Vilniaus universitetas

Doktorantas:
Paulius Vaitkevičius

Vadovas:
Dr. Virginijus Marcinkevičius

IV metų I pusmečio ataskaita
2022 m. kovo 24 d.

Mašininiu mokymusi grįstų atvirųjų šaltinių žvalgybos informacijos išskyrimo ir analizės metodai

Doktorantūros laikotarpis: 2018 - 2022

Vilniaus universitetas

TURINYS

1. Studijų plano vykdymas
2. Trumpas per pusmetį gautų mokslinių rezultatų pristatymas
3. Problemos apibrėžimas, tyrimo objektas, tikslai ir planuojami gauti rezultatai
4. Kito pusmečio darbo planas

STUDIJŲ PLANO VYKDYMAS

Vilniaus universitetas

Visų studijų planas, vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2018/2019)	1	1		1			
II (2019/2020)	1	3		1		1	Publikuota
III (2020/2021)	2		1		1	1	Įteikta - nepriimta
IV (2021/2022)			1		1		Įteikta į kitą žurnalą

Vilniaus universitetas

Ataskaitinio pusmečio darbo planas ir jo įvykdymas

Egzaminai		Dalyvavimas konferencijose		Publikacijos	
Planas	**Įvykdyta**	**Planas**	**Įvykdyta**	**Planas**	**Įvykdyta**
		Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.	Dar neįvykdyta	Empirinio tyrimo rezultatų publikavimas (recenzuojamame leidinyje, CA WoS su Impact Factor).	Publikacija parašyta ir įteikta recenzuojam leidiniui „Journal of Intelligent Information Systems “ (CA WoS su Impact Factor)

Vilniaus universitetas

Visų mokslinių tyrimų ir disertacijos rengimo etapai

1. Mokslinių tyrimų disertacijos tema apžvalga ir analizė
2. Mokslinio tyrimo vykdymas:
    1. Tyrimo metodikos sudarymas
    2. Teorinis tyrimas
    3. Empirinis tyrimas
    4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas
3. Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas
4. Daktaro disertacijos parengimas ir svarstymas padalinyje
5. Daktaro disertacijos gynimas

PROBLEMOS APIBRĖŽIMAS, TYRIMO OBJEKTAS, TIKSLAI IR PLANUOJAMI GAUTI REZULTATAI

Vilniaus universitetas

Tyrimo tikslas

Sukurti apsimetinėjimo atakoms atsparų metodą, grįstą giliaisiais neuroniniais tinklais ir natūralios kalbos apdorojimo algoritmais, kuris leistų efektyviai ir patikimai atpažinti **duomenų išviliojimo internete** (angl. „Phishing“) tinklapius.

Vilniaus universitetas

Tyrimo objektas

1. Mašininio mokymo ir giliojo mašininio mokymo algoritmai, skirti atpažinti duomenų išviliojimo internete tinklapius.
2. Atsparūs priešiškoms atakoms algoritmai (angl. „Adversarial Machine Learning“).

Vilniaus universitetas

Tyrimo uždaviniai

1. Atlikti literatūros analizę, išanalizuoti state-of-the-art algoritmus duomenų išviliojimo internete tinklapių atpažinimui.
2. Atkartoti *state-of-the-art* algoritmų rezultatus.
3. Sukurti duomenų rinkinius eksperimentų vykdymui.
4. Pasiūlyti naują efektyvesnį duomenų išviliojimo internete tinklapių atpažinimo metodą.
5. Atlikti eksperimentinius tyrimus, palyginant pasiūlytą metodą su *state-of-the-art* algoritmais.

Vilniaus universitetas

Planuojami rezultatai

1. Atlikta **literatūros analizė**, palyginant pažangiausius tyrimo srities algoritmus;
2. Atlikti **eksperimentiniai tyrimai**:
   - ✓ Mašininio mokymosi algoritmų efektyvumo palyginimas;
   - ✓ Giliojo mašininio mokymosi (GMM) algoritmų efektyvumo palyginimas;
   - ✓ GMM algoritmų (RNN, LSTM, GRU, CNN, kt.) efektyvumo tyrimai, naudojant natūralaus teksto apdorojimo technikas (N-grams, word embedings, kt.).
   - ✓ Naujo GMM algoritmo kūrimas, sprendžiant apibrėžtus uždavinius.
   - ✓ Pasiūlyto GMM algoritmo eksperimentinis tyrimas analizuojant jo efektyvumą;
   - ✓ Pasiūlyto GMM algoritmo atsparumo priešiškoms atakoms (angl. „Adversarial Machine Learning) eksperimentiniai tyrimai.

PER PUSMETĮ GAUTŲ MOKSLINIŲ REZULTATŲ PRISTATYMAS

KITO PUSMEČIO PLANAS

Tyrimų motyvacija

Vilniaus universitetas

2020 metais Sabir et al. [9] bei kiti [10, 5] pademonstravo, kad ML grįsti *state-of-the-art* sukčiavimo internete atpažinimo metodai (kurių deklaruojamas tikslumas $> 98\%$) yra smarkiai pažeidžiami (pažeidžiamumo sėkmė $> 66\%$).

[9] naudotų obfuskacijos technikų pavyzdžiai:

Type	Level	Obfuscation Method	Target *x* (https://store.steampowered.com/)
Domain	Char	Addition	https://store.steampowereda.com
		Insertion	https://store.steaompowered.com
		BitSquatting	https://store.steampowerel.com
		Homoglyph	http://https://store.steamp0wered.com
		Omission	https://store.seampowered.com
		SubDomain	https://store.steam.powered.com
		Hyphenation	https://store.st-eampowered.com
		CharacterSwap	https://store.steampowired.com
		Repetition	https://store.steaampowered.com
		Transpose	https://store.stemopowered.com
	Word	WordSubDomain	https://store.steampowered.ai-assisted.com
		WordHyphenation	https://store.ai-assisted-steampowered.com
		WordRepetition	https://store.steampowered-steampowered.com
		WordSwap	https://store.poweredsteam.com/
Path	Word	PathDm	https://store.steampowered-operated.com/steampowered
		PathExe	https://store.steampowered-operated.com/steampowered.exe
TLD	Word	TldReplace	https://store.steampowered.in.rs

Tyrimų tikslas

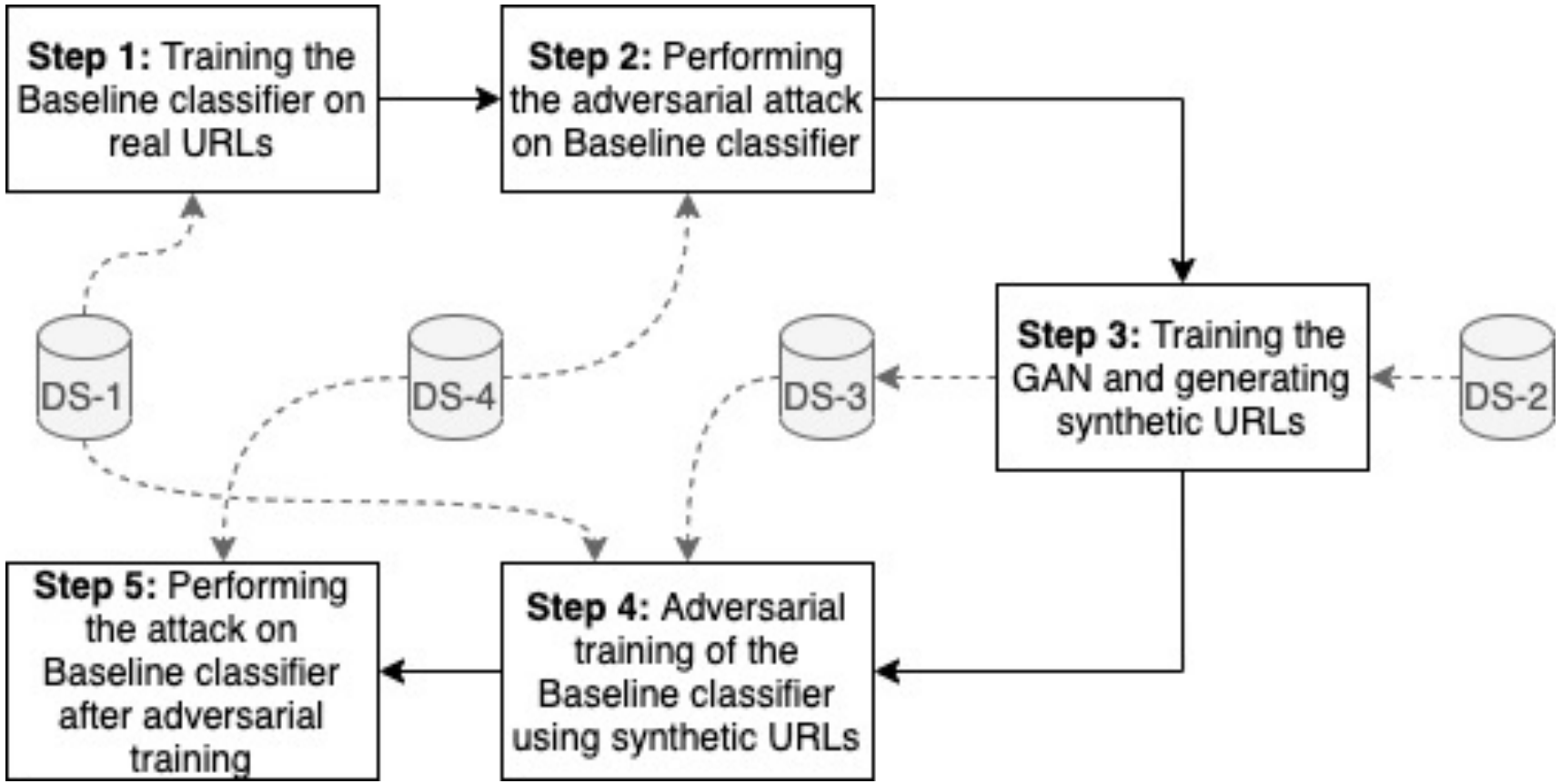
Atliktų tyrimų tikslas:

- sukurti apsimetinėjimo atakoms atsparų metodą, pasinaudojant ankstesnių RNN tyrimų patirtimi,
- ištestuoti metodo efektyvumą, naudojant Sabir et al. [9] pažeidžiamumo atakų įrankius,
- palyginti sukurtą metodą su šios srities *state-of-the-art* metodais,
- sugeneruoti didelę sintetinių mokymo ir testavimo duomenų aibę vėlesniam nuosavo atsparaus apsimetinėjimo atakoms sukčiavimo internete metodo kūrimui.

Nėra publikacijų, kur su generatyviniais modeliais būtų generuojami teksiniai fišingo URL adresai.

Eksperimento schema

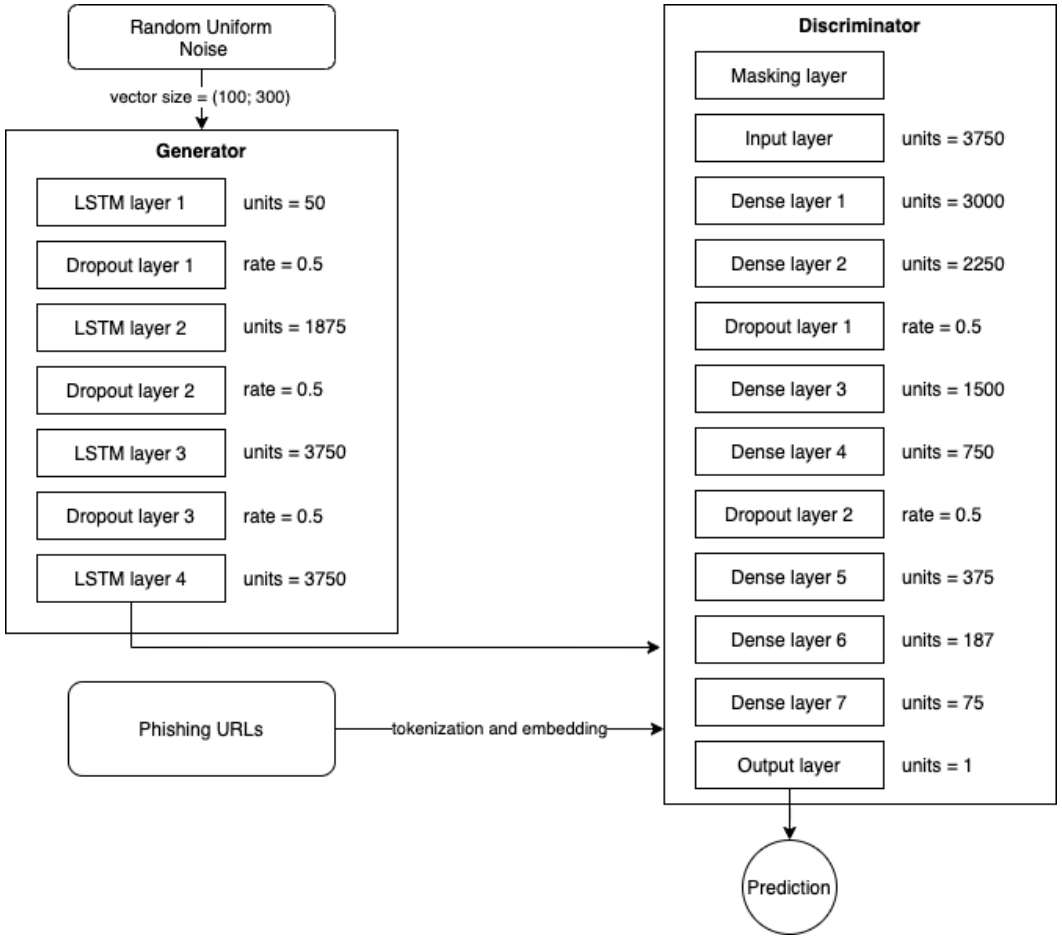
Vilniaus universitetas



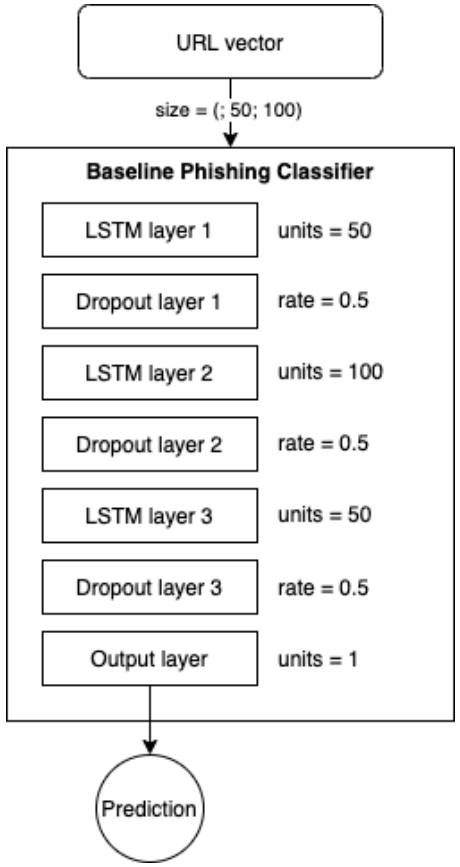
Metodai

Vilniaus universitetas

WGAN-GP



Baseline classifier



Vilniaus universitetas

Pirmo pusmečio rezultatų apibendrinimas

- Pakoreguota ir pakartotina įteikta publikacija
- Rašoma disertacija

Vilniaus universitetas

Kito pusmečio darbų planas

- Paskutiniojo tyrimo rezultatų pristatymas tarptautinėje konferencijoje
- Disertacijos rašymas

Vilniaus universitetas

AČIŪ UŽ DĖMESĮ

ORCID

Paulius Vaitkevičius
VU DMSTI doktorantas
+370 650 83623
paulius.vaitkevicius@mif.vu.lt