Vilniaus universitetas

LIETUVOS TEISMO EKSPERTIZĖS CENTRAS

Ataskaitinė informatikos krypties doktorantų konferencija 2022-09-30

Andrius Chaževskas (VU DMSTI doktorantas, Išmaniųjų technologijų tyrimų grupė)

Vilniaus universitetas

Darbo tema.

Teksto semantinės analizės ir mašininio mokymosi algoritmų taikymo slaptažodžių parinkimui tyrimas.

Application of text semantic analysis and machine learning algorithms for passwords guessing.

Darbo vadovas.

Prof. dr. Igoris Belovas.

Doktorantūros studijų laikotarpis.

2020 m. spalio mėn. 1 d. – 2024 m. rugsėjo mėn. 30 d..

Ataskaitinis laikotarpis.

2022 m. kovo mėn. 25 d. – 2022 m. rugsėjo mėn. 30 d..

Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2020/2021) Pirmas pusmetis	1	1		1 (L[1])			
I (2020/2021) Antras pusmetis	1	1	1 (L)	1 (T[2])		1 (KD/R[3])	Publikuota
II (2021/2022) Pirmas pusmetis	1	1		1 (L)	1 (KD[4])		
II (2021/2022) Antras pusmetis	1	1	1 (L)	2 (T)	1 (KD / R)	1 (KD)	Publikuota
III (2022/2023) Pirmas pusmetis							
III (2022/2023) Antras pusmetis			1 (T)		1 (CA WoS[5])		
IV (2023/2024) Pirmas pusmetis							
IV (2023/2024) Antras pusmetis			1 (T)		1 (CA WoS)		

[1] Tarpinių rezultatų pristatymas konferencijoje Lietuvoje.

[2] Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.

[3] Tarpinių rezultatų publikavimas (recenzuojamoje konferencijos darbų medžiagoje).

[4] Mokslinių tyrimų disertacijos tema apžvalga (konferencijos darbų medžiagoje).

[5] Rezultatų publikavimas (recenzuojamame periodiniame leidinyje CA WoS su Impact Factor).

Vilniaus universitetas

Ataskaitinių metų darbo planas ir jo vykdymo suvestinė

Egzaminai	**Dalyvavimas konferencijose**	**Publikacijos**
Planas	**Įvykdyta**	**Būklė**
„Natūralios kalbos apdorojimas“.	„Natūralios kalbos apdorojimas“, 2022-09-26.	Išlaikytas.
Dalyvavimas konferencijose		
Planas	**Įvykdyta**	**Konferencijos tipas**
Tarpinių rezultatų pristatymas tarptautinėje konferencijoje.	Dalyvauta ir skaitytas pranešimas 2022 metų liepos 3-6 dienomis vykusioje tarptautinėje konferencijoje Euro 2022, Aalto universitete, Helsinkis, Suomija.	Tarpatutinė konferencija.
	Dalyvauta ir skaitytas pranešimas 2022 metų rugsėjo 8-10 dienomis vykusiame XVIII tarptautiniame kongrese „Kriminalistika ir teismo ekspertologija: mokslas, studijos, praktika“, Mykolo Riomerio universitete, Vilnius, Lietuva.	Tarptautinis kongresas.

Publikacijos			
Planas	**Įvykdyta**	**Būklė**	**Publikacijos tipas**
Mokslinių tyrimų disertacijos tema apžvalga (konferencijos darbų medžiagoje).	Chaževskas, Andrius; Belovas, Igoris; Marcinkevičius, Virginijus. Markov, Probabilistic and Rule-Based Password Guessing Methods: Survey And Comparison // XVIII tarptautinio kongreso „Kriminalistika ir teismo ekspertologija: mokslas, studijos, praktika“ vykusio Mykolo Riomerio universitete, Vilnius, Lietuva recenzuojamas periodinis leidinys.	Publikuota	Recenzuojamoje konferencijos darbų medžiagoje (be cituojamumo rodiklio).

Vilniaus universitetas

Kvalifikacijos kėlimas

- Darbas su VU Kauno fakulteto III k. studentais, dėstomas kursas “Skaitmeninio turinio teisminė analizė” (2022 pavasaris).
- Bendrieji gebėjimai:

  Mokymai doktorantams apie intelektinės nuosavybės apsaugą,

  EEML vasaros mokykla.

Vilnius University
Faculty of Mathematics and Informatics

CERTIFICATE OF ATTENDENCE

No. 110000-01-15

This certificate is presented to

Andrius Chazevskas
VILNIUS UNIVERSITY

For attending the **Eastern European Machine Learning Summer School** on 6-14 of July, 2022, at Vilnius, Lithuania

This activity was awarded 4 ECTS.

Dean of the Faculty of Mathematics and Informatics, Vilnius University

prof. Paulius Drungilas

Matematikos ir informatikos fakultetas

The certificate was issued on 14 July, 2022.

Vilniaus universitetas

Visų mokslinių tyrimų ir disertacijos rengimo etapai

	Darbo pavadinimas	Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje): 1.1. Analitinės apžvalgos atlikimas. 1.2. Disertacijos tyrimo objekto detalizavimas. 1.3. Mokslinių problemų susietų su tyrimo objektu identifikavimas ir tyrimo tikslo suformavimas.	2020 m. spalio mėn. – 2021 m. rugsėjo mėn.	Atlikta, apibendrinti rezultatai mokslinėje ataskaitoje.
	Mokslinio tyrimo vykdymas:		
2.	**2.1.Tyrimo metodikos sudarymas:** 2.1.1. Uždavinių, skirtų tyrimo tikslui pasiekti, suformulavimas. 2.1.2. Tyrimo metodikos išsikeltiems uždaviniams spręsti parinkimas. 2.1.3. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.	2021 m. spalio mėn. – 2022 m. sausio mėn.	Atlikta, apibendrinti rezultatai mokslinėje ataskaitoje.

Visų mokslinių tyrimų ir disertacijos rengimo etapai

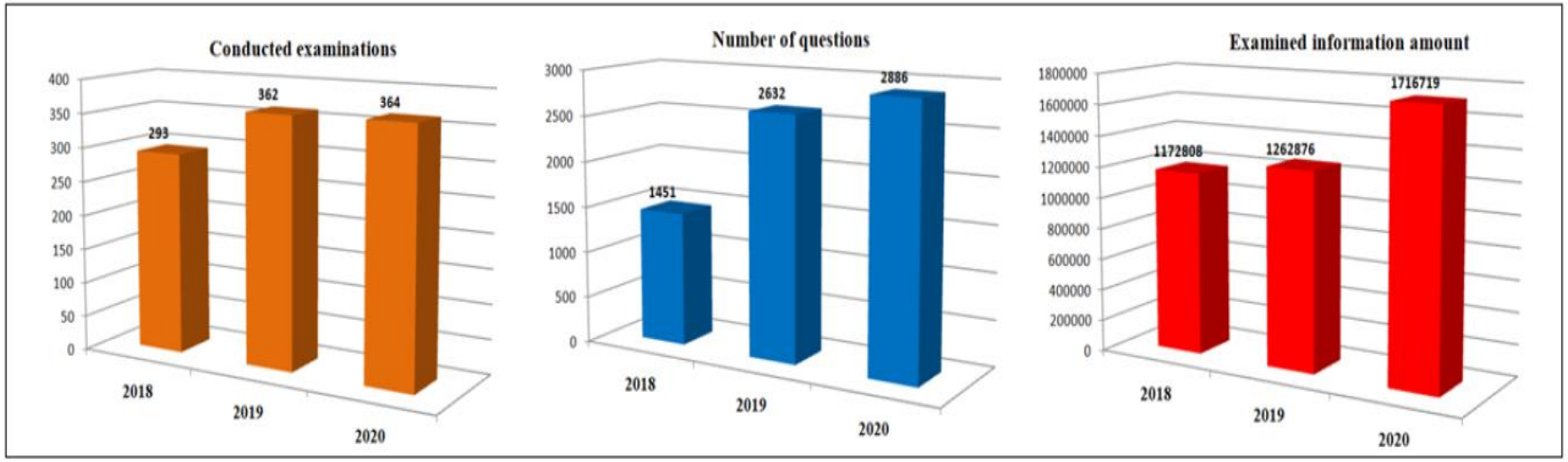
	Darbo pavadinimas	Atlikimo terminai	Pastabos
	2.2. Teorinis tyrimas: 2.2.1. Mašininio mokymosi metodų naudojamų automatizuotame slaptažodžių parinkime tyrimas. 2.2.2. Semantinės slaptažodžių analizės ir šablonų parinkimo metodų tyrimas. 2.2.3.Slaptažodžių parinkimo algoritmų taikant semantinę analizę tyrimas.	2022 m. sausio mėn. – 2022 m. rugsėjo mėn.	Vykdomas apibendrinti rezultatai mokslinėje ataskaitoje.
	2.3. Empirinis tyrimas: 2.3.1. Skirtingų algoritmų palyginimas. 2.3.2. Įgyvendintų algoritmų modifikacijos, ar naujų algoritmų kūrimas, sprendžiant apibrėžtus uždavinius. 2.3.3. Sukurtų modifikacijų eksperimentinis tyrimas analizuojant jų efektyvumą	2022 m. spalio mėn. – 2023 m. gegužės mėn.	Vykdomas - pradėtas skirtingų algoritmų (metodų) palyginimas.
	2.4. Gautų rezultatų analizė ir apibendrinimas	2023 m. birželio mėn. – 2023 m. rugsėjo mėn.	

Visų mokslinių tyrimų ir disertacijos rengimo etapai

	Darbo pavadinimas	Atlikimo terminai	Pastabos
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų ir kt.) parengimas: 3.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas. 3.2. Analitinės disertacijos dalies parengimas. 3.3. Teorinės disertacijos dalies parengimas. 3.4. Eksperimentinės disertacijos dalies parengimas. 3.5. Bendrųjų išvadų formulavimas.	2023 m. spalio mėn. – 2024 m. gegužės mėn.	
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2024 m. birželio mėn.	
5.	Daktaro disertacijos gynimas	2024 m. rugsėjo mėn.	

Ekspertiniai tyrimai

- Teisminės ekspertizės (susijusios su IT) Lietuvoje.
- Pagrindiniai užsakovai.
- Tiriamieji objektai.
- Tyrimų statistika.



Vilniaus universitetas

Problemos

Kaip ištirti šifruotą informaciją?

Slaptažodžių parinkimo metodai:

- **Žodynų taikymas;**
- **Nutekintų slaptažodžių duomenų bazių panaudojimas;**
- **Pilno perrinkimo atakos („brute-force“);**
- **Kombinuotos (mišrios) slaptažodžių parinkimo atakos, skirtinguose etapuose naudojant žodynų ir „brute force“ atakas.**

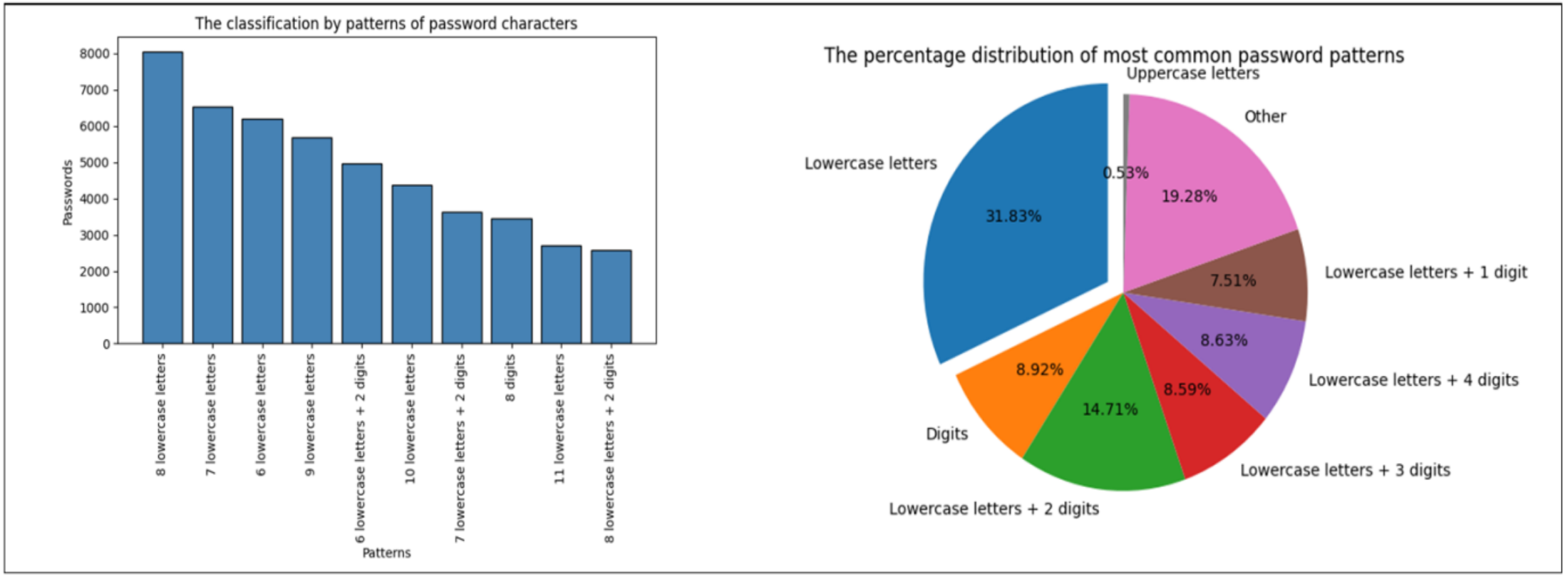
Slaptažodžių parinkimo priemonės:

- **Laboratorijos aparatūrinė įranga;**
- **Laboratorijos programinė įranga.**

Laikas (kiek galime skirti laiko ir resursų parinkti slaptažodį).

Objektai

Pagrindiniai tyrimo objektai yra: lietuviškas **slaptažodis** (ir jį atitinkantis šablonas, būdingas mūsų regiono vartotojams), bei mašininio mokymosi **algoritmas** jo parinkimui.



Tikslai ir uždaviniai

Disertacijos tikslas:

Sukurti naują arba modifikuoti (patobulinti) jau esamą slaptažodžių parinkimo metodą, adaptuotą mūsų regiono vartotojų slaptažodžių ypatybėms, pritaikytą teisminės IT ekspertizės ir tyrimo uždaviniams.

Disertacijos uždaviniai:

•Atlikti naujausių slaptažodžių parinkimo metodų apžvalgą, siekiant nustatyti tinkamiausius lietuviškų slaptažodžių parinkimui.

•Atlikti eksperimentus su empirinėmis duomenimis taikant atrinktus (pirmajame žingsnyje) slaptažodžių parinkimo metodus, siekiant rasti efektyviausius.

•Pasiūlyti metodą, naudojantį kontekstinę vartotojo informaciją, ištirti jo veikimą ir palyginti gautus rezultatus su alternatyviais metodais.

•Pasiūlyti neuroninių tinklų taikymų grįstą metodą, naudojantį kontekstinę naudotojo informaciją ir slaptažodžių šablonus, ištirti jo veikimą ir palyginti gautus rezultatus su alternatyviais metodais.

Slaptažodžių parinkimo metodai

Metodas	Algoritmas/programa
Taisyklėmis pagrįstas slaptažodžių parinkimo metodas	John the Ripper Hashcat
Markovo grandinės	OMEN
PCFG (angl. k. probabilistic context-free grammars) - tikimybiniai gramatikos taisyklių rinkiniai.	PCFG cracker
Pasikartojantys neuroniniai tinklai - Recurrent Neural Networks (RNNs).	RNN
Generatyviniai besivaržantys tinklai - Generative adversarial networks (GAN).	PassGan

Empiriniai duomenys

Duomenų bazės pavadinimas	Bendras slaptažodžių skaičius	Unikalių slaptažodžių skaičius	Šaltinis
LT1	110303	97657	https://raidforums.com
LT2	157617	114861	https://raidforums.com
LT3	645973	470298	https://raidforums.com
Rockyou			http://downloads.skullsecurity.org/passwords
Lietuvių k. žodynas	83258		https://github.com/giekaton/lithuanian-words-txt
Anglų k. žodynas	466550		https://github.com/dwyl/english-words
Lietuviškų slapt. rinkinys	523755		https://github.com/lexcor/LT-SecList

Atlikti eksperimentai

Taisyklėmis grįsto metodo taikymo rezultatai.

Stage	Password candidates	Guessed passwords	Guessed percent	Dictionaries and Rules
1	83258	2320	2.02 %	LitDict
2	83258	3517	3.06 %	LitDictAscii
3	549808	4614	4.02 %	LitDictAscii, EnDict
4	2748628	7221	6.28 %	LitDictAscii, EnDict, Basic rule set
5	42334773	19020	16.55 %	LitDictAscii, EnDict, Best64 rule set
6	82664027	46328	40.32 %	LitDictAscii, EnDict, Dataset, Best64 rule set
7	3864339263	56086	48.82 %	LitDictAscii, EnDict, Dataset, Best64, Specific rule sets
8	32206829929	75443	65.66 %	LitDictAscii, EnDict, Dataset, Rockyou-30000 rule set
9	106374865246	80104	69.72 %	LitDictAscii, EnDict, Dataset, Dive rule set

Atlikti eksperimentai

OMEN ir PCFG taikymo rezultatai.

Password candidates	OMEN guessed passwords	Guessed percent
10^5	2605	2.27 %
10^6	8186	7.12 %
10^7	19254	16.76 %
10^8	34129	29.71 %
10^9	42674	37.14 %
$5 \cdot 10^9$	50537	43.99 %

Password candidates	PCFG guessed passwords	Guessed percent
10^5	12109	10.54 %
10^6	25740	22.40 %
10^7	42633	37.11 %
10^8	51797	45.08 %
10^9	57771	50.28 %
$5 \cdot 10^9$	60707	52.84 %

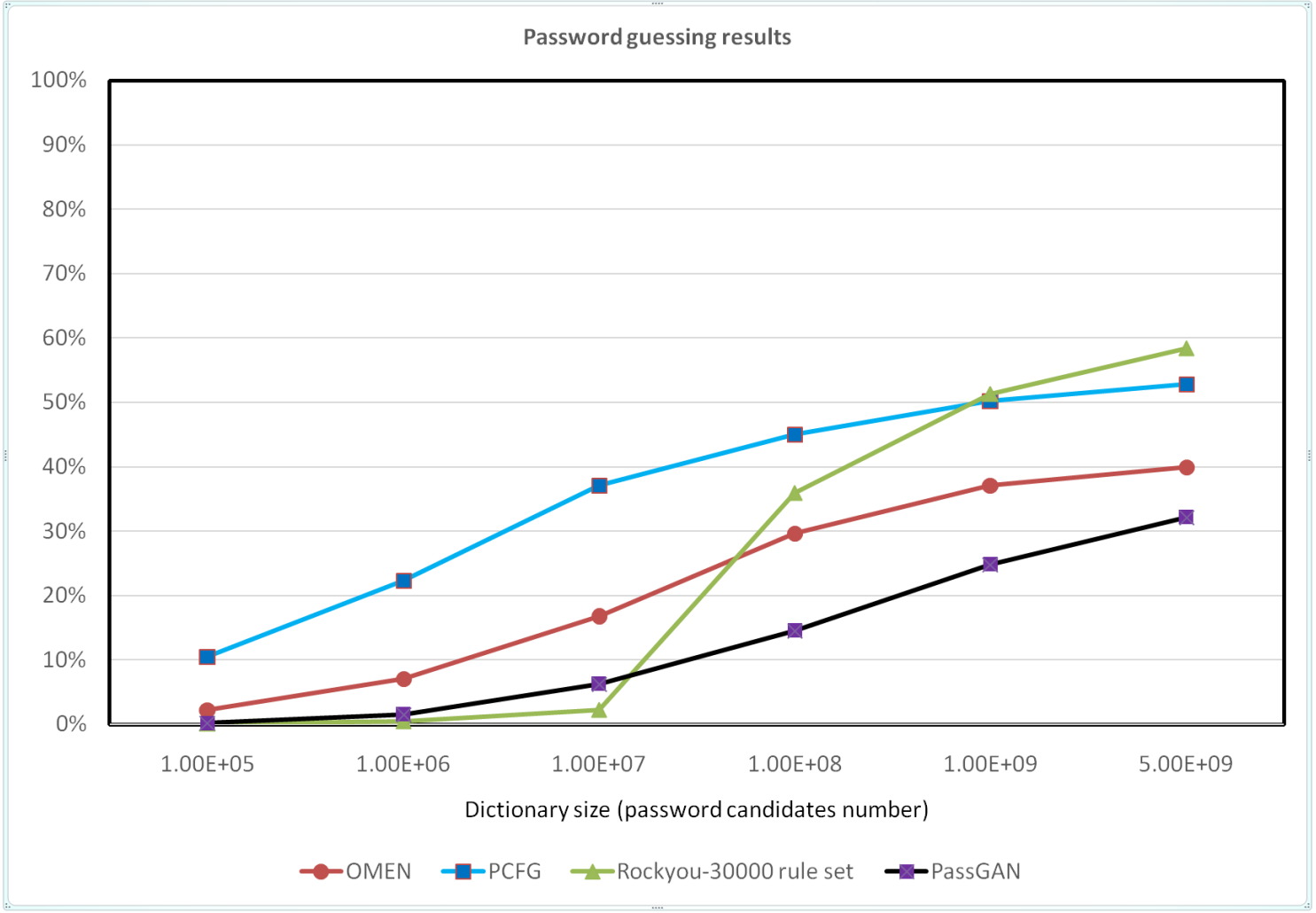
Atlikti eksperimentai

Rockyou taisyklių ir PassGan taikymo rezultatai.

Password candidates	PassGAN guessed passwords	Guessed percent
10^5	253	0.22 %
10^6	1846	1.61 %
10^7	7245	6.31 %
10^8	16788	14.61 %
10^9	28609	24.90 %
$5 \cdot 10^9$	36979	32.19 %

Password candidates	Rockyou rules guessed passwords	Guessed percent
10^5	112	0.10 %
10^6	536	0.47 %
10^7	2565	2.24 %
10^8	41274	35.92 %
10^9	58965	51.32 %
$5 \cdot 10^9$	67162	58.46 %

Palyginimas



Vilniaus universitetas

Kito pusmečio darbo planas

1. Tęsti eksperimentus pagal sudarytą teorinį ir empirinį tyrimų planą atlikimas.
2. Gautų rezultatų analizė. Publikacijos (CA WoS) ruošimas.

Vilniaus universitetas

Ačiū už dėmesį

Andrius Chaževskas

VU DMSTI doktorantas
Andrius.Chazevskas@mif.stud.vu.lt