
**Bāzes pieejas priekšrocības un trūkumi**

| *Priekšrocības* | *Trūkumi* |
|---|---|
| 1. Uzvedības profila automātiskas atjaunināšanas iespēja, saņemot jaunus datus par lietotāja darbībām.<br>2. Algoritma struktūra pieļauj vienkārši veikt liela profilu apjoma vienlaicīgu paralēlu apstrādi. Tas ļauj palielināt sistēmas veiktspēju, neradot liekas grūtības, ja tas ir nepieciešams, palielinoties apkalpojamo lietotāju skaitam.<br>3. Bāzes algoritma darbības nodrošināšanai nepastāv prasība pēc liela ieejas datu apjoma, minimāli pietiekoša ir trīs tipu datu pieejamība:<br>• lietotāja unikālais identifikators,<br>• lietotāja lomas identifikators,<br>• lietotāja veiktās transakcijas identifikators.<br>4. Bāzes algoritmu ir iespējams vienkārši paplašināt ar jauniem metriku tipiem. | 1. LUP izveide ir grūti formalizējams un laikietilpīgs uzdevums.<br>2. LUP atjaunināšanas uzdevumā liela problēma ir lietotāja profila atjaunināšanas procesa aktivizācijas kritēriju noteikšana.<br>3. Dažādu lietotāju uzvedības datiem, arī tad ja lietotājiem ir vienādas lomas sistēmā, ir statistisko raksturlielumu atšķirības, kas padara rezultējošo modeli par pārāk vispārinātu.<br>4. Nosakot LUP robežlielumus, kuru pārsniegšana tiek pasludināta par anomālu uzvedību (iebrukumu), ir jāņem vērā reālās problēmsfēras specifika un tos nevar formalizēt vispārīgā veidā.<br>5. Iespēja slēpt ļaunprātīgu rīcību, kombinējot „labus" un „sliktus" pieprasījumus.<br>6. Pastāv sistēmas papildus noslodze, kas rodas katras transakcijas analīzes sistēmas funkcionēšanas dēļ reāllaikā.<br>7. Dati, kuri tiek uzkrāti sistēmā, var saturēt konfidenciālu informāciju. |

**Bāzes pieejas uzlabošana.** Lai palielinātu anomalitātes atklāšanas efektivitāti un likvidētu bāzes pieejas galvenos trūkumus, tika piedāvāts pāriet no vispārīga lietotāju klases uzvedības profila uz personīgu profilu katram lietotājam un dinamisku anomalitātes līmeņa sliekšņa noteikšanu.

***Lietotāja uzvedības personīgs adaptīvs profils.*** Viens no galvenajiem lietotāju grupas kopējā LUP profila trūkumiem ir tas, ka liela lietotāju apjoma uzvedība (3. attēls), arī tiem lietotājiem, kuriem sistēmā ir vienādas lomas, var būt ļoti atšķirīga. Rezultātā izveidotais LUP apraksta kāda vidējā lomas pārstāvja uzvedību.

Lai palielinātu katra lietotāja uzvedības analīzes precizitāti, šajā pētījumā ir piedāvāts izmantot Lietotāja Uzvedības Personīgu Adaptīvu Profilu (LUPAP) [72], kurš satur informāciju tikai par viena lietotāja uzvedības specifiku un īpatnībām (4. attēls).



3. att. Kopējais profils

4. att. Personīgais profils

Profila adaptīvo dabu nodrošina **pastāvīgas profila atjaunināšanas** procedūra, kas balstās uz lietotāja aktivitātēm konkrētā brīdī. LUPAP atjaunināšana tiek realizēta, kad lietotājs noslēdz darba sesiju, turklāt ievērojot noteikumu, ka šīs sesijas laikā lietotāja darbībās nedrīkst būt atklāta nekāda anomāla uzvedība.

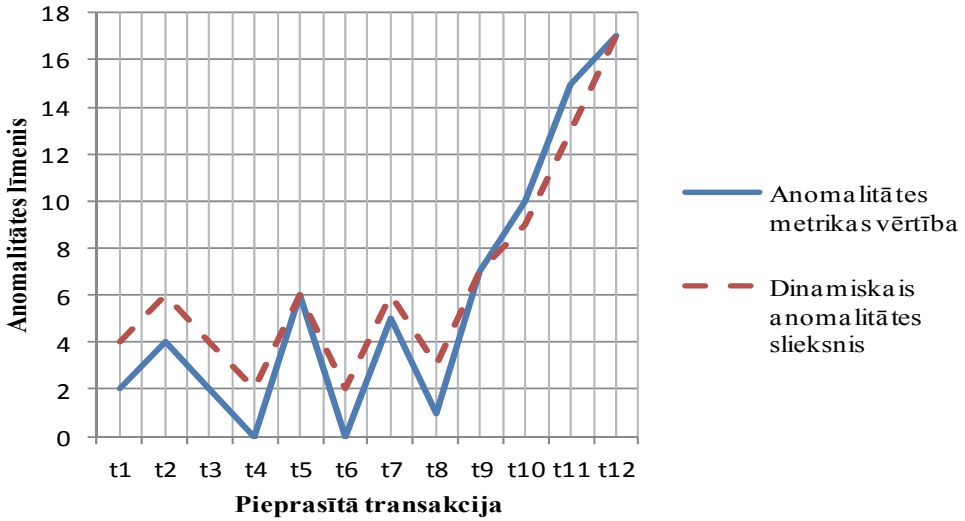
Pie galvenajām LUPAP priekšrocībām var minēt lielu konkrētā lietotāja anomālas uzvedības atklāšanas precizitāti, pastāvīgu adaptēšanos mērķa lietotāja uzvedībai, kā arī iespēju vienam lietotājam elastīgi pielāgot dažādus profilus.

***Anomalitātes līmeņa dinamisks slieksnis.*** Bāzes algoritmā izmantotā pieeja, kad anomalitātes slieksni nosaka konstantes *r* fiksēta vērtība, nav veiksmīgākā visām realizācijām. Pieaugot vispārējai metrikas vērtībai, tās līmenis var pārsniegt uzstādīto slieksni, un tālāk arī tipiska lietotāja uzvedība tiks atzīta par anomālu. Rezultātā iestājas šāda situācija:

- pastāvīgas *„normālas”* uzvedības rezultātā tiek nodrošināta anomalitātes metrikas vērtību stacionāra daba,
- pēc tam, iegūstot zīmīgi anomālus pieprasījumus, metrikas līmenis samazinās,
- tad seko liels „normālo” pieprasījumu apjoms, kas atkal nodrošina stacionāru metrikas vērtību, bet jau citā, zemākā, līmenī.

Šīs problēmas risināšanai tiek piedāvāta pieeja, izmantojot **dinamisku slieksni** — *r**. Šajā gadījumā vērtība *r** tiek izskaitļota katru reizi, kad tiek izskaitļota moduļa atgriežamā atbilde. Citiem vārdiem sakot, *r** ir dinamiska konkrētā brīža anomalitātes līmeņa sliekšņa vērtība katram saņemtajam pieprasījumam.

Anomalitātes sliekšņa vērtības izmaiņu dinamika kādam testa lietotāja uzvedības gadījumam ir parādīta 5. attēlā.




5. att. Anomalitātes sliekšņa izmaiņu dinamika

## Trešā nodaļa. Lietotāja uzvedības profila efektivitātes novērtējumu apskats

Pastāv dažādas informācijas sistēmu efektivitātes novērtēšanas pieejas [98, 19]. Atkarībā no apskatāmā uzdevuma specifikas, tiek izmantoti vai nu vispārīgi efektivitātes rādītāji (**uzticamība, ticamība** un **drošība**) vai atsevišķu rādītāju kopums (kas apraksta sistēmas **pragmatisko, tehnisko, tehnoloģisko** un **ekspluatācijas efektivitāti**). Atkarībā no izvēlētā pamata kritērija un rādītāju sistēmas tiek izmantotas dažādas matemātiskas metodes sarežģītu sistēmu efektivitātes novērtēšanai, jo sevišķi informācijas teorijas un varbūtību teorijas metodes.

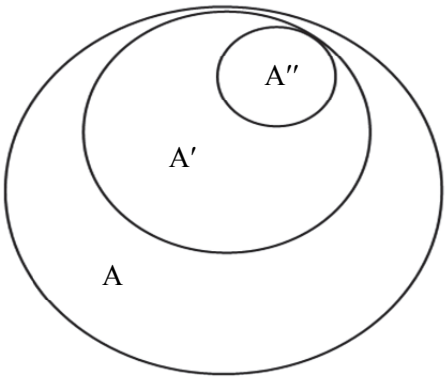
**Lietotāja uzvedības profila kvalitātes novērtēšanas varbūtiskie raksturlielumi.** Izmantojot varbūtisko pieeju, pats novērtējamais objekts tiek apskatīts kā *„melnā kaste”* [29]. Objektu salīdzināšana tiek veikta netieši, salīdzinot statistiskos parametrus, kas tiek iegūti no izejošās datu plūsmas.

Attiecībā uz LUP šāda pieeja var tikt realizēta, salīdzinot dažādu metrikas īpašību raksturlielumus, kuri tiek iegūti no profiliem, kas apmācīti, izmantojot atšķirīgas kopas [72]. Ieviešot lielas izmaiņas apmācības kopā, būtu jāvar novērot arī lielas atšķirības anomālas uzvedības atklāšanas rezultējošajā kvalitātē.

Izmantojot šādu pieeju, profilus var apskatīt gan pēc to atšķirībām no kopējā klases profila, gan pēc to statistiskajām īpašībām, salīdzinot dispersijas un standartnovirzes tām metriku vērtību kopām, kas iegūtas no atšķirīgiem profiliem.

Veiksmīgākai vizualizācijai 6. attēlā shematiski parādīti interešu līmeņi. Pieņemsim, ka $A$ šajā gadījumā apzīmē visu iespējamo pieprasījumu tipu pilno kopu mērķa sistēmā. Tad $A'$ ir tikai lietotāju daļas interešu apakškopa, kas raksturo lietotājus, kurus vieno viena un tā pati loma šajā sistēmā. Bet $A''$ ir $A'$ apakškopa, kas apzīmē viena konkrēta lietotāja specifiskās

intereses. Ir skaidrs, ka profilam, kas ņem vērā $A''$ grupas intereses, būs citi statistiskie raksturlielumi, kas būs atšķirīgi no $A'$ (sk. 6. att.).




6. att. Interešu līmeņu iekļaušana

Pieņemot interešu gradāciju šādā interpretācijā, ir iespējams pāriet uz **LUP kvalitātes novērtējuma varbūtiskajiem raksturlielumiem**. LUP kvalitāti šajā gadījumā var novērtēt, izmantojot izteiksmi *„1 — profila slēdziena kļūdas varbūtība”* [50]. Turklāt konkrētajai uzvedībai labāk der tas profils, kurš uzrāda mazāku abu tipu kļūdas varbūtību (kad *„anomāla”* uzvedība tiek noteikta par *„normālu”* un otrādi):

$$Q(\text{LUP}) = 1 - (p(Err_{fpr}) + p(Err_{fnr})),$$

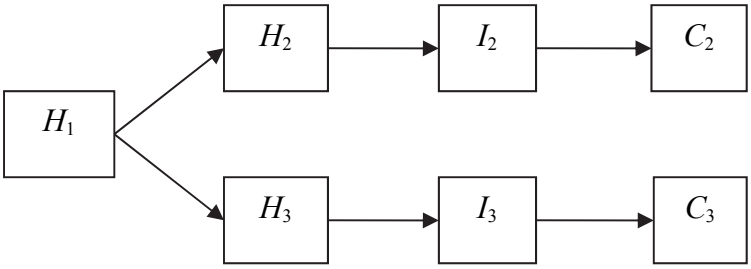
kur $Q$ — vispārīgs profila derīguma apzīmējums,

$p(Err_{fpr})$ — pirmā tipa kļūdas varbūtība (*False positive error rate*),

$p(Err_{fnr})$ — otrā tipa kļūdas varbūtība (*False negative error rate*).

Izmantojot šādu pieeju, ir iespējams attēlot kopīgā LUP efektivitāti, salīdzinot to ar LUPAP, kad, iekļaujoties kopīgā profilā, var veikt daudz lielāku dažādu darbību skaitu, kuras tiks uzskatītas par *„normālām”*, bet personīgais profils apraksta ierobežotu apakškopu, kurā iekļauties svešam cilvēkam ir daudz sarežģītāk.

**Lietotāja uzvedības profila kvalitātes novērtēšanas informācijas kritēriji.** Galvenais novērtējuma raksturlielums, izmantojot informācijas kritērijus, ir *informācijas caurlaides spēja* [19, 90, 4, 34]. Jo tā ir lielāka, jo efektīvāk sistēma darbojas, jo vairāk informācijas tā izgūst no katra saņemtā ziņojuma. Tāpēc ir pietiekami salīdzināt anomālās aktivitātes atklāšanas efektivitātes rādītājus, izmantojot kopīgo LUP un LUPAP. Šajā gadījumā par novērtējuma mēru tiek pieņemta entropija, kas raksturo katru ienākošo paziņojumu par lietotāja veikto darbību, kas ļauj salīdzināt šos rādītājus, izmantojot atšķirīgas pieejas (7. att.).




7. att. Efektivitātes salīdzināšana, izmantojot entropijas atšķirības

Attēlā sistēmas darbības atšķirīgo stāvokļu un tos aprakstošo raksturlielumu definēšanai izmantoti šādi apzīmējumi:

$H_1$ — sistēmas sākotnējā entropija, neizmantojot LUP,

$H_2$ — sistēmas entropija pēc kopīgā LUP ieviešanas,

$H_3$ — sistēmas entropija pēc LUPAP ieviešanas,

$I_j$ — vidējais informācijas apjoms, nomainot vienu profilu pret citu ($I_2 = H_1 - H_2; I_3 = H_1 - H_3$),

$C_i = I_j \setminus \tau$ — sistēmas informācijas caurlaides spēja, transakcijas $\tau$ analīzei izmantojot profilu *j*.

Rezultātā sistēma ar augstāko C vērtību tiek atzīta par efektīvāko, jo tās informācijas caurlaides spēja ir lielāka:

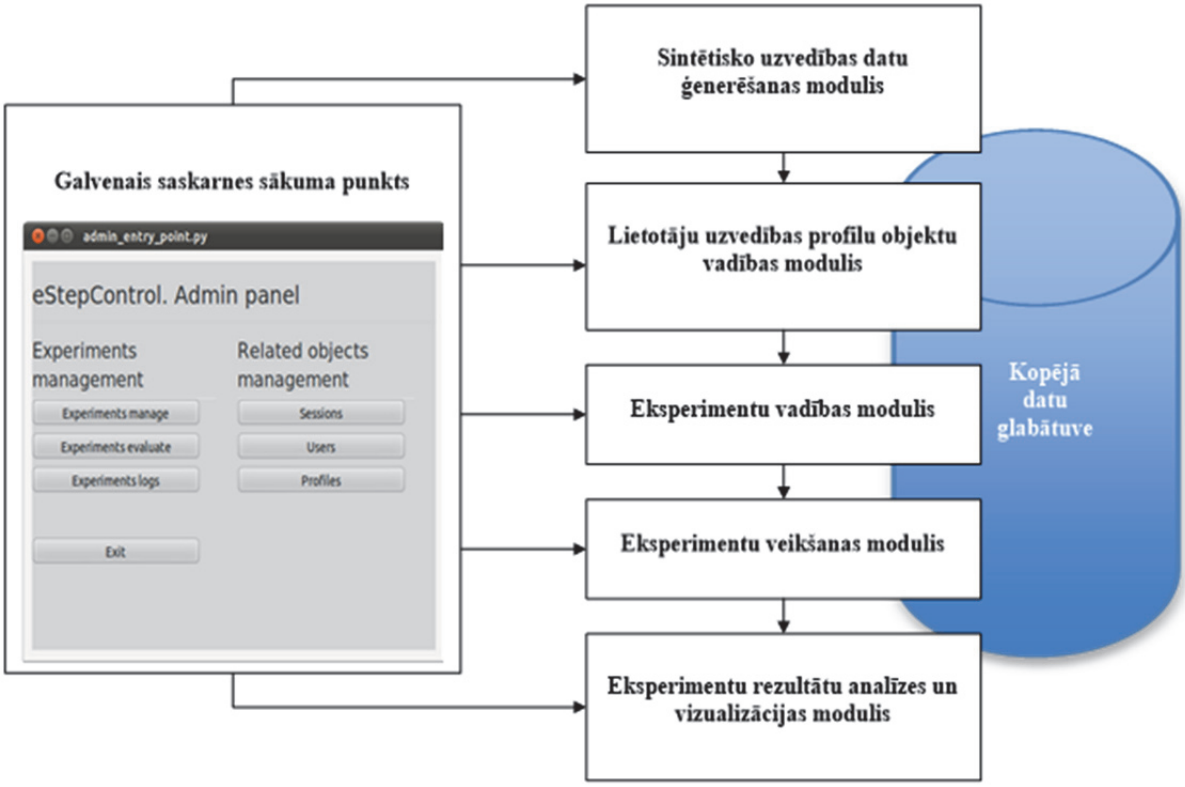
$$E = \begin{cases} C_2, & C_2 > C_3, \\ C_3, & \text{citos gadījumos,} \end{cases}$$

kur $E$ — efektīvākais profils.

## Ceturtā nodaļa. Eksperimentālās sistēmas programmiskā realizācija

Lai demonstrētu, ka teorētiski darboties spējīga sistēma spēj risināt reālus uzdevumus, tika izveidota programmatūras platforma eksperimentu veikšanai. Visas teorētiski aprakstītās operācijas tika realizētas, izmantojot programmēšanas valodas *Python* [59] un *PHP* [30], kā arī tika izveidota infrastruktūra eksperimentu veikšanai.

Eksperimentāla pētījuma veikšanai ar izstrādāto teorētisko pieeju, tika izveidota programmatūras sistēma. Tās vispārīga struktūra parādīta 8. attēlā. Galvenais ieejas punkts attēlo pieejamu eksperimentu veikšanas izvēlni. Pati sistēma sadalīta vairākos loģiskos moduļos, un katrs no tiem realizē savu nepieciešamās funkcionalitātes daļu.




8. att. Eksperimentālās sistēmas vispārīga struktūra

**Sesiju ģenerēšanas modulis** izveido lietotāju uzvedības sesijas, kuras raksturo dažādi nepieciešamie raksturlielumi. Uzvedības sesiju ģenerēšana var tikt veikta divos režīmos: manuālā un automātiskā.

**Uzvedības profilu vadības modulis** nodrošina LUP programmisko eksemplāru izveidi, apmācību un dzēšanu.

**Eksperimentu vadības modulis** paredzēts eksperimentu tiešai inicializācijai. Katram eksperimentam ir savs mērķis, sava izmantoto uzvedības profilu kopa, izmantoto sesiju kopa un iekšējo parametru vērtību kopa.

**Eksperimentu veikšanas modulis** paredzēts eksperimentu veikšanas procesa vadīšanai. Pēc eksperimenta izveidošanas to var veikt vairakkārt, tā iekšējie parametri nemainās, bet apmācībai un analīzei tiek izmantoti jauni dati par uzvedību.

**Rezultātu vizualizācijas modulis** veic rezultātu izvadi diagrammu veidā vai sākotnējā formā, lai tos varētu apstrādāt citos programmlīdzekļos, kā MS Excel [10] 10 vai Statistica [44]. Tipisks eksperimenta rezultāts var būt anomalitātes līmeņa vērtības izmaiņu dinamika, un šajā gadījumā modulis izvadīs tās diagrammu.

**Testa problēmsfēras ierobežojumi.** Problēmsfēra, kurā sistēmu izmanto, izvirza diezgan stingras prasības pret katras transakcijas apstrādes ātrumu: tika izvēlēts augšējais 100 modeļu apstrādes laika slieksnis, kas atbilst 500 milisekundēm.

**Sistēmas realizācijas dalītā modeļa galvenie raksturlielumi.** Praktiski visi serveri, kas realizēti, izmantojot *Python* valodu, izmanto vienu no tālāk minētajām ienākošo pieprasījumu apstrādes pieejām:

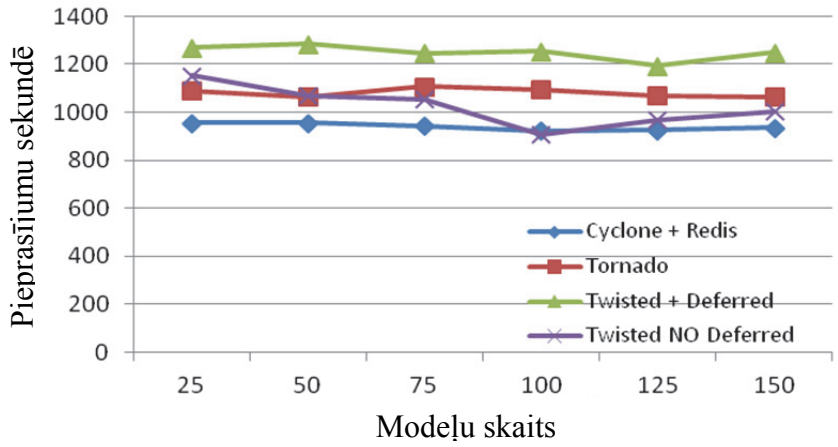
- sistēmas procesu izmantošana,
- sistēmas pavedienu (*system threads*) izmantošana,
- atliktā (*deferred*) pieprasījumu apstrāde.

Atliktā pieeja (*deferred*) [92, 81, 35] jāapskata sīkāk. Tās būtība slēpjas tajā, ka, saņemot pieprasījumu, tiek izsaukts modulis, kas atbild par tā apstrādi, tam tiek piešķirta funkcija — notikuma apdarinātājs *„Skaitļošana pabeigta”*, un pēc tā serveris „aizmirst” par saņemto pieprasījumu un netērē resursus tā uzturēšanai. Pēc kāda laika, kad pieprasījums ir pilnībā apstrādāts, serveris saņem signālu par notikuma *„Skaitļošana pabeigta”* iestāšanos un skaitļošanas rezultātu, tiek izsaukta iepriekš piešķirtā apstrādes funkcija, kas apkalpo katro atbildi.

Darbā ir veikti eksperimenti ar visām trim iepriekš minētajām ienākošo pieprasījumu apstrādes pieejām.

**Pieprasījuma apkalpošanas ātruma testēšana.** Par kandidātiem uz galvenā servera, kas apkalpo sistēmā ienākošos pieprasījumus anomalitātes līmeņa skaitļošanai, lomu tika izvēlēti trīs populāri *Python* serveri: *Twisted* [29], *Tornado* [102] un *Cyclone* [20].

Uz 9. attēlā parādīts apstrādāto pieprasījumu skaits sekundē, visiem šajā testēšanā izmantotajiem serveriem, izmantojot reālu modeļa ielādi un ar to veicot noteiktu operāciju daudzumu. Diagramma uzskatāmi parāda, ka atliktā (*deferred*) pieeja uzrāda vislabākos rezultātus.




9. att. Dažādu serveru reakcijas laiks

**Izstrādājamās sistēmas ieviešanas metodika.** Reālu uzdevumu apstākļos ne vienmēr var skaidri saprast, cik efektīvi izstrādājamo pieeju var izmantot dažādu mērķa sistēmu

ietvaros. Informācijas sistēmu uzbūves pieejas var stipri atšķirties, atkarībā no prasībām pret to struktūru, pieejamības līmeni un datu drošību. Rezultātā pastāv liels skaits **sarežģītu** informācijas sistēmu, kurām ir atšķirīgas arhitektūras, realizācijas metodes un platformas, bet tajās visās ir svarīgi atrast iespēju atklāt to lietotāju anomālas uzvedības gadījumus.

Kā sarežģītas informācijas sistēmas organizators var noteikt, ka šajā pētījumā izstrādātā pieeja anomālas aktivitātes atklāšanai ir izmantojama viņa sistēmā?

Tālāk tekstā sniegti mērķa IS galvenie raksturlielumi, kas aprakstītu sistēmu, kurā nepieciešams un ir pamatojums ieviest sistēmu, kas līdzīga šajā darbā izstrādātajai.

- Sensitīvu datu klātbūtne. Ja sistēma nedarbojas ar šādiem datiem, tad zīmīgi samazinās nepieciešamība ieviest šādu sarežģītu pieeju.
- Mērķa IS dalīta struktūra, kas realizēta, balstoties uz vienotas integrācijas platformas. Ja mērķa IS sastāv no nehomogēniem blokiem, kuri nav apvienoti kopējā loģiskā shēmā, tad izveidotās sistēmas ieviešanas iespējas var būt ierobežotas.
- Saglabāšanas iespējas un pieejas datiem par lietotāju pieprasījumiem esamība. Ja šādi dati nav pieejami, nav pamatinformācijas uzvedības profilu apmācībai.
- Drošības pārvaldības moduļa esamība mērķa IS. Izstrādātā pieeja ļauj novērtēt lietotāja pieprasījumu anomalitātes līmeni, bet par darbībām, kas jāveic, ja tiek atklāts iebrucējs, lemj ārpus izstrādātās sistēmas esošs drošības modulis.

Šādas funkcionalitātes ieviešanai, noteiktā laika posmā ir jābūt pieejamam budžetam, lai segtu izdevumus programmatūras un aparatūras nodrošināšanai. Rezultātā rodas prasība izveidotās sistēmas ārējo interfeisu unifikācijai, lai tās ieviešana dažādās mērķa platformās ietvertu vien to uzstādīšanu un nepieciešamo apkārtējās vides programmatūras servisu pieejamības testēšanu.

Izveidotajai lietotāja pieprasījuma anomalitātes līmeņa analīzes sistēmai ir nepieciešami vismaz trīs tipu ieejas dati:

- lietotāja unikālais identifikators,
- viņa lomas identifikators,
- viņa veiktās transakcijas identifikators.

Turklāt sistēmas apmācībai ir nepieciešami dati par lietotāja iepriekšējām sesijām darbā ar sistēmu. Rezultātā sistēmai ir jārealizē vismaz četri iekšējie interfeisi sadarbībai ar mērķa sistēmu, lai nodotu datus par katru no iepriekš minētajiem identifikatoriem. Atkarīgi no tā, cik pieejami ir nepieciešamie dati, mērķa sistēmas var iedalīt trīs grupās (skatīt 2. tabulu). No mērķa sistēmas tipa ir atkarīgas metodes, kas izmantotas tās mijiedarbībā ar anomāliju atklāšanas sistēmu. Iespējamie varianti tāpat norādīti 2. tabulā.

2. tabula

**Publisko saskarņu struktūra atkarībā no mērķa sistēmas tipa**

| *Mērķa sistēmas struktūras tips* | *Datu pieejamība* | *Sistēmas publisko saskarņu arhitektūra* |
|---|---|---|
| *Pilnībā pieejama* | *Pilnībā* | Izmanto standarta saskarnes. |
| *Daļēji pieejama* | *Daļēji* | Pieejamiem datiem izmanto standarta saskarnes, bet nepieejamiem datiem ir nepieciešama specifisku nestandarta saskarņu realizācija.<br>Izmanto nepieciešamo datu agregāciju no vairākiem mērķa sistēmas avotiem vai trūkstošo datu emulāciju, izmantojot konkrētajai mērķa sistēmai vispiemērotāko loģiku. |
| *Nepieejama* | *Nepieejami* | Visas saskarnes jārealizē nestandarta savienojošo skriptu veidā. |
| *Specifiski pieejama* | *Mērķa sistēma nenodrošina visus programmatūras vai aparatūras raksturlielumus, kas nepieciešami sistēmas funkcionēšanai* | Ja nav pieejami programmatūras vai aparatūras resursi, par to tiek informēts mērķa sistēmas administrators. Tālāk iespējami šādi darbības scenāriji:<br>• sistēmas darbības pārtraukšana,<br>• trūkstošo komponentu automātiskas uzstādīšanas mēģinājums,<br>• trūkstošo mezglu programmiska emulācija,<br>• nepieciešamo programmatūras komponentu aizvietojošu bibliotēku izmantošana (piemēram, ja nav pieejams *Redis*, izmanto citu *NoSQL* glabātuvi). |

Izstrādātās sistēmas ieviešanai nepieciešamo kritēriju noteikšana ļauj viegli noteikt, vai ir nepieciešams to ieviest attiecīgos dažādu mērķa IS specifikas apstākļos. Mērķa IS datu pieejamības līmeņu nošķiršana ļāva aprakstīt operāciju kopas, kas nepieciešamas izveidotās sistēmas ieviešanai katrā no tiem.

## Piektā nodaļa. Eksperimenti ar lietotāja uzvedības profiliem

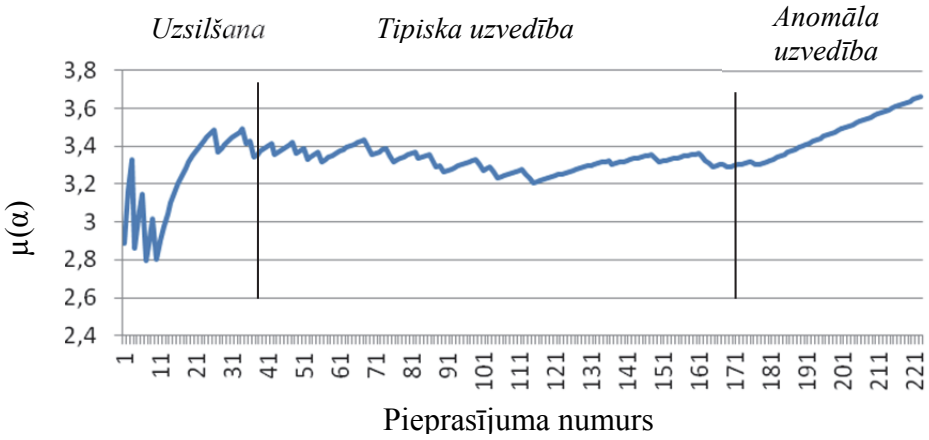
Eksperimenti ir galvenais šī pētījuma instruments. Katra veikto eksperimentu kopu ir veltīta svarīgai raksturlielumu daļai. Visu veikto eksperimentu kopējais uzdevums ir LUP un LUPAP raksturlielumu salīdzinošā apraksta saturiska papildināšana ar mērķi parādīt otrās pieejas priekšrocības.

**Pirmā eksperimentu sērija.** Sākotnēji eksperimentu veikšanai tika izveidota 1. versijas programmatūras realizācija. Tās pamata uzdevums bija bāzes algoritma izmantošanas iespējas principiāla novērtēšana, strādājot ar anomālas uzvedības atklāšanu. Pirmajā versijā tika realizētas testa datu kopu iegūšanas un uzglabāšanas iespējas, LUP izveide, kā arī mākslīgo transakciju kopu ģenerēšana un to anomalitātes līmeņa izskaitļošana.

Profils tiek izmantots analīzes režīmā, un anomalitātes līmeņa izmaiņu dinamika katram ienākošajam pieprasījumam tiek izvadīta diagrammas veidā (sk. 10. att.).

Šajā eksperimenta etapā atkarībā no procesa posma sistēma var atrasties vienā no trīs tālāk minētajiem stāvokļiem:

- ***sākuma stāvoklis (uzsilšana)*** — šajā etapā lietotāja trase nesatur pietiekami daudz informācijas, lai to efektīvi analizētu, un metrikas vērtības var būt neprecīzas un izteikti izkaisītas plašā diapazonā,
- ***normāla uzvedība*** — šajā etapā metrikas vērtībai jābūt stacionārā režīmā,
- ***anomāla uzvedība*** — šajā etapā metrikas vērtībai ir pastāvīgi jāpieaug.




10. att. Metrikas dinamika, izmantojot parametrus $w = 2$ un $Z = 1{,}5$

Metrikas vērtības izmaiņu dinamika pirmās eksperimentu sērijas gaitā atbilst teorētiski paredzamajai uzvedībai.

**Otrā eksperimentu sērija.** Šīs sadaļas mērķis ir formāli pierādīt, ka LUPAP būs efektīvāks (anomālas aktivitātes atklāšanas ziņā) par kopējo LUP. Divu profilu efektivitātes salīdzināšanai katrs no tiem tika apmācīts ar divām atšķirīgām datu kopām, kurām ir atšķirīgi interešu prioritāšu sadalījumi un kuras izveidotas, izmantojot vienādas algoritma iekšējo parametru vērtības. Katrs profils tika izmantots anomālas aktivitātes atklāšanai, izmantojot visus iespējamos *profils/dati* pāru variantus (sk. 3. tabulu).

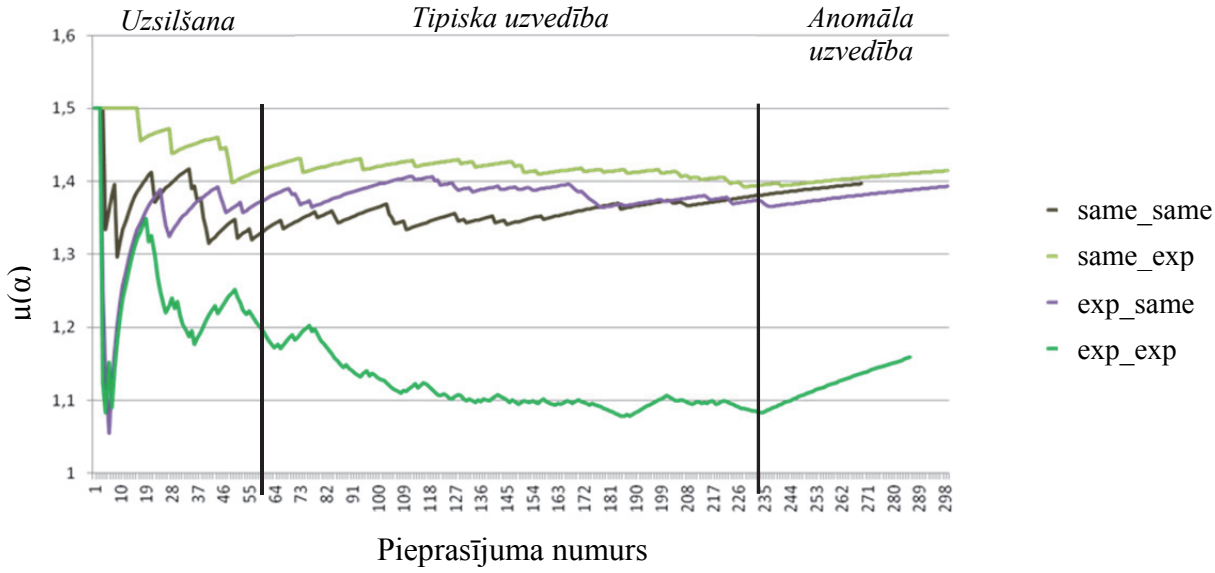
Piemēram, *same_exp* norāda, ka tiek izmantots profils, kas apmācīts, izmantojot vienmērīgi sadalītas intereses, bet analizē transakcijas ar eksponenciāli sadalītām interesēm.

3. tabula

**Publisko saskarņu struktūra atkarībā no mērķa sistēmas tipa**

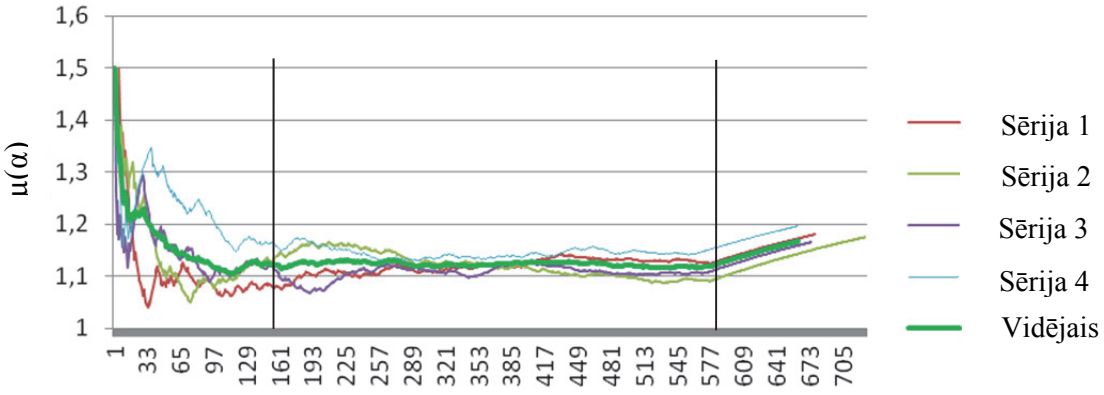
| *Profils \ Uzvedība* | *Lietotāju klases uzvedība* | *Individuāla uzvedība* |
|---|---|---|
| *Lietotāju klases uzvedība* | same_same | same_exp |
| *Individuāla uzvedība* | exp_same | exp_exp |

Apkopojoša diagramma par visu eksperimentu rezultātiem ir sniegta 11. attēlā. No tā ir redzams, ka *same_same, same_exp* un *exp_same* gadījumos profilu uzvedības atšķirības nav zīmīgas, bet tikai eksperiments *exp_exp* norāda uz krasi atšķirīgu uzvedību. Šāds rezultāts apstiprina hipotēzi par to, ka individuāls personīgais profils būs labāks klasifikators tieši tā konkrētā lietotāja uzvedībai, par kuru ievāktie dati tika izmantoti profila apmācībai.




11. att. Eksperimentu rezultātus apkopojoša diagramma

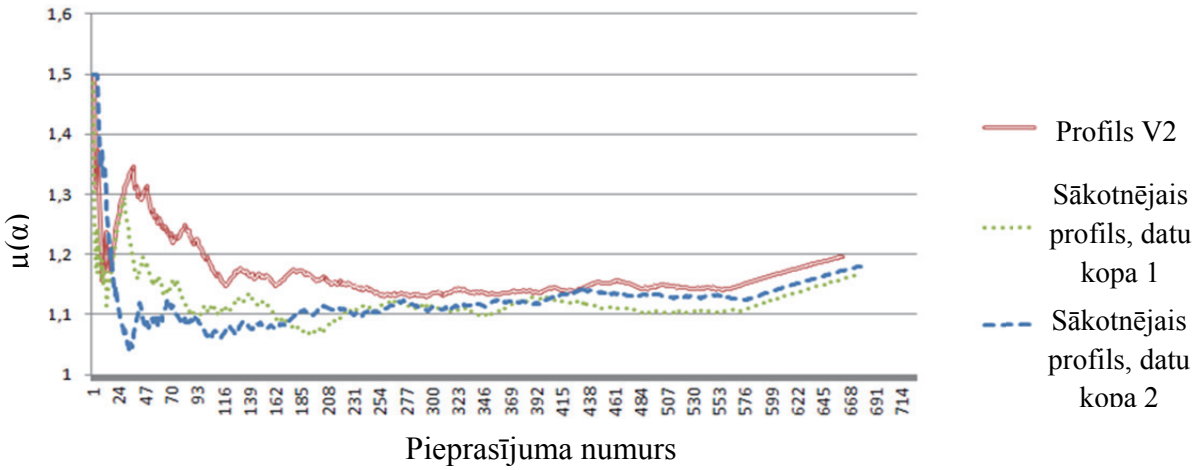
**Trešā eksperimentu sērija.** Trešās eksperimentu sērijas rezultāti (12. attēlā) norāda, ka laika gaitā iekšējie svari veiksmīgi līdzsvaro metrikas vērtības, nodrošinot stacionāru režīmu, ja lielāka pieprasījumu apjoma apmērā ir pastāvīga *„normāla"* lietotāja uzvedība.




12. att. Liela *„normālu"* darbību apjoma analīzes rezultāti, izmantojot profilu *exp_exp*

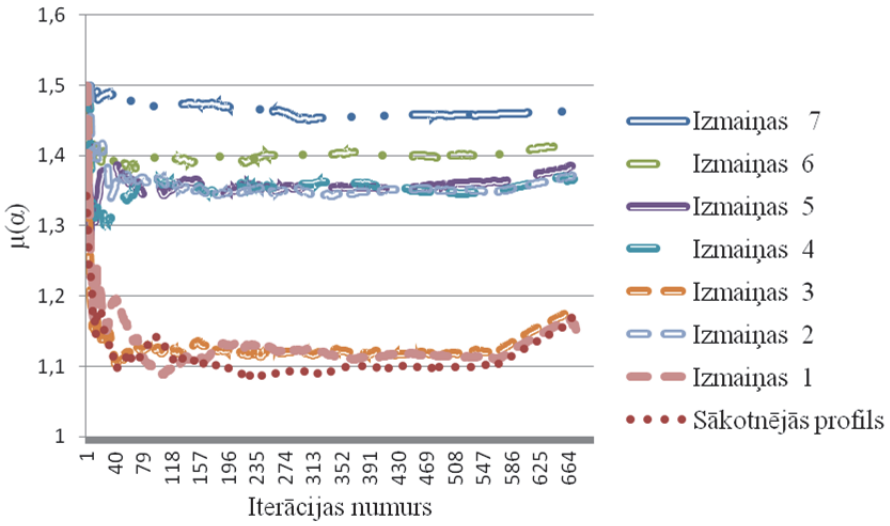
**Ceturtā eksperimentu sērija.** Ceturtās eksperimentu sērijas rezultāti ļauj secināt, ka pastāv minimāla atkarība no konkrētām vērtībām (13. attēls), kas atbilst apmācībā

izmantotajām transakciju kopām, jo profila izveidē galvenais faktors ir tieši lietotāja **uzvedības** specifika, bet viņa pieprasīto transakciju **secībai** nav svarīgas nozīmes.




13. att. Profili, kas izveidoti, izmantojot vienādus datus

**Piektā eksperimentu sērija.** Pēc piektās eksperimentu sērijas rezultātiem var spriest, ka, palielinoties prioritāšu apjomam lietotāja uzvedībā, metrikas vērtības tāpat attiecīgi izmainās. Turklāt — jo lielāka prioritāšu dažādība, jo lielāka ir arī anomalitātes līmeņa metrikas vērtību atšķirība (14. attēls).




14. att. Metriku apkopotās vērtības trešā eksperimenta etapiem

## Kopsavilkums un secinājumi

**Rezultātu analīze un diskusija**. Šī pētījuma galvenais rezultāts ir izstrādātā metodika efektīvai anomālijas līmeņa noteikšanai elektroniskas informācijas sistēmas lietotāja uzvedībā. Izveidotās pieejas efektivitāte atbilst sākotnēji noteiktajiem ierobežojumiem.

Lietotāja uzvedības personīgā adaptīvā profila (LUPAP) ieviešana ļāva palielināt bāzes pieejas precizitāti anomālas uzvedības atklāšanas procesā.

Efektivitātes rādītāju novērtēšanai un pierādīšanai tika veiktas dažādu eksperimentu sērijas, kas apskata dažādus izveidotās pieejas raksturlielumus.

Pēc pētījuma rezultātiem var spriest, ka darba sākumā izvirzītā galvenā hipotēze, kas saistīta ar lietotāja personīgo interešu mērķa sistēmā analīzes nepieciešamību, ir apstiprināta, un LUPAP ir labāks lietotāja uzvedības tipa klasifikators par LUP un līdz ar to spēj efektīvāk noteikt lietotāja uzvedības anomalitātes līmeni.

Galvenās hipotēzes ietvaros tāpat tika apstiprinātas arī sākotnēji izvirzītās pakārtotās hipotēzes:

**Jā** — Markova ķēdes grafu ir vērtīgi izmantot, lai atspoguļotu informāciju par lietotāja uzvedības īpatnībām.

**Jā** — pastāv reāla iespēja salīdzināt lietotāja uzvedības profilu efektivitāti.

**Jā** — pastāv iespēja ar programmatūras palīdzību ģenerēt datus par lietotāja uzvedību, ņemot vērtā viņa atšķirīgās intereses mērķa sistēmas ietvaros.

Galvenie rezultāti, kas sasniegti pētījuma un promocijas darba izstrādes ietvaros, var tikt formulēti tā, kā norādīts tālāk.

1) Ir izanalizētas pastāvošās mūsdienīgās pieejas informācijas sistēmas lietotāju anomālas uzvedības atklāšanai. Pētījumu gaitā ir secināts, ka neviena no apskatītajām pieejām nenodrošina pilnīgu visu izvirzīto prasību realizāciju.
2) Ir izpētītas un izanalizētas pastāvošās lietotāju uzvedības formalizācijas izveides metodes, kas balstās uz dažādu veidu Baijesa tīklu, inženierontoloģiju un mobilo aģentu izmantošanas.
3) Ir izpētītas un demonstrētas Markova ķēžu izmantošanas iespējas lietotāja uzvedības profila formalizācijas uzdevumos, kas apliecina izvirzītās zemākā līmeņa hipotēzes patiesumu.
4) Izstrādāta bāzes pieeja lietotāju kopējā uzvedības profila izveidei un izmantošanai.
5) Lai uzlabotu bāzes pieejas efektivitāti, nosakot iebrukumus un uzvedības anomalitāti, ir izstrādāta metode, kas izmanto lietotāja uzvedības personīgo adaptīvo profilu.
6) Ir izstrādāta metodika uzvedības profilu efektivitātes novērtēšanai, kas ļauj veikt personīgā adaptīvā un kopējā lietotāju uzvedības profilu salīdzinošo analīzi.
7) Izstrādātā anomalitātes līmeņa dinamiskā sliekšņa metode ļauj uzlabot lietotāju uzvedības klasifikāciju divās grupās — *„normāla”* un *„anomāla”*.
8) Izmantojot izstrādāto lietotāju uzvedības anomalitātes atklāšanas metodiku, ir veikta LUPAP un LUP uzvedības profilu efektivitātes salīdzinoša eksperimentālā analīze.

9) Ir izstrādātas un programmatūras sistēmas veidā realizētas dažādas eksperimentu kopas, kas ļauj novērtēt dažādus uzvedības profilu raksturlielumus.

### Promocijas darba galvenie secinājumi

Pēc promocijas darba gaitā veiktā pētījuma un eksperimentu rezultātiem var izdarīt šādus galvenos secinājumus:

1) Uz Markova ķēžu pamata var ātri izveidot un efektīvi izmantot lietotāju uzvedības profilus. Lietotāja uzvedības specifika ļauj unikāli definēt uzvedības profila struktūru.
2) Saskaņā ar uzvedības profilu efektivitātes novērtēšanas metodiku ir iespējams salīdzināt dažādu lietotāju uzvedības profilu efektivitāti.
3) Ieviešot lietotāja uzvedības personīgo adaptīvo profilu, tika uzlabota uzvedības anomalitātes atklāšanas precizitāte.

Realizēto eksperimentu rezultāti norāda, ka lietotāja uzvedības personīgā adaptīvā profila metode ir efektīvāka, nekā kopējais lietotāju uzvedības profils.

## BIBLIOGRĀFISKAIS SARAKSTS


1. Alavi M., Leidner D. E. Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues // *MIS Quarterly*.— March 2001.— Vol. 25, No. 1.— P. 107–136.— Available on-line on: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.8885&rep=rep1&type=pdf. — Last accessed 2014.04.08.
2. Antoniou G., Harmelen F. Web Ontology Language: OWL // Handbook on Ontologies / S. Staab, R. Studer (Eds.).— Berlin: Springer-Verlag, 2004, p. 67–92 (Series: International Handbooks on Information Systems).
3. *ApacheBench*.— Copyright Adam Twiss, Zeus Technology Ltd., 1996.
4. Arndt C. *Information Measures: Information and its Description in Science and Engineering*.— Springer-Verlag, 2013.— ISBN 978-3-540-40855-0. (*Springer Series: Signals and Communication Technology*).
5. Bahder T. B. *Mathematica for Scientists and Engineers*.— Addison-Wesley Publ. Company, 1995.— ISBN-10: 0201540908.
6. Baum L. E., Petrie T. Statistical inference for probabilistic functions of finte state Markov chains // *Annls of Mathematical Statistics*.— 1966.— No. 37.— P. 1554–1563.

7. Beizer B. *Black-Box Testing*.— John Wiley, 1995.— ISBN: 0471120944, 9780471120940.

8. Bernard V. L. *A Guide to Microsoft Excel 2002 for Scientists and Engineers*.— St. Francis Xavier University Nova Scotia, Canada; 2003.— 320 p.— ISBN 0 7506 5613 1.

9. Bhuyan M. H., Bhattacharyya D. K., Kalita J. K. Network Anomaly Detection: Methods, Systems and Tools // *IEEE Communications Surveys & Tutorials*.— 2013.— Vol. 16, Issue 1.— P. 303–336.— ISSN:1553-877X.

10. Billo E. J. *Excel for Scientists and Engineers: Numerical Methods.* 1st ed.— Wiley-Interscience, 2007.— 480 p.— ISBN-13: 978-0471387343.

11. Bishop C. M.. Pattern Recognition and Machine Learning (*Information Science and Statistics*).— NY: Springer-Verlag, 2006. 2006.

12. Bongard M. *Pattern Recognition*.— SAMS, 2000.— ISBN: 0810491656.

13. Brandes U., Eiglsperger M., Lerner J., Pich C. *Graph Markup Language (GraphML)*.— CRC Press, LLC, 2004.

14. Brooks D. R. *An Introduction to PHP for Scientists and Engineers*.— Springer Science Media, 2008.— ISBN-10: 184800236X.

15. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // *ACM Computing Surveys*.— 2009.— No. 9.— P. 1–72.

16. Chang M., Mathiske B., Smith E., Chaudhuri A., Bebenita M., Gal A., Wimmer Ch., Franz M. *The Impact of Optional Type Information on JIT Compilation Of Dynamically Typed Languages //* 7th Dynamic Languages Symposium (DLS 2011), Portland, Oregon, ACM Press, ISBN 978-1-4503-0939-4, pp. 13–24; October 2011. doi:10.1145/2047849.2047853.

17. Chunfu J., Feng Y. An Intrusion Detection Method Based on Hierarchical Hidden Markov Models // *Wuhan University Jornual of Natural Sciences*.— 2007.— Vol. 12, No. 1.— P. 135–128.

18. Cost of Cyber Crime Study: United States Benchmark Study of U.S. Companies.— Ponemon Institute, October 2013.— Available on-line on: http://media.scmagazine.com/documents/54/2013_us_ccc_report_final_61_13455.pdf. — Last accessed 2014.02.15.

19. Cover T. M., Thomas J. A. Elements of Information Theory.— John Wiley & Sons, 1991.— Print ISBN 0-471-06259-6. Online ISBN 0-471-20061-1.

20. Cyclone.io project. web-page / Internet.— www.cyclone.io — Last accessed 2014.02.06.

21. DARPA Agent Markup Language + Ontology Interface Layer / Internet.— http://www.daml.org/2001/03/daml+oil-index.— Last accessed 2014.10.11.

22. Dasgupta D., Gonsalez F., Yallapu K., Gomez J., Yarramstetti R.. CIDS: An agent-based intrusion detection system // Computers & Security.— 2005.— Vol. 24.— P. 387–398.

23. Day J. D., Zimmermann H. The OSI Reference Model // Proceedings of the IEFJ2.— 1983.— Vol. 71, No. 12.

24. Downey A. B. How to think like a computer scientist. C++ Version. 2012 / Internet.— http://www.xplora.org/downloads/Knoppix/books/Open_Book_Project/thinkCScpp.pdf. — Last viewed 2014.09.30.

25. DTI 2002. Information Security Breaches Survey 2002. Technical report. Department of Trade & Industry, April 2002. URN 02/318.— Available on-line on: http://www.vicomsoft.com/downloads/learning/dti_security_survey.pdf.— Last accessed 2013.10.12.

26. Duval T., Jouga B., Roger L. The Mitnick Case: How Bayes Could Have Helped // IFIP International Federation for Information Processing.— 2005.— Vol. 194/2005.— P. 91–104.— DOI: 10.1007/0-387-31163-7_8.

27. EGEE — Enabling Grids for E Enabling Grids for E-science / Internet.— http://www.eu-egee.org/.— Last accessed 2012.04.24.

28. Elliotte R. H., Means W. S. XML in a Nutshell. 3rd Edition.— O'Reilly Media; 2009.— Print ISBN: 978-0-596-00764-5.

29. Fettig A. Twisted Network Programming Essentials.— O'Reilly Media, 2005.— 238 p.— Print ISBN: 978-0-596-10032-2. ISBN 10: 0-596-10032-9.

30. Fine S, Singer Y, Tishby N. The Hierarchical Hidden Markov Model: Analysis and Applications // J. Machine Learning.— 1998.— Vol. 32, No. 1.— P.41–62.

31. Foster I. The Grid: Blueprint for a New Computing Infrastructure.— Morgan Kaufmann Publishers, 1999.— ISBN 1-55860-475-8.

32. GILDA virtual laboratory / Internet.— https://gilda.ct.infn.it/.— Last accessed 2012.04.24.

33. Gray J. The Transaction Concept: Virtues and Limitations // Proceedings of the 7th International Conference on Very Large Databases, September 9–11, 1981, Cannes, France.— IEEE Press, 1981.— P. 144–154.

34. Gray R. M. Entropy and Information Theory. 2nd ed.— New York: Springer US, 2011.— 409 p.— ISBN 978-1-4419-7970-4 (Online).

35. Grinstead C. M., Snell J .L. Introduction to Probability. 2nd ed.— American Mathematical Society, 1997.— 510 p.— ISBN-10: 0-8218-9414-5, ISBN-13: 978-0-8218-9414-9.

36. Gunter T. D; Nicolas P. T. The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions // Journal of Medical Internet Research.— 2005. Vol. 7, No. 1.— Doi: 10.2196/jmir.7.1.e3. PMC 1550638. PMID 15829475.

37. Hahn B. H., Valentine D. T. Essential MATLAB for Engineers and Scientists. 4th edition.— Elsevier Academic Press, 2009.— 416 p. — ISBN: 978-0-12-374883-6.

38. Hahn B. H. Fortran 90 for Scientists and Engineers.— London: Butterworth-Heinemann, Cambridge University Press, 1994. ISBN-10: 0-340-60034-9.

39. Harary F. Graph theory.— Addison-Wesley Publishing, 1969.

40. Hassanzadeh O. Introduction to Semantic Web Technologies & Linked Data, University of Toronto, CS 443: Database Management Systems — Winter 2011.— Available on-line on: http://www.cs.toronto.edu/~oktie/slides/web-of-data-intro.pdf.— Last accessed 2014.04.08.

41. Hawkins D. M. The Problem of Overfitting // J. Chem. Inf. Comput. Sci.— 2004.— Vol. 44.— P. 1–12.

42. Haykin S. Neural Networks: A Comprehensive Foundation.— New Jersey: Prentice Hall, 1999.

43. EHR — Electronic Health Records: Manual for Developing Countries.— WHO Library Cataloguing in Publication Data; World Health Organization, 2006.— ISBN 92 9061 2177.— Available on-line on: http://www.wpro.who.int/publications/docs/EHRmanual.pdf.— Last accessed 2014.05.27.

44. Hill T., Lewicki P. STATISTICS: Methods and Applications.— StatSoft. Tulsa. 2007.— Available on-line on: http://www.statsoft.com/textbook/.

45. IMS — Learner Information Packaging Information Model Specification, Final Specification. Version 1.0 / Internet.— http://www.imsglobal.org/profiles/lipinfo01.html.— Last accessed 2012.04.24.

46. Isaza G., Castillo A., López M., Castillo L. Towards Ontology-Based Intelligent Model for Intrusion Detection and Prevention // Advances in Soft Computing.— 2009.— Vol. 63.— P. 109–116.— DOI: 10.1007/978-3-642-04091-7_14.

47. Ivanov Y, Bobick A. Recognition of Visual Activities and Interactions by Stochastic Parsing // J. IEEE Trans on Pattern Analysis and Machine Intelligence.— 2000.— Vol. 22, No. 8.— P. 852–872.

48. Jaiganesh V., Mangayarkarasi S., Dr. Sumathi P. Intrusion Detection Systems: A Survey and Analysis of Classification Techniques // International Journal of Advanced Research in Computer and Communication Engineering.— 2013.— Vol. 2, Issue 4.— P. 1629–1635.— ISSN (Print): 2319-5940, ISSN (Online): 2278-1021.

49. Jha S., Kruger L., Kurtx T., Lee Y., Smith A. A Filtering Approach To Anomaly and Masquerade Detection. 2005. Technical report, Univ of Wisconsin, Madison.

50. Jha S., Tan K., Maxion R. A. Markov Chains, Classifiers and Intrusion Detection // Computer Security Foundations Workshop (CSFW), 2001. Proceedings. 14th IEEE , vol. 1, pp. 206–219, DOI: 10.1109/CSFW.2001.930147. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7408 Last accessed 2014.02.12

51. Joaquim P. M. Applied Statistics Using SPSS, STATISTICA, MATLAB and R. 2nd ed.— Springer Press, 2007. ISBN 978-3-540-71972-4.

52. Johnson N. L., Kotz S., Balakrishnan N. Continuous univariate distributions. Vol. 1, 2nd ed.— New York: John Wiley & Sons, 1994.— ISBN 978-0-471-58495-7. (Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics).

53. Judea P. Causal inference in statistics: An overview // Statist. Surv.— 2009.— Vol. 3.— P. 96–146.— Doi:10.1214/09-SS057. http://projecteuclid.org/euclid.ssu/1255440554.

54. Kallenberg O. Foundations of Modern Probability, 2nd ed.— Springer-Verlag, 2002.— 650 p. (Springer Series in Statistics).— ISBN 0-387-95313-2.

55. J. Kopena and W. C. Regli, “DAMLJessKB: A tool for reasoning with the semantic web,” IEEE Intelligent Systems, Vol. 18, pp. 74–77, 2003.

56. Laskey, K. B., Alghamdi, G., Wang, X., Barbara, D., Shackleford, T., Wright, E., Fitzgerald, J. Detecting Threatening Behavior Using Bayesian Networks // Proceedings of the Conference on Behavioral Representation in Modeling and Simulation, 2004.

57. Lawson, T. A Conception of Ontology.— University of Cambridge, 2004.

58. Liepins G. E. and Vaccaro H. S. Anomaly Detection: Purpose and Framework // Proceedings of the 12th National Computer Security Conference, October 1989. P. 495–504.

59. Lucks J. B. Python — All a Scientist Needs.— Pycon, 2008.— arXiv:0803.1838v1.

60. M´etivier F. Scientific Relational Databases using MySQL and Python: Lecture notes.— Paris: Institut de physique du globe de Paris & Universit´e Paris Diderot, 2014.— 53 p.

61. Manavoglu E., Pavlov D., Lee C. Probabilistic User Behavior Models // Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03), November 2003, Melbourne, Florida.— IEEE, 2003, p. 203–210.

62. Markov A. A. Theory of Algorithms.— M.: 1954. [Translated by Jacques J. Schorr-Kon and PST staff] Imprint Moscow, Academy of Sciences of the USSR, 1954 [Jerusalem, Israel Program for Scientific Translations, 1961; available from Office of Technical Services, United States Department of Commerce] Added t.p. in Russian Translation of Works of the Mathematical Institute, Academy of Sciences of the USSR, v. 42. Original title: Teoriya algorifmov. [QA248.M2943 Dartmouth College library. U.S. Dept. of Commerce, Office of Technical Services, number OTS 60-51085.].

63. Maxfield B. Essential MATHCAD for Engineering, Science and Math.— London: Elsevier Academic Press, 2009.— ISBN: 978-0-12-374783-9.

64. Mea D. V. What is e-Health: The death of telemedicine? // Journal of Medical Internet Research.— 2001.— Vol. 3, No. 2.— Doi:10.2196/jmir.3.2.e22.

65. Millman, K. J., Aivazis M. Python for Scientists and Engineers. University of California, Berkeley // IEEE Computational Science & Engineering.— 2011. Vol. 13, Issue 2.— P. 9–12.— ISSN: 1521-9615.

66. Mun G. J., Kim Y. M., Kim D. K., Noh B. N. Network Intrusion Detection Using Statistical Probability Distribution // Computational Science and Its Applications — ICCSA 2006.— Springer-Verlag Berlin Heidelberg, 2006, p. 340–348, (Lecture Notes in Computer Science, Vol.3981).

67. Networkx / Internet.— http://networkx.lanl.gov/. — Last viewed at 2014–02–07.

68. Noy N. F., McGuinness D. L. Ontology Development 101: A Guide to Creating Your First Ontology. KSL Technical Report,— Stanford University, Stanford, CA, 94305, 2006.— Availible on-line: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.— Last accessed 2015.01.07.

69. Orwant J. Computer Science & Perl Programming: Best of TPJ. 1st edition.— O'Reilly Media: 2002.— ISBN-10: 0596003102.

70. Osipov P., Borisov A. Advantages of Deferred Approach for Time-Critical Tasks // Informatica.— 2014.— Vol. 25, No. 3.— P. 467–484.— DOI: http://dx.doi.org/10.15388/Informatica.2014.24. Cited in: ACM, DBLP, EBSCO, SCOPUS, INSPEC, IAOR, Cambridge Scientific Abstracts, Mathematical Reviews, MathSciNet, Science Citation Index Expanded, Web of Science.

71. Osipov P. A., Borisov A. N. System for anomalous activity detection based on Markov models // Automatic Control and Computer Sciences.— 2011.— Vol. 45, No. 2.— P. 46–60. Cited in: **SpringerLink, SCOPUS, Academis OneFile, DBLP, Inspec**.

72. Osipov P. A., Mrochko A. E. and Borisov A. N. Identification of Differences of User Behavior Profiles and User Class Templates // Automatic Control and Computer Sciences.— 2014.— Vol. 48, No. 2.— P. 65–79. Cited in: **SpringerLink, SCOPUS, Academic OneFile, DBLP, Inspec**.

73. Osipov P., Rinkevics A., Kuleshova G., Borisov A. Markov chains in the task of author's writing style profile construction // *Scientific Journal of Riga Technical University, Information Technology and Management Science*.— 2014.— Vol. 17, RTU, Riga, P. 119–125. Cited in: **EBSCO, Google Scholar, Ulrich's International Periodicals Directory, VINITI.**

74. Osipov P., Borisov A. Simulation of Typical Behavior User using Markov Models // Proceedings of 2011 Baltic Congress on Future Internet and Communications (BCFIC 2011), 15–18 February 2011, Riga, Latvia.— Riga: Transport and Telecommunication Institute, 2011.— P. 229–236.

75. Osipov P. A. Borisovs A. Identification of Transaction Types using standard Clinical Document Architecture // Proceedings of XVI International Youth Forum „Radio Electronics and Youth in XXI century", 17–19 April 2012, Kharkov, Ukraine. Vol. 6.— Kharkov: Kharkov National University of Radio Electronics, 2012.— P. 150–151.

76. Osipov P. A., Borisov A. N. eHealth System Anomaly Activity Detection, Based on User Behavior Model // Modeling and Analysis of Safety and Risk in Complex Systems: Proceedings of International Scientific School MA SR — 2011 (Saint-Peterburg, Russia, 28 June – 02 July 2011).— SPb.: SUAI, SPb., 2011.— P. 405–412.

77. Osipovs P., Borisovs A. Approaches to the Construction of Behavioural Patterns of Information System Users // Scientific Journal of Riga Technical University. Information Technology and Management Science.— 2012.— Vol. 15.— P.176–182. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

78. Osipovs P., Borisovs A. Approaches to the Creation of Behavioural Patterns of Information System Users // Scientific Journal of Riga Technical University. Information Technology and Management Science.— 2012.— Vol. 15.— P.58–64. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

79. Osipovs P., Borisovs A. Non-Signature-Based Methods for Anomaly Detection // Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.— 2010.— Vol. 44.— P. 106–110. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

80. Osipovs P., Borisovs A. Usage of Ontologies in Systems of Data Exchange // Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.— 2009.— Vol. 40.— P. 108–116. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

81. Osipovs P., Borisovs A. Use of Deferred approach in Scientific Applications // Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.— 2011.— Vol. 49.— P.139–144. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

82. Osipovs, P., Borisovs, A. Practice of Web Data Mining Methods Application // Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.— 2009.— Vol. 40.— P. 101–107. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

83. Pearl, J. (1985). Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning (UCLA Technical Report CSD-850017). *Proceedings of the 7th Conference of the Cognitive Science Society,* University of California, Irvine, CA. pp. 329–334. Retrieved 2009–05–01.

84. Phyo A. H., Furnell S. M. A Detection-Oriented Classification of Insider IT Misuse // Proceedings of the 3rd Security Conference, Las Vegas, USA, 14–15 April 2004.

85. Pinheiro C., Carlos A. R. Social Network Analysis in Telecommunications.— John Wiley & Sons, 2011.— 304 p.— ISBN 978-1-118-01094-5.

86. Pokorny J. NoSQL databases: a step to database scalability in web environment // International Journal of Web Information Systems.— 2013.— Vol. 9, No. 1.— P. 69–82.— DOI 10.1108/ 17440081311316398.

87. Prechelt L. An empirical comparison of C, C++, Java, Perl, Python, Rexx, and Tcl for a search/string-processing program. Fakultat fur Informatik. Universitat Karlsruhe. Technical Report 2000–5. March 10, 2000.— Available on-line on http://www.inf.fu-berlin.de/inst/ag-se/teaching/V-EMPIR-2014/doc/jccpprtTR.pdf.— Last accessed 2014.05.27.

88. Razmerita L. Modeling Behavior of Users in Adaptive and Semantic-enhanced Information Systems: The role of a User Ontology // Proceedings of 5th International Conference of Adaptive Hypermedia and Adaptive Web-Based Systems and Authoring of Adaptive and Adaptable Hypermedia workshop, P. 69–77. — 29 of July – 1 August 2008, Hanover. Internet. http://www.academia.edu/3134137/Modeling_behavior_of_users_in_adaptive_and_semantic-enhanced_information_systems_The_role_of_a_user_ontology Last accessed 2014.09.08

89. Redis / Internet.— http://redis.io/topics/faq.— Last accessed 2014.04.01.

90. Reza F.M. An Introduction to Information Theory.— New York: Dover Publications, Inc., 1994.— ISBN 0-486-68210-2.

91. RFC7159; The JavaScript Object Notation (JSON) Data Interchange Format / Internet.—
http://tools.ietf.org/html/rfc7159. — Last accessed 2013.05.03.

92. Samek M. Practical UML Statecharts in C/C++: Event-Driven Programming for Embedded Systems.— CRC Press, Newnes 2008.— 728 p.— ISBN-10: 0750687061; ISBN-13: 978-0750687065.

93. Scarfone K., Mell P. Guide to Intrusion Detection and Prevention Systems (IDPS): Recommendations of the National Institute of Standards and Technology. NIST Special Publication 800-94.— Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, 2007.

94. Seung-Hyun K., Kyong H .K., Jong K., Sung-Je H., Sangwan K. Workflow-Based Authorization Service in the Grid // Journal of Grid Computing.— 2004.— No. 2.— P. 43–55.

95. Shelestov, S. Skakun, N. Kussul. Agent-based approach to implementing a model of user behavior Grid-systems // Proceedings of Space Research Institute NASU-NSAU «Informatyka, kibernetyka ta obchyslyuval′na tekhnika», Issue. 9 (132).— Donetsk: DonNTU, 2008.— P. 8–14.— ISSN: 1996-1588.

96. Sheskin D. Handbook of Parametric and Nonparametric Statistical Procedures.— CRC Press, 2004.— P. 54.— ISBN 1584884401.

97. Shingo T., Susumu D., Shinji S. A user-oriented secure file system on the Grid // The 3rd IEEE/ACM Int. Symp. on Cluster Computing and the Grid (CCGrid 2003), Conference Report / Oral Presentation. May, 2003.

98. Shirai K. Interest rate risk modeling using extended lognormal distribution with variable volatility // Stochastic Modeling.— International Actuarial Association May 2010.— ISBN: 978-0-9813968-2-8.

99. Shneiderman B. The Relationship Between COBOL and Computer Science // American Federation of Information Processing Societies 1985.— AFIPS 0164-l 239/85/040348-352$01 .OO/OO.

100. Sriparasa S. S. JavaScript and JSON Essentials.— O'Reilly Media, 2013.— ISBN 10:1-78328-604-0.

101. Tabini M. PHP as a General-Purpose Language // Linux Journal.— Aug 18, 2004. Internet — http://www.linuxjournal.com/article/6627 Last accessed 2014.10.10

102. Tornadoweb. project web-page / Internet.— www.tornadoweb.org.— Last accessed 2014.02.06.

103. Turban E., Aronson J. E., Liang T. P. Decision Support Systems and Intelligent Systems. 7th ed.— Prentice Hall, 2005.

104. Undercoffer J., Pinkston J., Joshi A., Finin T. A Target-Centric Ontology for Intrusion Detection // IJCAI-03 Workshop on Ontologies and Distributed Systems, Morgan Kaufmann Pu, P. 47–58. — Acapulco, 9 August 2003. Internet. — http://citeseerx.ist.psu.edu/viewdoc/ summary?doi=10.1.1.13.727 Last accessed 2014.10.01.

105. Wood R. C Programming for Scientists and Engineers.— Penton Press, 2002.— ISBN 1 8571 8030 5.

106. Zinky J., Shapiro R., Siracuse S., Wright T. Experience with Dynamic Crosscutting in Cougaar // Lecture Notes in Computer Science.— 2010.— Vol. 4803.— P. 595–612.— DOI: 10.1007/978-3-540-76848-7_41.

107. Осипов П. А., Борисов А. Н. Система обнаружения аномальных действий на основе моделей Маркова // Автоматика и вычислительная техника.— 2011.— № 2.— С. 46–60. Cited in: **SpringerLink, Ulrich's International Periodicals Directory, VINITI**.

108. Осипов П. А., Мрочко А. Е., Борисов А. Н. Идентификация отличий профиля поведения пользователя и шаблона класса пользователей // Автоматика и вычислительная техника. — 2014. — № 2. — С.5–24. Cited in: **SpringerLink, Ulrich's International Periodicals Directory, VINITI**.