VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
Fakulta informačních technologií
Ústav inteligentních systémů

Ing. Maroš Barabas

Bezpečnostná analýza sieťového toku pomocou behaviorálnych signatúr

Tézy k dizertačnej práci

Školitel: doc. Dr. Ing. Petr Hanáček

Kľúčové slová

aproximácia, behaviorálna signatúra, bezpečnosť, detekcia, honeypot, IDS, klasifikácia sieťového toku, metriky, pretečenie zásobníku, sieťová analýza, sieťové útoky

Keywords

aproximation, behavioral signature, buffer overflow, detection, honeypot, IDS, metrics, network analysis, network attacks, network traffic classification, security

Obsah

1 Úvod

Internet v dnešnej dobe zasahuje do každodenného života viac ako 40 % svetovej populácie. Aktuálne smery vývoja informačných technológií, vznikajúca globálna informačná architektúra založená na Internete uľahčujúca výmenu služieb a tovaru spôsobujú, že bežné zariadenia v našom prostredí sú pripájané do Internetu. Tento trend sa postupne dostáva i do rôznych odvetví ako automobilový priemysel alebo bežné doplnky – chytré telefóny, hodinky, šperky, či dokonca i chytré oblečenie, ale napríklad i letecký priemysel pre ktorý je dôležitým prvkom pripojenie pasažierov k zábavnému systému lietadla a k Internetu. Ďalšími oblasťami sú priemyselná automatizácia, riadiace súčasti priemyselných systémov, využitie ICT technológií na skvalitnenie života v mestách, i do ľudských príbytkov – automatizácia systémov riadiacich prostredie domov alebo bytov, ako napr. vykurovanie, vzduchotechnika, osvetlenie, apod. S postupujúcimi technológiami sa mení i charakter hrozieb a odpoveďou na tieto hrozby je zvýšený záujem o bezpečnosť. Ako sa mení charakter informácií a zdrojov dostupných prostredníctvom Internetu, mení sa i povaha útočníkov. Jedným z často sa opakujúcich pojmov, poslednú dekádu, sú pokročilé perzistentné hrozby (z angl. *Advanced Persistent Threats*), alebo *APT* [49], ktoré označujú sofistikované techniky a vektory útokov, ktoré slúžia na obchádzanie bezpečnostných mechanizmov, skrytie činností útočníka a jeho nedetegované zotrvanie na napadnutých systémoch v čo najdlhšom čase, zväčša za účelom získania informačných aktív obete. Ako *APT* bývajú často označované útoky, ktoré sú sponzorované štátnymi zložkami, ale toto tvrdenie je ťažké dokázať [22]. Vytvorenie týchto techník je zložité a útočníci sa tak presúvajú z pozície jednotlivca do organizovaných skupín, ktoré môžu byť financované firmami, ale napríklad i vládami za účelom špionáže (často označovanej i ako kyber-šiponáž, z angl. *Cyber-espionage*) alebo za účelom útoku na iný štát s deštruktívnym úmyslom narušenia, prípadne zničenia (tiež známe pod pojmom *Cyberwar* alebo *Cyberwarfare*) [13].

1.1 Motivácia

V posledných rokoch sa objavujú stále častejšie zverejňované rozsiahle útoky, ktoré je možné označiť ako *APT*. Nižšie uvedené príklady útokov majú spoločné charakteristiky – sú cielené, sú sofistikované a ako možní útočníci sú väčšinou označované vlády, prípadne skupiny pracujúce pre vládne organizácie rôznych štátov. Príkladmi týchto útokov môžu byť malware *Stuxnet* [29], rozsiahla kyber-špionážna sieť známa pod označením *Red October*[46], útoky na britské a americké vládne úrady [37], úniky informácií zo spoločnosti Sony Corporation [7] apod. Aktuálny stav a trendy v bezpečnosti poukazujú na neschopnosť efektívne čeliť moderným kybernetickým útokom, odolnosť aplikácií, systémov, či infraštruktúr voči kybernetickým útokom je na stále nedostatočnej úrovni. Tento fakt vyplýva nielen z uvedených článkov popisujúcich bezpečnostné incidenty, ale i z rôznych renomovaných štúdií, ktoré sa zaoberajú vývojom kybernetickej bezpečnosti. Ako príklad je možné uviesť správy technologických spoločností zaoberajúcich sa počítačovou bezpečnosťou, ako McAfee, Symantec, Bit9, Cisco a ďalších [11], ktoré spoločne upozorňujú nielen na vysoký nárast počtu nových útokov, ale predovšetkým na ich sofistikovanosť a zameranie na neustále nové ciele a technológie.

1.2 Sieťové bezpečnostné technológie

Medzi základné sieťové bezpečnostné technológie, ktoré adresujú pokročilé sieťové útoky sú IDS (Intrusion Detection Systems) systémy [52, 41], ktoré môžu byť založené na rôznych metódach rozpoznávania sieťových anomálií a útokov. Metódy a techniky na detekciu útokov zo sieťového toku je možné rozdeliť na *detekciu založenú na signatúrach* a *detekciu založenú na anomáliách*. Signatúry môžu byť tvorené *bitovým reťazcom* alebo *pravidlami*. Tieto metódy ale majú veľkú nevýhodu v nutnosti vytvárania a udržovania databáz signatúr voči ktorým sa porovnáva sieťový tok. Presne špecifikovaná signatúra má výhodu v malom množstve chybných hlásení o útoku – false-positive, ale ich manuálne vytváranie je problematické a môže trvať príliš dlho. Z toho dôvodu vznikali metódy na automatické generovanie signatúr, ktoré je možné rozdeliť na *signatúry založené na exploite* a *signatúry založené na zraniteľnosti* a líšia sa procesom ich generovania. Automatické metódy na generovanie signatúr ale majú problémy pri určovaní, či ide naozaj o útok. Tento problém adresujú systémy s názvom honeypot, ktoré dokážu odhaliť útok na systém a v prípade útokov na pretečenie zásobníka – buffer overflow, vykazujú vysokú úspešnosť detekcie [35, 43]. Tieto metódy sa zakladajú na predpoklade, že na spôsobenie samotného pretečenia alokovanej pamäte a uloženie návratovej hodnoty adresy na kód, ktorý vkladá priamo útočník, musí byť návratová hodnota i útočníkov kód súčasťou vstupu od útočníka. Detekčné metódy buffer overflow útokov so sieťového toku sa potom zameriavajú na analýzu dát paketov [58, 21, 40], identifikáciu binárnych dát na vstupe sieťovej služby [30], prípadne modelovanie aplikačného protokolu s detekciou anomálií [55]. V prípade štúdií, ktoré sa zameriavajú na vylepšenie tradičných IDS systémov, sa za poslednú dekádu sústreďuje veľa vedeckých prác na automatizáciu generovania signatúr pre tieto systémy. Jedným z hlavných smerov týchto výskumov je automatizácia honeypot systémov [42], ktoré by dokázali vytvárať signatúry pre rôzne typy útokov a to hlavne bez zásahu a nutnej manuálnej analýzy človekom [26, 27, 28]. Pri automatickom vytváraní signatúr je dôraz kladený hlavne na kvalitu vygenerovaných signatúr. Kvalita signatúr priamo ovplyvňuje detekčné metódy, konkrétne ich presnosť v rozpoznaní (klasifikácii, vyhľadaní) útokov.

Okrem výskumných prác okolo detekcie týchto útokov na honeypot systémoch, vznikli i práce zameriavajúce sa na detekciu týchto útokov v rámci sieťového toku, ale väčšinou ide o detekciu na základe obsahu paketov, ktorá je v prípade šifrovaných spojení neúčinná a v prípade, že sa podoba útoku zmení (bitová reprezentácia) dajú sa tieto metódy úplne obísť. Nástroje, ktoré sa zameriavajú na detekciu útokov v rámci dátovej časti paketov majú veľmi vysokú úspešnosť detekcie nešifrovaných spojení (napr. HTTP, TELNET, DNS apod.) a to pomocou metód založených na štatistickej distribúcii bytov v aplikačnom protokole [57, 6, 56], prípadne na metódach strojové učenia ako neurónové siete [31], kohonenové siete [38], SVM [18, 12], na metódach PART [10], Random Forest [8], Grading Classifier [39], Adaboost [16] alebo IBK [1] a iných.

Princíp detekcií anomálií na základe atribútov paketu je v metódach strojového učenia, ktoré je možné rozdeliť na základe uvedených vedeckých prác na *štatistické, zhlukovacie, klasifikačné, učenie s učiteľom, porovnávacie* a *hybridné*. V prípade detekcie anomálií a útokov na základe atribútov paketov sa používajú rôzne techniky dolovania dát [15], stochastické modely a metódy založené na pravdepodobnostnom rozložení [48, 33, 34, 5], štatistické metódy [2, 53] apod. Tieto metódy pracujú len

s hlavičkami paketov a majú dobré výsledky v prípade šifrovanej komunikácie, binárnych protokoloch apod. V rámci tejto práce je predstavený prierez dvadsiatich rokov výskumu a vývoja metód a techník na detekciu sieťových útokov, ktoré prešli dlhým časovým obdobím, ktoré prinieslo radu zmien ako v útokoch na sieťové systémy, tak v použitých metódach. Je nutné ale podotknúť, že v oblasti použitých metód a techník existuje niekoľko problémov, ktoré vplývajú na ich účinnosť.

1.2.1 Vybrané problémy v detekčných metódach

Veľa sieťových detekčných nástrojov je odkázaná na nešifrovanú komunikáciu, použité metódy sa zameriavajú na vyhľadávanie podreťazcov v dátovej časti paketov analyzovaného spojenia a ich porovnávanie s množinou známych signatúr. Tento prístup nie je možný u šifrovanej komunikácie bez jej dešifrovania.

Ďalším probémom je neexistencia obecnej dátovej množiny pre porovnanie účinnosti prístupov publikovaných vo vedeckých prácach [51]. Existujúce dátové sady sú vytvárané v simulovaných prostrediach a neodzrkadľujú podmienky v reálnom prostredí, experimenty nad dátovými sadami nie sú dostatočne popísané a pomer validných záznamov a útokov je nevyvážený.

Jedným z hlavných problémov detekčných metód je ich veľmi úzke zameranie na konkrétny problém alebo oblasť a v porovnaní s ostatnými detekčnými metódami a nástrojmi sa tieto práce stávajú neporovnateľnými. Niektoré nástroje sa zameriavajú na inšpekciu dátovej časti paketov voči signatúram zachytených útokov. Tieto metódy sú účinné voči špecifickým útokom a je ťažké ich výsledky porovnávať s metódami, ktoré sú výrazne komplexnejšie (s rastúcou komplexnosťou – pokrytie širokého spektra typov útokov klesá účinnosť detekcie alebo rastie počet nesprávne detegovaných útokov) a to hlavne v prípade, že pri testovaní týchto úzko zameraných nástrojov je použitá iba určitá časť testovacej sady, ktorá vyhovuje danému experimentu.

Vo vedeckých článkoch sa používajú rôzne štatistické funkcie hodnotenia úspešnosti detekčných metód, ako napr. *senzitivita*, *špecifickosť*, *precíznosť*, *účinnosť* alebo *presnosť* klasifikácie. Tieto štatistické funkcie sa bežne používajú na určenie úspešnosti klasifikácie metód, ale často môžu byť náchylné na skreslenie štatistických dát, hlavne pri použití nerovnomerných vzoriek (napr. nepomer vzoriek útokov k validnej komunikácii). Pre lepšie výsledky vo vedeckých prácach a článkoch je tak možné použiť zvolenú štatistickú funkciu, ktorá nereflektuje najhorší parameter. Napríklad pri vysokom počte nesprávne klasifikovaných útokov je možné použiť *senzitivitu* klasifikácie, ktorá zobrazuje pomer správne klasifikovaných validných komunikácií ku všetkým validným komunikáciám.

1.3 Cieľ práce

Cieľom tejto práce je návrh autonómneho detekčného systému so zameraním na behaviorálnu analýzu sieťového toku a rozpoznanie útokov na základe detekcie anomálií v správaní v pozorovanom sieťovom toku. Normálne správanie obecne poukazuje na množinu charakteristík, ktoré je možné pozorovať a extrahovať pri bežnom fungovaní pozorovaného zdroja. Nevalidné je potom také, ktoré vybočuje z bežného modelu fungovania tým, že vykazuje anomálie v pozorovaných charakteristikách.

Vo výskumných prácach je možné pozorovať použitie základných štatistických metrík, kde detekcia je daná porovnávaním odchyliek od štatistického priemeru, prípadne

zhlukovaním štatistických atribútov alebo vyhľadávanie anomálií na základe pravdepodobnostnej príslušnosti do definovaných tried. Základné metriky uvádzané v literatúre dokážu dosiahnuť dobré výsledky v detekcii základných útokov, ale zlyhávajú pri ďaleko jednoduchších útokoch, ktoré nemajú volumetrický charakter a komplexnejších alebo zložitejších útokoch (napr. buffer overflow útok na službu so spustením príkazového riadku namiesto práce s dátami v rámci pôvodnej funkcionality služby). Tento problém je adresovaný nasledujúcimi cieľami:

- Vytvorenie sieťových detekčných metrík, ktoré dokážu pridať do signatúry spojenia parametre popisujúce správanie/priebeh danej komunikácie.
- Definícia kontextu spojenia pre popis okolitých spojení analyzovaného systému a detekciu komplikovanejších útokov.
- Zabezpečenie autonómnosti detekčného systému s možnosťou učenia od expertnej časti systému bez nutnosti zásahu človeka.

Jedným z cieľov je zapojenie systémov honeypot do procesu získavania informácií o aktuálne prebiehajúcich útokoch a využitie týchto informácií ako expertnej znalosti pri učení klasifikačných algoritmov k detekcii identifikovaných útokov zo sieťového toku. Vzhľadom na zložitosti procesu určenia podmienok kladených na vstupné dáta do učiaceho algoritmu, je potrebné vytvoriť model minimalizovaného sieťového protokolu a následne sady metrík pre popis sieťových spojení. Výstup zo sady metrík by mal tvoriť signatúru sieťového spojenia, ktorá bude využitá ako vstupný vektor pri učení identifikovaných útokov i následne pri analýze sieťového toku. Hlavnou požiadavkou na formát signatúry sieťového spojenia je jej konštantná dĺžka pre všetky analyzované komunikácie, ktorá je dôležitá pre rad klasifikačných algoritmov. Signatúra by ďalej mala čo najvernejšie sprostredkovať behaviorálne aspekty komunikácie modelovaním správania účastníkov komunikácie využitím atribútov sieťového toku. Predpokladom je vysoká účinnosť detekcie sieťových útokov typu buffer overflow pomocou koeficientov aproximačných funkcií priebehu spojenia.

2 Metriky paketových sietí

V tejto práci je definovaný minimalistický sieťových paketový protokol, ktorý spĺňa len základné požiadavky pre prenos informácií a nad týmto minimalistickým protokolom sú definované základné mechanizmy, ktoré sú potrebné pri analýze spojení a detekcii útokov. Tento sieťový protokol je založený na blokoch informácií posielaných medzi zdrojom a cieľom, ktoré sú nazývané pakety. Paket je definovaný ako n-tica $p = (id, len, src, dst, data)$, kde id, len, sú celé čísla, src a dst sú bitové reťazce konečnej dĺžky a dátová časť aplikačnej vrstvy $data$ je postupnosť bitov. Prvky a n-tice p sú obecne nazývané atribúty paketu. Množina U_C je tvorená všetkými zaznamenanými paketmi zasielanými medzi určeným zdrojom a cieľom a postupnosť C, ktorá je tvorená množinou paketov U_C s definovaným usporiadaním podľa indexu paketu id.

Nad definovaným spojením je vytvorená charakteristika spojenia, ktorá je tvorená množinou skalárnych atribútov a množinou variabilných atribútov (reprezentovaných vektorom) $X_C = (S, V)$. Ďalej je definovaný *skalárny atribút spojenia*, ktorý je v rámci všetkých paketov daného spojenia rovnaký (nemenný) a nadobúda tak v spojení konštantnú hodnotu. Skalárne atribúty daného spojenia sú zaradené do množiny S podľa

nasledujúcej definície. Majme množinu A obsahujúcu hodnoty atribútu a pre všetky prvky p_i,

$$A = \{a_i | a_i \in p_i, \forall p_i \in C\},$$

kde a_i je hodnota atribútu a paketu p_i. Atribút a je v rámci spojenia konštantný a je prvkom množiny S práve vtedy, ak platí

$$\forall a_i, a_j \in A : a_i = a_j \rightarrow a \in S.$$

Ak je hodnota atribútu a pre aspoň jedno $p_i \in C$ rozdielna, budeme hovoriť o *variabilnom atribúte spojenia* a budeme definovať vektor v,

$$\exists a_i, a_j \in A : a_i \neq a_j \rightarrow v = (a_1, a_2, \ldots, a_n),$$

kde $v \in V$ a $n = |C|$ je veľkosť spojenia.

Skalárna časť charakteristiky spojenia je tvorená takými atribútmi paketov, ktoré sú konštantné v rámci celej komunikácie. Variabilná časť je tvorená množinou vektorov, kde každý vektor odpovedá jednému atribútu paketu, ktorý sa v rámci spojenia mení (napr. veľkosť paketu). Charakteristika spojenia tak popisuje vlastnosti daného spojenia. Do charakteristiky spojenia je pre úplnosť pridaný čas ako skalárne a variabilné atribúty daného spojenia a smer jednotlivých paketov ako nominálny variabilný atribút.

Čas ako skalárna hodnota je do charakteristiky spojenia X_C vložený ako atribút času prvého paketu (začiatok spojenia) a posledného paketu (ukončenie spojenia), ktorý je 0-vý v prípade, že spojenie stále trvá (je živé). Čas je do charakteristiky spojenia X_C vložený ako variabilný atribút v_t:

$$v_t = (t_0, t_1, \ldots, t_n),$$

kde t_i je rozdiel času príchodu paketu p_i a p_{i-1}, pre $0 < i \leq |v_t|, p_i \in C$, $v \in V$ a $t_0 = 0$.

Do charakteristiky spojenia X_C je vložený variabilný atribút v_d, ktorý sa skladá zo zložiek nadobúdajúcich hodnôt 0 alebo 1. Hodnoty označujú smer paketu:

$$v_d = (d_0, d_1, \ldots, d_n),$$

kde d_i je smer paketu p_i, $p_i \in C$ a $v_d \in V$ a platí

$$d_i = \begin{cases} 0 & \text{ak } p_i \in C^{OUT} \\ 1 & \text{ak } p_i \in C^{IN} \end{cases}$$

Nad vlastnosťami, resp. charakteristikou každého spojenia je podľa definícií metrík spojenia vytvorená sada metrík a ich výstupov. Metrika je funkcia, ktorej výstupom je číslo alebo postupnosť čísiel konštantnej dĺžky, ktoré reprezentujú jednu vlastnosť analyzovaného spojenia.

Pre každý skalárny atribút $s \in S, S \in X_C$, ktorý je reprezentovaný číslom, je výstupom metriky φ číslo reprezentované hodnotou tohoto atribútu, ktorý je konštantný pre všetky pakety spojenia C a patrí do množiny hodnôt M_C:

$$\varphi(s) = m, m \in M_C.$$

Pre každý variabilný ordinálny atribút $v \in V$, ktorý je reprezentovaný vektorom, je definovaná množina výstupov funkcií f_i,

$$M_C \supset \{m_i | m_i = f_i(v), v \in V\},$$

Táto množina je tvorená výstupmi funkcií, ktorých vstupnou hodnotou je vektor v a výstupom (hodnotou metriky) je číslo. Príkladom takýchto funkcií sú štatistické funkcie *priemer, medián, minimum, maximum, súčet, štandardná odchylka.*

Množina hodnôt metrík M_C tvorených týmito funkciami môže byť doplnená ľubovoľnými ďalšími funkciami. Pre nominálne variabilné atribúty ale nemajú uvedené štatistické funkcie zmysel (napríklad pri atribúte *poradové číslo id* nemá zmysel počítať uvedené štatistické funkcie, pretože daný atribút je nominálny) a preto je potrebné zaviesť štatistické funkcie i pre variabilné nominálne atribúty.

Pre každý variabilný nominálny atribút, ktorý je reprezentovaný nominálnou hodnotou, definujeme množinu vektorov, kde každý vektor reprezentuje prvky s hodnotou c.

$$v_c = \{x_i | x_i = c, x_i \in v\},$$

kde x_i je komponent (zložka) variabilného atribútu reprezentovaného vektorom $v \in V$. Ak sú komponenty x_i vektoru v reprezentované postupnosťou bitov, potom pre každý bit $j \in x_i$, $j = 0, 1, \ldots M$, $M = |x_i|$ definujme komponentný vektor v_j, ktorý je definovaný nasledovne:

$$v_j = (x_0^j, x_1^j, \ldots x_N^j), \forall x_i \in v, N = |v|,$$

kde x_i^j je j-tý člen postupnosti bitov v komponente x_i vektoru v. Potom pre každý komponentný vektor v_j je možné definovať metriky $f(v_j)$.

Pre uvedené definície je vytvorená metrika $f(v)$ – počet, definovaná ako funkcia f_n podľa predpisu

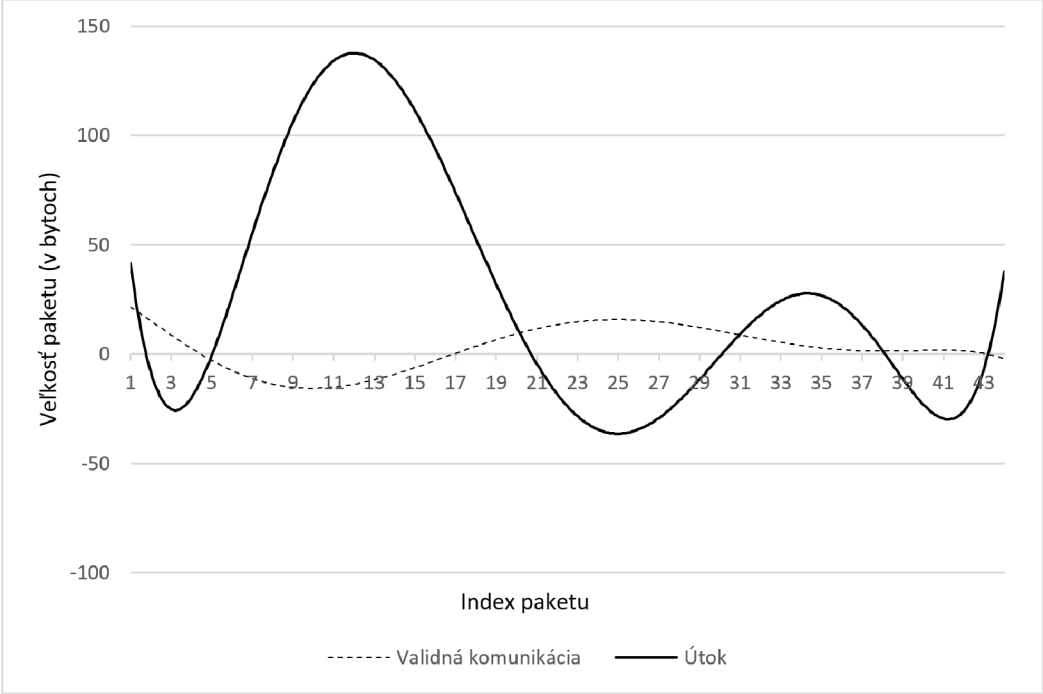
$$f_n = \sum_{i=1}^{N} x_i,$$

kde x_i je zložka vstupného vektoru, t.j. vektory v_c alebo komponentné vektory v_j. Dôležitým faktorom pri popise spojenia a následnej analýze je samotný priebeh spojenia. Vzhľadom na to, že každé spojenie môže mať iný počet paketov, je potrebné atribúty s variabilnými hodnotami spojenia aproximovať. Aproximácie tak transformujú dvojrozmerný priebeh spojenia na jeden rozmer (napr. koeficienty) a vytvárajú behaviorálnu signatúru spojenia (modelujúcu správanie účastníkov komunikácie v rámci sieťového toku). V práci sú použíte aproximačné funkcie polynómami rôznych stupňov a diskrétna Fourierova transformácia. Na obrázku 1 je zobrazené porovnanie validnej komunikácie a útoku (aproximácie veľkosti paketov v definičnom obore indexov paketov) na ilustrácii fitovania veľkostí paketov pomocou uvedenej polynomiálnej aproximácie (index paketu odpovedá jeho poradovému číslu).

Ak je daný atribút a_j variabilným atribútom spojenia, potom vstupnými atribútmi $(a_1, a_2, \ldots a_n)$, $a_i \in p$, definujeme funkciu f v indexovej rovine (definičným oborom funkcie je index paketu v spojení) a následne jej aproximáciu $P_n(f)$ polynómom n-tého stupňa. Výstupom je n členov polynómu, ktorý definujeme ako vektor v charakterizujúci spojenie C:

$$v = (p_1, p_2, \ldots, p_n)$$

kde p_i je i-tý člen polynómu $P_n(f)$, $v \in X_C$.



Obr. 1: Porovnanie polynomiálnych aproximácií 6. stupňa validnej komunikácie a útoku.

Ďalej definujeme funkciu f v indexovej rovine (definičným oborom funkcie je index/poradie paketu v spojení a hodnota $f(i)$ je hodnota atribútu a s indexom i) a následne jej aproximáciu $F(f)$ Fourierovou transformáciou. Vzhľadom na to, že vstup je postupnosť paketov, ide o diskrétnu Fourierovu transformáciu (zápis v exponenciálnom tvare):

$$F(n) = \sum_{k=0}^{N-1} f(k)e^{-ink2\pi/N}, n = 0, 1, \ldots N-1,$$

kde $F(n)$ je n-tý koeficient Fourierovej transformácie, N je počet vstupných atribútov funkcie f, ktoré sú stanovené pri definícii signatúry a v prípade menších počtov paketov v spojení sú hodnoty $a_i, |v| < i < N$ doplnené 0. Výstupom Fourierovej transformácie je $N/2+1$ koeficientov (druhá polovica koeficientov je symetrická k prvej polovici a nie je potrebný ich výpočet):

$$a + bi = F(n), n = (0 \ldots 10),$$

kde a je reálna a b je imaginárna časť koeficientu $F(n)$.

Pri experimentoch bol použitý zápis koeficientov pomocou uhla (φ) a veľkosti vektoru (c) tvoreného reálnou a imaginárnou zložkou koeficientu. Tento prevod je možný riešením sústavy rovníc

$$b = a \tan \varphi,$$

$$c = \sqrt{a^2 + b^2},$$

kde $(\varphi, c) \in M_C$.

2.1 Dátová časť

Pre charakteristiku spojenia a definíciu metrík sú používané len hlavičky paketov analyzovaného spojenia. Ďalšou veľmi dôležitou časťou pri klasifikácii sieťových útokov je dátová časť paketu, ktorá obsahuje dáta vyššieho protokolu. Vzhľadom na informácie získané pri zbieraní dát, je väčšina dnešných protokolov šifrovaná. Vzhľadom na fakt, že šifrované dáta bez príslušnej znalosti kľúča a následnom dešifrovaní obsahu paketov je analýza takýchto dát zbytočná a mohla by znehodnotiť klasifikačný proces, v rámci tejto práce nebude dátová časť paketov z týchto dôvodov súčasťou analýzy.

2.2 Rozšírenie protokolu na TCP/IPv4

V tejto práci je definovaný minimalistický sieťový paketový protokol, ktorý je použitý pri definícii sieťového spojenia, ktoré je reprezentované postupnosťou paketov medzi zdrojom a cieľom. Nad týmto spojením je definovaná jeho charakteristika, ktorá je daná množinou skalárnych a variabilných atribútov paketov tohto spojenia. Skalárne atribúty sú také, ktoré majú v rámci všetkých paketov spojenia konštantné hodnoty a u variabilných atribútov sa tieto hodnoty v rámci spojenia menia. V rámci práce sú predmetom analýzy a detekčného systému iba protokoly TCP a IPv4. Po aplikovaní pravidiel pre vytvorenie charakteristiky spojenia nad TCP/IP protokolom verzie 4 dostaneme pre analyzované spojenie nasledujúcu charakteristiku spojenia:

- **Skalárne atribúty protokolu**: verzia protokolu IP, DSCP, ECN, identifikácia skupín fragmentov, Time to Live, protokol, zdrojová adresa, cieľová IP adresa, zdrojový port, cieľový port.
- **Skalárne časové atribúty**: čas príchodu prvého paketu, čas príchodu posledného paketu.
- **Variabilné atribúty**: IHL, dĺžka IP paketu, IP atribúty, ofset fragmentu, kontrolný súčet hlavičky IP, IP rozšírené možnosti – Options, Sekvenčné číslo, číslo potvrdenia, potvrdzovacie číslo, ofset dát, rezervované bity, TCP atribúty, veľkosť okna, kontrolný súčet hlavičky TCP, urgent pointer, TCP rozšírené možnosti – Options.
- **Variabilný časový atribút**: časový rozdiel príchodu paketu od príchodu predchádzajúceho paketu.
- **Smer paketu**: nominálna hodnota určujúca smer paketu.

Vzhľadom na veľké množstvo atribútov vytvorených z TCP/IP protokolov, je potrebné tieto dve množiny skalárnych a variabilných atribútov zmenšiť. Vzhľadom na povahu informácií, ktoré jednotlivé atribúty nesú je logické niektoré z nich odstrániť už pri návrhu tohoto systému. Sú to hlavne atribúty, ktoré sú za každých okolností konštantné, *konkrétne verzia protokolu IP (konštantná hodnota pre protokol IPv4), protokol (bude konštantný s hodnotou 6 – TCP), Rezervované bity (vždy 0).*

Ďalej sa medzi variabilnými atribútmi nachádzajú také, ktoré z určitého pohľadu nedávajú pri analýze zmysel a to z dôvodu ich účelu pre potreby riadenia TCP alebo IP protokolu ako napr. kontrolné súčty a pod. Sú to konkrétne: *kontrolný súčet hlavičky IP* a *hlavičky TCP*, *ofset fragmentu* a *ofset dát*, *IP atribúty* a *IP identifikácia.* Uvedené atribúty okrem kontrolných súčtov sa týkajú IP fragmentácie, ktorá v prípade

analýzy môže spôsobovať problémy (informácia o veľkosti dát je sploštená do maximálnej veľkosti MTU) a preto pred analýzou a vytvorením signatúry spojenia je potrebné vykonať defragmentáciu paketov a to i pre IP fragmentáciu aj pre TCP segmentáciu. Po defragmentácii sú tieto informácie zbytočné a z atribútov pre vytvorenie signatúry sú zahodené. S hodnotami Options u IP i TCP protokole sa počíta ako s ordinálnymi hodnotami.

Pre každý variabilný atribút sú spočítané funkcie minimum, maximum, medián, súčet, priemer a štandardná odchýlka. Pre každý variabilný atribút sú spočítané funkcie pre aproximáciu atribútu (vektoru hodnôt) polynómom n-tého stupňa a to pre prichádzajúcu, odchádzajúcu komunikáciu a oba smery a Fourierovou transformáciou pre oba smery. V prípade napríklad polynómu 3-tieho a 5-teho stupňa, a uchovávanými 10-timi koeficientmi Fourierovej transformácie je pridaných (polynóm n-tého stupňa má n+1 koeficientov) 56 hodnôt do signatúry na jeden variabilný atribút, čo celkovo činí *593 hodnôt* pre TCP/IP protokol verzie 4. Pre metriky pridávajúce čas a smer paketov je pridaných 6 hodnôt pre smer a 58 hodnôt pre čas. Celkový počet hodnôt základných metrík je tak *657*. Celková teoretická veľkosť signatúry je cca 1902 bytov na spojenie (veľkosť sa môže líšiť od implementácie). V prípade analýzy veľkých dátových tokov je možné prispôsobiť signatúry voľbou vhodnejších metrík, napr. vynechať aproximáciu Fourierovou transformáciou, čím sa ušetrí až 46 % (880 bytov na spojenie) celkovej veľkosti.

2.3 Kontext spojenia

Kontext spojenia je možné chápať ako informáciu o aktuálnom stave prostredia spojenom s daným analyzovaným spojením. V prípade, že nás bude zaujímať iba dvojica aktérov, ktorí vystupujú v rámci danej komunikácie, teda zdroj a cieľ komunikácie, budeme hovoriť o *lokálnom kontexte* K_L spojenia C:

$$K_L = \{s; src_{IP} \in add(s) \wedge dst_{IP} \in add(s)\}, src_{IP}, dst_{IP} \in C,$$

kde $add(s)$ je množina adries spojenia (konkrétne zdroj a cieľ), src_{IP} je zdrojová adresa a dst_{IP} je cieľová adresa spojenia C. Lokálny kontext tak obsahuje všetky signatúry spojení medzi aktérmi analyzovaného spojenia. Pri analýze IPv4 protokolu sa jedná iba o IP adresy aktérov bez portov, pretože porty (minimálne zdrojový) sú vo väčšine prípadov v spojení generované dynamicky. Pre globálny kontext je definícia adekvátna.*Globálny kontext* K_G *spojenia* C je rozšírenie lokálneho kontextu o všetky aktuálne signatúry spojení oboch aktérov:

$$K_G = \{s; src_{IP} \in add(s) \vee dst_{IP} \in add(s)\}, src_{IP}, dst_{IP} \in C.$$

Kotext spojenia je použitý pre detekciu pokročilých útokov. Tieto útoky sú nad rámec tohoto textu.

V práci sú ďalej predstavené základné algoritmy extrakcie rysov *analýza základných komponent (PCA)* [24] a metóda *Forward Selection* [59, 32] a boli popísané základné metódy použité na analýzu signatúr spojení pri detekcii sieťových útokov, ktoré boli v práci použité v experimentoch. Veľmi zjednodušene je v práci popísaný algoritmus *Support Vector Machine* [17, 9], *metóda rozhodovacieho stromu* [54] a *bayesovský klasifikátor* [25, 19], ktoré je možné použiť na klasifikáciu vstupných dát reprezentovaných

definovanou signatúrou a kontextom spojenia do tried validnej komunikácie a triedy útokov. Popísanými algoritmami nie je obmedzený výber klasifikačných algoritmov pre detekčný systém, je možné použiť ďalšie metódy a techniky. Uvedené algoritmy sú súčasťou architektúry detekčného systému, ktorý s využitím definovanej signatúry sieťových spojení a kontextu spojenia zdrojového a cieľového systému, vytvára základ pre detekčný sieťový systém.

3 Architektúra detekčného systému

Architektúra celého systému pre detekciu útokov zo sieťového toku sa zameriava na architekúru TCP/IPv4 a pozostáva z troch logických vrstiev. Prvá vrstva je *extrahovacia*, v ktorej zo sieťového toku, ktorý je rozdelený na spojenia je následne pre každé spojenie vytvorená charakteristika popisujúca dané spojenie. V druhej vrstve, *popisnej*, sú z charakteristík extrahované hodnoty metrík a tie pre dané spojenie tvoria signatúru. Tento proces je v oboch vrstvách identický, ako pre získanie signatúry spojenia z útoku na honeypot systém, tak i v prípade klasifikácie sieťového toku. V prípade signatúry získanej z honeypot systémov je táto signatúra využitá pre učiaci proces klasifikačných algoritmov. V popisnej vrstve je pre danú signatúru vytvorený kontext spojenia. Samotná klasifikácia spolu s analýzou a rozhodovacím procesom tvorí vrstvu *klasifikačnú*.

3.1 Extrahovacia vrstva

Vstupom detekčného systému i extrahovacej vrstvy je záznam sieťového toku, ktorý je predmetom analýzy. v rámci extrahovacej vrstvy je sieťový tok zaznamenávaný pomocou PCAP [23] a pre účely dodatočnej analýzy môže byť ukladaný (ide o kompletný záznam sieťového toku). Všetok zachytený tok (prúd paketov) je následne rozdelený do jednotlivých spojení, ktoré sú identifikované na základe definície spojenia a rozšírení pre účely TCP/IP architektúry. V prípade honeypot systému nasadeného do monitorovanej siete je prichádzajúca komunikácia nahrávaná priamo na sieťovom zásobníku operačného systému honeypotu. V prípade, že dôjde k útoku na služby honeypot systému, sieťová komunikácia (záznam útoku) je zaznamenaná a odoslaná k extrakcii spojenia. Ďalšie procesy so záznamom útoku sú identické s analyzovanou komunikáciou.

3.2 Popisná vrstva

Pre každé spojenie je vytvorená charakteristika spojenia, z tejto charakteristiky sú extrahované výstupy metrík a tie pre dané spojenie tvoria sadu, ktorá je nazvaná signatúra spojenia. Táto signatúra je základným popisom analyzovaného spojenia, prípadne zaznamenaného útoku a môže byť predmetom rozšírenia systému o nové metriky. Nad rámec signatúry je možné z charakteristiky spojenia vybrať ďalšie informácie, napr. hodnota, či je dané spojenie ukončené, identifikácia typu služby (napr. port, NBAR), pravdepodobnosť, že ide o šifrovanú komunikáciu, výstupy rôznych logických pravidiel (napr. na počet komunikácií apod.), ďalšie štatistické a aproximačné funkcie, atď.

3.3 Klasifikačná vrstva

Vstupom do klasifikačnej vrstvy je vytvorená signatúra spojenia, ktorá spolu s globálnym i lokálnym kontextom spojenia je klasifikovaná voči už vytvoreným modelom analyzovaných spojení a záznamov útokov. Samotná klasifikácia na validnú a nevalidnú komunikáciu (útok), obsahuje klasifikátory, ktoré sú rozdelené do vrstiev. Jednotlivé vrstvy klasifikátorov sú riadené tzv. arbitrom, ktorý rozhoduje o zapojení ďalšej vrstvy klasifikátora na základe definovaných pravidiel. V prípade honeypot systému je vytvorená signatúra záznamu útoku použitá ako vstup do klasifikátorov a honeypot vystupuje ako učiteľ s expertnou znalosťou. V prípade, že honeypot deteguje pretečenie zásobníka, vytvorená signatúra sieťového toku sa použije pre učiaci proces a jednotlivé modely klasifikátorov sa upravia. Pre urýchlenie procesu je možné vytvorenú signatúru porovnať s už existujúcou bázou záznamov útokov, aby nedochádzalo k učeniu a úpravám modelov vždy, keď dôjde k útoku na známu zraniteľnosť v honeypot systéme. Signatúry zachytených útokov, ktoré nie sú známe môžu byť označené a pripravené na manuálnu analýzu pre detekciu prípadných zero-day zraniteľností. Výstupom klasifikačného procesu je pravdepodobnostné ohodnotenie príslušnosti daného spojenia do kategórie útokov. Tento výstup je daný kombináciou ohodnotení (pravdepodobností) jednotlivých klasifikátorov. Ich zapojenie do procesu klasifikácie riadi arbiter. Účelom arbitra je riadenie klasifikačnej časti systému a obsluha základných funkcionalít detekčnej časti. Medzi hlavné činnosti patrí určovanie váh pri klasifikačnom procese na základe ktorých sa vyhodnocuje, či ide o útok alebo validnú komunikáciu, proces riadenia incidentu pri jeho vytvorení, vyhodnocovanie detekčného procesu a rozhodovanie nad ďalšími klasifikačnými úlohami. V prípade, že analyzovaná komunikácia je vyhodnotená ako podozrivá, môže byť ďalej určená na ďalšiu analýzu inou klasifikačnou vrstvou, ale i prioritizácia úloh a spúšťanie procesov učenia klasifikačných metód v prípade aktualizácie databáz útokov z honeypot systémov.

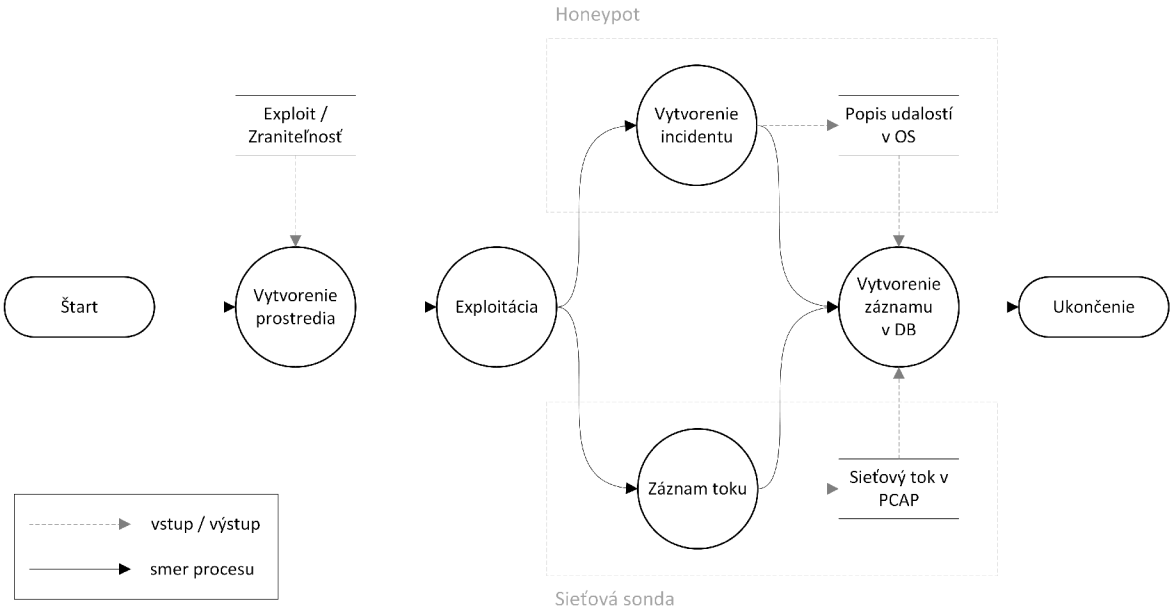
4 Experimenty

Navrhnutý systém bol v rámci overenia konceptu čiastočne experimentálne implementovaný v podobe jednoduchých programov pre spracovanie vstupných dát definovanými metrikami do dátovej sady charakterizujúcej namerané dáta. Táto sada hodnôt následne slúžila pre experimenty rôznymi metódami strojového učenia pre určenie kombinácie metód pre najlepšie vlastnosti detekcie a overenie konceptu systému simuláciami útokov a detekčné schopnosti v rámci existujúcich dátových množín. Dáta použité pri experimentoch boli získané z niekoľkých zdrojov. V rámci tejto kapitoly sú popísané zdroje dát pre experimenty, priebeh jednotlivých experimentov a dosiahnuté výsledky v nadväznosti na state-of-the-art.

4.1 Laboratórne prostredie

Pre fázu analýzy (a z časti i implementáciu) bolo vytvorené laboratórne prostredie, ktoré slúžilo na vytvorenie simulácie útokov a ich detekciu. Toto prostredie bolo vytvorené za účelom laboratórnych pokusov, dôraz bol kladený na izoláciu prostredia a procesov do takej miery, aby zaznamenané dáta pri simulovaných útokoch boli v čo najmenšej miere ovplyvnené okolím. Cieľom bolo vytvorenie dátovej sady reprezentujúcej buffer overflow útoky na sieťové služby, zaznamenanie úplného sieťového toku

útoku, overenie jeho detekcie na honeypot systéme a vytvorenie charakteristiky spojenia na základe definovaných metrík. Celý priebeh experimentu v laboratórnom prostredí je uvedený na obrázku 2. Proces na strane honeypot systému a na strane sieťovej sondy po ukončenie celého procesu bol zautomatizovaný. Príprava prostredia s výberom a inštaláciou zraniteľného programu a príprave exploitu so samotnou exploitáciou bol manuálny.



Obr. 2: Priebeh experimentu od vytvorenia prostredia po zaznamenanie zachyteného útoku.

V rámci laboratórneho prostredia bol vytvorený referenčný obraz systému pre každý testovaný operačný systém. Každý z programov získaných z internetových zdrojov pre účely simulácie útokov obsahuje zraniteľnosť buffer overflow, ktorá bola vybraná pre fázu experimentov z dôvodu jej rozšírenosti a nebezpečnosti vzhľadom na riziko týchto útokov (útočník po úspešnom útoku môže získať práva zraniteľnej služby a prístup do operačného systému). V rámci experimentov bolo nájdených 294 zraniteľností na pretečenie zásobníka s verziami programov a existujúcim exploitom. V rámci týchto zraniteľností bolo získaných celkovo 57 programov pre Windows XP, Windows 2000 a Linux. V rámci experimentov boli vykonané útoky podľa existujúceho exploitu (zväčša použité nástroje Metasploit, nástroje systému Backtrack a exploity z databáz exploitdb a podobných) a po úspešnej exploitácii služby bol tok spojený s útokom zaznamenaný do databázy pre ďalšiu analýzu.

Výsledky simulácií boli neskôr použité pri návrhu detekčného systému a experimenty s rôznymi sadami metrík. Spočiatku boli získané dáta použité i na experimenty s rôznymi klasifikačnými metódami a metódami strojového učenia, ale nazbierané vzorky dát neboli pre tieto úlohy dostatočne kvalitné (hlavne kvôli malému počtu úspešných útokov, nízkej diverzite útokov a homogénnosti laboratórneho prostredia). Popis experimentov, dátová sada a podrobný popis metrík sú uvedené v rámci článku [4]. V rámci experimentov so simulovanými útokmi v laboratórnom prostredí boli použité základné štatistické metriky, ktoré boli doplnené aproximačnými metódami (gaussové krivky, polynomiálna aproximácia a Fourierova transformácia), celkovo 169 metrík. Na

overenie klasifikačných algoritmov nad experimentálnou sadou dát bol použitý program *RapidMiner* [45] a to s metódami uvedenými v tabuľke 1.

V tabuľke 1 sú uvedené výsledky jednotlivých klasifikačných algoritmov. Celková účinnosť (čiže pomer správne identifikovaných útokov voči všetkým správne identifikovaným) algoritmov je veľmi nízka z dôvodu vysokého false-positive spôsobeného veľkým počtom vzoriek validných spojení voči vzorkám záznamov útokov. Toto tvrdenie dokazuje vysoká špecifickosť (pomer správne identifikovaných validných spojení voči všetkým záznamom validných spojení), ktorá je daná vysokým počtom záznamov validných spojení (priemerne 16-násobok k počtom záznamov útokov). Pri porovnaní klasifikačných algoritmov je kvalitatívne najlepším ukazovateľom presnosť klasifikácie (pomer správne identifikovaných vzoriek voči všetkým vzorkám), kde najlepšie výsledky dosiahol, so skoro 89 % presnosťou, SVM algoritmus. Tento výsledok je pre nízky pomer záznamov útokov voči všetkým záznamom očakávaný.

Klasifikačná metóda	SVM	Rozhodovací strom	Bayesovský klasifikátor s PCA s automatickým počtom komponentov	Bayesovský klasifikátor s diskretizáciou ordinálnych parametrov	Naivný bayesovský klasifikátor	Naivný bayesovský klasifikátor a PCA s fixným počtom komponentov
Účinnosť	41,67 %	25,00 %	16,67 %	25,00 %	8,33 %	8,33 %
Špecifickosť	96,09 %	94,97 %	94,48 %	94,97 %	96,13 %	96,73 %
Precíznosť	41,67 %	25,00 %	16,67 %	25,00 %	14,29 %	16,67 %
Presnosť	89,85 %	87,82 %	87,82 %	87,82 %	76,14 %	75,63 %

Tabuľka 1: Prehľad výsledkov experimentov s klasifikačnými metódami nad laboratórnymi dátami.

Pre neuspokojivé výsledky so simulovanou dátovou sadou sa výskum uberal dvoma smermi, v prvom išlo o nájdenie vhodnej verejnej dátovej sady, ktorá by spĺňala predpoklady pre kvalitný vstup testovania klasifikačných algoritmov a záznam reálnej komunikácie v prostredí kampusu VUT.

4.2 Výskumné databázy

Jednou z dôležitých častí pri posúdení metód klasifikácie sú dáta, ktoré sú jednoznačné (existuje znalosť o každom spojení, či ide o útok alebo validnú komunikáciu) a zároveň autentické (dáta nevykazujú skreslenie na základe ich pôvodu). V rámci výskumu *1999 KDD Cup* [20], ktorý sa uskutočnil v roku 1999 bola vytvorená databáza sieťových tokov, ktoré obsahujú útoky na sieťové služby. Táto databáza bola určená pre

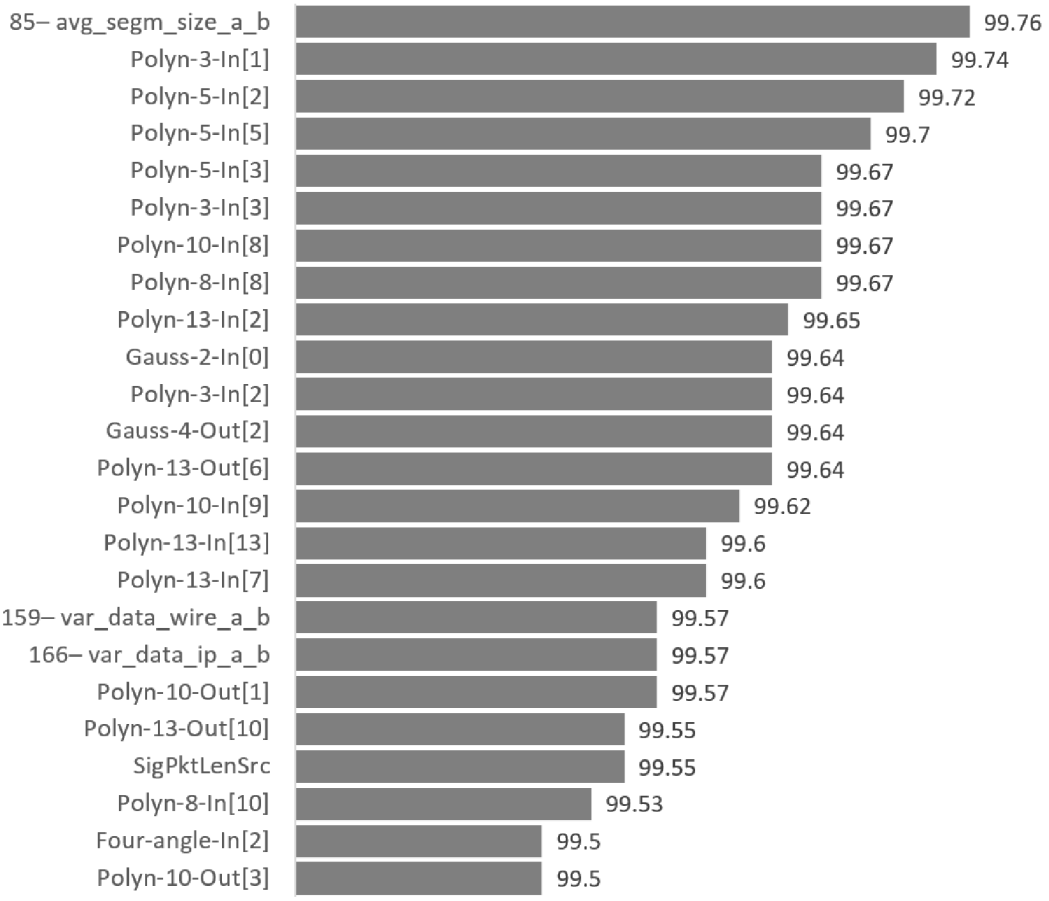
výskumné účely sieťových detekčných algoritmov a je široko najpoužívanejšia databáza vo výskumných projektoch [50]. I napriek tomu, že veľa expertov na systémy pre detekciu prienikov do siete uvádza, že väčšina aktuálnych útokov sú variantmi už známych útokov a signatúry pôvodných útokov sú postačujúce pre ich detekciu, práve projekt *1999 KDD Cup*, tzv. *Classifier Learning Contest* vyvrátil toto tvrdenie [14]. Cieľom ďalšieho výskumu bolo nadviazať na tento projekt, vytvoriť a poskytnúť databázu obdobnú databáze HoAH [36], ale s rozdielnou úrovňou detailu. Zmyslom vytvorenia takejto databázy bolo vytvoriť množinu dát, ktorá bude obsahovať i nové, zložitejšie útoky s maximálnou možnou úrovňou detailu (v ideálnom prípade bez straty informácie). V roku 2009 bola vytvorená sada CDX 2009 [47], ktorá obsahovala záznamy útokov na vytvorené virtualizované prostredie. Databáza CDX 2009 bola vybraná pre experimenty z dôvodu najvernejšej simulácie útokov a reálneho prostredia [47] oproti KDD Cup a DARPA dátovým sadám.

Klasifikácia pomocou definovaných metrík bola vykonaná bayesovským klasifikátorom s 5-násobnou krížovou validáciou pre každý experiment s výberom rysov pomocou SFFS. Na záver boli porovnané výsledky oboch výskumných prác a prehľad najlepších metrík a diskriminátorov. V rámci experimentu boli dosiahnuté zaujímavé výsledky s aproximačnými funkciami. Najlepšia celkovú presnosť detekcie bola dosiahnutá pomocou metrík polynomiálnej aproximácie a to hlavne pri aproximácii prichádzajúceho smeru (od zdroja k cieľu) polynómom 3. a 5. stupňa. Dobré výsledky boli dosiahnuté i aproximáciou gaussovými krivkami a Fourierovými koeficientami. Medzi experimenty s klasifikačnými metódami bol zaradený i algoritmus rozhodovacieho stromu. Experimenty s algoritmom rozhodovacieho stromu viedli k výsledkom, ktoré dokázali klasifikovať vstupy získané z dátovej sady CDX 2009 až s celkovou presnosťou 99,76 %. Celkové dosiahnuté F-score u najlepšieho (podľa účinnosti klasifikácie) stromu bolo 83 %. Celková dosiahnutá presnosť bayesovského klasifikátoru bola 99,83 % a F-score 87,81 %.

4.3 Zhodnotenie výsledkov experimentov

Navrhnutý systém bol experimentálne čiastočne implementovaný v podobe skriptov a modelov nástroja RapidMiner [45]. Vzhľadom na potreby vstupnej dátovej sady sieťového toku, ktorá by obsahovala dostatočné množstvo útokov a validnej komunikácie, ktorá by zodpovedala parametrom reálnej komunikácie, bolo vytvorené laboratórne prostredie a simulované útoky na služby honeypot systémov. Tieto dáta boli použité pri experimentoch s klasifikačnými metódami, ale použitá dátová sada nemala potrebné kvalitatívne parametre pre uspokojivé výsledky experimentov. Použité honeypoty boli z tohto dôvodu nasadené v sieti Vysokého učení technického v Brně pre zber potrebných záznamov útokov a sieťová sonda pre zber reálnych validných komunikácií. Tento projekt ale nepriniesol dostatočnú diverzitu útokov a bol ukončený. Použité honeypot systémy ďalej pokračovali v zbere útokov, ktoré sa používali pre identifikáciu infikovaných počítačov v sieti (a následné upovedomenie majiteľa o detegovanej infekcii systému zo strany Kolejí a menz VUT v Brně). Pre ďalšie experimenty boli použité verejne dostupné databázy a navrhnutá behaviorálna signatúra bola použitá pri klasifikácii rôznymi metódami ako napr. rozhodovacím stromom alebo bayesovským klasifikátorom pre overenie jej detekčnej schopnosti. Ďalej bol experimentami potvrdený predpoklad, že aproximácia priebehu spojenia ako simulácia správania analyzovaného spojenia a jeho aktérov bude vo výsledkoch experimentov dosahovať vysokú

úspešnosť klasifikácie. Najlepšie dosiahnuté výsledky týchto experimentov sú zhrnuté v nasledujúcej kapitole.



Obr. 3: Prehľad najlepších metrík a diskriminátorov zoradených podľa celkovej presnosti klasifikácie (nad 99,50 %).

4.3.1 Dosiahnuté výsledky

V práci bolo vykonaných niekoľko experimentov nad dátami vytvorenými v laboratórnom prostredí a dátami z verejných dátových sád, z ktorých pre finálne experimenty bola vybraná dátová sada CDX 2009. Výsledky testov v laboratórnom prostredí neboli z dôvodu nízkej kvality nazbieraných dát uspokojivé, naopak výsledky s dátami CDX 2009 sú blízke výsledkom dosiahnutým v obdobných výskumných prácach.

Pri experimentoch s oboma vstupnými dátovými sadami bolo použitých niekoľko klasifikačných algoritmov. Experimenty s algoritmom rozhodovacieho stromu dosiahli klasifikačnú presnosť **99,76 %** s celkovým F-score **83 %** a najlepší bayesovský klasifikátor **99,83 %** presnosť s celkovým F-score **87,81 %**. Najlepšie rysy (a tým i najlepšie metriky) pre detekciu buffer overflow útokov zo sieťového toku, vyšli aproximačné funkcie. V najlepších 24 rysoch podľa celkovej presnosti klasifikácie nad 99,5 % (zobrazené na obrázku 3) bayesovským klasifikátorom sa nachádzajú polynómy až 17 krát (najviac 13. stupňa, ktorý má zastúpenie 5 krát). Vzhľadom na smer komunikácie je v rámci aproximácie polynómami vo väčšej miere zastúpená aproximácia prichádzajúcej (smer

od zdroja k cieľu) komunikácie a to z dôvodu prítomnosti veľkého paketu nesúceho obsah útoku na prijímajúci buffer služby.

5 Záver

V tejto práci bol predstavený stručný úvod do problematiky sieťovej bezpečnosti a základné rozdelenie sieťových bezpečnostných technológií. V úvode boli vytýčené ciele práce a požiadavky na návrh autonómneho detekčného systému so zameraním na behaviorálnu analýzu sieťového toku a rozpoznanie anomálií. Hlavným cieľom práce je vytvorenie konceptu detekčného systému, ktorý pomocou definovaných metrík redukuje sieťový tok na signatúry spojení s dôrazom na autonómnosť systému pomocou vytvorenia expertnej znalosti honeypot systému, ako učiteľa a nezávislosť na technologických aspektoch analyzovaných dát (ako napr. šifrovanie, použité protokoly, technológie, či prostredie).

5.1 Prínos práce

Prínosy práce je možné sumarizovať pomocou definovaného cieľa práce v úvode kapitoly, ale celkový prínos je možné rozdeliť do niekoľkých bodov, ktoré lepšie poskytnú prehľad o dosiahnutých výsledkoch.

Jedným z hlavných prínosov práce je zakomponovanie behaviorálnej charakteristiky spojenia na zlepšenie detekcie špecifických sieťových útokov (so zameraním na buffer overflow útoky). Tento cieľ bol dosiahnutý vytvorením špecifikácie signatúry spojenia, ktorá obsahuje koeficienty aproximácií priebehu spojenia variabilných atribútov charakteristiky sieťového toku. Celkové výsledky naznačujú vysokú mieru klasifikácie týchto útokov práve pomocou behaviorálnej časti signatúry až do 99,83 % presnosti klasifikácie. Práca poskytuje ucelený formalizovaný prístup k definícii charakteristiky sieťového toku, metrík ako funkcií, ktoré z definovanej charakteristiky vytvárajú signatúru spojenia, modelu detekčného systému a to s dôraznosťou na exaktnosť definícií a rozšíriteľnosť systému o ďalšie funkcie a metódy. Tento formálny model obsahuje všetky atribúty sieťového toku a je možné nad ním vystavať ďalšie detekčné mechanizmy, podobne je možná redukcia atribútov, prípadne metrík pre minimalizáciu veľkosti signatúry a zrýchlenie klasifikačných algoritmov. Práca ďalej prináša ďalší pohľad na detekciu komplikovanejších útokov pomocou kontextu spojenia, ktorým je možné modelovať správanie analyzovaných uzlov (systémov) v sieti počas určitého časového obdobia, väzby medzi sieťovými prvkami a pozorovať tak vzory správania analyzovaného spojenia alebo systému. Ďalším prínosom je zakomponovanie autonómnosti ako prvku detekčného systému pomocou nasadeného honeypot systému. Vytvorenie väzby expertnej znalosti honeypot systému v roli učiteľa klasifikačných algoritmov vytvára autonómnosť systému pri detekcii i neznámych útokov a možnosť samostatného učenia (v zmysle bez zásahu človeka) na základe poznatkov zbieraných z honeypot systémov. V neposlednom rade bola počas práce s vedeckými prácami a publikáciami o sieťových detekčných systémoch a metódach vypracovaná štúdia, ktorá sumarizuje dosiahnuté výsledky za posledných 20 rokov s poukázaním na problémy, ktorým čelia výskumné práce v tejto oblasti.

Popisované výsledky práce nadväzujú a rozširujú aktuálny stav bádania v oblasti sieťovej bezpečnosti o behaviorálne aspekty komunikácie pomocou aproximačných funkcií a tým zlepšujú detekčné schopnosti pre špecifické útoky. Práca ďalej prináša nové

využitie honeypot systémov, ktoré vystupujú ako expertné systémy zahrnuté do kontinuálneho procesu automatického učenia detekčných algoritmov, čím zvyšujú autonómnosť celého systému. Jedným z prínosov je i vytvorenie formalizovaného prístupu k definícii charakteristiky spojenia, metrík a behaviorálnej signatúry, na ktorom je možné stavať ďalšie rozšírenia detekčných techník.

5.2 Zhodnotenie

Koncept detekčného systému s využitím aproximácií priebehu spojení je podľa uvedených výsledkov experimentov zaujímavým riešením a to i z pohľadu konceptu honeypotov, ktoré nie sú bežne používanou technológiou v detekčných systémoch.

Najlepšie uvedené výsledky obdobných analyzovaných výskumných projektov dosiahli systémy *AHM-NID* (99,60 %), alebo *Octopus-IIDS* (97,40 %), ale len s analýzou obsahov paketov, ktorá nie je vždy možná. Medzi riešenia analyzujúce atribúty paketov je možné zaradiť výskumnú prácu Moore [34, 33, 3], ktorý dosiahol 95 % presnosť nad vlastnou dátovou sadou. Navrhnutý systém pomocou behaviorálnej signatúry dosiahol v experimentoch celkovú presnosť klasifikácie 99,83 %. Nie je účelom tejto práce tvrdiť, že systém je v účinnosti detekcie najlepší alebo, že navrhnutý koncept je ten správny, ale dosiahnuté výsledky poukazujú na zaujímavý prístup analýzy sieťového toku.

5.3 Budúcnosť

Práca predstavuje koncept detekčného systému, ktorý predstavuje základ pre ďalšie výskumné práce, rozšírenia detekčných techník, zlepšenie modelu z pohľadu optimalizácie apod.

Aktuálny koncept systému pracuje nad TCP/IPv4 architektúrou. V minulosti ale boli zaznamenané útoky na protokoly pracujúce nad UDP. Príkladom môže byť zraniteľná služba *Sentinel LM* v programe *Sentinel License Manager 7.2.0.2*, ktorá umožňuje útočníkovi spustiť vlastný vytvorený kód na vzdialenom systéme pomocou zaslania veľkého dátového toku na UDP port 5093. Sieťový tok na UDP protokole je problematický pri identifikácii začiatku a konca spojenia, ale aj charakteristika spojenia môže byť pri tej istej službe vždy iná, keďže k doručeniu niektorých paketov nemusí počas spojenia dôjsť. UDP spojenia sú väčšinou služby, ako VoIP alebo streamovacie služby (video, hudba), ktoré môžu trvať dlhý časový interval (prípadne môžu byť z pohľadu detekčného systému nekonečné), čo vytvára novú požiadavku detekčného systému a to analýza spojení počas ich priebehu.

Vytvorená signatúra spojenia obsahuje definovanými sieťovými metrikami spracovanú kompletnú charakteristiku spojenia. Tá bola pri rozšírení modelu na architektúru TCP/IPv4 zmenšená o niektoré nepotrebné atribúty TCP a IP paketov. I tak ale signatúra stále obsahuje 657 hodnôt metrík a dosahuje veľkosť 1,9 kB na spojenie, čo je v prípade nasadenia systému na vysoko-rýchlostné siete veľké číslo. Pri optimalizácii (čo najväčšom zmenšení výslednej signatúry) by ale nemalo dochádzať k znižovaniu kvality signatúry (strate informácií). Jednou z možností je i optimalizácia požitých algoritmov, spracovanie tokov alebo aproximačných algoritmov presunom do špecializovaného hardware.

Vzhľadom na neexistenciu vhodnej databázy útokov, ktorá by spĺňala podmienky, ktoré sú na ňu kladené klasifikačnými experimentami, vznikla po vytvorení laboratórneho prostredia snaha o vytvorenie komplexnej verejnej databázy, ktorá by slúžila

všetkým výskumných skupinám a organizáciám pre experimenty s detekčnými nástrojmi a metódami. Táto databáza by mala byť automatizovaná s možnosťou dodania zraniteľného programu a exploitu pre simulovanie útoku a zber dát. Využitím tejto databázy je možné v rámci podmienok stanoviť napr. povinnosť výskumníka zverejniť úplné dosiahnuté výsledky tak, aby bolo možné porovnať účinnosť zvolených metód verejnosťou.

Jedným z problematických súčastí analýzy sieťového toku v reálnom alebo skoro reálnom čase, sú neukončené spojenia. Podľa dostupnej literatúry (RFC 793 [44]) je spojenie ukončené až po 5-tich minútach v prípade, že v tomto čase nedôjde k prijatiu ďalšieho paketu. Tento časový interval (i v prípade, že by bol menší) môže spôsobiť veľké výkonnostné problémy udržiavania celých spojení v pamäti analyzátora a zdržovanie analýzy a následného vyhodnotenia po dobu väčšiu ako 5 minút. To platí i pre neukončené spojenia. Jedným z riešení je vytvorenie kontinuálneho procesu, ktorý po veľmi krátkom intervale (v závislosti na danej sieti, rádovo v sekundách) spúšťa proces vytvorenia charakteristiky a signatúry spojenia, ktoré budú zaradené do analýzy a v prípade, že príde ďalší paket, ktorý patrí do daného spojenia, bude charakteristika i daná signatúra aktualizovaná a proces analýzy reštartovaný. Tento prístup, ani iné prístupy pre vyriešenie tohoto problému neboli doteraz otestované.

Ďalší výskum nad navrhnutým konceptom detekčného nástroja sa môže vzťahovať na využitie, prípadne úpravu lokálneho a globálneho kontextu spojenia, zdrojového a cieľového systému pre lepšiu detekciu pokročilých útokov. S tým súvisia abstraktnejšie úrovne detekčných metód, ako napríklad korelačné pravidlá, či modelovanie psychologického profilu systémov (rozpoznanie útoku na základe zmeny správania človeka). V práci ďalej chýba zapojenie kontextu do experimentov, či možnosť analyzovať špecifické protokoly, ktoré sú na vzostupe obdobných výskumných projektov (napr. protokoly riadiacich systémov ICS). Jednou z ďalších oblastí, ktoré stoja za pozornosť v budúcnosti projektu je pridanie možnosti aktívneho zabránenia prienikov do siete, prípadne zefektívnenie detekcie a zameranie sa na abstraktnejšie metódy (korelácie, modelovanie vzťahov apod).

Literatúra

[1] Aha, D. W.; Kibler, D.; Albert, M. K.: Instance-based learning algorithms. *Machine learning*, ročník 6, č. 1, 1991: s. 37–66.

[2] Anderson, D.; Lunt, T. F.; Javitz, H.; a kol.: *Detecting unusual program behavior using the statistical component of the Next-generation Intrusion Detection Expert System (NIDES)*. SRI International, Computer Science Laboratory, 1995.

[3] Auld, T.; Moore, A. W.; Gull, S. F.: Bayesian neural networks for internet traffic classification. *Neural Networks, IEEE Transactions on*, ročník 18, č. 1, 2007: s. 223–239.

[4] Barabas, M.; Homoliak, I.; Drozd, M.; a kol.: Automated Malware Detection Based on Novel Network Behavioral Signatures. *International Journal of Engineering and Technology*, ročník 5, č. 2, 2013: str. 249.

[5] Barbara, D.; Wu, N.; Jajodia, S.: Detecting Novel Network Intrusions Using Bayes Estimators. In *SDM*, SIAM, 2001, s. 1–17.

[6] Bolzoni, D.; Zambon, E.; Etalle, S.; a kol.: Poseidon: A 2-tier anomaly-based intrusion detection system. *arXiv preprint cs/0511043*, 2005.

[7] Bonner, L.: Cyber Risk: How the 2011 Sony Data Breach and the Need for Cyber Risk Insurance Policies Should Direct the Federal Response to Rising Data Breaches. *Wash. UJL & Pol'y*, ročník 40, 2012: str. 257.

[8] Breiman, L.: Random forests. *Machine learning*, ročník 45, č. 1, 2001: s. 5–32.

[9] Burges, C. J.: A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, ročník 2, č. 2, 1998: s. 121–167.

[10] Cao, Y.; Wu, J.: Projective ART for clustering data sets in high dimensional spaces. *Neural networks*, ročník 15, č. 1, 2002: s. 105–120.

[11] Cisco Systems, I.: Cisco 2014 Annual Security Report. 2014.

[12] Cortes, C.; Vapnik, V.: Support-vector networks. *Machine learning*, ročník 20, č. 3, 1995: s. 273–297.

[13] Cyberwar: War in the Fifth Domain.
URL `http://www.economist.com/node/16478792`

[14] Elkan, C.: Results of the KDD'99 classifier learning. *ACM SIGKDD Explorations Newsletter*, ročník 1, č. 2, 2000: s. 63–64.

[15] Ertoz, L.; Eilertson, E.; Lazarevic, A.; a kol.: Detection and summarization of novel network attacks using data mining. *Minnesota INtrusion Detection System (MINDS) Technical Report*, 2003.

[16] Freund, Y.; Schapire, R. E.: A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, ročník 55, č. 1, 1997: s. 119–139.

[17] Gunn, S. R.; a kol.: Support vector machines for classification and regression. *ISIS technical report*, ročník 14, 1998.

[18] Haijun, X.; Fang, P.; Ling, W.; a kol.: Ad hoc-based feature selection and support vector machine classifier for intrusion detection. In *Grey Systems and Intelligent Services, 2007. GSIS 2007. IEEE International Conference on*, IEEE, 2007, s. 1117–1121.

[19] Han, J.; Kamber, M.: Data Mining: Concepts and Techniques. *University of Illinois at Urbana-Champaign,*, 2006.

[20] Hettich, S.; Bay, S.: Kdd cup 1999 data. *The UCI KD Archive, Irvine, CA: University of California, Department of Information and Computer Science*, 1999.

[21] Hsu, F.-H.; Chiueh, T.-c.: CTCP: a transparent centralized tcp/ip architecture for network security. In *Computer Security Applications Conference, 2004. 20th Annual*, IEEE, 2004, s. 335–344.

[22] ISACA: Advanced Persistent Threat Awareness Study Results. Technická správa, ISACA, 2014.

[23] Jacobson, V.; Leres, C.; McCanne, S.: libpcap, Lawrence Berkeley Laboratory, Berkeley, CA. *Initial public release June*, 1994.

[24] Jauregui, J.: Principal component analysis with linear algebra. 2012.

[25] Kantardzic, M.: *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.

[26] Khosravifar, B.; Bentahar, J.: An experience improving intrusion detection systems false alarm ratio by using honeypot. In *Advanced Information Networking and Applications, 2008. AINA 2008. 22nd International Conference on*, IEEE, 2008, s. 997–1004.

[27] Kim, H.-A.; Karp, B.: Autograph: Toward Automated, Distributed Worm Signature Detection. In *USENIX security symposium*, ročník 286, San Diego, CA, 2004.

[28] Kreibich, C.; Crowcroft, J.: Honeycomb: creating intrusion detection signatures using honeypots. *ACM SIGCOMM Computer Communication Review*, ročník 34, č. 1, 2004: s. 51–56.

[29] Kushner, D.: The real story of stuxnet. *IEEE Spectrum*, ročník 50, č. 3, 2013: s. 48–53.

[30] Liang, Z.; Sekar, R.: Automatic generation of buffer overflow attack signatures: An approach based on program behavior models. In *Computer Security Applications Conference, 21st Annual*, IEEE, 2005, s. 10–pp.

[31] Mafra, P. M.; Moll, V.; da Silva Fraga, J.; a kol.: Octopus-IIDS: An anomaly based intelligent intrusion detection system. In *Computers and Communications (ISCC), 2010 IEEE Symposium on*, IEEE, 2010, s. 405–410.

[32] Marill, T.; Green, D. M.: On the effectiveness of receptors in recognition systems. *Information Theory, IEEE Transactions on*, ročník 9, č. 1, 1963: s. 11–17.

[33] Moore, A.; Zuev, D.; Crogan, M.: *Discriminators for use in flow-based classification*. Queen Mary and Westfield College, Department of Computer Science, 2005.

[34] Moore, A. W.; Zuev, D.: Internet traffic classification using bayesian analysis techniques. In *ACM SIGMETRICS Performance Evaluation Review*, ročník 33, ACM, 2005, s. 50–60.

[35] Newsome, J.; Song, D.: Dynamic taint analysis for automatic detection, analysis, and signature generation of exploits on commodity software. 2005.

[36] NoAH FP6 EU Project. 2008.
URL `http://www.fp6-noah.org/index.html`

[37] Norton-Taylor, R.: Titan Rain: How Chinese Hackers Targeted Whitehall. *The Guardian*, ročník 5, 2007.

[38] Oja, E.; Kaski, S.: *Kohonen maps*. Elsevier, 1999.

[39] Panda, M.; Abraham, A.; Patra, M. R.: A hybrid intelligent approach for network intrusion detection. *Procedia Engineering*, ročník 30, 2012: s. 1–9.

[40] Pasupulati, A.; Coit, J.; Levitt, K.; a kol.: Buttercup: On network-based detection of polymorphic buffer overflow vulnerabilities. In *Network Operations and Management Symposium, 2004. NOMS 2004. IEEE/IFIP*, ročník 1, IEEE, 2004, s. 235–248.

[41] Paxson, V.: Bro: a system for detecting network intruders in real-time. *Computer networks*, ročník 31, č. 23, 1999: s. 2435–2463.

[42] Peter, E.; Schiller, T.: A practical guide to honeypots. *Washington Univerity*, 2011.

[43] Portokalidis, G.; Slowinska, A.; Bos, H.: Argos: an emulator for fingerprinting zero-day attacks for advertised honeypots with automatic signature generation. In *ACM SIGOPS Operating Systems Review*, ročník 40, ACM, 2006, s. 15–27.

[44] Postel, J.: Transmission Control Protocol. RFC 793, IETF, September 1981.
URL `https://tools.ietf.org/html/rfc793`

[45] RapidMiner.
URL `https://rapidminer.com/`

[46] "Red October"Diplomatic Cyber Attacks Investigation.
URL `https://securelist.com/analysis/publications/36740/red-october- diplomatic-cyber-attacks-investigation/`

[47] Sangster, B.; O'Connor, T.; Cook, T.; a kol.: Toward Instrumenting Network Warfare Competitions to Generate Labeled Datasets. In *CSET*, 2009.

[48] Shin, S.; Lee, S.; Kim, H.; a kol.: Advanced probabilistic approach for network intrusion forecasting and detection. *Expert Systems with Applications*, ročník 40, č. 1, 2013: s. 315–322.

[49] Tankard, C.: Advanced Persistent threats and how to monitor and deter them. *Network security*, ročník 2011, č. 8, 2011: s. 16–19.

[50] Tavallaee, M.; Bagheri, E.; Lu, W.; a kol.: A detailed analysis of the KDD CUP 99 data set. In *Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications 2009*, 2009.

[51] Tavallaee, M.; Stakhanova, N.; Ghorbani, A. A.: Toward credible evaluation of anomaly-based intrusion-detection methods. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, ročník 40, č. 5, 2010: s. 516–524.

[52] THE SNORT PROJECT. Snort, The OpenSource Network Intrusion Detection System.
URL `http://www.snort.org/`

[53] Tzur-David, S.; Avissar, H.; Dolev, D.; a kol.: SPADE: Statistical Packet Acceptance Defense Engine. In *High Performance Switching and Routing (HPSR), 2010 International Conference on*, IEEE, 2010, s. 119–126.

[54] Utgoff, P. E.: Incremental induction of decision trees. *Machine learning*, ročník 4, č. 2, 1989: s. 161–186.

[55] Wang, H. J.; Guo, C.; Simon, D. R.; a kol.: Shield: Vulnerability-driven network filters for preventing known vulnerability exploits. *ACM SIGCOMM Computer Communication Review*, ročník 34, č. 4, 2004: s. 193–204.

[56] Wang, K.; Parekh, J. J.; Stolfo, S. J.: Anagram: A content anomaly detector resistant to mimicry attack. In *Recent Advances in Intrusion Detection*, Springer, 2006, s. 226–248.

[57] Wang, K.; Stolfo, S. J.: Anomalous payload-based network intrusion detection. In *Recent Advances in Intrusion Detection*, Springer, 2004, s. 203–222.

[58] Wang, L.; Li, Z.; Chen, Y.; a kol.: Thwarting zero-day polymorphic worms with network-level length-based signature generation. *IEEE/ACM Transactions on Networking (TON)*, ročník 18, č. 1, 2010: s. 53–66.

[59] Whitney, A. W.: A direct method of nonparametric measurement selection. *Computers, IEEE Transactions on*, ročník 100, č. 9, 1971: s. 1100–1103.

Životopis

Osobné údaje

Meno:	Maroš Barabas
Národnosť	slovenská
Dátum narodenia	4. júna 1985
E-mail:	`ibarabas@fit.vutbr.cz`
Web:	`www.fit.vutbr.cz/ ibarabas`
Linkedin:	`https://www.linkedin.com/in/mbarabas`

Vzdelanie

od 2009	Fakulta informačních technologií VUT v Brně, doktorský študijny program Výpočetní technika a informatika
2007 – 2009	Fakulta informačních technologií VUT v Brně, magisterský študijny program Informační technologie, DP: Nástroj pro snazší zabezpečení počítačů s OS Linux
2004 – 2007	Fakulta informačních technologií VUT v Brně, bakalársky študijny program Informační technologie, BP: GUI pro konfiguraci FTP serveru
2000 – 2004	Gymnázium Pierra de Coubertina, Piešťany, Slovensko

Kariéra

od 2014	Head of Security Technologies Division, AEC a.s., vedúci tímu bezpečnostných technológií
2012 – 2014	Senior IT bezpečnostní konzultant, AEC spol. s r.o., Počítačova bezpečnost, audit operačních systémů, penetrační testování, pokročilé bezpečnostní technologie
2011 – 2013	Technicko-věděcký pracovník, Vysoké učení technické v Brně, Fakulta informačních technologií, Automatizované zpracování útoků, FR-TI1/037 (spoluriešiteľ)
2006 – 2011	Software Engineer, Red Hat Inc., Udržování a vývoj FTP serverů a klientů, Dizajn a vývoj projektov `Sectool`, `OpenSCAP`, `SCAP-Workbench`

Certifikácie

Certified Ethical Hacker, EC-Council, Certified Ethical Hacker 7, 2013
Red Hat Certified Technician, Red Hat Certified System Administrator, 2007

Publikácie

2 kapitoly do knihy
2 publikácie v DBLP
5 publikácií v Scopus
4 publikácie v recenzovanom neimpaktovanom priodiku (DSM 1211-8737)
7 publikácií na medzinárodnej konferencii
5 citácií

Produkty

Vysoce interaktivní honeypot s taint analýzou, software, 2013

Projekty

Spolehlivost a bezpečnost v IT, VUT v Brně, FIT-S-14-2486, 2014-2016
Pokročilé bezpečné, spolehlivé a adaptivní IT, VUT v Brně, FIT-S-11-1, 2011-2013
Automatizované zpracování útoků, MPO ČR - TIP, FR-TI1/037, 2009-2013

Ďalšie aktivity

Podaný patent v rámci projektu MPO ČR - TIP, FR-TI1/037, 2009-2013 (zatiaľ neprijatý)

Abstrakt

Táto práca sa zameriava na popis aktuálneho stavu bádania v detekčných metódach sieťových útokov a následne vylepšenie schopnosti detekcie špecifických útokov vytvorením formálneho popisu sieťových metrík, ktoré aproximujú priebeh sieťového spojenia a vytvárajú signatúru založenú na behaviorálnej charakteristike analyzovaného spojenia. Cieľom práce nie je prevencia voči aktuálne prebiehajúcim útokom, ani reakcia na tieto útoky, dôraz sa kladie na analýzu spojenia, získania čo najviac informácií a vytvorenie základu detekčného systému, ktorý dokáže minimalizovať veľkosť dát zbieraných zo siete s ponechaním najdôležitejších informácií pre nasledujúcu analýzu. Hlavným cieľom práce je vytvorenie konceptu detekčného systému, ktorý pomocou definovaných metrík redukuje sieťový tok na signatúry spojení s dôrazom na behaviorálne aspekty komunikácie, zvyšuje autonómnosť detekčného systému pomocou vytvorenia expertnej znalosti z honeypot systému, s podmienkou nezávislosti na technologických aspektoch analyzovaných dát (ako napr. šifrovanie, použité protokoly, technológie, či prostredie). Vytvorenie konceptu expertnej znalosti honeypot systému v roli učiteľa klasifikačných algoritmov vytvára autonómnosť systému pri detekcii i neznámych útokov a možnosť samostatného učenia (v zmysle bez zásahu človeka) na základe poznatkov zbieraných z útokov na tieto systémy. V práci je predstavený postup vytvorenia laboratórneho prostredia a experimenty s definovanou signatúrou spojenia nad získanými dátami i nad prevzatou testovacou databázou. Dosiahnuté výsledky sú v závere porovnané s aktuálnym prehľadom sieťových detekčných systémov s vyzdvihnutím prínosu navrhnutých aproximačných metód priebehu analyzovaného spojenia.