METINĖ ATASKAITINĖ INFORMATIKOS KRYPTIES
DOKTORANTŲ KONFERENCIJA 2020 M. SPALIO 22 D.

ATASKAITA

DOKTORANTAS VIKTORAS BULAVAS
INFORMATIKA (N009)

VADOVAS: PROF. HABIL. DR. GINTAUTAS DZEMYDA

KONSULTANTAS: DR. VIRGINIJUS MARCINKEVIČIUS

DOKTORANTŪROS LAIKOTARPIS 2017 M. - 2021 M.

Disertacijos tyrimo objektas, tikslai ir planuojami gauti rezultatai

- Preliminari disertacijos tema ir tyrimo objektas:
  - **Mašininio mokymo metodų taikymas ankstyvajam kibernetinių incidentų aptikimui**
- Tyrimo tikslai:
  - Sukurti arba patobulinti mašininiu mokymusi grįstą metodą, skirtą ankstyvajam kibernetinių incidentų aptikimui
- Planuojami gauti rezultatai:
  - Panaudoti parinktus metodus, siekiant prognozuoti bei valdyti ankstyvąjį kibernetinių incidentų etapą

Konferencijos

- 2020 m. lapkričio 28 – 30 d. 11-oje „Duomenų analizės metodai programų sistemoms“ konferencijoje, Druskininkai, stendinis pranešimas „Fractal Dimension of Network Traffic as a Feature for Intrusion Detection“.

DAMSS

CERTIFICATE OF PARTICIPATION

This is certify that

Viktoras Bulavas

Vilnius University

has attended the 11th international workshop

Data Analysis Methods for Software Systems

held in Druskininkai, Lithuania, 28 - 30 November 2019
and has presented a paper

Fractal Dimensionality of Network Traffic as a Feature for Intrusion Detection

Program Committee

Dr. Saulius Maskeliūnas

Prof. Gintautas Dzemyda

Druskininkai, 29 November 2019

Bendrųjų gebėjimų mokymai

- Dalyvauta VU doktorantų bendrųjų gebėjimų mokymuose „Academic writing“, 0,94 ECTS kredito, kurie vyko 2019 m. gruodžio mėn. 9-12 dienomis Fizinių ir technologijos mokslų centre.

VILNIAUS UNIVERSITETO

PAŽYMĖJIMAS

Nr. MVG-MID-2019-407

Viktoras Bulavas

2019/2020 akademiniais metais dalyvavo šiuose VU doktorantų bendrųjų gebėjimų mokymuose:

Academic writing (0,94 ECTS). Mokymai vyko 2019 m. gruodžio mėn. 9 -12 dienomis Fizinių ir technologijos mokslų centre.

Mokslo ir inovacijų departamento direktorė Vida Lapinskaitė

2019 m. gruodžio 19 d.

Pristatymai DMSTI

- Metinė ataskaitinė informatikos krypties doktorantų konferencija 2019 m. spalio 30 d.
- DMSTI 2020-10-22, „Mašininio mokymo algoritmų efektyvumo palyginimas sprendžiant ankstyvojo kibernetinių incidentų aptikimo uždavinį“

Publikacijos

- Bulavas, Viktoras. Fractal dimensionality of network traffic as a feature for intrusion detection // 11th international workshop on data analysis methods for software systems (DAMSS 2019), Druskininkai, Lithuania, November 28-30, 2019 / Lithuanian Computer Society, Vilnius University Institute of Data Science and Digital Technologies, Lithuanian Academy of Sciences. Vilnius : Vilnius University Press, 2019. ISBN 9786090703243. eISBN 9786090703250. p. 16. DOI: 10.15388/Proceedings.2019.8.

Trečiųjų mokslo metų darbo planas

3. Empirinis tyrimas (2019 m. birželis – 2020 m. gegužė):
3.1. Skirtingų algoritmų palyginimas.
3.2. Įgyvendintų algoritmų modifikacijos, 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).

Ketvirtųjų mokslo metų darbo planas

5. Atskirų daktaro disertacijos dalių (analizės rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas (2020 m. spalis– 2021 m. gegužė):
5.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas.
5.2. Analitinės disertacijos dalies parengimas.
5.3. Teorinės disertacijos dalies parengimas.
5.4. Eksperimentinės disertacijos dalies parengimas.
5.5. Bendrųjų išvadų formulavimas.
6. Daktaro disertacijos parengimas ir svarstymas padalinyje (2021 m. birželis).
7. Daktaro disertacijos pateikimas gynimui (2021 m. rugsėjis).

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

Vilniaus universitetas

MAŠININIO MOKYMO METODŲ EFEKTYVUMO TYRIMAS SPRENDŽIANT ANKSTYVOJO KIBERNETINIŲ INCIDENTŲ APTIKIMO UŽDAVINĮ

Vykdyti tyrimai

1. Papildomos sintetinės dedamosios, Fraktalinės dimensijos įvedimas ir panaudojimas nustatant DOS ir DDOS atakas.
2. Duomenų šaltinio dedamųjų svarbos nustatymas.
3. Mašininio mokymo algoritmų diegimas ir parametrų, reikalingų efektyvumui padidinti, lyginant su kitais autoriais, paieška.

1. Fraktalinės dimensijos įvedimas

- Daugiamačiai kibernetinio saugumo duomenys redukuoti Kim, Reddy and Vannucci pasiūlytais metodais, sukuriant dvimates vizualizacijas, kurių kaita laike teikia kibernetinio įsilaužimo indikacijas.
- Realizuotos DOS, DDOS, Brute Force –Web, Brute Force –XSS ir SQL Inject atakų fraktalinės dimensijos.

Rezultatai ir išvados

- Įvertinus statistinių algoritmų sudėtingumą, galima apibendrinti, kad fraktalinė dimensija gali būti įvesta ir taikoma realaus laiko uždavinyje nustatant DOS ir DDOS atakas. Tačiau klasikiniai statistiniai dažniai duoda greitesnį rezultatą.
- Įvertinta gautų taškinių vaizdų Hausdorfo fraktalinė dimensija (sudėtingumas) ir aproksimuojančio (Higuchi, 1988) Box-Counting algoritmo sudėtingumas $O(n^2)$.
- Tačiau paprasta min/max statistika veikia greičiau.

2. Duomenų šaltinio dedamųjų tyrimas

- Išbandyti dar 3 skirtingi būdai parenkant požymius:
  - Filtravimas (koreliacijų analizė).
  - Reikšmingiausių savybių atrinkimas (KBest).
  - Rekursinis reikšmingiausių savybių atrinkimas (RFE).

Rezultatai

- Geriausias tikslumas atrenkant savybes (paliekant tą patį savybių skaičių skirtingais metodais) pasiektas panaudojat Kbest.
- Duomenų šaltinyje nustatyta tiek nereikšmingų tiek stipriai koreliuotų savybių.
- Su 20 reikšmingų savybių pasiekiamas duomenų šaltinio autorių publikacijoje nurodytas tikslumas.

3. Mašininio mokymo algoritmų taikymas

- Python aplinkoje su CIC-IDS-2017 duomenimis realizuoti 7 lyginamojo duomenų šaltinio autorių publikuoto tyrimo mašinų mokymo algoritmai (KNN, RFC, CART, Adaboost, Gaussian Naive Bayes, QDA, MLP), ir palygintas gautas efektyvumas.
- Papildomai tam pačiam uždaviniui Python Keras –Tensorflow 2.1 aplinkoje realizuoti ANN ir CNN neuroniniai tinklai.
- Siekiant pagerinti rezultatus, panaudojant tinklelio paieškos (‚Grid Search‘) metodą atrinkti skirtingų algoritmų mokymo hiperparametrai.
- Rezultatai patikrinti panaudojus kryžminio patikrinimo ('cross-validation’) ir validavimo rinkinio atidėjimo metodus.
- Suskaičiuoti algoritmų tikslumo (accuracy), klasių prognozės tikslumo (precision), jautrumo (recall), harmoninio tikslumo F1, Hamming Loss ir Jaccart Score, subalansuoto tikslumo rodikliai.

Rezultatai (1 iš 2)

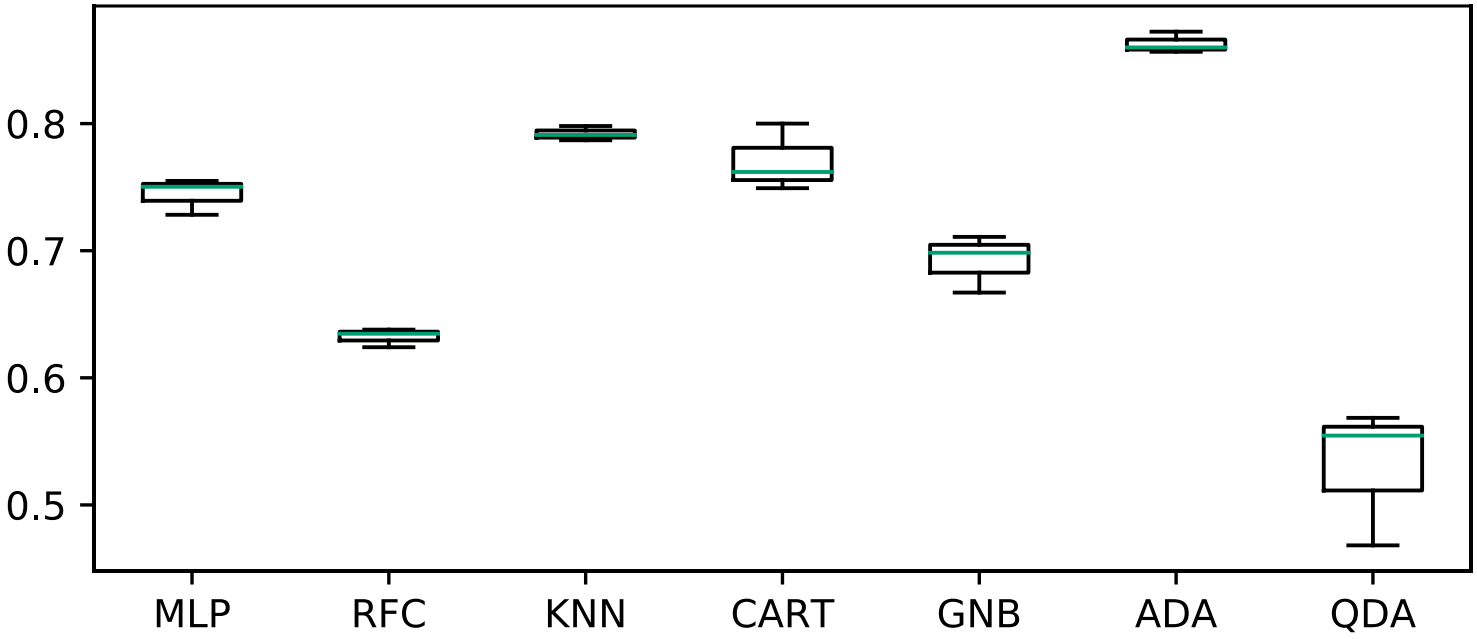
Algorithm	Precision		Recall		F1		Time (s)	
	PUB	EKSP	PUB	EKSP	PUB	EKSP	PUB	EKSP
KNN	0,96	0,989	0,96	0,989	0,96	0,985	1908	897
RFC	0,98	0,991	0,97	0,993	0,97	0,992	74	36
ID3 vs. CART	0,98	0,997	0,98	0,998	0,98	0,998	235	7
Adaboost	0,77	0,999	0,84	0,999	0,77	0,999	1126	278
Naive-Bayes	0,88	0,82	0,04	0,87	0,04	0,84	15	4
ANN	-	0,986	-	0,986	-	0,796	-	294
CNN	-	0,994	-	0,994	-	0,865	-	320

Here PUB: as published by dataset authors in Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018

Rezultatai (2 iš 2)

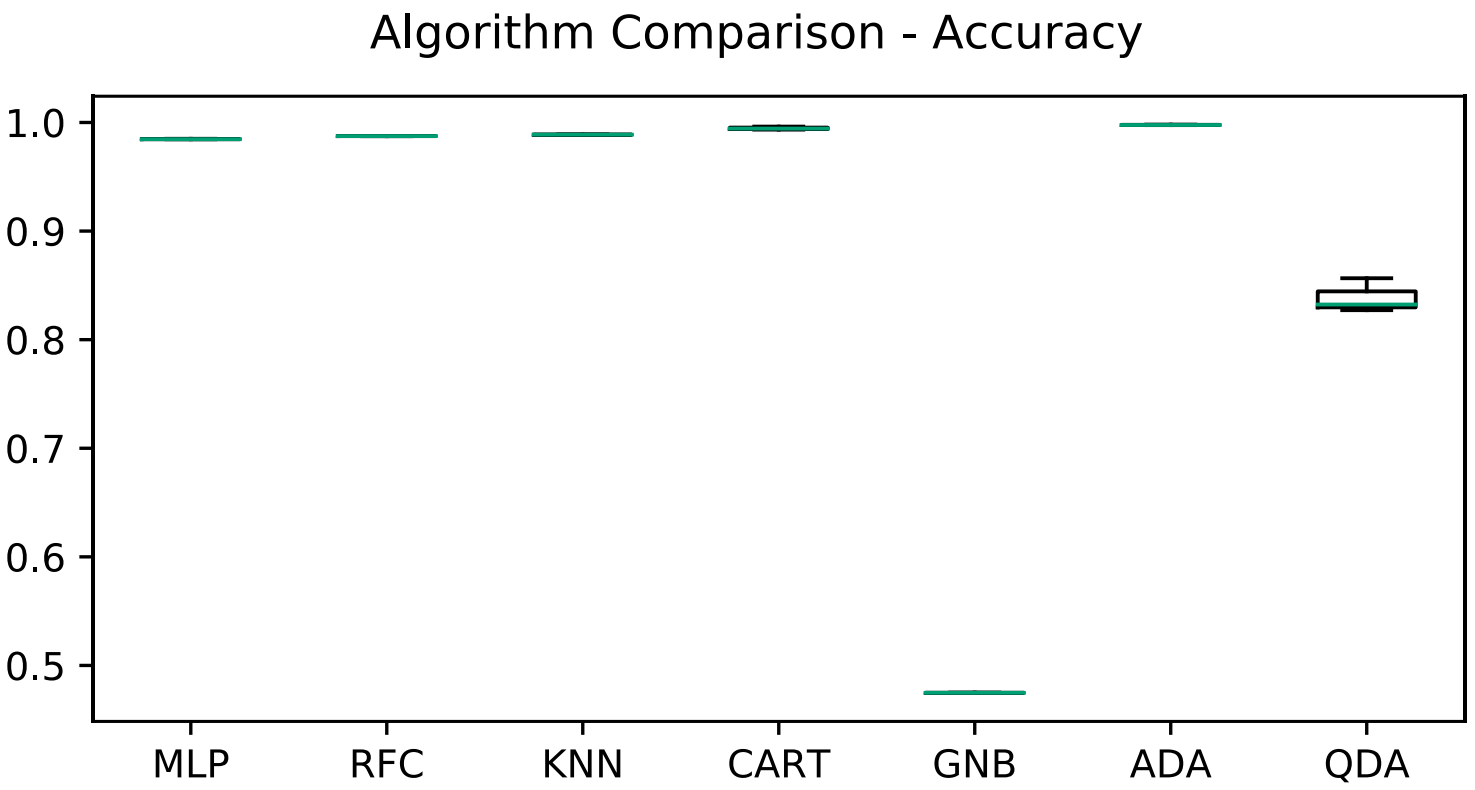
Algorithm	Accuracy	BAS	Hamming	Jackart
KNN	0,988	0,792	0,002	0,978
RFC	0,987	0,632	0,011	0,975
CART	0,997	0,770	0,003	0,989
Adaboost	0,998	0,862	0,001	0,995
Naive-Bayes	0,789	0,692	0,210	0,734
ANN	0,986		0,013	0,987
CNN	0,997		0,002	0,992
MLP	0,984	0,744	0,015	0,970
QDA	0,839	0,530	0,161	0,726

Subalansuotas tikslumo indeksas



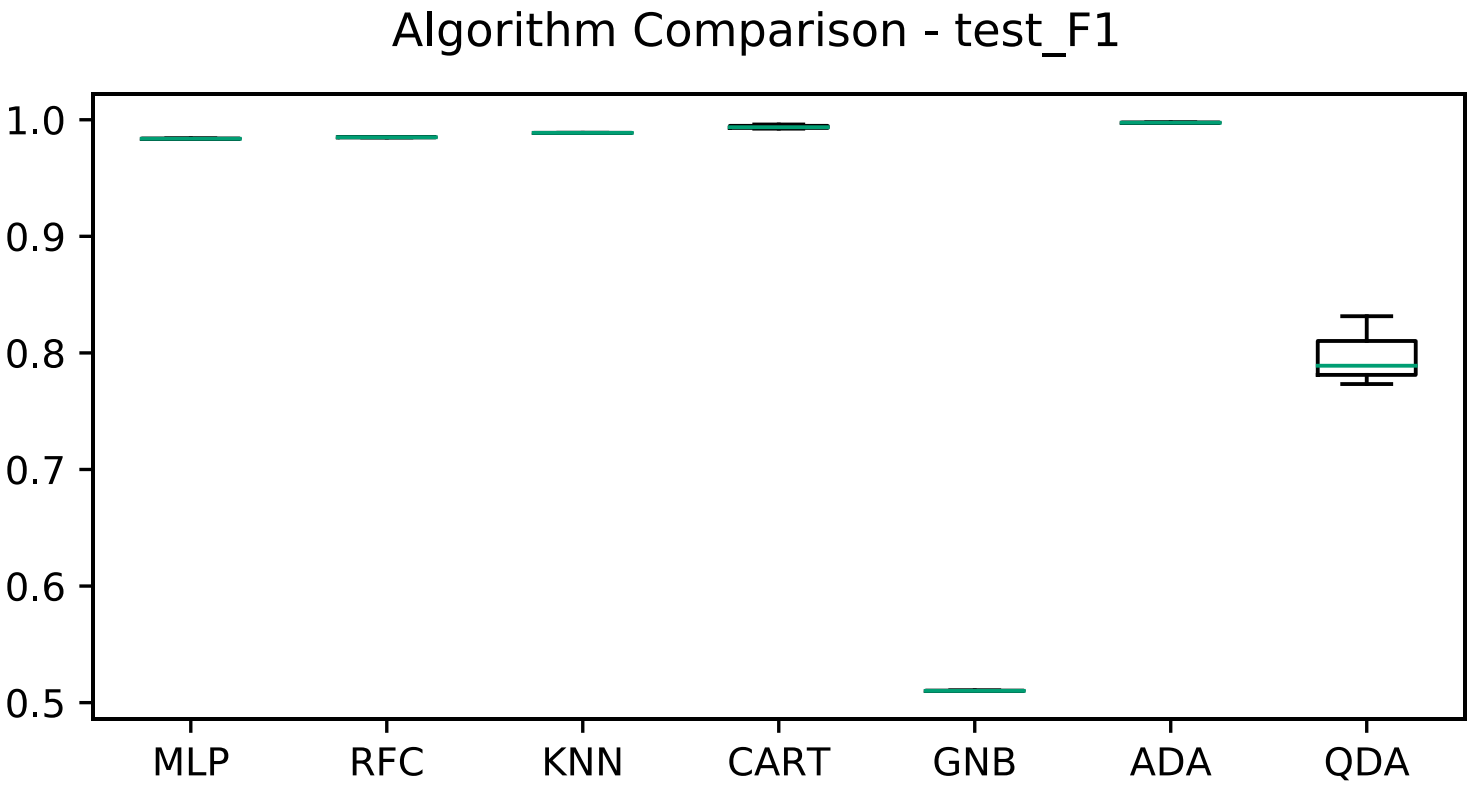
Pastaba: kuo daugiau - tuo geriau

Tikslumas



Kuo daugiau tuo geriau

F1 harmoninis rodiklis



Kuo daugiau tuo geriau

Adaboost rezultatai (1 iš 2)

Class	precision	recall	f1-score	support
BENIGN	0,998	0,999	0,999	85175
Bot	0,916	0,839	0,876	391
DDoS	1	1	1	25603
DoS GoldenEye	0,998	0,993	0,996	2057
DoS Hulk	0,999	0,998	0,999	34569
DoS Slowhttptest	0,995	0,991	0,993	1046
DoS slowloris	0,996	0,992	0,994	1077
FTP-Patator	1	0,998	0,999	1187
Heartbleed	1	1	1	6

Adaboost rezultatai (2 iš 2)

Class	precision	recall	f1-score	support
Infiltration	1	0,857	0,923	7
PortScan	0,998	1	0,999	18164
SSH-Patator	1	0,997	0,998	644
Web Attack-Brute Force	0,708	0,973	0,819	294
Web Attack-Sql Injection	1	0,667	0,8	6
Web Attack-XSS	0,556	0,038	0,072	130
Accuracy			0,998	170356
Macro avg	0,944	0,889	0,898	170356
Weighted avg	0,998	0,998	0,997	170356

KNN rezultatai (1 iš 2)

Class	precision	recall	f1-score	support
BENIGN	0,996	0,987	0,991	85175
Bot	0,853	0,834	0,843	391
DDoS	0,999	0,998	0,998	25603
DoS GoldenEye	0,995	0,994	0,994	2057
DoS Hulk	0,997	0,999	0,998	34569
DoS Slowhttptest	0,988	0,988	0,988	1046
DoS slowloris	0,992	0,989	0,99	1077
FTP-Patator	0,999	0,997	0,998	1187
Heartbleed	1	1	1	6

KNN rezultatai (2 iš 2)

Class	precision	recall	f1-score	support
Infiltration	0,5	0,143	0,222	7
PortScan	0,95	0,987	0,968	18164
SSH-Patator	0,981	0,986	0,984	644
Web Attack-Brute Force	0,713	0,82	0,763	294
Web Attack-Sql Injection	1	0,5	0,667	6
Web Attack-XSS	0,398	0,254	0,31	130
Accuracy			0,990	170356
Macro avg	0,891	0,832	0,848	170356
Weighted avg	0,990	0,990	0,990	170356

Rezultatai ir išvados

- Panaudojus hiper-parametrus, atrinktus GridSearch, panaudojant klasikinius mašininio mokymo metodus pavyko pasiekti geresnių rezultatų, nei duomenų šaltinio autorių publikacijoje.
- Papildomai sumodeliuotų ANN ir CNN rezultatai geresni, nei duomenų šaltinio autorių gautieji klasikiniais algoritmais, bet kol kas blogesni, nei šiame tyrime atliktų skaičiavimų.
- Vertinant subalansuoto tikslumo rodiklį galima teigti, kad praktiniams įsilaužimo nustatymo uždaviniams tiksliausiais Adaboost ansamblis.

Ketvirtųjų mokslo metų darbo planas

5. Atskirų daktaro disertacijos dalių (analizės rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas (2020 m. spalis– 2021 m. gegužė):
5.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas.
5.2. Analitinės disertacijos dalies parengimas.
5.3. Teorinės disertacijos dalies parengimas.
5.4. Eksperimentinės disertacijos dalies parengimas.
5.5. Bendrųjų išvadų formulavimas.
6. Daktaro disertacijos parengimas ir svarstymas padalinyje (2021 m. birželis).
7. Daktaro disertacijos pateikimas gynimui (2021 m. rugsėjis).

- Planuojama parengti dvi mokslines tyrimų publikacijas (recenzuojamuose leidiniuose, WoS su Impact Factor).
- Erasmus+ (savanoriška) praktika Gdansko technologijų universiteto Elektronikos, telekomunikacijų ir informatikos fakultete, 2019 rugsėjo 25 – 2020 sausio 31 d.

AČIŪ UŽ DĖMESĮ!

Viktoras Bulavas

E-mail: viktoras.Bulavas [eta] stud.mii.vu.lt

- Laikas klausimams!