PROJEKT

Multitenantné operačné centrum kybernetickej bezpečnosti riešené ako otvorená cloudová služba s prvkami strojového učenia

V17 Výsledky a vyhodnotenie testov

KPB5 – Realizácia riešenia

typ – protokol

mesiac odovzdania – M24

Financované Európskou úniou
NextGenerationEU

PLÁN OBNOVY

ÚRAD VLÁDY
SLOVENSKEJ REPUBLIKY

Obsah

1. Strojové učenie a umelá inteligencia v oblasti kybernetickej bezpečnosti 3

1.1. Detekcia nežiadúcich emailových správ (Spam) 3

1.1.1. Vyhodnocovanie experimentov 3

1.1.2. Výsledky 5

1.1.3. Zhodnotenie testov 6

1.2. Detekcia DoS/DDoS útokov 6

1.1.4. DoS (Denial of Service) 6

1.1.5. DDoS (Distributed Denial of Service) 7

1.1.6. Analýza štatistických parametrov na vybraných DDoS útokoch 8

1.1.7. Variabilnosť, šikmosť a špicatosť 10

1.1.8. Detekčné hodnoty pre autokoreláciu a regresiu 13

1.1.9. Reakcia divergencie na DDoS útok 14

1.1.10. Využitie 3σ-Tunela pre rozpoznanie DDOS útokov 15

1.1.11. Detekčné funkcie pre strojové rozpoznanie DDoS útokov 17

1.1.12. Zhodnotenie testov 19

1.3. Detekcia škodlivého kódu (Malware) 19

1.1.13. Príprava a postup pri testovaní 20

1.1.14. Výsledky testov 20

1.1.15. Zhodnotenie testov 21

1.4. Detekcia phishing 22

1.1.16. Príprava a postup pri testovaní 22

1.1.17. Výsledky testov 22

1.1.18. Zhodnotenie testov 27

1.5. Detekcia pokročilých perzistentných hrozieb APT 28

1. STROJOVÉ UČENIE A UMELÁ INTELIGENCIA V OBLASTI KYBERNETICKEJ BEZPEČNOSTI

S nástupom informačného veku sa kybernetická bezpečnosť stala naliehavým problémom v dnešnej spoločnosti a zručnosťou, ktorá je v priemysle veľmi žiadaná. Podniky, vlády aj individuálni používatelia sú všetci vystavení riziku bezpečnostných útokov a narušení. Základným cieľom kybernetickej bezpečnosti je chrániť používateľov a ich údaje. Kybernetická bezpečnosť je mnohostranný problém, od vysoko technických oblastí (kryptografia a sieťové útoky) až po oblasti, s ktorými sa stretávame s používateľmi (detekcia nenávistných prejavov alebo podvodných transakcií s kreditnými kartami). Pomáha predchádzať poškodeniu citlivých informácií, predchádzať finančným podvodom a stratám a chrániť používateľov a ich zariadenia pred škodlivými aktérmi. Veľká časť analytických procesov, vyšetrovania a detekcie kybernetickej bezpečnosti je v súčasnosti riadená strojovým učením ML (Machine Learning) a „inteligentnými“ systémami.

V nasledujúcom texte popisujeme výsledky testov pre nasledujúce definované bezpečnostné hrozby:

- spam
- DoS/DDoS útoky
- malware
- phishing
- pokročilé perzistentné hrozby APT

1.1. Detekcia nežiadúcich emailových správ (Spam)

Spamové e-maily (nežiaduce správy) tvoria približne 60 % globálnej e-mailovej prevádzky. Okrem toho, že softvér na detekciu spamu sa od prvej spamovej správy v roku 1978 výrazne zdokonalil, každý, kto má e-mailový účet, vie, že spam je naďalej časovo náročný a drahý problém.

Z širšieho množstva bežne používaných metód a algoritmov pre detekciu a rozlišovanie medzi neškodnými/vyžiadanými emailovými správami (angl. ham) a nežiadúcimi/nevyžiadanými správami (angl. spam) sme vybrali a otestovali nasledovné metódy:

- Decision Tree (rozhodovací strom)
- SVM
- XGBoost
- Naive Bayes
- Logistická regresia

V snahe vytvoriť dostatočne robustný a univerzálny systém sme pre naše experimenty použili nasledovné voľne dostupné datasety:

- SpamAssassin
- SMS Spam

1.1.1. Vyhodnocovanie experimentov

Pre vyhodnocovanie experimentov boli použité nasledovné metriky.

Konfúzna matica

Konfúzna matica je jedným z najpopulárnejších spôsobov vyhodnocovania akéhokoľvek klasifikačného modelu. Hoci matica sama o sebe nie je metrikou, maticovú reprezentáciu možno použiť na definovanie rôznych metrík, z ktorých všetky sa stanú dôležitými v nejakom konkrétnom prípade alebo scenári. Konfúznu maticu je možné vytvoriť pre binárnu klasifikáciu, rovnako ako aj pre klasifikačný model s viacerými triedami.

Konfúzna matica sa vytvorí porovnaním predikovaného označenia triedy dátového bodu s jeho skutočným označením triedy. Toto porovnanie sa zopakuje pre celý súbor dát (dataset) a výsledky tohto porovnania sa zostavia v maticovom alebo tabuľkovom formáte. Táto výsledná matica predstavuje konfúznu maticu (obr. nižšie).

		Skutočná trieda	
		Kladná	Záporná
Predikovaná trieda	Kladná	Počet správne kladných (TP)	Počet nesprávne kladných (FP)
	Záporná	Počet nesprávne záporných (FN)	Počet správne záporných (TN)

Obr. 1: Konfúzna matica

Vo všeobecnosti máme zvyčajne 2 triedy – kladnú/pozitívnu a zápornú/negatívnu triedu. Na základe tejto štruktúry môžeme jasne vidieť štyri pojmy dôležitosti.

- True Positive (TP) – Počet správne kladných: Toto je suma celkového počtu príkladov (dátových bodov z datasetu) z pozitívnej triedy, kde sa skutočné označenie triedy rovnalo predpovedanému označeniu triedy, t. j. celkovým počtom prípadov, v ktorých sme správne predpovedali označenie pozitívnej triedy pomocou nášho modelu.
- False positive (FP) – Počet nesprávne kladných: Toto je počet celkových prípadov z negatívnej triedy, kde ich model nesprávne klasifikoval tým, že ich predpovedal ako pozitívne. Odtiaľ pochádza názov, falošne pozitívny.
- True Negative (FN) – Počet nesprávne záporných: Toto je počet celkového počtu príkladov z negatívnej triedy, kde sa skutočné označenie triedy rovnalo predpovedanému označeniu triedy, t. j. celkový počet prípadov, v ktorých sme správne predpovedali označenie negatívnej triedy pomocou nášho modelu.
- False negatívne (FN) – Počet nesprávne záporných: Toto je počet celkového počtu prípadov z pozitívnej triedy, kde ich náš model nesprávne klasifikoval tým, že ich predpovedal ako negatívne. Odtiaľ pochádza názov, falošne negatívny.

Accuracy

Táto metrika je jednou z najpopulárnejších spôsobov merania resp. vyhodnocovania efektívnosti ľubovoľného klasifikátora. Je definovaná ako celková presnosť alebo podiel správnych predpovedí modelu. Vzťah pre výpočet z konfúznej matice je nasledovný:

$$A = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision

Precision (presnosť), známa aj ako kladná prediktívna hodnota, je ďalšou metrikou, ktorú možno odvodiť z konfúznej matice. Táto metrika je definovaná ako počet predikcií, ktoré sú skutočne správne alebo relevantné zo všetkých predikcií založených na kladnej triede. Vzorec pre túto metriku je nasledovný:

$$P = \frac{TP}{TP + FP}$$

Model s vysokou hodnotou presnosti identifikuje vyšší podiel kladnej triedy v porovnaní s modelom s nižšou presnosťou. Táto metrika sa stáva dôležitou v prípadoch, keď sa viac zaujímame o nájdenie maximálneho počtu kladných tried, aj keď sa celková presnosť zníži.

Recall

Recall (úplnosť) je metrika modelu na identifikáciu percenta relevantných dátových bodov. Je definovaná ako počet inštancií kladnej triedy, ktoré boli správne predikované. Označuje sa aj ako miera pokrytia alebo citlivosť. Vzťah pre vypočítanie tejto metriky je nasledovný:

$$R = \frac{TP}{TP + FN}$$

Táto metrika sa stáva dôležitým ukazovateľom efektívnosti klasifikátora v scenároch, v ktorých chceme zachytiť čo najväčší počet prípadov konkrétnej triedy, aj keď to spôsobí zvýšenie falošných detekcií/predikcií. Ako príklad možno uviesť prípad bankového podvodu, model s vysokou hodnotou metriky recall nám poskytne vyšší počet prípadov potenciálnych podvodov. Rovnako to však zvýši vyvolanie poplachov pre väčšinu podozrivých prípadov.

F1-Score

Existujú prípady, v ktorých chceme dosiahnuť vyváženú optimalizáciu presnosti a úplnosti. *F1-Score* je metrika, ktorá je harmonickým priemerom metrík *precision* a *recall* a pomáha nám optimalizovať klasifikátor pre vyváženú presnosť a úplnosť predikcie. Vzorec pre výpočet *F1-Score* je nasledovný:

$$F = \frac{2.P.R}{P + R}$$

1.1.2. Výsledky

SpamAssassin

Metóda	Accuracy	Precision	Recall	F1-score	Processing time (sec.)
Decision tree	0.9761	0. 9776	0.9774	0.9774	1.4
SVM	0.9928	0.9929	0.9928	0.9929	11.9
XGBoost	**0.9928**	**0.9929**	**0.9928**	**0.9929**	**3.3**
Log. Regression	0.9833	0.9833	0.9833	0.9833	1.2
Naive Bayes	0.8438	0.8624	0.8439	0.8330	1.3

Tab.1: Výsledky testov pre dataset SpamAssassin

SMS Spam

Metóda	Accuracy	Precision	Recall	F1-score	Processing time (sec.)
Decision tree	0.7585	0.7624	0.7585	0.7604	0.001
SVM	0.7706	0.7722	0.7706	0.7714	0.005
XGBoost	0.7826	0.7757	0.7826	0.7791	0.001
Log. Regression	**0.7887**	**0.7709**	**0.7888**	**0.7795**	**0.001**
Naive Bayes	0.7785	0.7733	0.7785	0.7759	0.001

Tab. 2: Výsledky testov pre dataset SMS Spam

1.1.3. Zhodnotenie testov

Na základe vykonaných experimentov možno zhodnotiť, že najvhodnejšia metóda pre detekciu emailových správ typu spam z hľadiska efektívnosti je metóda XGBoost. Pre dataset „SpamAssassin“ dosahuje úspešnosťou detekcie spamu úroveň 99%. Táto hodnota je takmer identická pre metriky Precision aj Recall, takže je veľmi dobre pokrytá detekcia falošne pozitívnych aj falošne negatívnych detekcií správ typu spam. Pre tento dataset dosahuje takmer identické výsledky aj metóda SVM, avšak čas vyhodnocovania v prípade algoritmu SVM je približne 4x väčší oproti XGBoost. Napriek tomu sa jedná stále o niekoľko sekundový proces, takže možno nasadiť obe metódy pre real-time aplikácie. Pre dataset „SMS Spam“ úspešnosť detekcie spamu výrazne klesla na hodnotu ~78%. Najlepšie výsledky dosiahla metóda Logistická regresia, hoci rozdiel medzi metódami XGBoost a Logistickou regresiou predstavuje len desatiny percent. Z hľadiska času potrebného pre predikciu dát, algoritmus SVM opäť vyžaduje najdlhší interval. Na základe vykonaných experimentov teda možno sumarizovať, že najvhodnejšou metódou na detekciu správ typu spam je algoritmus **XGBoost**.

1.2. Detekcia DoS/DDoS útokov

Útoky typu „denial-of-service“ (DoS) a „distributed denial-of-service“ (DDoS) sa zameriavajú na narušenie normálneho fungovania cieľového servera, služby alebo siete tým, že ju zahltia záplavou internetovej prevádzky. Útoky DoS zvyčajne pochádzajú z jedného zdroja, zatiaľ čo útoky DDoS sú spustené z viacerých napadnutých zariadení, často rozptýlených globálne. Toto množstvo zdrojov sťažuje zastavenie útokov DDoS a môžu mať výraznejší dopad.

Tieto útoky fungujú tak, že preťažia cieľ väčším počtom požiadaviek, ako dokáže spracovať, čo spôsobí výrazné spomalenie alebo úplné zrútenie systému a odmietnutie služby legitímnym používateľom. Medzi bežné metódy patrí zneužívanie zraniteľností servera alebo používanie siete botov, známych ako botnet, na zaplavenie cieľa prevádzkou. Motivácie týchto útokov sa môžu líšiť od jednoduchého vandalizmu a osobných záští až po vydieranie a politický aktivizmus.

1.1.4. DoS (Denial of Service)

Skratka DoS v preklade znamená útok odmietnutia nejakej služby. Tento typ útoku sa používa na zahltenie servera. V tomto prípade útočník využíva zariadenie, ktorým sa pokúša narušiť normálnu prevádzku cieľového servera, služby alebo siete. Veľmi rýchlo sa stali hrozbou kybernetickej bezpečnosti. Sú jednoduché avšak ich dopad môže byť pre jedincov ale

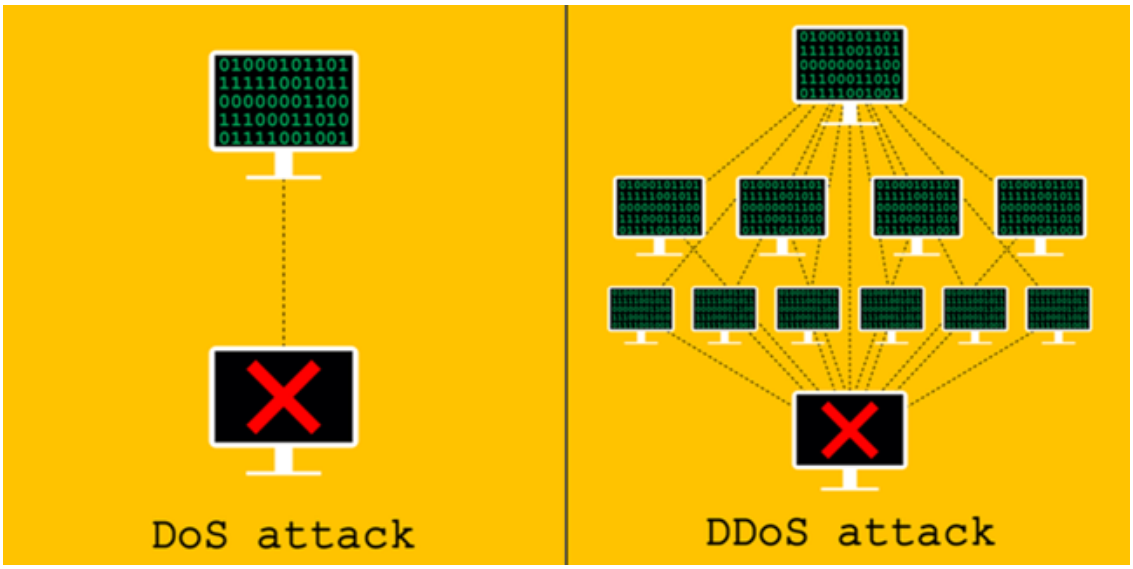
aj spoločnosti devastujúci.

DoS útoky sa často používajú pre vypnutie zariadení a sietí a existuje veľa rôznych spôsobov využitia spomínaných DoS útokov. Patria sem napr.:

- Pretečenie vyrovnávacej pamäte (angl. buffer overflow): Vyrovnávacia pamäť je dočasné úložisko pre ukladanie dát. Ak program alebo systémový proces umiestni viac údajov ako bolo vopred určené, ďalšie údaje potom pretečú a môže sa stať, že niektoré z nich potom uniknú do iných vyrovnávacích pamätí. To má potom za následok poškodenie alebo prepísanie ďalších údajov. Útočník dokáže na základe tohto princípu poškodiť súbory, zmeniť alebo odhaliť rôzne informácie.
- ICMP útok (angl. ICMP attack alebo Ping of Death): Tento typ útoku využíva ICMP protokol pričom sa využívajú ICMP echo-requesty tzv. pingy. Všeobecne sa echo-requesty a echo-reply používajú na ping sieťového zariadenia s cieľom overiť konektivitu medzi zariadením a odosielateľom. Ak by útočník zahltil cieľ paketmi požiadaviek, sieť musí odpovedať rovnakým počtom paketov, čo spôsobí nedostupnosť zariadenia pre iné prichádzajúce požiadavky.
- UDP útok (angl. UDP attack): Útok je založený na posielaní UDP paketov so zdrojovou IP adresou, ktorú môže útočník sfalšovať a snahou je zahltiť náhodné porty na cieľovom servery. Výhodou UDP protokolu je, že nepotrebuje nadviazať spojenie, ako to je v prípade protokolu TCP. V prípade, že cieľový server príjme UDP paket na porte, ktorý nie je využívaný žiadnou aplikáciou, informuje o tom odosielateľa. Pokiaľ útočník odošle dostatočné množstvo paketov na takýto port, môže byť preťažený.
- SYN útok (angl. SYN flood): Toto je útok, ktorý vzniká pri vytváraní TCP spojenia medzi 2 zariadeniami. TCP využíva trojfázové potrasenie rúk (angl. 3-way handshake). Nadviazanie spojenia začína odoslaním správy SYN. Ak je server aktívny odošle sa SYN/ACK správa späť ku klientovi, ktorí chcel nadviazať spojenie. Server sa snaží odpovedať dokým nevyprší čas na nadviazanie komunikácie, pričom si údaje uchováva v pamäti. Ak je klientom útočník s falošnou a nedostupnou zdrojovou adresou, zaslanie veľkého množstva takýchto paketov môže spôsobiť vyčerpanie pamäte, ktorá je určená pre TCP spojenia a všetky ďalšie pokusy o komunikáciu budú ďalej ignorované.
- Útok slzy (angl. Teardrop attack): Počas tohto útoku sa útočník snaží posielať fragmentované pakety a sieť sa potom snaží skompilovať tieto dáta do ich pôvodného stavu. Útočník môže využiť zraniteľnosť zastaraného operačného systému. Je bežné, že mnohé spoločnosti stále využívajú staršie verzie systémov z dôvodu opätovného spúšťania starej aplikácie a keďže niektoré verzie operačných systémov ako sú napr.: Linux či Windows obsahujú chybu pri spätnom zostavení fragmentovaného paketu, server môže byť práve kvôli tejto zraniteľnosti napadnutý.

1.1.5. DDoS (Distributed Denial of Service)

Na rozdiel od jednoduchého DoS útoku, DDoS útok (Distributed Denial of Service) naberá na sile, pretože útočník využíva viacero systémov naraz a zvyšuje tak závažnosť cielených útokov. Útočník môže použiť malware, ktorý odošle po sieti do zariadení so zámerom ich infikovať. Stanú sa tak súčasťou tzv. botnetu a podľa príkazov útočníka pomocou riadiaceho servera títo boti potom jednotne útočia na cielený server. DDoS útoky je veľmi ťažké odhaliť, pretože útočiaci boti môžu byť rozmiestnený na rôznych miestach po svete. Hlavný a kľúčový rozdiel medzi DoS a DDoS útokom je, že pri DDoS s použitím botov môže útočník posielať obrovské objemy prevádzky.



Obr. 2: DoS a DDos útoky

DDoS útoky sa rozdeľujú do 3 hlavných kategórií:

1. Objemový útok (angl. Volumetric attack): Tento typ využíva zraniteľné služby ako sú napr.: NTP, DNS či SSDP. Útočník sa snaží zahltiť cieľ veľkým objemom paketov, na ktoré sa cieľ snaží odpovedať, až kým nedôjde ku kolapsu na strane cielenej sieťovej infraštruktúry. Zvyčajne to je širokopásmový útok, ktorý dokáže presiahnuť 100 GB či dokonca niekoľko TB za sekundu a je okamžite očividný pre poskytovateľov pripojenia. Na detekciu objemových útokov sa osvedčilo monitorovanie siete, analyzovanie toku a prietoková telemetrická analýza.
2. Protokolový útok (angl. Protocol attack): Tento typ útoku sa snaží vyčerpať zdroje servera alebo jeho systémov a spolieha sa na zraniteľnosti internetových komunikačných protokolov. Medzi protokolové útoky patria napr.: SYN útok, UDP útok, ICMP útok. Tieto útoky sa považujú za dostatočne ochromujúce na cielenú infraštruktúru vďaka frekvencii trvania. V roku 2018 útočníci využili BGP protokol na presmerovanie prevádzky, ktorá bola určená pre službu „MyEtherWallet“, čo však bola iba zásterka pre krádež kryptomenových peňaženiek. Trvalo to približne 2 hodiny.
3. Útok na aplikačnú vrstvu (angl. Application layer attack): Tento útok sa zameriava na konkrétnu aplikáciu a jej zraniteľnosti, aby nebola schopná spracovávať, odosielať či prijímať dáta. Útoky sú slabé a pomalé a sústreďujú sa na požiadavky GET a POST. Cieľom je potom ochromiť webový server a meria sa v požiadavkách za sekundu (angl. Requests per second - Rps). Sú nízke a slabé preto, aby boli v súlade s protokolom.

Existuje niekoľko príznakov, ktoré nám dokážu pomôcť pri odhalení, či je naša sieť alebo aplikácia alebo celá infraštruktúra pod DDoS útokom. Môže to byť náhly nárast webovej aktivity na serveri, znížený výkon siete alebo náhla neschopnosť komunikovať s aplikáciou.

1.1.6. Analýza štatistických parametrov na vybraných DDoS útokoch

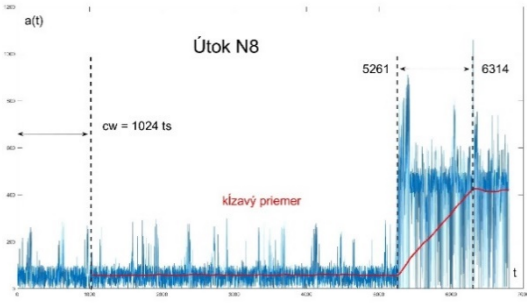
Na rôznych záznamoch DDoS útokov budeme prezentovať reakciu vybraných štatistických parametrov. Vynechali sme množstvo experimentov, ktoré sme realizovali na rôznych simulovaných scenároch. Na týchto scenároch vybrané štatistické parameter fungovali „vzorovo“ a my sme nadobudli prvotnú predstavu o efektívnosti použitia jednotlivých parametrov pri rozpoznaní zmeny a nárastu simulovanej prevádzky. Pri nasadení na

rozpoznanie reálneho útoku reakcie štatistických parametrov už neboli také jednoznačné. Pre prezentovanie sme vybrali osem rôznych záznamov DDoS útokov.

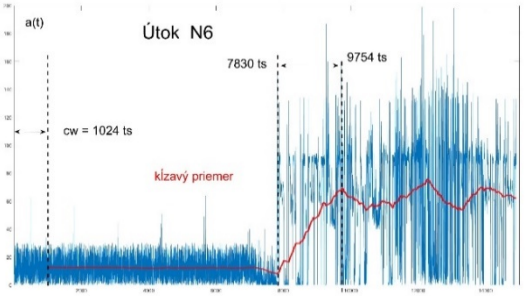
Prvé štyri záznamy predstavujú štandardné DDoS útoky (N-normal), pri ktorých sa parametre chovajú obdobne ako ma simulovaných dátach. Ďalšie dva záznamy predstavujú na prvý pohľad tiež štandardné útoky, avšak parametre sa pri nich chovajú podstatne odlišným neurčitým spôsobom (S-special). V poradí siedmy záznam je zástupcom tzv. Low Rate (LR) DDoS útokov, čiže útokov s nízkou priemernou intenzitou toku. Posledný záznam je útok, ktorý sa podľa nášho názoru nedá detegovať uvedenými štatistickými parametrami (P-ako problém). Pri záznamoch sme zároveň s počtom prírastkov paketov a(t) zobrazili aj ich kĺzavý priemer m(t) s výpočtovým oknom cw = 1024 ts.

Záznamy sú navzorkované do daných časových okien (ts-time slot), väčšinou ts = 1ms. Pre postupný výpočet parametrov sme použili posun tzv. výpočtového okna (cw – compute window) o jeden ts. Pri prevádzaní experimentov sme pre veľkosti výpočtových okien kvôli odhadu Hurstovho exponentu použili rôzne mocniny dvojky (256,512,1024,2048,4096). Po subjektívnom hodnotení priebehu hodnôt skúmaných parametrov sme sa rozhodli ďalej zaoberať hlavne s výpočtovým oknom o veľkosti cw = 1024 ts. Pri cw = 512 ts dochádzalo k významnému „rozvlneniu“ niektorých parametroch, pri cw = 2048 sa významne zväčšila doba výpočtu parametrov, hlavne Hurstovho exponentu.

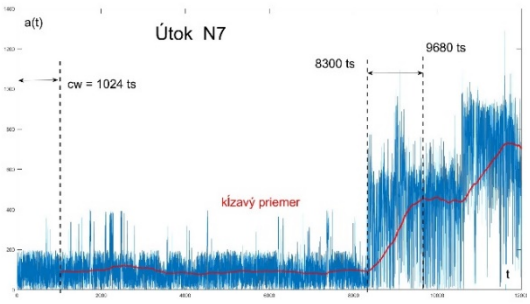
Na grafoch záznamov DDoS útokov sme zobrazili aj priebeh kĺzavého priemeru a zobrazili časový interval, v ktorom kĺzavý priemer m(t) nadobudol od začiatku útoku svoje lokálne maximum. Tento interval využijeme pri subjektívnom hodnotení ostatných štatistických koeficientov. Na všetkých záznamoch DDoS útokov môžeme pozorovať lineárny rast kĺzavého priemeru. Hľadáme koeficienty, ktorým pri začiatku útoku začnú rásť hodnoty rýchlejšie než lineárne.



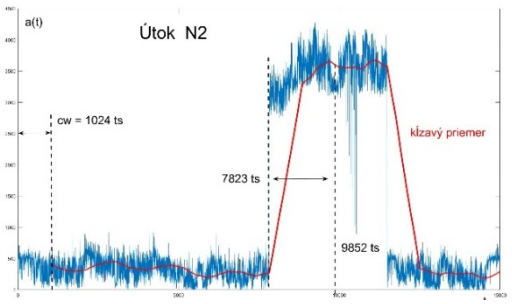
Obr. 3: Útok N8 a kĺzavý priemer



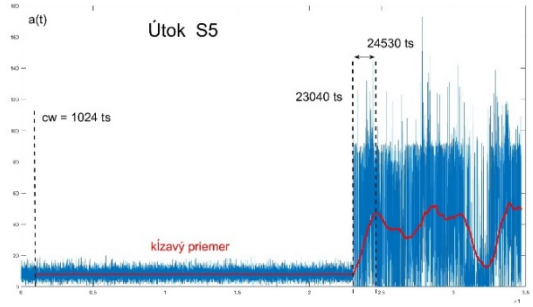
Obr. 4: Útok N6 a kĺzavý priemer



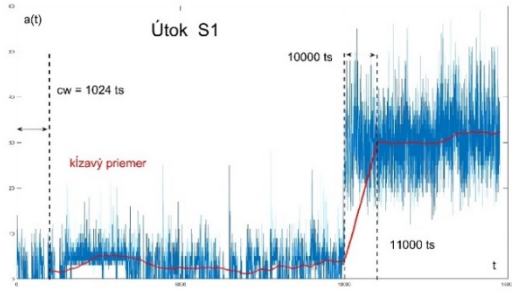
Obr. 5: Útok N7 a kĺzavý priemer



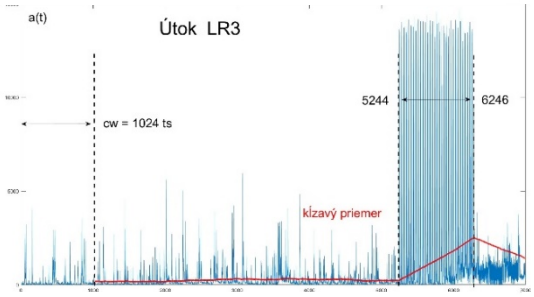
Obr. 6: Útok N2 a kĺzavý priemer



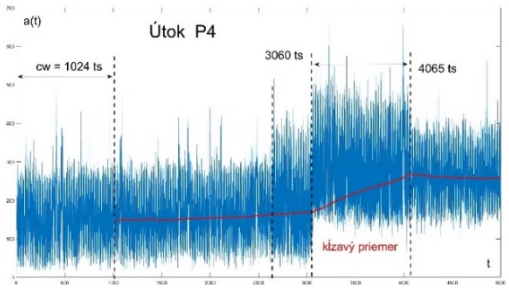
Obr. 7: Útok S5 a kĺzavý priemer



Obr. 8: Útok S1 a kĺzavý priemer



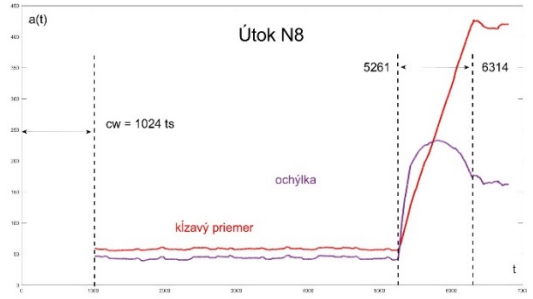
Obr. 9: Útok LR3 a kĺzavý priemer



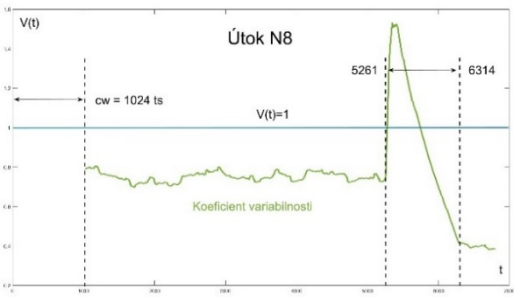
Obr. 10: Útok P4 a kĺzavý priemer

1.1.7. Variabilnosť, šikmosť a špicatosť

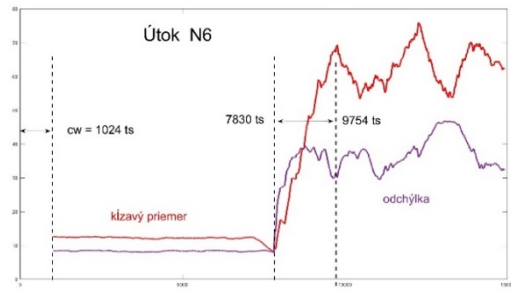
Ako prvými sa budeme zaoberať základnými štatistickými momentami, resp. koeficientami variabilnosti V(t), šikmosti K(t) a špicatosti Skw(t). Koeficient variabilnosti predstavuje podiel smerodajnej odchýlky a strednej hodnoty prírastkov v danom časovom okne cw. Variabilnosť vyjadruje mieru rozptylu prírastkov a(i) okolo strednej hodnoty. V prvom grafe zobrazíme priebeh kĺzavého priemeru a smerodajnej odchýlky, na druhom grafe sú hodnoty koeficientu variabilnosti:



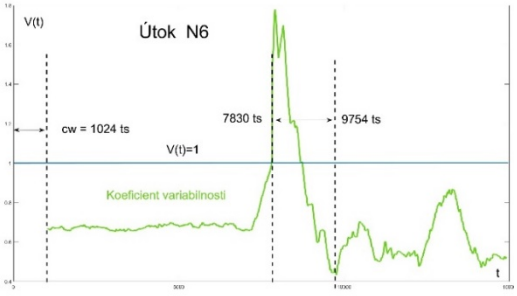
Obr. 11: Útok N8, priemer a odchýlka



Obr. 12: Útok N8, koeficient variabilnosti

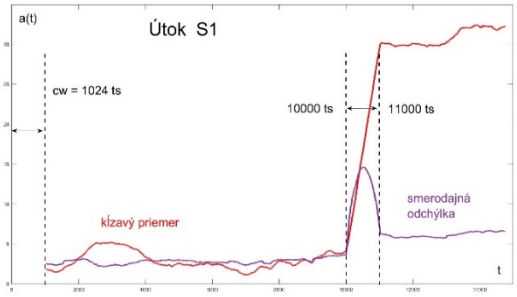


Obr. 13: Útok N6, priemer a odchýlka

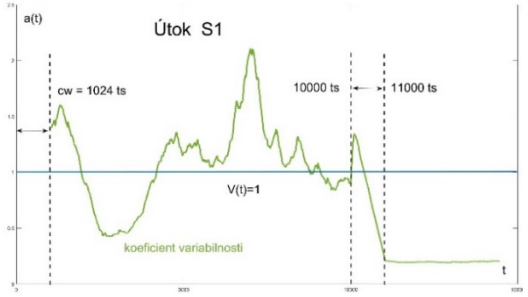


Obr. 14: Útok N6, koeficient variabilnosti

V prípade štandardných DDoS útokov (N2,N6,N7,N8) koeficient variabilnosti reagoval rovnako. Štandardná prevádzka predstavovala stacionárny tok s odchýlkou menšou ako stredná intenzita toku. Okamihu nástupu útoku odchýlka zareagovala rýchlejším nárastom hodnôt než bol lineárny rast kĺzavého priemeru, preto podiel odchýlky a priemeru nadobudol v okamihu útoku hodnotu V(tA) = 1. Tento fakt by sa dal využiť pre rýchlu a jednoduchú detekciu pri takomto type útoku. To ale nie je pravidlo, na obr.15 pri útoku S1 vidíme, že počas štandardnej prevádzky je narušená stacionarita a koeficient niekoľko krát pre útokom prekročil hodnotu V(tA) = 1.

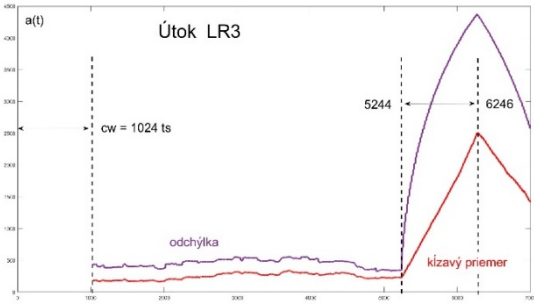


Obr. 15: Útok S1, priemer a odchýlka

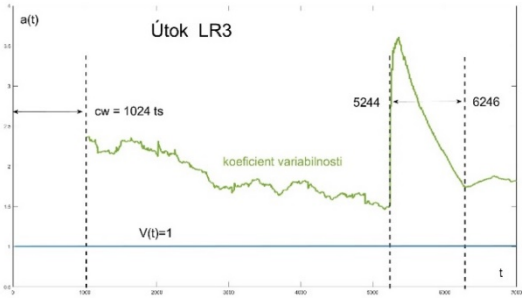


Obr. 16: Útok S1, koeficient variabilnosti

Na obr. 17 a 18 sú grafy priemeru, odchýlky a koeficientu variabilnosti pre tzv. Low Rate DDoS útok LR3.



Obr. 17: Útok LR3, priemer a odchýlka



Obr. 18: Útok LR3, koeficient variabilnosti

V prípade LR3 je počas bežnej prevádzky odchýlka väčšia než stredná intenzita toku, preto koeficient variabilnosti je počas celého priebehu V(t)>1. Hodnota 1 sa teraz nedá použiť pre rýchlu detekciu. Ale v okamihu útoku hodnoty odchýlky opäť narástli rýchlejšie než lineárny rast strednej intenzity, čo sa prejavilo na výraznom skokovom náraste koeficientu variabilnosti. Táto zmena by sa mala dať rozpoznať pomocou predikčných metód.

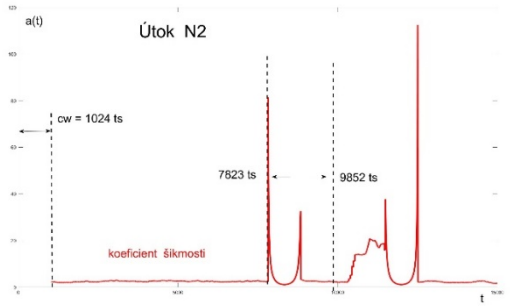
Ďalšími jednoduchými štatistickými parametrami sú koeficient šikmosti a koeficient špicatosti. Ide vlastne o tretí a štvrtý centrálny moment vyškálovaný smerodajnou odchýlkou.

$$\mu_3 = \frac{1}{\sigma^3} E[X - EX]^3 \qquad \mu_4 = \frac{1}{\sigma^4} E[X - EX]^4 \qquad \sigma^2 = E[X - EX]^2$$

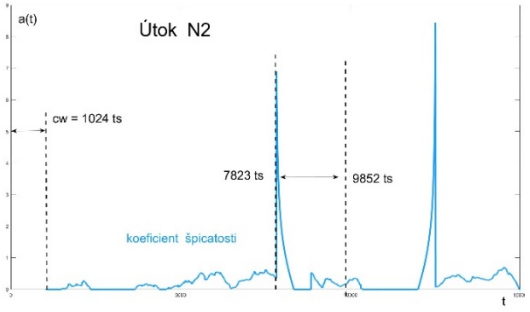
V našom prípade realizácie premennej X sú hodnoty prírastkov a(t) a koeficienty popisujú vlastnosti rozdelenia pravdepodobnosti prírastkov v danom výpočtovom okne cw. Pri postupnom načítavaní útočnej prevádzky do vyhodnocovaného okna cw sme predpokladali výraznú zmenu rozdelenia pravdepodobnosti a tým pádom aj zmenu koeficientov. Tento

predpoklad sa potvrdil skoro u všetkých analyzovaných tokov. Obidva štatistické koeficienty zareagovali výrazným nárastom svojich hodnôt hneď na začiatku DDoS útoku.

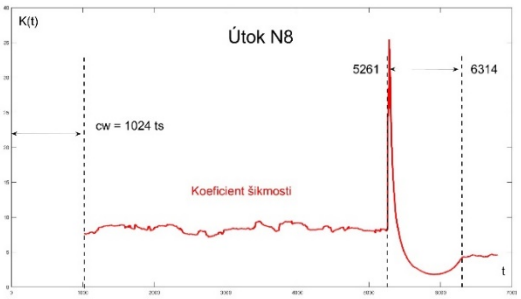
Ako prvé uvádzame grafy koeficientov aplikovaných na vybraných dvoch štandardných útokoch a následne niektoré problematické situácie.



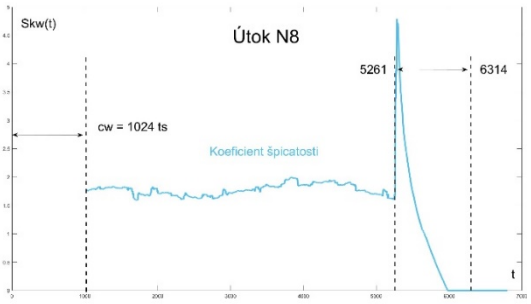
Obr. 19: Útok N2, koeficient šikmosti



Obr. 20: Útok N2, koeficient špicatosti



Obr. 21: Útok N8, koeficient šikmosti

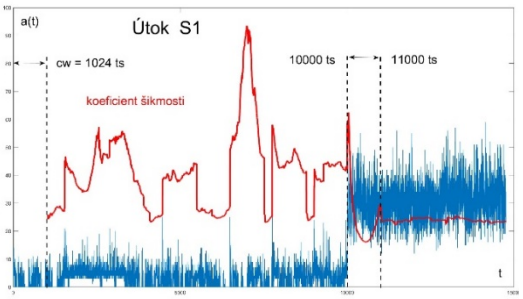


Obr. 22: Útok N8, koeficient špicatosti

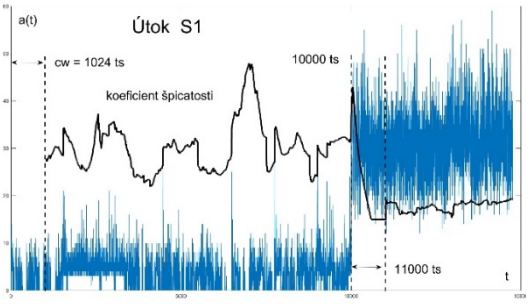
Pri štandardných útokoch typu N koeficienty šikmosti s špicatosti sa počas výpočtu na štandardnej prevádzke vďaka stacionárnosti len mierne vlnia, v okamihu nástupu DDoS útoku sa ich hodnoty výrazne zväčšia. Predpokladáme, že tento výrazný skok budeme vedieť strojovo rozpoznať pomocou vhodných predikčných metód.

Podstatne nepriaznivejšia situácia nastala pri neštandardných útokoch, napr. S1. Vďaka nestacionarite bežnej prevádzky a častému výskytu vysokých peakov obidva koeficienty silne kmitajú, vďaka čomu sa skok pri nástupe útoku stráca v celkovom priebehu toku hodnôt obidvoch koeficientov. Nástup DDoS útoku je v takomto prípade strojovo nerozpoznateľný.

Pre vizualizáciu reakcie koeficientov na zmenu charakteru bežnej prevádzky a na výskyt peakov sme zobrazili hodnoty koeficientov do spoločného grafu spolu s prírastkami toku S1.

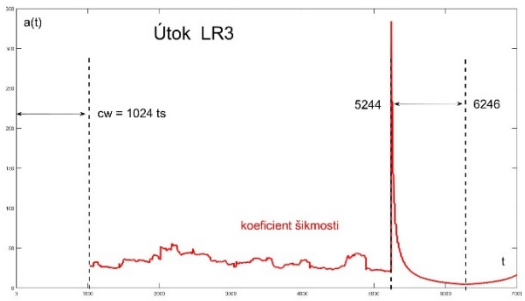


Obr. 23: Útok S1, koeficient šikmosti

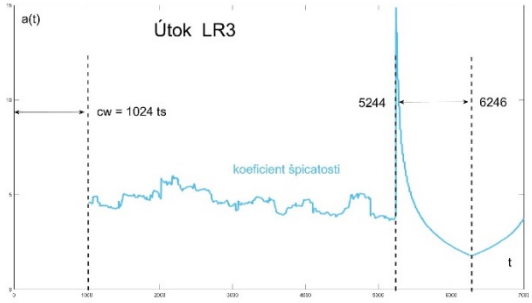


Obr. 24: Útok S1, koeficient špicatosti

Reakcia koeficientov pri útoku s nízkou strednou intenzitou LR3 dopadla rovnako pozitívne ako pri štandardných útokoch typu N.



Obr. 25: Útok LR3, koeficient šikmosti



Obr. 26: Útok LR3, koeficient špicatosti

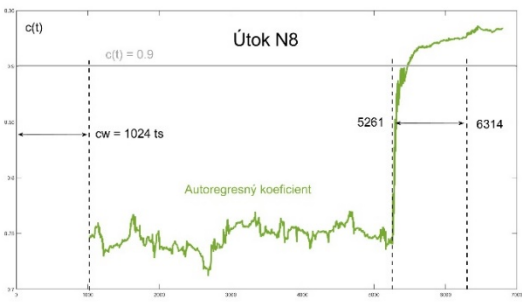
Opäť vidíme možnosť strojovo rozpoznať výrazný nárast hodnôt koeficientov pri začiatku útoku od hodnôt popisujúcich bežnú prevádzku.

1.1.8. Detekčné hodnoty pre autokoreláciu a regresiu

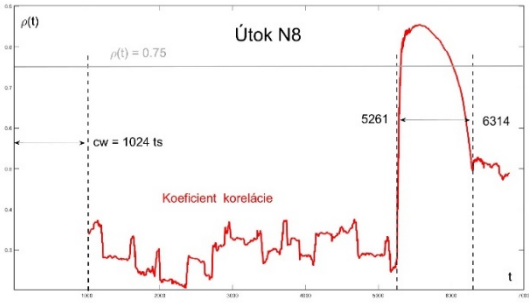
Ďalšie dva koeficienty popisujú autoregresnú a korelačnú závislosť medzi dvoma za sebou idúcimi cw oknami, resp. medzi náhodnými premennými X(t) a X(t-1), ktorých realizácie sú prírastky toku a(t) v uvažovaných oknách.

$$X(t) = c(t).X(t-1) + \varepsilon(t), \qquad \rho(t) = \frac{E[(X(t) - EX(t))(X(t-1) - EX(t-1))]}{\sigma(t)\sigma(t-1)}$$

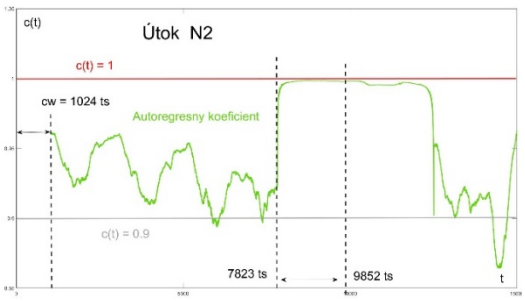
Veľkou výhodou tých koeficientov je, že ich hodnoty sa pohybujú v intervale (-1,1). V okamihu načítavania útoku do vyhodnocovaných cw okien pri väčšine útokoch začali ich hodnoty rýchlo rásť ku hornej hranici 1. Z tohto dôvodu sa je možne pre obidva koeficienty nastaviť určitú medznú hodnotu (threshold), ktorej prekročenie by signalizovalo začiatok DDoS útoku. Čím nižšia je táto hodnota, tým včasnejšia je signalizácia útoku, na druhú stranu moc nízke útoky môžu spôsobiť falošné hlásenia. Na základe prevedených experimentov a v snahe identifikovať čo najväčšie množstvo DDoS útokov bez falošných hlásení sme nastavili hranicu pre autoregresný koeficient na c(t) = 0.9 a korelačný koeficient na ρ(t) = 0.75. Na nasledujúcich grafoch sme opäť zobrazili vývoj koeficientov pre útoky N8, N2, S1 a LR3. Útok N8 predstavuje chovanie koeficientov pri väčšine štandardných útokov, na N2 vidíme štandardný útok na ktorom ale autokorelačný koeficient má všetky hodnoty väčšie ako 0.9 a detekcia sa nedá použiť. Pri neštandardnom útoku S1 obidva koeficienty pri útoku presiahli detekčné hladiny. Pre útok s nízkou intenzitou LR3 je táto metóda nepoužiteľná.



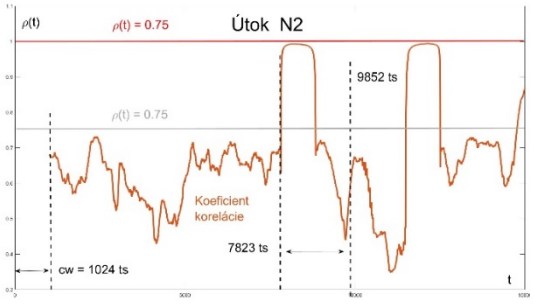
Obr. 27: Útok N8, Autoregresný koeficient



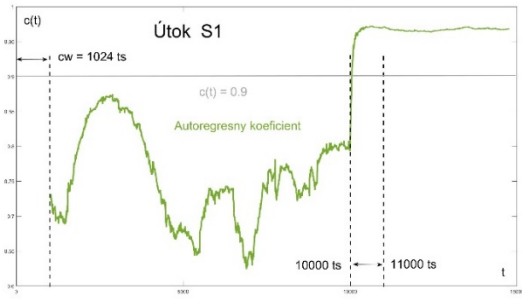
Obr. 28: Útok N8, koeficient Korelácie



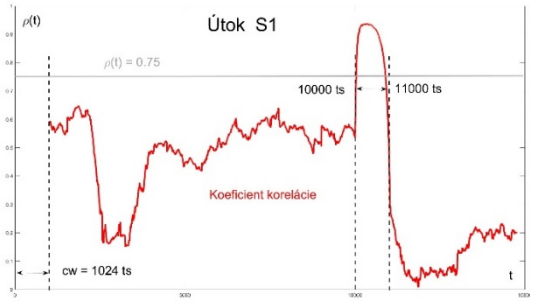
Obr. 29: Útok N2, Autoregresný koeficient



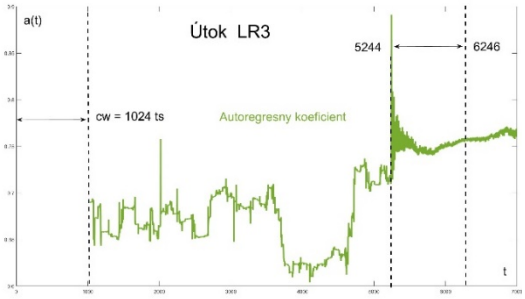
Obr. 30: Útok N2, koeficient Korelácie



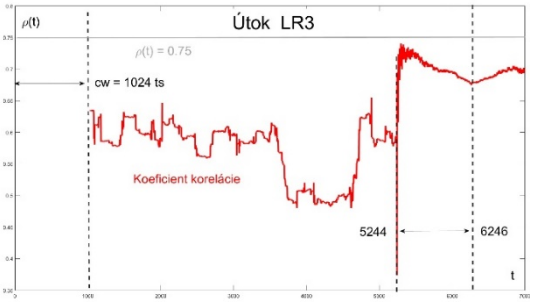
Obr. 31: Útok S1, Autoregresný koeficient



Obr. 32: Útok S1, koeficient Korelácie



Obr. 33: Útok LR3, Autoregresný koeficient



Obr. 34: Útok LR3, koeficient Korelácie

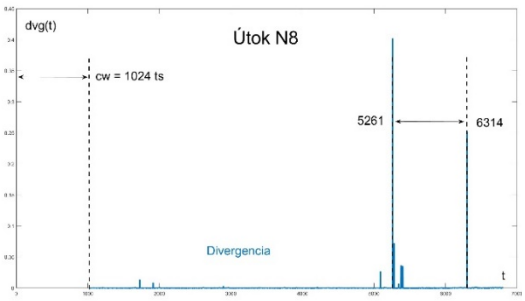
Pre väčšinu útokov sa dá predpokladať, že prekročenie detekčných hodnôt pri autoregresii a korelácii sa bude dať v kombinácii s ďalšími koeficientami využiť pre skorú jednoduchú detekciu záplavového útoku.

1.1.9. Reakcia divergencie na DDoS útok

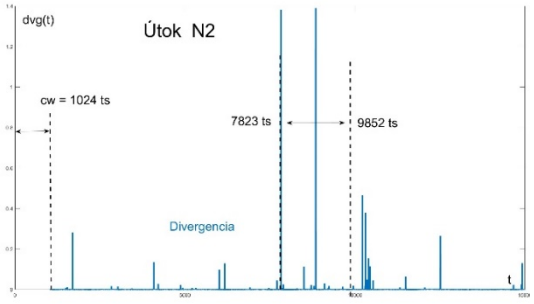
Posledným analyzovaným parametrom Kullback-Leibler divergencia istým spôsobom porovnáva rozdelenie pravdepodobnosti P(.) prírastkov a(t) v za sebou idúcich prekrytých výpočtových oknách cw.

$$dvg(t) = \sum_{k=1}^{n} P(t-1) ln(\mathrm{P(t-1)/P(t)})$$

Výhodou divergencie je, že na sledovaných vzorkách vo väčšine prípadoch reagovala okamžite na nástup DDoS útoku vysokým impulzom (peakom). Nevýhodou je, že mala tendenciu takto reagovať aj na peaky v bežnej prevádzke, čo by mohlo spôsobovať množstvo falošných hlásení.

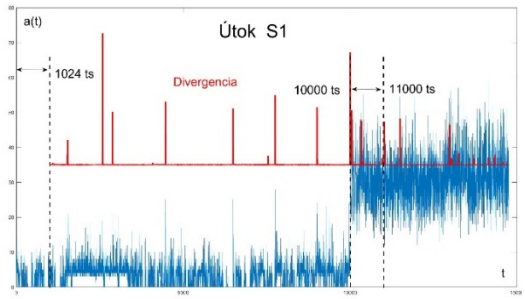


Obr. 35: Útok N8, Divergencia

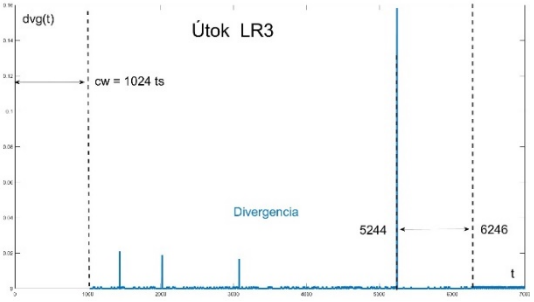


Obr. 36: Útok N2, Divergencia

Ďalšou nevýhodou použitia divergencie je fakt, že vysoký impulz pri nástupu útoku trvá veľmi krátko oproti ostatným parametrom, preto jej aplikáciu v kombinácii s ostatnými parametrami vidíme ako problematickú. Použitie divergencie sme preto odložili do budúcnosti.



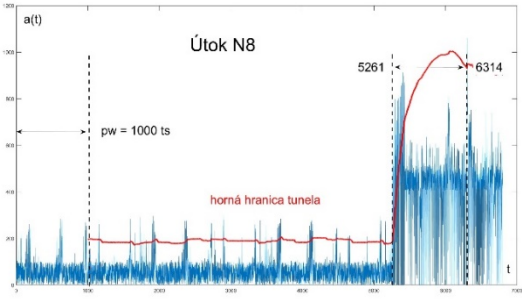
Obr. 37 Útok S1, Divergencia



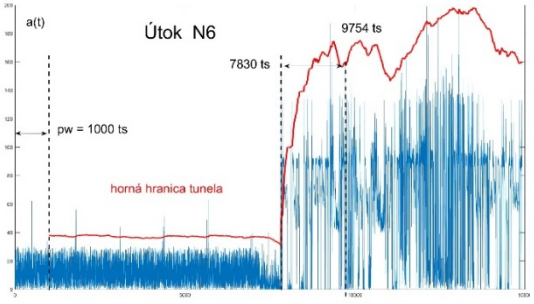
Obr. 38: Útok LR3, Divergencia

1.1.10. Využitie 3σ-Tunela pre rozpoznanie DDOS útokov

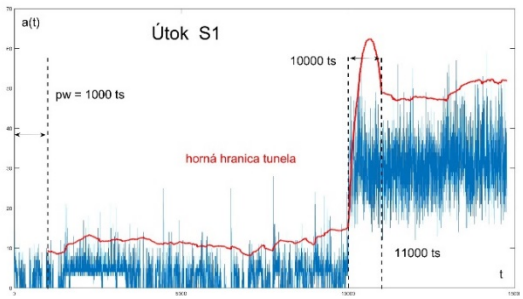
Na základe experimentov sme pre výpočet priemeru a odchýlky zvolili 1000 predchádzajúcich hodnôt, a šírku interval sme zvolili (EX-3σ, EX+3σ) (pri menšej veľkosti intervalu dochádzalo k častým falošným hláseniam). Túto metódu sme nazvali predikčný 3σ-tunel. Vzniká prirodzená otázka, prečo nepoužiť predikčný tunel priamo na prírastky IP toku a(t) a nezdržiavať sa výpočtom štatistických koeficientov. Po prevedení experimentov sme zistili, že síce prekročenie hornej hranice tunela deteguje relatívne skoro nástup DDoS útoku, ale zároveň dochádza k častému hláseniu falošného poplachu. Pri zväčšení násobnosti smerodajnej odchýlky sa síce počet falošných hlásení zmenšil ale zároveň sa zmenšila schopnosť detegovať útok. Na nasledujúcich obrázkoch si ukážeme na vybraných útokoch len hornú hranicu 3σ-Tunela.



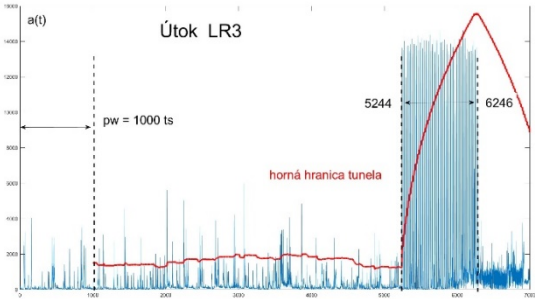
Obr. 39: Útok N8, horná hranica 3σ-Tunela



Obr. 40: Útok N6, horná hranica 3σ-Tunela

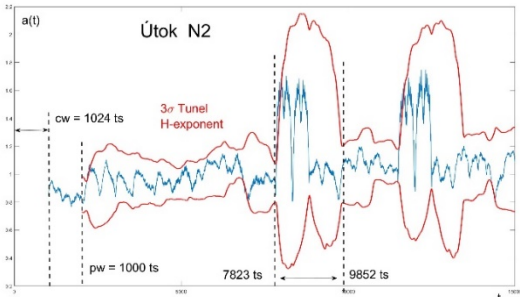


Obr. 41: Útok S1, horná hranica 3σ-Tunela

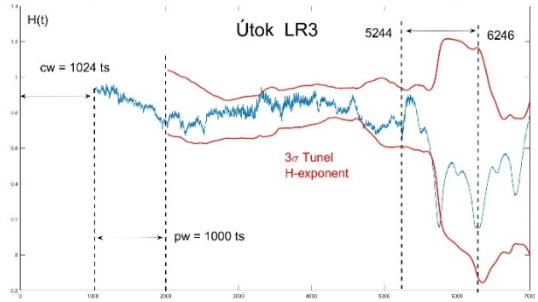


Obr. 42: Útok LR3, horná hranica 3σ-Tunela

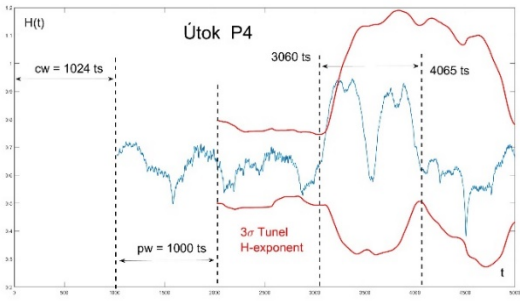
Na nasledujúcich grafoch budeme demonštrovať chovanie sa 3σ-Tunela na parametre s najväčšou variabilitou z pomedzi uvažovaných štatistických koeficientov, a to na Hurstovom exponente (odhadnutého RS štatistikou s odstránením autoregresie AR1).



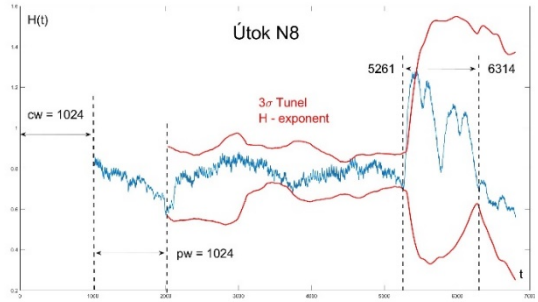
Obr. 43: N2, 3σ-Tunel pre H-exponent



Obr. 44: LR3, 3σ-Tunel pre H-exponent



Obr. 45: P4, 3σ-Tunel pre H-exponent



Obr. 46: N8, 3σ-Tunel pre H-exponent

Vo väčšine prípadoch (až na útok LR3) 3σ-Tunel skoro rozpoznal začiatok DDoS útok s minimálnym, niekedy až nulovým počtom falošných hlásení.

Ako sme už uviedli, na základe experimentov sme nastavili veľkosť predikčného okna na 1000 ts a šírku intervalu pre 1001 hodnotu na trojnásobok smerodajnej odchýlky. Pri takomto nastavení sme vo väčšine experimentov (rôzne koeficienty a rôzne útoky) získali skoré rozpoznanie začiatku útoku s minimálnym počtom falošných hlásení. Na predchádzajúcich obrázkoch sú to 3 prípady zo 4. Takto nastavený tunel má dve užitočné vlastnosti:

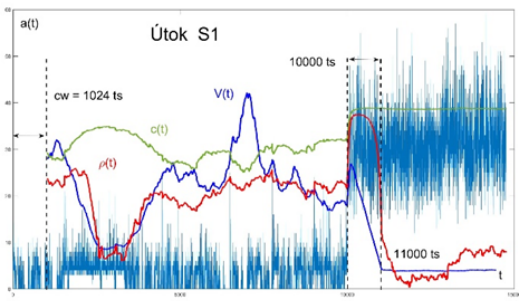
- Má dostatočne „dlhú pamäť“ na to, aby dokázal pokryť lokálnu variabilnosť štatistického koeficientu a nespôsoboval veľké množstvo falošných hlásení
- Má dostatočne „dlhú pamäť“ na to, aby nedokázal zareagovať na relatívne veľký nárast hodnôt skúmaného koeficient pri začiatku DDoS útoku (veľký vzhľadom na relatívnu variabilnosť)

Samozrejme, aby sme mohli vysloviť nejaké relevantnejšie závery a odporúčania, je potrebné previesť množstvo experimentov na záznamov rôznych DDoS útokov.

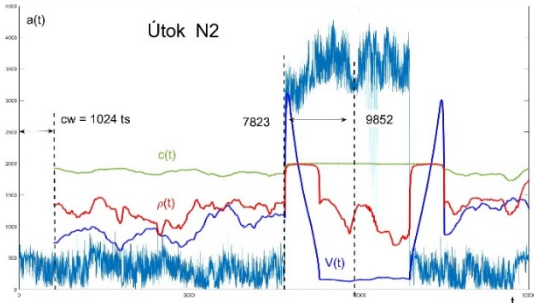
1.1.11. Detekčné funkcie pre strojové rozpoznanie DDoS útokov

V predchádzajúcich kapitolách sme uviedli dva spôsoby rozpoznania DDoS útokov pomocou štatistických koeficientov. Prvým spôsobom bolo určenie detekčných hodnôt pre vhodné parametre, 1,00 pre variabilnosť V(t) a H-exponent, 0,90 pre autoregresný koeficient c(t) a 0,75 pre korelačný koeficient ρ(t). Druhý spôsob pomocou predikčného tunela bol analyzovaný v predchádzajúcej časti.

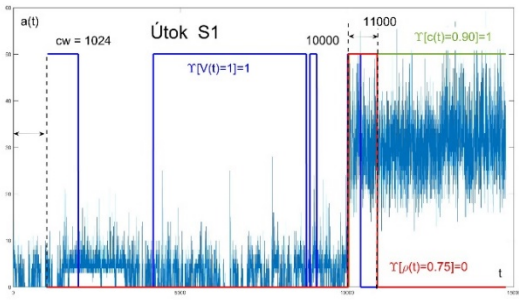
Pri využití hraničných hodnôt sme zaviedli dvojhodnotovú logickú funkciu Y[.], ktorá vyhodnocuje prekročenie hraničnej hodnoty pre daný štatistický koeficient. Nasledujúcich grafoch si ukážeme variabilnosť V(t), autoregresný koeficient c(t) a korelačný koeficient ρ(t) a priebeh ich funkcie Y[.] pre útoky S1a N2.



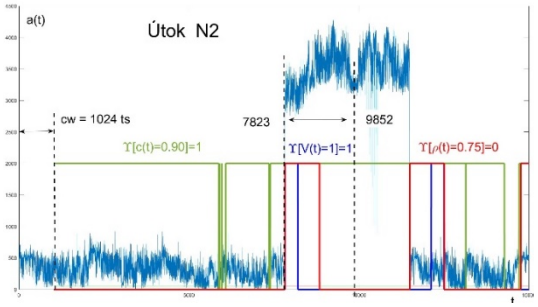
Obr. 47: S1, koeficienty V(t), c(t), ρ(t)



Obr. 48: N2, koeficienty V(t), c(t), ρ(t)



Obr. 49: S1, priebeh funkcií Y[.]

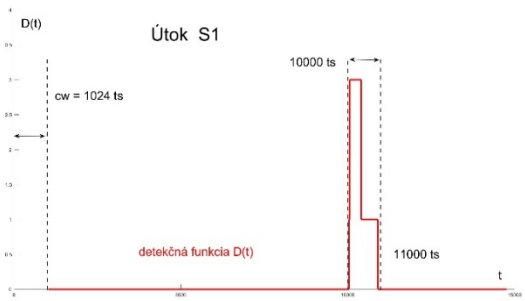


Obr. 50: N2, priebeh funkcií Y[.]

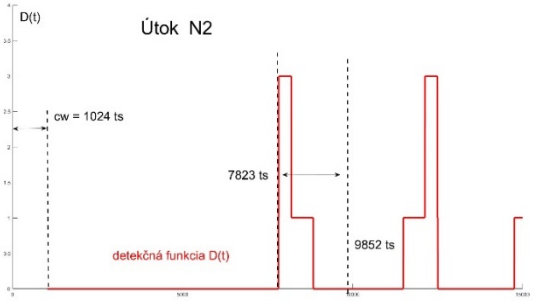
Množiny hodnôt, na ktorých koeficienty prekračujú svoje detekčné hranice, nazveme detekčné intervaly. Na obr. 49 a 50 vidíme, ako sa tieto intervaly navzájom prekrývajú. Aby sme eliminovali falošné hlásenia a zároveň dosiahli včasné rozpoznanie útokov, zavedieme pre sledované koeficienty novú detekčnú funkciu D(t), pričom pre zjednodušenie označíme napr. pre koeficient V(t) funkciu Y[V(t)>1,00] ako YV(t):

$$D(t) = Y_V(t).Y_C(t) + Y_V(t).Y_\rho(t) + Y_C(t).Y_\rho(t)$$

Na nasledujúcich obrázkoch vidíme priebeh detekčnej funkcie pre útoky S1 a N2.



Obr. 51: S1, priebeh detekčnej funkcie D(



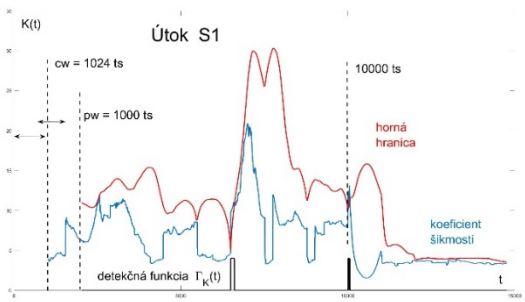
Obr. 52: N2, priebeh detekčnej funkcie D(t)

Vďaka tvaru funkcie D(t) sme eliminovali eventuálne falošné hlásenia a zároveň podľa hodnôt 1 alebo 3 vieme rozpoznať, a akom čase detegovali útok dva alebo všetky 3 štatistické parametre. V tabuľke 3 uvádzame vyhodnotenie použitia detekčnej funkcie D(t) pre analyzované útoky. Prvá hodnota predstavuje počet falošných hlásení, druhá určuje dobu, kedy detegovali útok minimálne dva koeficienty a tretia hodnota označuje dobu, kedy detegovali útok všetky tri štatistické koeficienty V(t), c(t) a ρ(t).

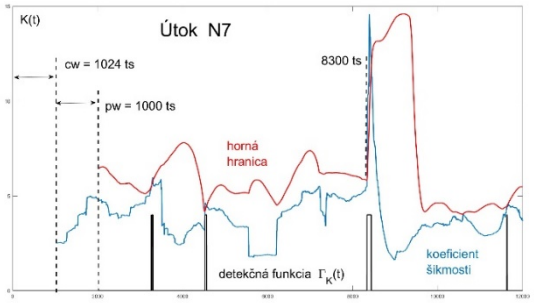
útok	S1	N2	LR3	S5	N6	N7	N8
D(t)	0 – 34 - 52	0 - 5 - 5	x	x	0 – 75 - 119	0 - 228 - 480	0 - 47 - 210

Tab. 3: Vyhodnotenie rozpoznania DDoS útoku pomocou detekčnej funkcie D(t)

Rozpoznanie útoku podľa hraničnej hodnoty bolo účinné hlavne pri štandardných útokoch, pričom vo viacerých prípadoch dva koeficienty prekročili svoje hraničné hodnoty podstatne skôr než všetky tri. Pri neštandardných útokoch ako LR3, P4 a S5 žiadne dva detekčné intervaly, ak vôbec vznikli, nemali spoločný prienik, a preto útok nebol rozpoznaný. Rovnakú metodiku budeme aplikovať aj na rozpoznanie útoku pomocou predikčného tunela. V predchádzajúcej kapitole sme určili, pri zohľadnení doby rozpoznania útoku a počtu falošných hlásení, prvé tri najefektívnejšie štatistické koeficienty: šikmosť K(t), špicatosť Skw(t) a variabilnosť V(t). Keďže sme vybrali koeficienty, ktorým sa pri nástupe DDoS útoku výrazne zväčšia hodnoty, preto sa budeme zaoberať len hornou hranicou predikčného tunela. Dobu prekročenia hornej hranice vyhodnotíme ako detekčný interval a popíšeme ho dvojhodnotovou funkciou Γ[.], pričom napr. pre koeficient špicatosti označíme ΓK(t) = Γ[K(t)>Hr(t)], kde Hr(.) je horná hranica predikčného intervalu pre koeficient K(.).



Obr. 53: S1, priebeh funkcie Γ(t)



Obr. 54: N7, priebeh funkcie Γ(t)

Podobne ako pri predchádzajúcej metóde, pri rozpoznávaní pomocou hraničnej hodnoty, aj teraz zavedieme detekčnú funkciu, ktorá bude kombináciou identifikátorov Γ(t):

$$DP(t) = ΓK(t).ΓSkw(t) + ΓK(t).ΓV(t) + ΓSkw(t).ΓV(t)$$

V tabuľke tab.2 uvádzame vyhodnotenie použitia detekčnej funkcie DP(t).

útok	S1	N2	LR3	S5	N6	N7	N8
$D_P(t)$	1 – 6 - 28	0 - 1 - 1	1 - 1 - 1	1 - 18 - 41	1 - 12 - 12	1 - 39 - 49	0 - 1 - 1

Tab. 4: Vyhodnotenie rozpoznania DDoS útoku pomocou detekčnej funkcie D(t)

1.1.12. Zhodnotenie testov

Na základe doteraz prevedených experimentoch vidíme využitie vybraných štatistických koeficientov má podstatne väčší potenciál pri rýchlom rozpoznaní DDoS útokov (FAD – Fast Attack Detection) než štandardné sledovanie priemernej a špičkovej intenzity. Pri DDoS útoku je silný predpoklad, že sa tieto intenzity výrazne zvýšia. Avšak priemerná intenzita rastie len lineárne v závislosti od veľkosti kĺzavého okna, čo vedie k neskorej detekcii útoku, a rozpoznávanie útoku pomocou nastavenia nejakého tresholdu pre špičkovú rýchlosť môže viesť k častým falošným hláseniam vďaka výskytu píkov v IP toku.

Pri prechode štandardnej IP prevádzky do DDoS útoku sa výrazne zmení nie len priemerná a špičková intenzita, a zmení sa celková pravdepodobnostná štruktúra záplavového toku. Na zachytenie tejto zmeny sme využili prezentované štatistické koeficienty. Tie môžeme rozdeliť na dva typy, koeficienty, ktoré pri nástupu útoku prekročili nejakú detekčnú hodnotu a koeficienty, ktoré na útok reagovali výraznou zmenou svojich hodnôt resp. impulzom. Niektoré koeficienty dokonca patria do obidvoch skupín. Na základe tohto delenia sme vytvorili:

- Metódu použitia detekčných hraníc (MDH)
- Metóda použitia predikčných tunelov (MPT)

Porovnaním tabuliek 3 a 4 vidíme, že MPT rozpoznala každý uvažovaný typ útoku a to podstatne skôr ako MDH, avšak pri väčšine útokov signalizovala jedno falošné hlásenie. Predpokladáme, že tento problém by sa mohol odstrániť viacerými spôsobmi, napríklad pridaním ďalšieho vhodného detekčného koeficientu, ale nejakým typom vyhladenia hodnôt koeficientu pred použitím predikčného tunela, napr. exponenciálnym alebo Holtovým vyhladením.

1.3. Detekcia škodlivého kódu (Malware)

Malvér, skratka pre škodlivý softvér, zahŕňa rôzne typy škodlivého softvéru vrátane vírusov, červov, trójskych koní a ransomvéru. Každý typ má jedinečné vlastnosti a metódy infekcie a poškodenia. Vírusy sa napríklad pripájajú k čistým súborom a nekontrolovateľne sa šíria, čím poškodzujú súborové systémy. Červy sa naopak samy replikujú bez ľudského zásahu a často zneužívajú zraniteľnosti siete. Trójske kone klamú používateľov tým, že sa maskujú ako legitímny softvér, čím získavajú neoprávnený prístup k systémom.

Ransomvér predstavuje obzvlášť hrozivú formu malvéru. Šifruje údaje obete a požaduje výkupné za dešifrovací kľúč. Tento typ malvéru môže spôsobiť značné prevádzkové narušenia a finančné straty. Dopad malvéru sa neobmedzuje len na jednotlivých používateľov; rozširuje sa aj na veľké organizácie a môže viesť k masívnym únikom údajov, finančným stratám a poškodeniu reputácie.

Prevencia a zmierňovanie útokov malvéru zahŕňa kombináciu robustného antivírusového softvéru, vzdelávania používateľov a dodržiavania osvedčených postupov v

oblasti kybernetickej bezpečnosti, ako sú pravidelné aktualizácie softvéru a opatrné sťahovanie súborov alebo aplikácií z internetu. Pochopenie rôznych foriem škodlivého softvéru je kľúčové pre účinné opatrenia v oblasti kybernetickej bezpečnosti, pretože umožňuje jednotlivcom a organizáciám rozpoznať potenciálne hrozby a podľa toho reagovať.

Zo širšieho spektra malvéru sme sa rozhodli zamerať v našich testoch len na typ ***„trójsky kôň“***. Testy boli vykonané na voľne dostupnom datasete *Trojan Detection* pre nasledovné metódy:

- Random Forest (náhodný les)
- AdaBoost
- Naive Bayes
- neurónová sieť typu FNN

1.1.13. Príprava a postup pri testovaní

Práca s modelmi pozostávala z nasledujúcich krokov.

1. Predspracovanie dát (data preprocessing)
   - Nahratie datasetu pre detekciu trojanov pomocou knižnice pandas.
   - Čistenie dát (data cleaning): vynechanie nerelevantných a duplicitných záznamov, podľa potreby doplnenie chýbajúcich dát.
   - Kódovanie kategorických dát (label encoding): prevod kategorických dát ako sú identifikátor toku (Flow ID), zdrojová a cieľová IP adresa (Source IP, Destination IP), časová pečiatka (Timestamp) – na numerické hodnoty.
   - Kódovanie návestí tried (Class Label Encoding): Prevod označenia triedy ("Trojan" a "Benign") na binárny tvar (1 resp 0).
   - Výber príznakov (Feature Selection): Použiť vzájomné informácie na výber príznakov, ktoré sú najdôležitejšie pre klasifikáciu
2. Rozdelenie dát na trénovacie a testovacie
   - Rozdelenie dát na trénovaciu a testovaciu množinu (typicky býva 80% trénovacích a 20% testovacích dát). Na rozdelenie sa použila funkcia using train_test_split z knižnice scikit-learn.
3. Výpočty algoritmami strojového učenia (Machine Learning): Náhodný les (Random Forest), AdaBoost (adaptive boosting), Gaussian Naive Bayes a Hlboké učenie s doprednou neurónovou sieťou. Nasleduje popis použitých algoritmov strojového učenia.

Analýza a vyhodnocovanie testov prebiehalo rovnako ako v prípade predchádzajúcich typov kybernetických hrozieb na základe definovaných metrík ako *Precision, Recall, F-1 Score, Accuracy* kvantifikovaných z *konfúznej matice*. Samotné výpočty prebiehali v testovacom prostredí Google Colab.

1.1.14. Výsledky testov

Výsledky testov sú prehľadne zobrazené v tabuľke 5.

Metóda	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Processing time (sec.)
Random Forest classifier (50 estimators)	99.99	0.999967	0.999967	0.999967	9 sek
skúšať rôzne počty stromov: 10	99.99				ďalšie spolu 60 sek.

20	99.9934				
30	99.9967				
40	100.000				
50	99.9934				
60	99.9967				
70	99.9967				
80	99.9967				
90	99.9934				
AdaBoostClassifier (100 estimators)	99.9639				20 sek.
cross-validation to obtain a more robust estimate	99.9626				60 sek.
GaussianNB Classifier					20
gnb.score	91.4984				
cross_val_score	90.1080				
using reciver operating characteristic score	92.1831				
weighted avg		92.1035	91.4985	91.4487	
Deep learning (Feed forward neural network) – 10 epoch	98.3947				55 sek. samotný trénning
weighted avg		98.3953	98.3947	98.3947	

Tab. 5: Výsledky testov pre detekciu malware

1.1.15. Zhodnotenie testov

Na základe vykonaných experimentov možno zhodnotiť, že najvhodnejšie metódy pre detekciu malvéru typu „trojský kôň“ z prevádzky v sieti sú metódy

- Random Forest: úspešnosť 99,99%, čas vyhodnocovania ~9 sek.
- AdaBoost: úspešnosť 99,96%, čas vyhodnocovania ~19 sek.
- neurónová sieť typu FNN: úspešnosť 98,39%, čas vyhodnocovania ~15 sek.

Z nich Random Forest vykonala výpočty dvakrát rýchlejšie. Ďalšia v poradí neurónová sieť typu FNN (Feed forward neural network) má ešte vysokú úspešnosť (98,39%) ale čas trénovania siete bol dlhší (55 sek.). Najrýchlejšou metódou bola GaussianNB Classifier – avšak dosiahla najhoršie výsledky (91.49%). Za najvhodnejšou metódou na detekciu malvéru typu „trojský kôň“ analyzovaním prevádzky považujeme algoritmus Random Forest.

1.4. Detekcia phishing

Phishing je klamlivá kybernetická taktika, ktorú útočníci používajú na získanie citlivých informácií od nič netušiacich obetí. Táto metóda zvyčajne zahŕňa odosielanie podvodných e-mailov alebo správ, ktoré napodobňujú legitímne zdroje, ako sú banky, vládne agentúry alebo známe spoločnosti. Cieľom je oklamať príjemcov a prinútiť ich prezradiť osobné údaje, ako sú prihlasovacie údaje, čísla kreditných kariet alebo iné citlivé informácie.

Tieto útoky môžu byť vysoko sofistikované a využívajú techniky sociálneho inžinierstva na vytvorenie pocitu naliehavosti alebo strachu, čo núti obete rýchlo reagovať bez spochybňovania pravosti žiadosti. Phishingové správy často obsahujú odkazy na falošné webové stránky, ktoré sa veľmi podobajú legitímnym, kde obete nevedomky zadávajú svoje citlivé informácie.

1.1.16. Príprava a postup pri testovaní

Testovanie prebiehalo na datasete, ktorý pozostával zo súborom s 57 000 logov, ktoré boli získané z prevádzky na FRI UNIZA, pôvodne obsahujúcim približne 160 premenných, z ktorých bolo identifikovaných 41 relevantných pre identifikáciu phishingu. Viaceré z týchto premenných boli kategorické, a preto bolo nevyhnutné ich predspracovať kombináciou one-hot encodingu a hashing tricku, čím sa dataset rozšíril na 2 009 premenných vhodných pre klastrovanie.

Pred klastrovaním boli všetky premenné štandardizované pomocou StandardScaleru, aby sa odstránili rozdiely v rozsahu hodnôt. Následne bola aplikovaná redukcia dimenzie PCA na 50 komponentov, čo umožnilo efektívnu prácu s algoritmami klastrovania pri zachovaní variability dát.

Tento krok bol nevyhnutný a časovo náročný, pričom má zásadný význam pre prípravu modelu strojového učenia. Logy pochádzajúce z rôznych systémov totiž obsahujú odlišné premenné, a preto bude pri tvorbe finálneho modelu na identifikáciu phishingu veľkou výzvou vyvinúť robustný skript, schopný spracovať aj neočakávané alebo nové premenné.

Zo širšieho spektra metód a algoritmov AI/ML, ktoré sa používajú pre detekciu phishing-u, sme sa rozhodli otestovať nasledovné klastrovacie metódy:

- K-means
- HDBSCAN

1.1.17. Výsledky testov

k-means

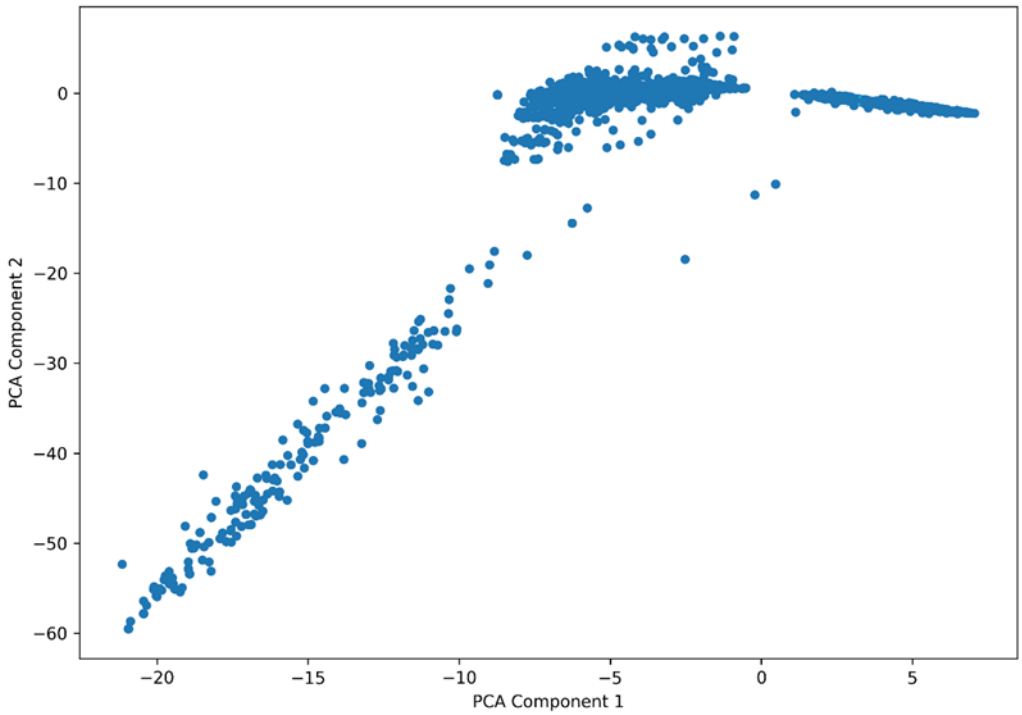
Pri použití algoritmu K-Means na náš dataset logov sme najprv predpokladali 3 klastre. Výsledné rozdelenie bolo veľmi nerovnomerné:

Cluster 0: 57 695 bodov

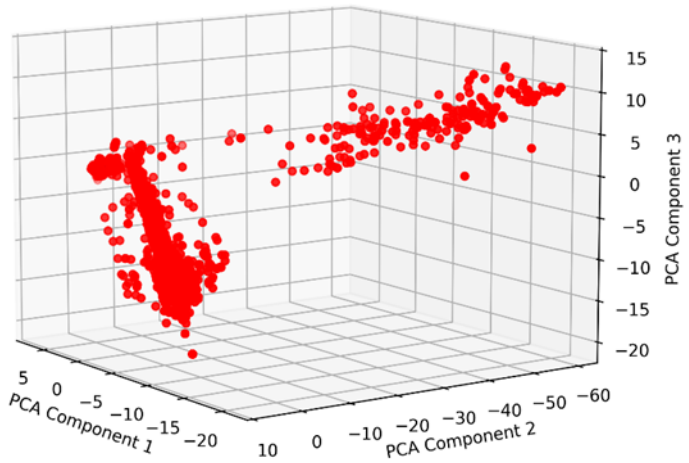
Cluster 1: 2 body

Cluster 2: 24 bodov

Výsledky K-Means sú vizualizované na obrázkoch 55 a 56.



Obr. 55: 2-D vizualizácia k-means



Obr. 56: 3-D vizualizácia k-means

Zmena predpokladaného počtu klastrov (napr. na 4 alebo 5) nepriniesla lepšie rozdelenie dát. Niektoré nové klastre obsahovali len 1–2 body, pričom hlavný klaster stále zachytával väčšinu dát. Toto nerovnomerné rozdelenie odráža prirodzenú štruktúru datasetu, kde je väčšina logov bežného typu a iba veľmi malé množstvo logov predstavuje podozrivé alebo phishingové udalosti.

K-Means priradí každý bod do nejakého klastru, bez možnosti označiť body ako body ležiace mimo zhluku alebo klastru (angl. outlier). V dôsledku toho sú extrémne alebo neštandardné hodnoty automaticky zahrnuté do niektorého z klastrov, často do malých alebo marginálnych klastrov. Tento fakt môže znížiť schopnosť K-Means samostatne detegovať netypické alebo podozrivé logy.

Celkovo sa ukazuje, že K-Means je vhodný pre rýchle a deterministické rozdelenie dát do predom stanoveného počtu klastrov, avšak jeho schopnosť zachytiť jemnejšie rozdiely alebo identifikovať odchýlky v logoch je obmedzená. Nerovnomerné rozdelenie v tomto prípade korešponduje s prirodzenou nerovnováhou dát, kde je počet phishingových logov veľmi nízky.

HDBSCAN

Algoritmus HDBSCAN bol aplikovaný na dataset logov s parametrami min_cluster_size=3000 a min_samples=20. Tieto nastavenia umožnili identifikovať 6 hlavných klastrov a zároveň vyznačiť odľahlé hodnoty (outliery), ktoré nepripadali do žiadneho klastru.

Výsledné rozdelenie počtu logov v jednotlivých klastroch bolo nasledovné:

Outliery (label = -1): 14 913 bodov

Cluster 0: 13 493 bodov

Cluster 1: 6 702 bodov

Cluster 2: 7 704 bodov

Cluster 3: 7 891 bodov

Cluster 4: 3 363 bodov

Cluster 5: 3 655 bodov

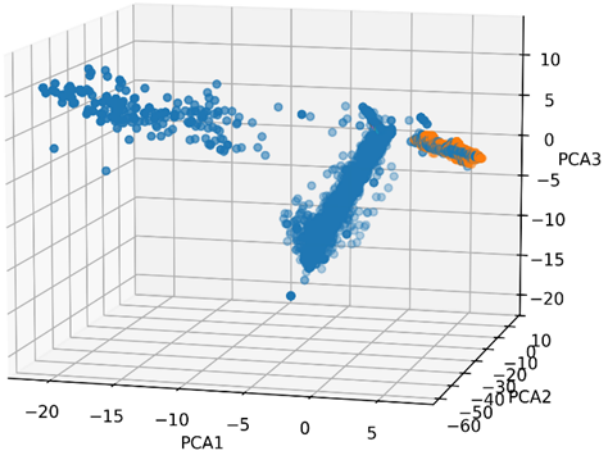
Interpretácia výsledkov

1) Outliery (-1):
   - algoritmus označil 14 913 logov ako odľahlé, ktoré nepatrili do žiadneho hustého klastru
   - tieto logy pravdepodobne obsahujú netypické alebo potenciálne podozrivé udalosti, čo je výhodou pre detekciu phishingu alebo anomálií

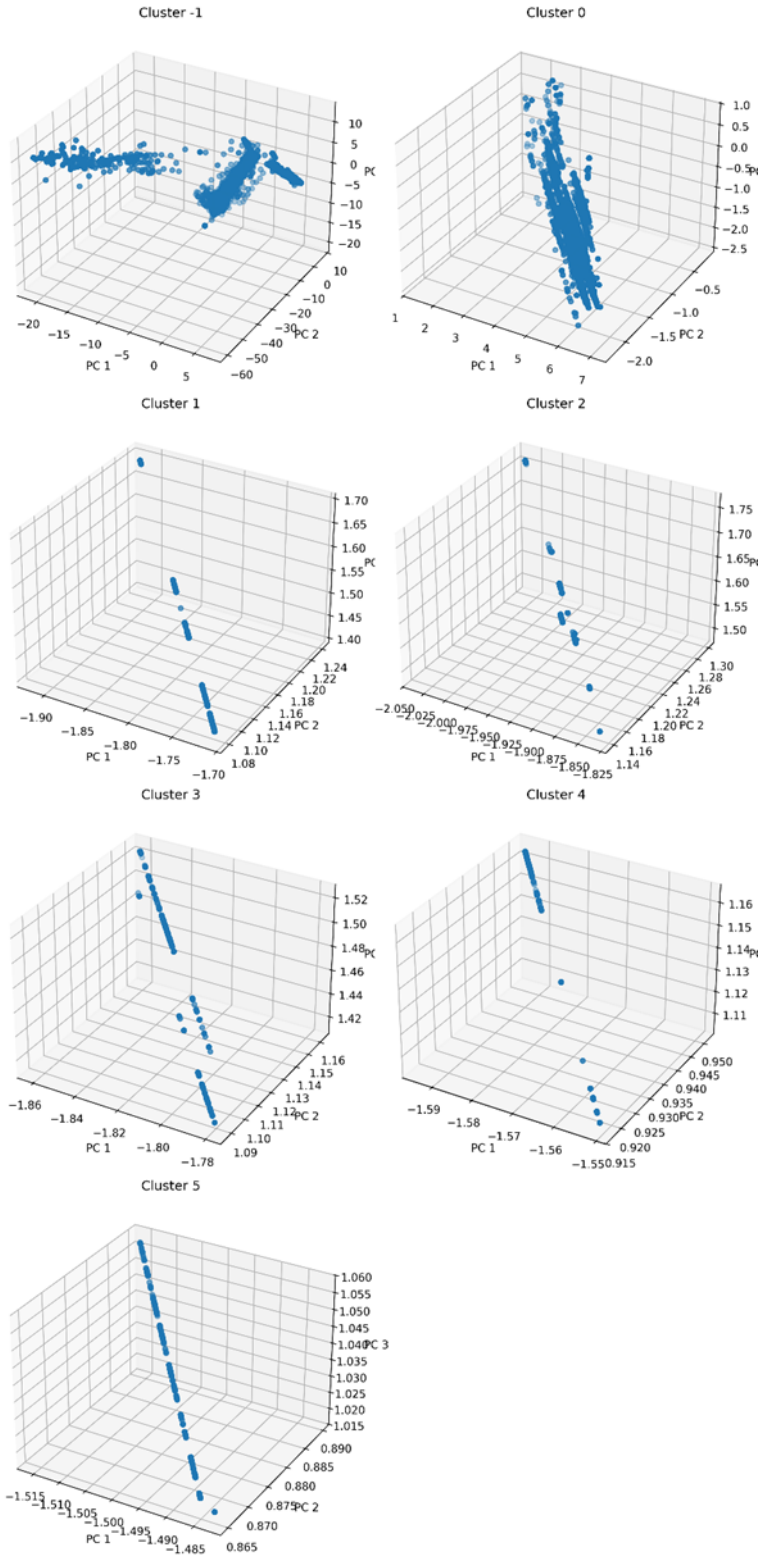
2) Hlavné klastre (0–5):
   - klastre sú pomerne rovnomerne zastúpené, pričom najväčšie klastre obsahujú 13 493 a 7 891 bodov a menšie klastre 3 363 a 3 655 bodov
   - toto rovnomernejšie rozdelenie v porovnaní s K-Means umožňuje lepšie zachytiť jemnejšie rozdiely medzi logmi a identifikovať štruktúru hustoty v dátach

3) Význam parametrov:
   - min_cluster_size=3000 zabezpečilo, že vznikajú len významné klastre, ktoré obsahujú dostatočný počet bodov
   - min_samples=20 určil citlivosť na hustotu a prispel k detekcii outlierov

Výsledky HDBSCAN sú vizualizované na obrázkoch 57 a 58.



Obr. 57: 3-D vizualizácia HDBSCAN



Obr. 58: 3-D vizualizácia jednotlivých klastrov pri HDBSCAN

Na obrázku 58 vidíme, že logy priradené do jednotlivých klastrov sú rozložené v rovnobežných rovinách, zatiaľ čo logy označené ako odľahlé hodnoty (outliery) sa nachádzajú mimo týchto hlavných rovín. Tento vizuálny efekt naznačuje, že HDBSCAN dokáže efektívne rozlíšiť typické správanie od netypických alebo podozrivých udalostí, na rozdiel od K-Means, ktorý priraďuje každý bod do nejakého klastra.

V tomto kontexte predstavuje využitie vzdialenosti logu od roviny hlavného klastra ako pravdepodobnostného skóre phishingu nový a inovatívny spôsob interpretácie dát. Tento prístup umožňuje kvantifikovať mieru odchýlky jednotlivých logov a potenciálne identifikovať podozrivé alebo škodlivé udalosti s jemnejšou granularitou než bežné binárne označenie outlierov.

Zároveň je potrebné tento koncept v rámci ďalšieho výskumu overiť, aby sa potvrdila jeho účinnosť pri automatizovanej detekcii anomálií a phishingových útokov, a aby bolo možné vyvinúť robustný systém, ktorý kombinuje vzdialenostné skóre s ďalšími charakteristikami logov (napr. čas, IP adresy, typ udalosti) pre presnejšiu identifikáciu podozrivých aktivít.

1.1.18. Zhodnotenie testov

HDBSCAN sa preukázal ako vhodný nástroj pre analýzu logov s nerovnomernou distribúciou, keďže dokáže:

- identifikovať husté oblasti logov
- označiť odľahlé hodnoty (outliery), ktoré môžu predstavovať phishing alebo iné podozrivé udalosti
- zachovať rovnomernejšie rozdelenie klastrov v porovnaní s K-Means, čo uľahčuje interpretáciu výsledkov a ďalšiu analýzu
- nový a inovatívny prístup: ak je nový log priradený do niektorého klastru, zvyčajne nejde o anomáliu ani potenciálny útok. Ak však log nie je zaradený do žiadneho klastra, jeho miera závažnosti sa môže kvantifikovať pomocou vzdialenosti od najbližšieho klastru, čím vzniká pravdepodobnostné skóre phishingu. Tento prístup umožňuje jemnejšie kvantifikovať riziko podozrivých udalostí a podporuje automatizovanú detekciu anomálií.

Celkovo HDBSCAN poskytuje robustný a flexibilný prístup ku klastrovaniu sieťových logov, najmä pri veľkých a heterogénnych datasetoch, pričom nový koncept pravdepodobnostného skórovania outlierov predstavuje sľubnú oblasť pre ďalší výskum a overenie jeho účinnosti.

K-Means je vhodný, ak je potrebný rýchly a deterministický rozklad, avšak nedeteguje outliery a nerovnomerne rozdeľuje body v nerovnomerne hustých dátach.

HDBSCAN je vhodný pre robustnú analýzu logov, pretože dokáže detegovať husté oblasti aj odľahlé hodnoty, čo je významnou výhodou pri identifikácii phishingu a iných anomálií v sieťových logoch. Na rozdiel od algoritmov, ktoré priraďujú každý bod do klastru, HDBSCAN umožňuje označiť netypické alebo odľahlé logy (outliery), ktoré nespĺňajú kritériá hustoty potrebné pre vznik klastra. Týmto spôsobom algoritmus efektívne filtruje šum a sústredí sa na logy, ktoré reprezentujú opakovateľné a významné vzory správania.

V rámci pilotného testovania sa HDBSCAN preukázal ako vhodnejší na zachytenie jemnejších rozdielov medzi logmi, ktoré by mohli byť indikátormi podozrivých alebo neštandardných udalostí. Algoritmus dokáže rozdeliť dáta do klastrov podľa hustoty bodov, pričom zachováva štruktúru hierarchie a umožňuje identifikovať menšie, ale významné skupiny logov, ktoré by pri použití klasického K-Means mohli byť „stratené“ v masívnych klastroch.

Hlavnou prednosťou HDBSCAN voči K-Means je teda jeho schopnosť:

- automaticky určiť počet klastrov podľa hustoty dát
- robustne identifikovať outliery, ktoré môžu indikovať anomálie alebo phishingové aktivity
- zachovať jemné rozdiely medzi logmi, čo zlepšuje interpretáciu a analytickú hodnotu výsledkov.

Celkovo HDBSCAN poskytuje flexibilný a robustný prístup k analýze logov, ktorý zohľadňuje prirodzenú variabilitu dát a umožňuje efektívnu detekciu podozrivých udalostí bez nutnosti predom stanoveného počtu klastrov.

1.5. Detekcia pokročilých perzistentných hrozieb APT

Pokročilé pretrvávajúce hrozby (APT) predstavujú kategóriu kybernetických útokov, pri ktorých neoprávnený používateľ získa prístup k sieti a zostane dlhší čas nezistený. Tieto útoky sú zvyčajne zamerané na krádež údajov alebo monitorovanie sieťovej aktivity, a nie na spôsobenie okamžitých škôd. APT sú často sponzorované národnými štátmi alebo veľkými zločineckými organizáciami, čo ich robí vysoko sofistikovanými a náročnými na odhalenie a odstránenie.

APT zahŕňajú zdĺhavý, nenápadný a zložitý proces infiltrácie. Útočníci používajú rôzne metódy na získanie vstupu, vrátane spear-phishingu, zneužívania zraniteľností alebo používania škodlivého softvéru. Po vniknutí si vytvoria oporné miesto, často vytvárajú zadné vrátka a používajú techniky, aby zostali nezistení. Konečným cieľom APT nie je okamžité narušenie, ale dlhodobá špionáž, krádež údajov alebo získanie strategických výhod.

Boj proti APT si vyžaduje mnohostranný prístup vrátane implementácie viacvrstvových bezpečnostných obranných opatrení, neustáleho monitorovania sieťovej aktivity, pravidelných bezpečnostných auditov a školenia zamestnancov na rozpoznanie potenciálnych hrozieb. Včasná detekcia a rýchla reakcia sú nevyhnutné na zmiernenie dopadu APT. Organizácie musia zaujať postoj neustálej ostražitosti a pripravenosti reagovať na tieto pokročilé hrozby.