

HODNOTENIE KVALITY HESIEL PROSTREDNÍCTVOM DETEKCIE ANOMÁLIÍ

Dominik Fronc

Školiteľ: doc. RNDr. Martin Stanek Phd.

CIEĽ BAKALÁRSKEJ PRÁCE

- Implementácia hodnotenia kvality hesiel algoritmom izolačného lesa
- Vyskúšanie 2 prístupov (anomálie – slabé/silné heslá)
- Porovnanie s používanými metódami

MOTIVÁCIA

Registrácia na *zoznam.sk* :

Heslo: password

Heslo:*

●●●●●●●●

Zopakuj heslo:*

●●●●●●●●

Nemáte dostatočne silné heslo, skúste vytvoriť dlhšie heslo

Heslo musí mať minimálne 8 a maximálne 16 znakov a musí obsahovať kombináciu aspoň 3 znakov: veľkých a malých písmen, číslíc a a špeciálnych znakov (~!@#\$\$%&*). Povinný údaj.

Heslo: P@ssword123

Heslo:*

●●●●●●●●●●

Zopakuj heslo:*

●●●●●●●●●●

Veľmi silné heslo

PASSWORD STRENGTH ESTIMATORS

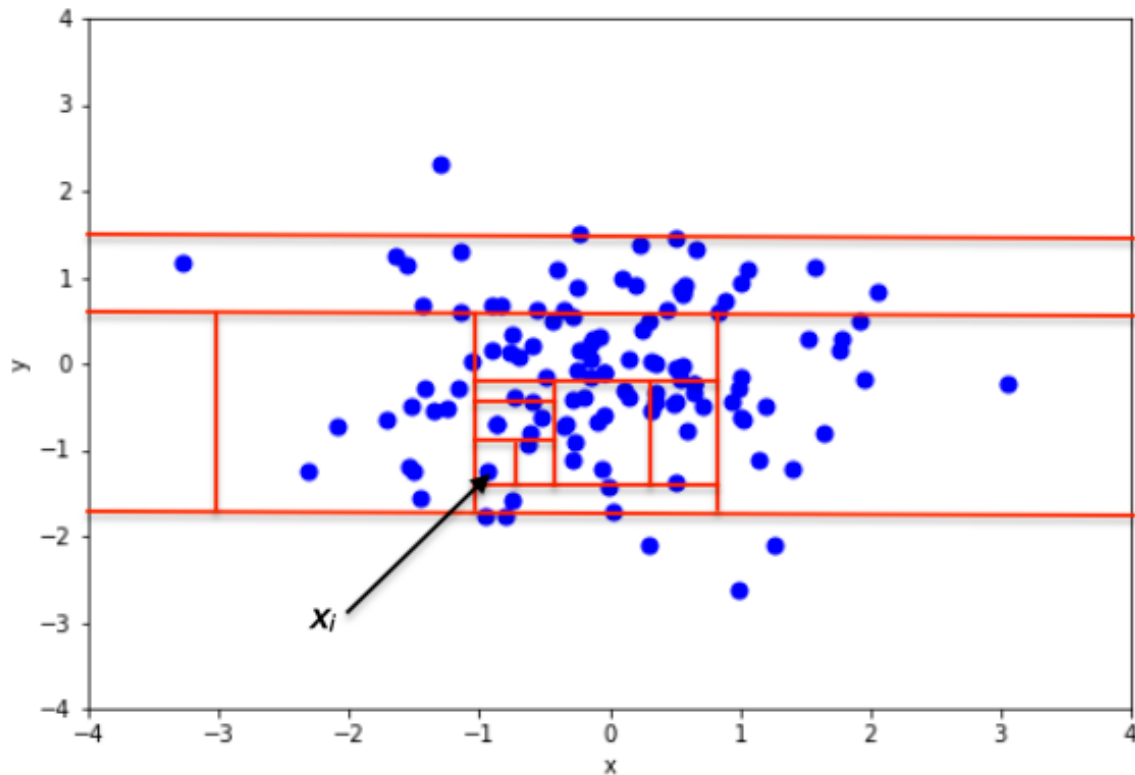
Existujúce riešenia:

- ZXCVCBN
- LPSE
- PESRank
- Nástroj využívajúci NN

Password Guessability Service (PGS)

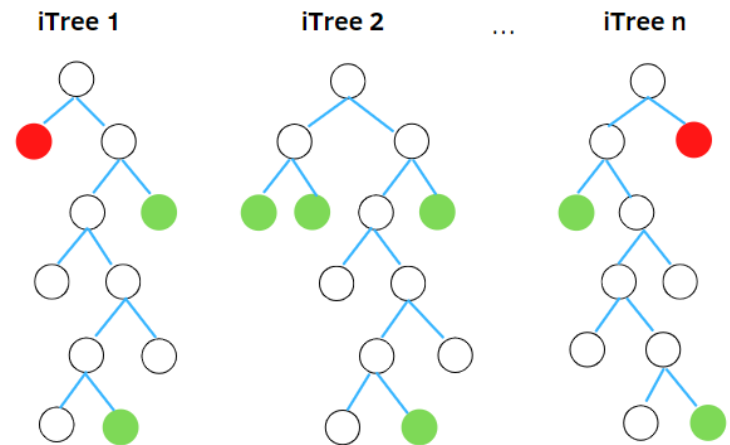
- Služba vyvinutá výskumnou skupinou na Carnegie Mellon University
- Hodnotí kvalitu hesiel
- PCFG, HashCat, NN ...

IZOLAČNÝ LES



https://en.wikipedia.org/wiki/Isolation_forest#/media/File:Isolating_a_Non-Anomalous_Point.png

https://cdn-images-1.medium.com/max/852/1*D78QLbcwXesyhquuofnOg.png



- Bod v dátach – vektor: (a_1, a_2, \dots, a_k)
- Každému bodu pridelené skóre
- Hyperparametre
- Anomálie – silné / slabé heslá

DÁTOVÉ SÚBORY

Dátový súbor	Počet hesiel	Dátum zverejnenia	Zdroj
RockYou	14 328 556	2009	[6]
Crackstation	63 442 240	2019	[4]
Carders.cc	1 902	2010	[3]
Facebook Phished	2 430	2010	[3]
Hak5	2 351	2009	[3]
Hotmail	8 930	neznámy	[3]
Phpbb	184 344	2009	[3]

METRIKY

- Dĺžka hesla
- Počet malých písmen
- Počet veľkých písmen
- Počet číslic
- Počet špeciálnych symbolov
- Počet znakových skupín
- Počet rôznych použitých znakov
- Počet nesequenčných prechodov
- Priemerný počet výskytov dvojíc znakov v hesle

TESTOVACIE A TRÉNOVACIE SADY

Sada	Kvalita hesiel	Počet hesiel	Heslá
Trénovacia sada 1	slabé	10 000	SecList [1]
Trénovacia sada 2	slabé	30 000	Crackstation 10^6
Trénovacia sada 3	silné	30 000	CrackStation 10^{14}
Trénovacia sada 4	silné	10 000	generované reťazce
Testovacia sada 1	celý rozsah	3 534	CrackStation výber
Testovacia sada 2	celý rozsah	11 281	Hotmail + Hak5

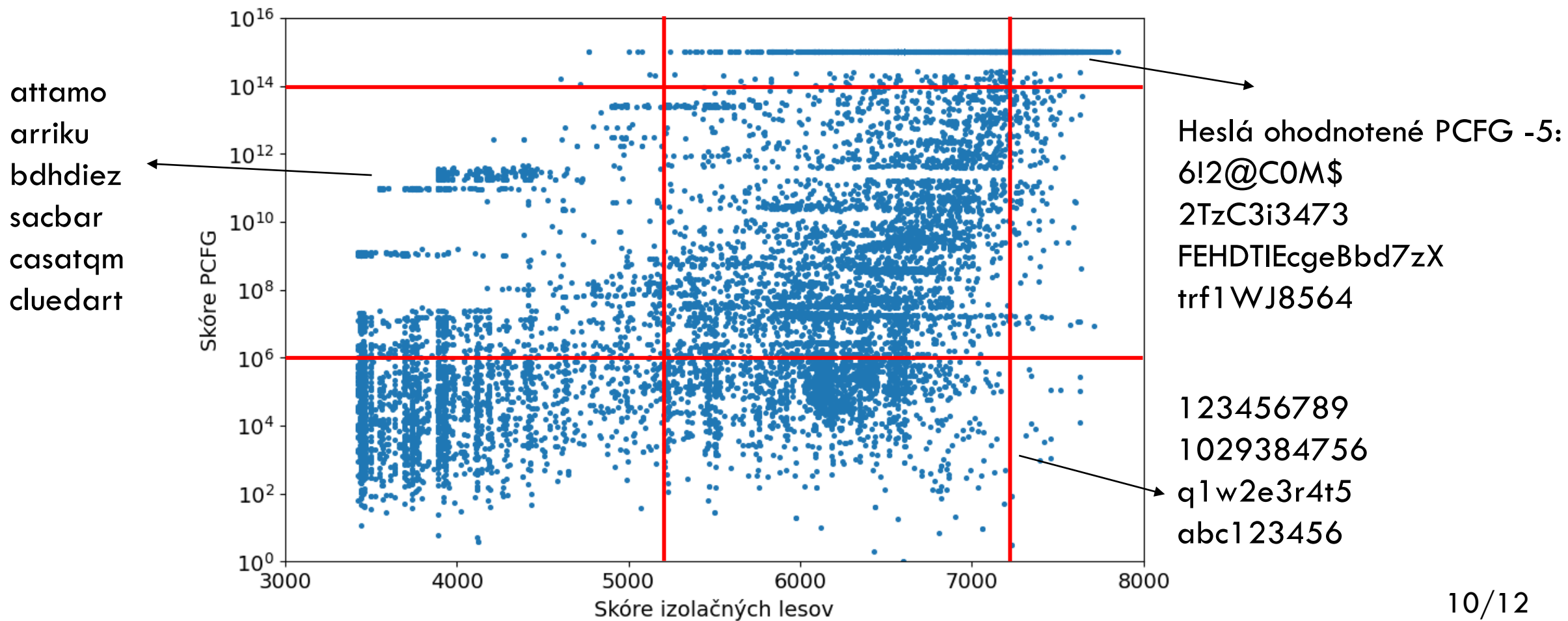
POROVANIE PRESNOSTI

- Izolačný les – skóre anomálie
- PCFG – počet pokusov
- Kategorizovanie hesiel do kategórií: slabé, priemerné, silné
- Hraničné hodnoty PCFG: 10^6 , 10^{14}
- Hľadanie hraníc izolačného lesa cez optimalizačnú úlohu
- Presnosť modelu

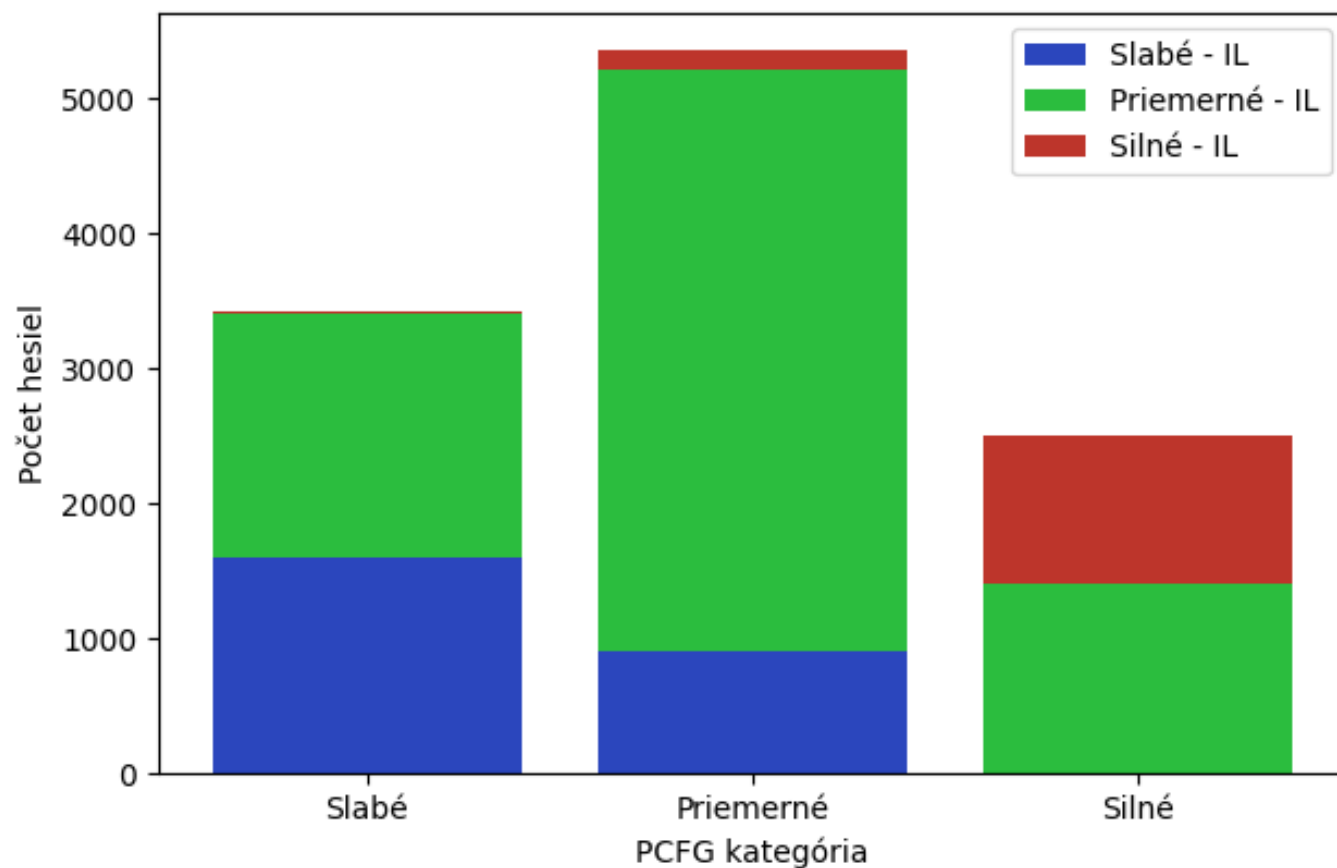
POROVANIE TRÉNOVACÍCH SÁD

Aspekt	Sady slabých hesiel		Sady silných hesiel	
	Sada 1	Sada 2	Sada 3	Sada 4
Presnosť modelu [%]	62,82	53,90	47,55	53,09
Nadhodnotené heslá [%]	16,55	24,06	30,29	14,75
Podhodnotené heslá [%]	20,63	22,05	22,16	32,16
Slabé heslá [počet]	2 777	1 183	29	3 815
Priemerné heslá [počet]	7 107	9 607	11 252	6 703
Silné heslá [počet]	1 397	491	0	763

POROVNANIE S PCFG

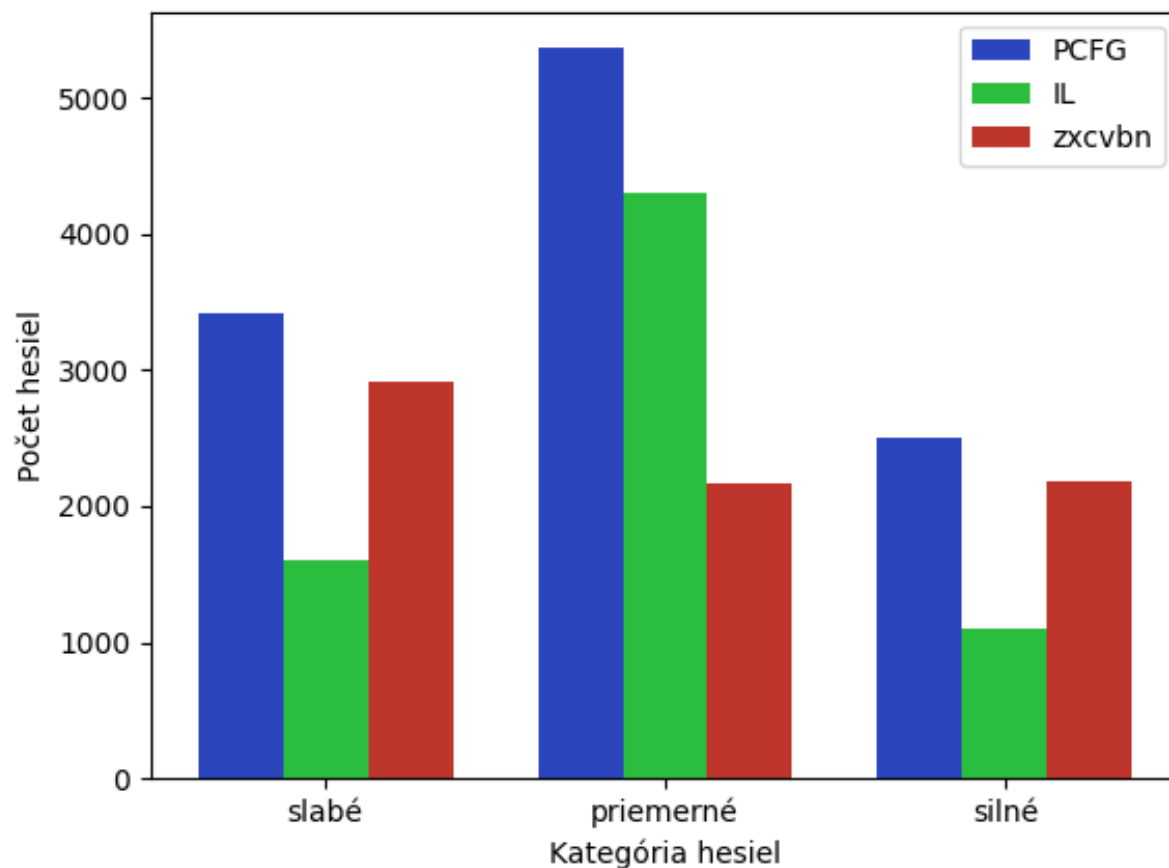


POROVNANIE S PCFG



- IL – izolačný les
- 62,82% presnosť
- podhodnotenú vs. nadhodnotenú heslá
- Nadhodnotenú: a123456789, 1a2s3d4f5g, 14725836, sarah123, carla20 atď.
- Podhodnotenú: maltuauH, carlosll, amiteamo, alapage, tooozkh, 4kitten atď.

POROVNANIE S ZXCVCBN



- IL: 62,82% presnosť
- Zxcvbn: 63,66% presnosť
- Zxcvbn – dobrá identifikácia slabých a silných, horšia priemerných
- IL – dobrá kategorizácia priemerných, horšie kategorizovanie hesiel z postupností a slovníkových hesiel

MOŽNÉ VYLEPŠENIA

- Použitie slovníka
- Zachytávanie hesiel z postupností
- Minimálna dĺžka
- Funkcia vysvetlenia kategorizácie hesla

ĎAKUJEM ZA POZORNOST