METINĖ ATASKAITINĖ INFORMATIKOS KRYPTIES
DOKTORANTŲ KONFERENCIJA 2019 M. SPALIO 30 D.

ATASKAITA

DOKTORANTAS VIKTORAS BULAVAS
INFORMATIKA (N009)

VADOVAS: PROF. HABIL. DR. GINTAUTAS DZEMYDA

KONSULTANTAS: DR. VIRGINIJUS MARCINKEVIČIUS

DOKTORANTŪROS LAIKOTARPIS 2017 M. - 2021 M.

DMSTI-DS-N009-19-11

Tyrimo objektas, tikslai ir planuojami gauti rezultatai

- Preliminari disertacijos tema ir tyrimo objektas:
  - **Mašininio mokymo metodų taikymas ankstyvajam kibernetinių incidentų aptikimui**
- Tyrimo tikslai:
  - Gauti naujos informacijos apie tinkamus ankstyvojo anomalijų aptikimui mašininio mokymosi metodus
- Planuojami gauti rezultatai:
  - Panaudoti parinktus metodus, siekiant prognozuoti bei valdyti ankstyvąjį kibernetinių incidentų etapą

Antrųjų mokslo metų darbo planas

- Teorinis tyrimas (2018 m. lapkritis – 2019 m. gegužė)
- Empirinis tyrimas (2019 m. birželis – 2020 m. gegužė)
- Egzaminai:
  - Atpažinimo teorija – 9 kreditai, išlaikyta
  - Optimizavimo teorija, algoritmų sudėtingumas – 7 kreditai, išlaikyta
- Dalyvavimas mokslinėje konferencijoje
- Parengti vieną mokslinę tyrimų publikaciją konferencijos darbų medžiagoje.

Konferencijos (1)

- 2018 m. spalio 12 dieną konferencijoje „The 59th International Scientific Conference on Information Technology and Management Science of Riga Technical University“, Rygoje, pristatytas pranešimas „Investigation of network intrusion detection using data visualization methods“.

RĪGAS TEHNISKĀ UNIVERSITĀTE

IEEE Advancing Technology for Humanity

CERTIFICATE OF PARTICIPATION

issued to

Viktoras Bulavas,

Vilnius University
Lithuania

in recognition of his/her participation in

The 59th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS'2018)

held at Riga Technical University, Latvia on October 10-12, 2018

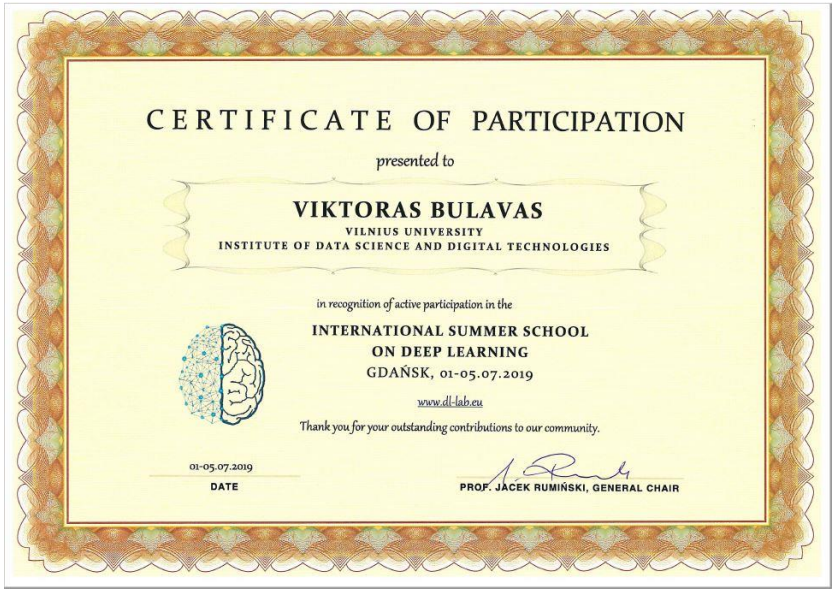
Presentation title: **Investigation of Network Intrusion Detection Using Data Visualization Methods**
by Viktoras Bulavas

Prof. Dr. Jānis Grabis
ITMS'2018 General Chair
Riga Technical University, Latvia

IEEE Conference record number: 45466

Konferencijos (2)

- 2019 m. liepos 1 – liepos 5 d. tarptautinė gilaus mašininio mokymosi vasaros stovykla ISSDL 2019, Gdansko technologijų universitete.



CERTIFICATE OF PARTICIPATION

presented to

VIKTORAS BULAVAS
VILNIUS UNIVERSITY
INSTITUTE OF DATA SCIENCE AND DIGITAL TECHNOLOGIES

in recognition of active participation in the

INTERNATIONAL SUMMER SCHOOL ON DEEP LEARNING
GDAŃSK, 01-05.07.2019

www.dl-lab.eu

Thank you for your outstanding contributions to our community.

01-05.07.2019
DATE

PROF. JACEK RUMIŃSKI, GENERAL CHAIR

Publikacijos

- Bulavas, Viktoras. Investigation of network intrusion detection using data visualization methods // 59th International scientific conference on information technology and management science of Riga Technical University (ITMS), October 10-12, 2018, Riga, Latvia : proceedings. Piscataway : NJ : IEEE, 2018, art. no. 8552977. eISBN 9781728100982. p. [1-6]. DOI: 10.1109/ITMS.2018.8552977.
- Paskelbta IEEE Xplore Digital Library 2018 gruodžio 18 d., prieinama per Scopus, IEEE Xplore Digital Library ir Google Scholar

Trečiųjų mokslo metų darbo planas

3. Empirinis tyrimas (2019 m. birželis – 2020 m. gegužė):
3.1. Skirtingų algoritmų palyginimas.
3.2. Įgyvendintų algoritmų modifikacijos, ar naujų algoritmų kūrimas, sprendžiant ankstyvo kibernetinių incidentų įspėjimo uždavinį.
3.3. Sukurtų modifikacijų eksperimentinis tyrimas analizuojant jų efektyvumą.

Trečiųjų mokslo metų darbo planas

4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas (2020 m. birželis – 2020 m. rugsėjis):
4.1. Teorinio tyrimo apibendrinimas.
4.2. Empirinio tyrimo apibendrinimas.
4.3. Rezultatų apibendrinimas, esminių rezultatų išskyrimas bei išvadų parengimas.

- Planuojama parengti vieną mokslinę tyrimų publikaciją (recenzuojamame leidinyje, WoS su Impact Factor).

ČIA TRUMPAI PRISTATOMI EKSPERIMENTAI, KURIUOS ATLIKAU IKI PRISTATYMO PAVASARĮ - DMSTI 2019-01-28, „MAŠININIO MOKYMO METODŲ TAIKYMAS ANKSTYVAJAM KIBERNETINIŲ INCIDENTŲ APTIKIMUI“

REZULTATAI JAU PRISTATYTI PUBLIKACIJOJE.

Tyrime naudoti duomenys

- Kanados kibernetinio saugumo tyrimų instituto NSL-KDD duomenų rinkinys (https://www.unb.ca/cic/datasets/nsl.html)
- **Reference:** M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” *Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 2009.

Principal Components Analysis

- Attention in this analysis is drawn to linear projection, in particular Principal Component Analysis (PCA), as introduced by Hotelling, helping to select the most informative dimensions for intrusion detection.
- Brauckhoff, Salamatian and May proposed implementing PCA method for anomaly detection and raised question of right number of Principal Components for analysis.
- A value of 10 Principal Components parameter was chosen, as recommended by Keerthi and Surendiran.
- PCA, together with Decision Tree, can be successfully used for traffic feature extraction and intrusion classification.

Principal Component Analysis of NSL-KDD

Fig. 5. Principal Component PC1-PC2

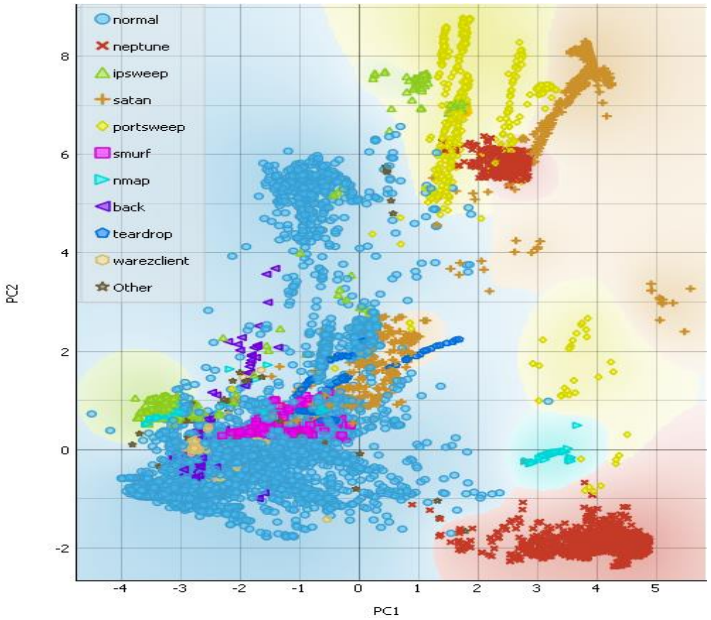
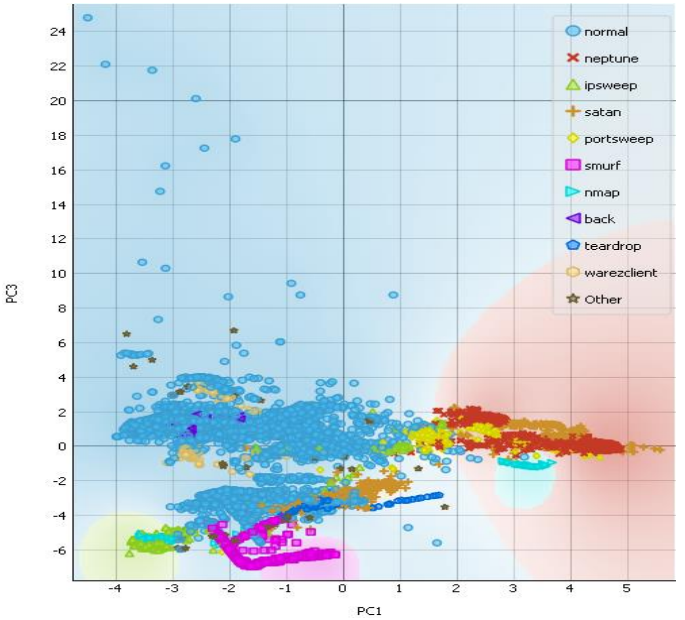


Fig.6. Principal Component PC1-PC3



Decision Tree for NSL-KDD

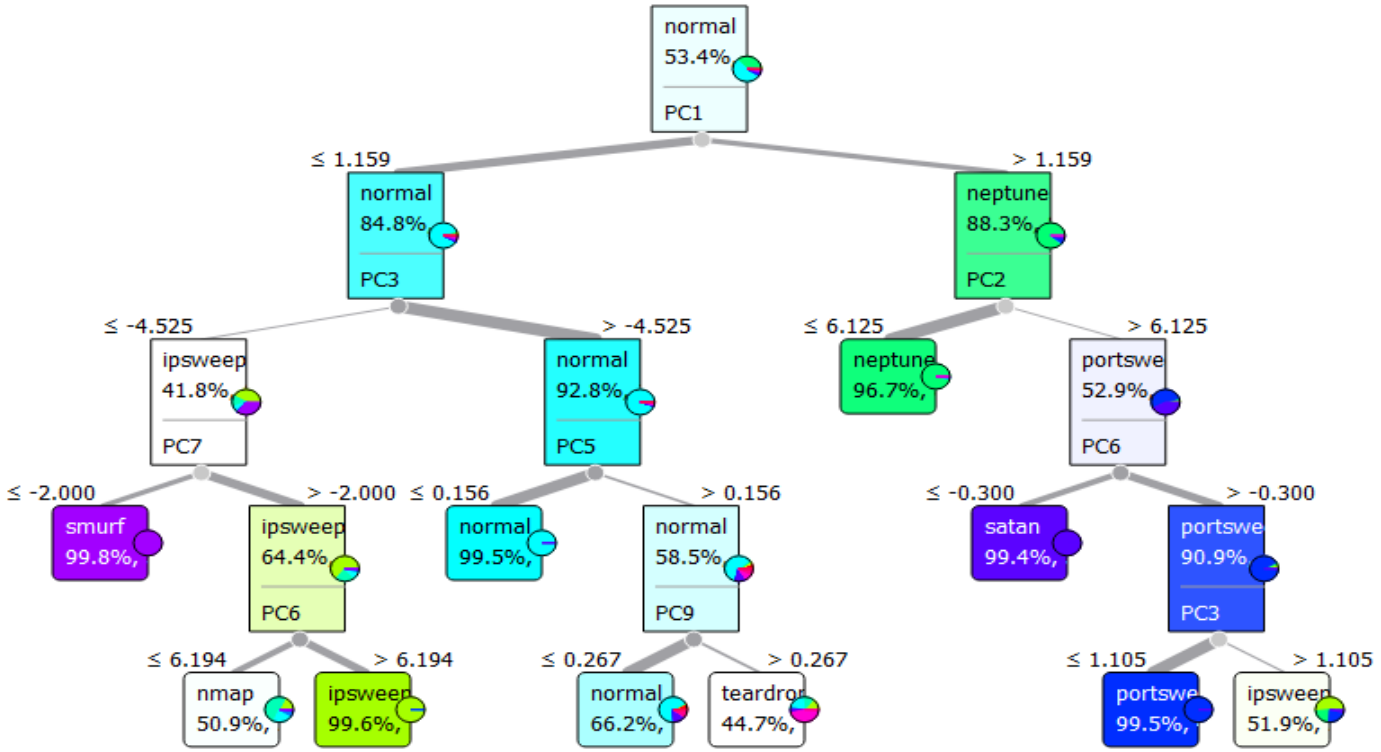


Fig.8. Decision Tree for NSL-KDD data using Orange 3 software.

Findings

- Investigation in this research demonstrates, that combination of PCA and Decision Tree methods allows classification of intrusions such as:
  - smurf,
  - satan,
  - neptune,
  - portsweep,
  - ipsweep
- with probabilities higher than 95% with depth of tree set to 4 and PCA components set to 10.
- Nevertheless, nmap and teardrop intrusions are classified purely, therefore deeper Decision Tree is needed to increase classification accuracy.

ČIA TRUMPAI PRISTATOMI EKSPERIMENTAI, KURIUOS ATLIKAU PO PRISTATYMO PAVASARĮ - DMSTI 2019-01-28, „MAŠININIO MOKYMO METODŲ TAIKYMAS ANKSTYVAJAM KIBERNETINIŲ INCIDENTŲ APTIKIMUI“

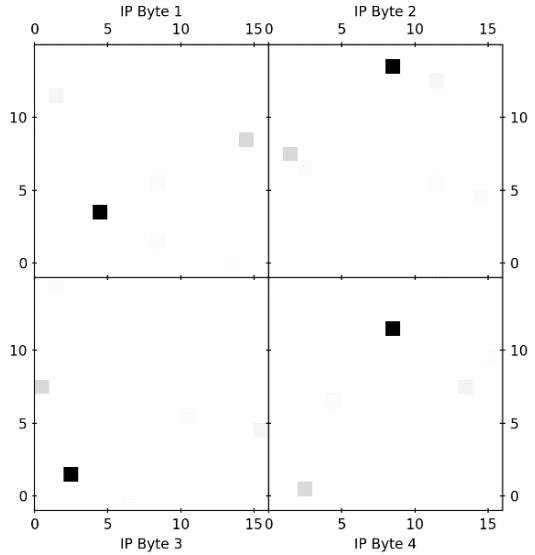
Tyrime naudoti duomenys

- Naudotas atviros prieigos kibernetinės saugos tinklo duomenų šaltinis CIC IDS 2018, kuriame sukaupti 80 tinklo parametrų.
  - I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” in Proceedings of the 4th International Conference on Information Systems Security and Privacy, 2018, no. January, pp. 108–116.
- Vilniaus universitete surinkti vienos darbo vietos duomenys.

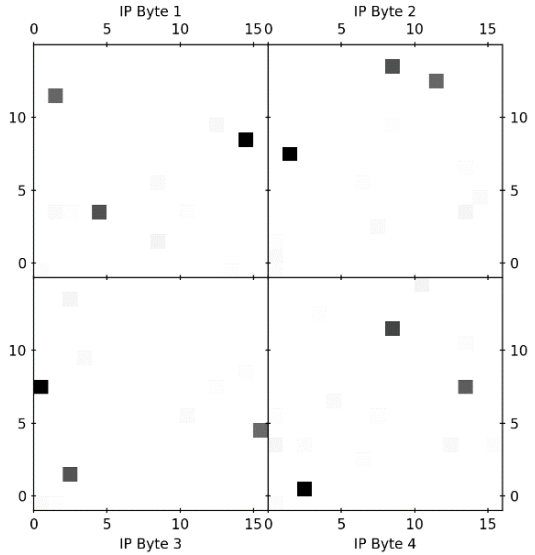
Tyrime atlikti darbai

- Šiame tyrime buvo įvairias metodais nagrinėjami bei grafiškai atvaizduoti rinkinyje CSE-CIC-IDS2018 surinktų DOS, DDOS, Brute Force –Web, Brute Force –XSS ir SQL Inject atakų ir ją sekusių įsilaužimų duomenys.
- Daugiamačiai kibernetinio saugumo duomenys redukuoti Kim, Reddy and Vannucci pasiūlytais metodais, sukuriant dvimates vizualizacijas, kurių kaita laike teikia kibernetinio įsilaužimo indikacijas.

Srauto vizualizacija Kim, Reddy ir Vannucci pasiūlytais metodais

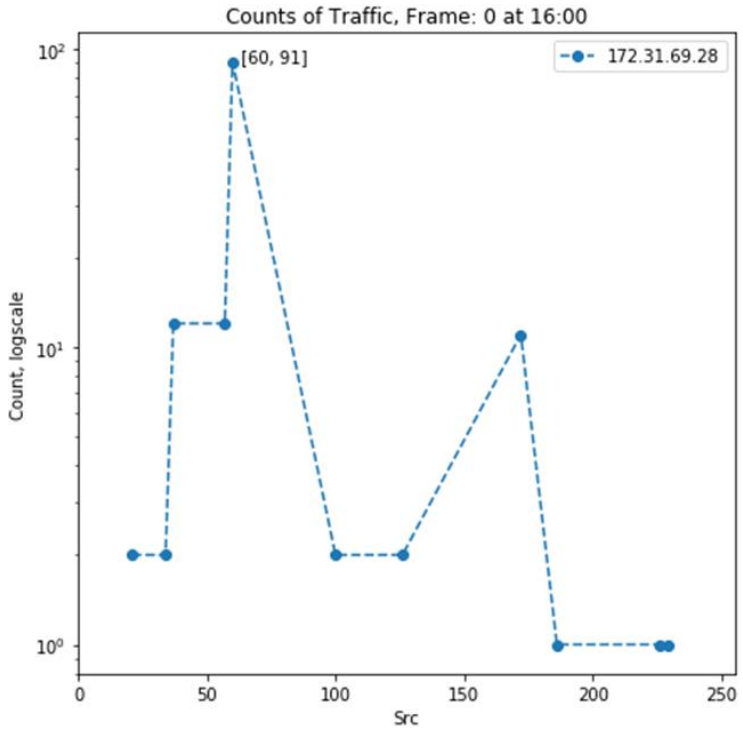


Pav. 1 a. Prieš DOS ataką

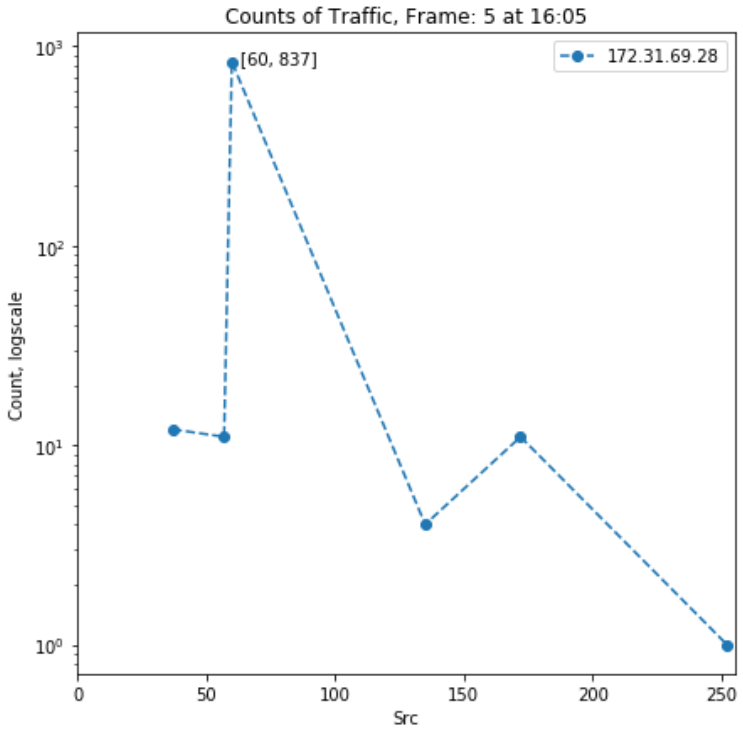


Pav. 1 b. Per DOS ataką

Įsilaužimo dinamika paprastų dažnių kalba

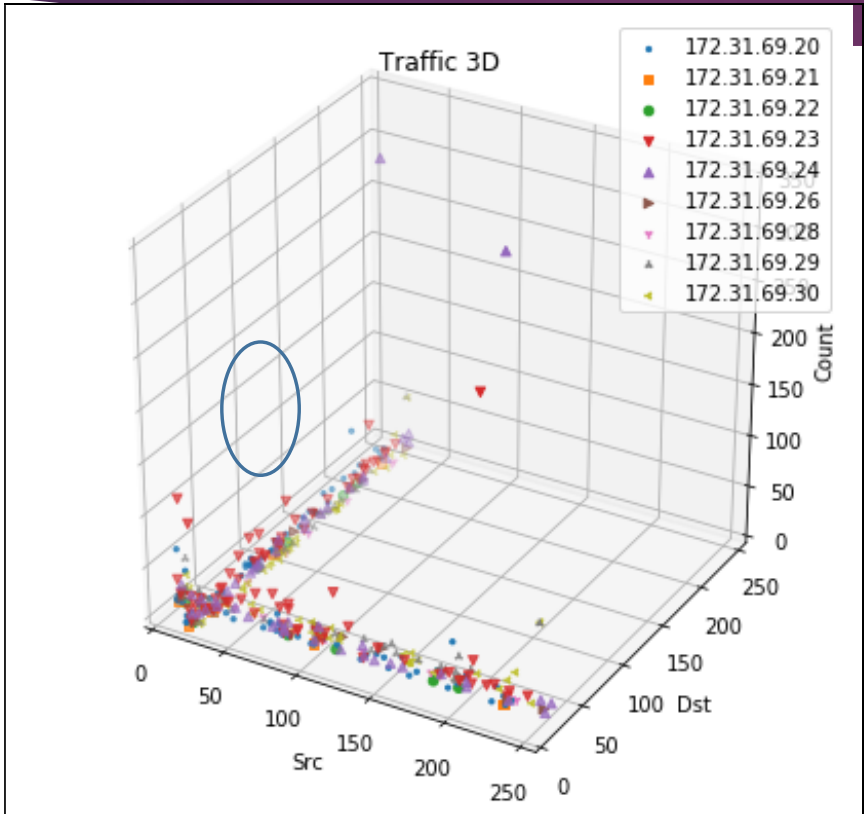


Pav. 9 a. Prieš DOS ataką (maksimumas-91 sesijos per 5 minutes iš puolančio adreso neženkliai išsiskiria iš kito srauto.)

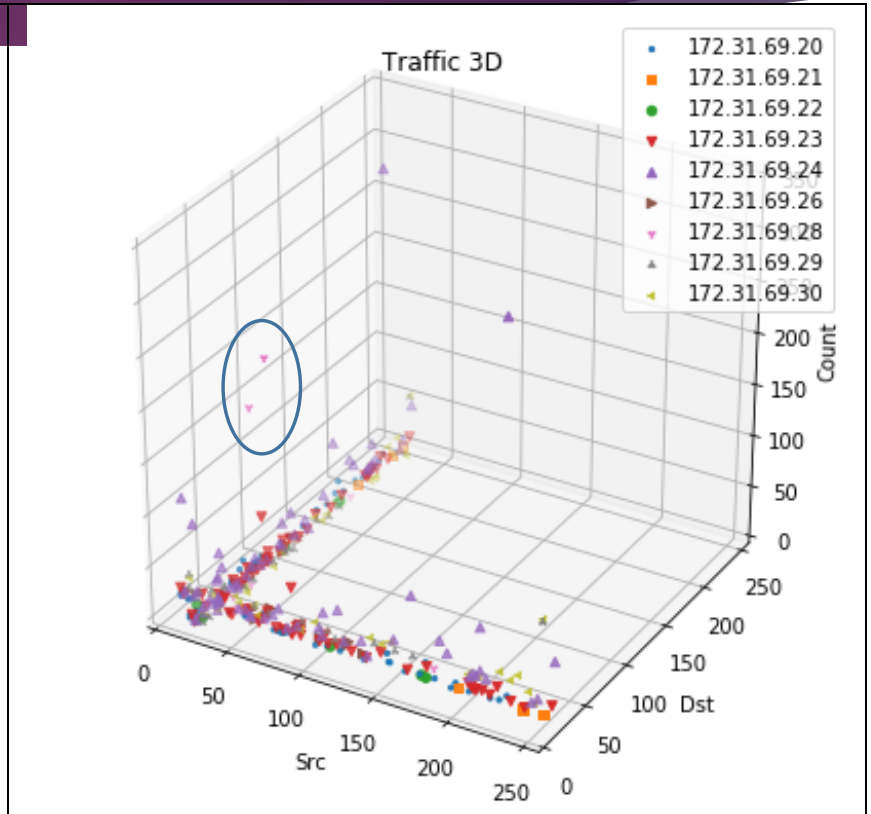


Pav. 9 b. Per DOS ataką (maksimumas – 837 sesijos per 5 minutes ženkliai išsiskiria iš kito srauto)

Anomalijų vizualizacija

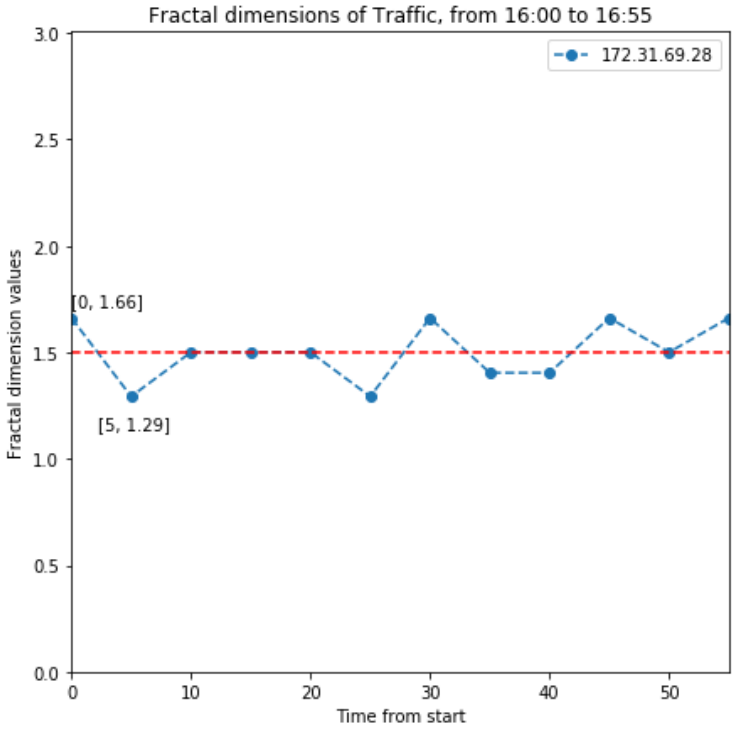


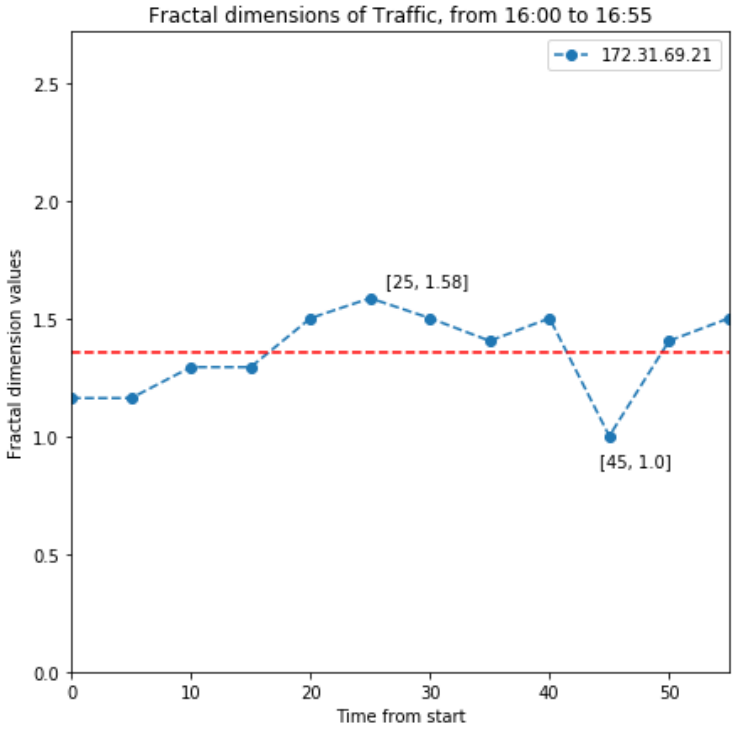
pav. 1a Srautas 16:02



pav. 1 Srautas 16:04

Higuchi Box-Count metodu susikaičiuota fraktalinė dimensija





Rezultatai ir išvados

- Tyrimo metu nagrinėtas kibernetinio įsilaužimo nustatymo statistinio ir mašinų mokymosi algoritmų sudėtingumas.
- Įvertinus statistinių algoritmų sudėtingumą, galima apibendrinti, kad nagrinėti metodai yra taikytini realaus laiko uždavinyje nustatant DOS ir DDOS atakas.
- Įvertinta gautų taškinių vaizdų Hausdorfo fraktalinė dimensija (sudėtingumas) ir aproksimuojančio (Higuchi, 1988) Box-Counting algoritmo sudėtingumas $O(n^2)$.
- Įvertinant dvimačio vaizdo retumą, galima teigti, kad griežtesni asimptotiniai Hausdorfo dimensijos skaičiavimo aproksimuojančiu Higuchi Box-Counting algoritmu vertinimai galimai būtų $O(n^k)$, kur $k \geq \log_2 3$ ir $n \leq 2$, kadangi n nėra tokie dideli, kad būtų išlošiama skaičiuojant efektyvesniais prie didelių n, nei Karatsubos daugybos skaičiavimo algoritmais. Tačiau n log n auga lėčiau nei n^k, kai n>3 ir k> 1, todėl paprasta statistika veikia greičiau.

Rezultatai ir išvados

- Shapo verčių metodu (Lundberg and Lee, 2017) buvo atliekamas pasirinktais mašinų mokymo metodais (SVM, Random Forest, KNN) paskaičiuotų duomenų požymių svarbos nustatymas, parodęs, kad nėra vienareikšmiai dominuojančių dimensijų, o ženklus dimensijų sumažinimas blogina prognozių kokybę, kas buvo patvirtinta ir ankstesniais autoriaus pirminių komponenčių dimensijų redukavimo tyrimais (Bulavas, 2018).

AČIŪ UŽ DĖMESĮ!

Viktoras Bulavas

E-mail: viktoras.bulavas@itpc.vu.lt