Detekcia sieťových anomálií a bezpečnostných incidentov s využitím DNS dát

Michal Kováčik
Výpočetní technika a informatika, 2. ročník, prezenční studium
Školitel: Jan Kořenek

Fakulta informačních technologií, Vysoké učení technické v Brně
Božetěchova 1/2, 612 66 Brno

ikovacik@fit.vutbr.cz

Abstrakt. Služba DNS je kritická pre normálne fungovanie Internetu a taktiež množstva dostupných služieb. Väčšina komunikácie na Internete totiž využíva v istej fáze práve DNS. Okrem jej základnej úlohy sa často stáva terčom zneužitia pri množstve rôznych škodlivých aktivít. Táto práca sa zaoberá nežiaducimi aktivitami spájajúcimi sa so službou DNS a jej zneužitím, ktoré sú priblížené spolu s mojim vlastným prístupom k ich detekcii. Najvýznamnejšou časťou práce je kapitola o dizertačnej práci, ktorá špecifikuje vytýčené ciele, približuje spôsob ich dosiahnutia a súčasný stav.

Kľúčové slová. detekce anomálií, bezpečnostní incidenty, DNS útoky, monitorování provozu

1 Úvod

Požiadavky na správu a bezpečnosť počítačových sietí neustále rastú spolu s ich rozvojom. Vysoká dôležitosť sa kladie hlavne dostupnosti služieb a diskrétnosti prenášaných informácií. Rozvíjajú sa však aj útoky a ich počet má stúpajúcu tendenciu. Tento narastajúci trend potvrdzujú aj spoločnosti ako napríklad NSFOCUS[1] alebo Symantec[2] zaoberajúce sa internetovou bezpečnosťou vo svojich výročných správach. Sila a počet útokov na DNS alebo zneužívajúcich službu DNS sa za posledné roky pravidelne takmer zdvojnásoboval, čo dokazuje stúpajúcu popularitu zahrnutia tejto služby do útokov. Monitorovanie sietí za účelom detekcie a zamedzenia sieťových anomálií si vyžaduje stále viac pozornosti.

Služba DNS (Domain Name System)[3] je z pohľadu štruktúry hierarchický systém doménových mien. Hlavnou funkciou služby je preklad doménových mien na IP adresy a opačne, vykonávaný rezolúciou. V skutočnosti služba pracuje s množstvom rôznych dotazov a je možné ju považovať za distribuovanú databázu sieťových informácií, ktorej uzlami sú menné servery. Protokol pracuje na jednoduchom princípe dotazu a odpovede a komunikácia sa vyznačuje symetrickosťou. To znamená že by mala existovať odpoveď na každý zaslaný dotaz, čo však v praxi kvôli protokolu UDP nie je možné zaručiť. Protokol DNS samotný nepoužíva šifrovanie a jeho autentifikácia pomocou zdrojovej IP adresy, portu a transakčného ID je veľmi jednoduchá.

Dôležitosť DNS je zrejmá aj útočníkom, ktorí protokol používajú za nedovolenými účelmi na škodlivé aktivity, prípadne zneužívajú vlastnosti DNS. Bezpečnostné opatrenia v množstve sietí bývajú k DNS

[1] spoločnosť zaoberajúca sa medzinárodnou webovou a sieťovou bezpečnosťou <http://www.nsfocus.com.au/>
[2] spoločnosť poskytujúca bezpečnostné sieťové riešenia <http://www.symantec.com/>
[3] <https://www.ietf.org/rfc/rfc1034.txt>, <https://www.ietf.org/rfc/rfc1035.txt>

prevádzke veľmi benevolentné, čo je obrovskou výhodou pre útočníkov, pre ktorých môže byť DNS prístupovou cestou aj do sietí s vysokým zabezpečením, ktoré sú konfigurované veľmi prísne voči ostatným službám. Útočníci využívajú rôzne techniky ako napríklad častá zmena doménového mena pre vyhnutie sa blokovaniu prístupu, podvrhnutie odpovede na dotaz klienta, zneužitie protokolu na posielanie odlišného typu dát a podobne. DNS komunikácia prebieha tiež medzi stanicami v botnetom.

Nasledujúca kapitola 2 sa venuje problematike zdrojových dát. Kapitola 3 sa venuje konkrétnym DNS anomáliám a ich detekcii. V ďalšej kapitole 4 sa nachádza formulácia cieľa mojej dizertačnej práce, spolu so spôsobmi jeho dosiahnutia. Záverečná kapitola 5 je súhrom tohto príspevku.

2 Monitoring a zdrojové dáta

Dôležitým faktorom pre voľbu detekčnej metódy je typ dostupných zdrojových dát. Na základe dostupného typu zdrojových dát je potom možné odhadovať presnosť a rýchlosť detekčnej metódy. V dnešnej dobe je veľmi populárnym riešením používanie tokových dát (NetFlow[4]). Tento spôsob monitorovania sa pre DNS, ktorý je aplikačným protokolom, javí pri niektorých typoch anomálií ako nedostatočný. Pri DNS sú vo väčšine prípadov veľmi dôležité dáta z položiek aplikačnej vrstvy, ktoré flow dáta neobsahujú. Najideálnejším riešením by samozrejme bolo zaznamenávanie celých paketov (Deep packet inspection), no analýza takýchto paketov by vyžadovala obrovské nároky na výpočtový výkon a rovnako obrovský priestor pre ukladanie zaznamenaných dát. Dôležitou požiadavkou pri monitoringu DNS je však aj efektivita monitorovania a spracovania prevádzky. Nutnosťou je teda hľadanie kompromisu medzi monitorovaním tokov a kompletných paketov.

Na základe možností, ktoré ponúka protokol IPFIX[5] (Internet Protocol Flow Information eXport), by práve jeho použitie malo byť strednou cestou zahŕňajúcou efektívny monitoring ako aj možnosti analýzy vybraných položiek aplikačných protokolov. Zdrojom IPFIX dát, ktorý používam sú dáta z DNS pluginu [5] pre FlowMon Exportér od spoločnosti INVEA[6], ktorý som vyvíjal. Týmto spôsobom mám k dispozícií vybrané položky z aplikačnej vrstvy paketov DNS prevádzky.

3 DNS anomálie a detekčné metódy

Anomálie DNS je možné rozdeliť do kategórií podľa niekoľkých faktorov. V tejto kapitole sa zameriam iba na vybrané typy anomálií, niekoľko vybraných detekčných metód a vlastný prístup k nim v rámci mojej dizertačnej práce. Úplné rozdelenie, detailný popis jednotlivých anomálií a metód detekcie, a popis viacerých typov anomálií som zhrnul v tézach [6]. Ďalšie informácie som čerpal z [9].

3.1 DNS Amplification

Je najpopulárnejším z útokov, ktorý službu DNS zneužíva. Útok sa skladá z dvoch hlavných častí. Prvou je spoofing[7] zdrojovej IP adresy, druhou je vygenerovanie dotazu, ktorý spôsobí čo najväčšiu odpoveď. Vzhľadom k tomu, že sa pri tomto útoku generuje obrovské množstvo dotazov a zneužívané DNS servery odpovedajú mnohonásobne väčšími odpoveďami, je možné tento útok detekovať už pomocou tokových dát vo forme NetFlow.

Na detekciu útoku je možné použiť mnoho zaujímavých metód, ako príklad vyberiem metódu založenú na NetFlow dátach [1], ktorá funguje pomocou jednoduchých prahov. Metoda vyniká jednoduchosťou a rýchlosťou, jej presnosť však nie je ideálna, pretože generuje priveľa falošných poplachov. Na základe

[4] definovaný v <http://www.ietf.org/rfc/rfc3954.txt>
[5] definovaný v <http://www.ietf.org/rfc/rfc5101.txt>
[6] viac na <https://www.invea.com/sk/go/flowmon>
[7] podvrhnutie

tejto metódy som v spolupráci so združením CESNET implementoval vlastnú, ktorá detekuje útok na základe homogenity dotazov a odpovedí, asymetrickej veľkosti dotazov a odpovedí a početnosti dotazov. Pri relatívne zachovanej jednoduchosti bola dosiahnutá oveľa vyššia presnosť detekcie. Metóda je nasadená ako detekčný modul v systéme NEMEA [2]. Z ďalších prístupov k detekcii, ktoré som analyzoval je možné spomenúť detekciu na základe metódy podobnosti a entropie. Metódy sa ukázali ako úspešné a sú schopné detekovať útok, nevýhodou však je nutnosť dodania vhodných referenčných dát.

Ako možnú alternatívu detekcie amplifikačného útoku som skúmal súvislosť s položkami *DNSSEC OK* a *UDP payload size*, ktoré sú súčasťou rozšírenia EDNS0[8]. Obsah týchto položiek však nie je možné priamo spojiť s útokmi, keďže väčšina DNS prevádzky používajúca EDNS0 pracuje s hodnotami položiek, ktoré boli predpokladané v prítomnosti útoku. Pre zlepšenie presnosti detekcie a potvrdenie útoku, je možné použiť mnou publikovaný detektor podvrhnutých adries na sieti [7], čo priblížim v 4.1.

3.2 DNS tunneling

Hlavnou myšlienkou je zapuzdrenie dát do klasickej DNS prevádzky, ktorá nebýva nijako obmedzovaná. Takto je potom možné tunelovať akúkoľvek prevádzku a obchádzať firewaly, či platené prístupové body do siete. Tunelované pakety sa vyznačujú neobvyklou veľkosťou, veľkou dĺžkou doménového mena, veľkým počtom číslic v názve domény, ktorý býva navyše vygenerovaný.

Tunelovanie vzhľadom k prenášaným paketom mení charakter DNS prevádzky a detekcia je teda za istých okolností možná aj z tokových dát. Použiteľné sú napríklad metódy založené na entropii, podobne ako v [4], kde je takáto metóda použitá na analýzu histogramov veľkostí paketov. Okrem toho autori v tomto prístupe sledujú aj frekvenciu nekonformných paketov. Ďalšou je možnosť analýzy tokových dát štatistickými metódami. V tomto prípade je však nutné správne určiť parametre pre detekciu a tiež hraničné hodnoty pre anomálne správanie. Od toho sa potom odvíja celková presnosť metódy. Každá sieť má iné charakteristiky a preto je vždy najskôr nevyhnutné vykonať analýzu sieťovej prevádzky. Vhodnejšia sa javí analýza obsahu paketov pri ktorej množstvo metód zameriava na zmysluplnosť prenášaných dotazov a odpovedí. Najčastejšia je detekcia pomocou frekvenčnej analýzy v rôznych variantách. Zo všetkých spomeniem frekvenčnú analýzu na jednotlivých bigramoch [8].

Pri vlastnej analýze a detekcii tunelovania pomocou DNS som sa zameriaval v prvom rade na netypické typy odpovedí, ktoré sú používané. Často sa pre prenos používajú hlavne typy TXT, SRV alebo napríklad NULL. Ďalšou sledovanou vlastnosťou bola neprimeraná veľkosť paketov. Význačnou je aj dĺžka doménového mena, ktorá býva oproti bežnej prevádzke dvoj- až troj-násobná. Použitím frekvenčnej analýzy doménového mena je detekcia veľmi úspešná, čo je bohužiaľ na úkor rýchlosti detekcie. Generované doménové mená majú na rozdiel od skutočných približne rovnomerné rozloženie znakov, čo nezodpovedá žiadnemu bežnému jazyku.

3.3 Cache poisoning

Jedná sa o podvrhnutie obsahu cache záznamu na serveri za účelom presmerovania. Detekcia je možná aj pomocou štatistickej analýzy DNS, no problémom zostáva generovanie množstva falošných poplachov.

Autori v [4] používajú pre detekciu algoritmus pracujúci s NetFlow, ktorý používa IP adresy zdrojov a cieľov, čísla portov, časy medzi príchodmi jednotlivých paketov a postupnosť udalostí. Algoritmus zaznamenáva prichádzajúce dotazy a odpovede a na základe ich postupnosti a početnosti je schopný generovať poplach pri útoku.

Pri vlastných experimentoch som sa zameral na detekciu pomocou krátkej histórie. Metóda sa zameriava na pokusy o uhádnutie transakčného ID a používam v nej transakčné ID dotazu, znenie dotazu, zdrojovú a cieľovú IP adresu a zdrojový port. Unikátne kombinácie dotazov sa zaznamenávajú a uchovávajú. Po príchode zodpovedajúcej odpovede je dotaz odstránený z histórie. Pokiaľ sa líši v transakčnom ID,

[8] Extension mechanisms for DNS <http://www.ietf.org/rfc/rfc2671.txt>

môže sa jednať o narodeninový útok, ktorým je cache poisoning sprevádzaný. Upozornenie sa však hlási až po obdržaní viac ako jedného paketu s rôznym ID, aby sa predchádzalo falošným poplachom. Problémom metódy je efektívne ukladanie histórie v prípade, že je počet dotazov väčší ako počet odpovedí, v tomto prípade môže nekontrolovane rásť množstvo záznamov pre porovnanie. Taktiež má algoritmus problém s niektorými anomálnymi prejavmi v DNS prevádzke, ktoré ale nesúvisia s cache poisoning.

3.4 Škodlivé domény

So škodlivými doménami sa spája používanie techniky fast-flux, ktorá dovoľuje zneužiť vlastnosti DNS na sťaženie zablokovania domén. Pre tento typ anomálneho správania obsahuje NetFlow nedostatočnú informáciu pre detekciu a jedinou možnosťou je v tomto prípade použitie formátu zdrojových dát obsahujúceho aj vybrané položky z aplikačnej vrstvy. Okrem úplných paketových dát sa ideálne ponúka IPFIX obohatený o aplikačné dáta, ktorý by obsahoval napríklad kľúčové položky ako TTL, dotazované doménové mená a podobne.

Autori v [10] sa zameriavajú na domény, na ktoré chodí abnormálny alebo koncentrovaný počet dotazov a na detekciu dotazov na neexistujúce doménové mená (NXDOMAIN). Detekcia odpovedí NXDOMAIN sa pritom javí ako pomerne úspešná. Okrem toho existuje viacero prác, ktorých výsledkom je reputačný systém na základe pasívnej analýzy DNS prevádzky. Jedným z nich je aj [3], kde autori extrahujú z DNS prevádzky 15 význačných príznakov na ktoré sa zameriavajú. Vhodným doplnkom každej metódy na detekciu domén je kontrola voči Blacklistom.

Pri vlastných experimentoch som sa zameral na niekoľko spôsobov určenia škodlivých domén. Analyzované domény predspracovávam rozdelením na jednotlivé úrovne domén a vynechaním častí kratších ako štyri znaky. Takto rozdelené doménové meno je podrobené frekvenčnej analýze. Navyše sa experimentálne snažím pracovať s analýzou skladby slov, ktorá pozostáva z niekoľkých častí. Prvou časťou je analýza dĺžky časti doménového mena, ktorá má hraničnú hodnotu priradenú na základe priemernej dĺžky doménového mena v normálnej prevádzke. Druhou časťou je detektor počtu samohlások, ktorý porovnáva počet samohlások voči počtu písmen. V tretej časti sa sleduje počet opakujúcich sa písmen v názve domény voči jej dĺžke. Posledná štvrtá časť analyzuje počet číslic v doménovom mene.

4 Ciele dizertačnej práce

Moja dizertačná práca sa zameriava na pokrytie nedostatkov existujúcich metód a tým o dosiahnutie lepších výsledkov v oblasti detekcie. Jednotlivé metódy pracujú s rôznym typom vstupných dát, prípadne využívajú iba podmnožinu dostupných informácií. Rôzne vstupné dáta často vedú k rôznym stupňom efektivity a presnosti pri detekcii. Z toho dôvodu v rámci mojej práce, využívam spoločne zdrojové dáta vo formáte NetFlow (tokové), IPFIX (obohatené o aplikačnú vrstvu) a plné paketové dáta. Pritom sa snažím nájsť čo najlepšiu rovnováhu v ich súčinnosti pre potreby posilnenia bezpečnosti počítačových sietí. Rovnako sa v rámci práce snažím o čo najlepšiu efektivitu detekčných metód a ich univerzálnosť. Cieľ mojej dizertačnej práce som formuloval vo vlastných tézach [6] a jeho znenie je:

S využitím kombinácie a korelácie zdrojových DNS dát s kompletným obsahom paketov (Deep packet inspection) a NetFlow/IPFIX dát (IP Flow monitoring) zefektívniť detekciu anomálií a bezpečnostných incidentov v DNS dátach s ohľadom na jej rýchlosť a presnosť.

Hlavný cieľ, ktorý som formuloval je možné rozčleniť na niekoľko menších cieľov, esenciálnych pre jeho dosiahnutie:

1. Analýza dostupných zdrojových DNS dát pomocou rôznych variant korelácie.

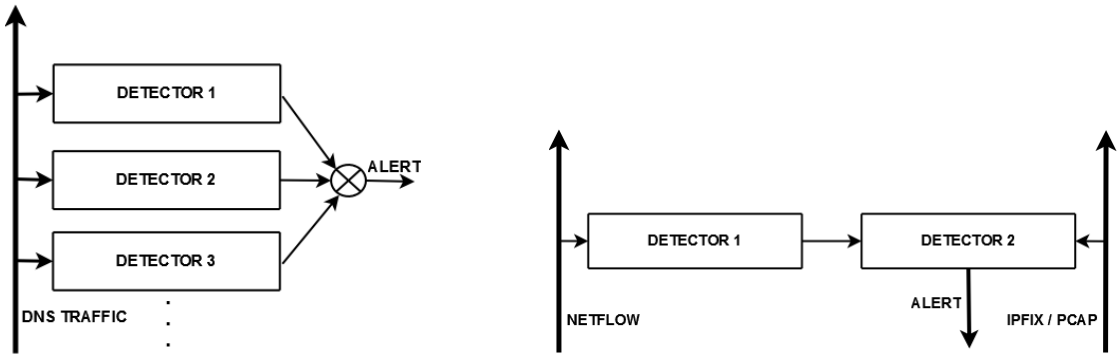
2. Určenie kľúčových metrík potrebných pre detekciu jednotlivých typov anomálií.
3. Návrh optimalizovaných detekčných metód.
4. Návrh vhodného spôsobu kombinácie výsledkov z jednotlivých detekčných metód.
5. Implementácia systému realizujúceho vybrané metódy.
6. Experimentálne vyhodnotenie dosiahnutých výsledkov.

4.1 Spôsob riešenia

Hlavnou myšlienkou práce je využitie rôznych typov dát spolu s DNS dátami za účelom vytvorenia sady detektorov pre rôzne typy anomálií, ktorých efektivita a presnosť bude vyššia než pri obyčajných detektoroch. Pre každú anomáliu môže koexistovať niekoľko detektorov, ktoré navzájom spolupracujú. Možné spôsoby spolupráce sú načrtnuté na Obrázku 1.

Ukážka vľavo na obrázku predstavuje spoluprácu na princípe potvrdenia incidentu a teda spresnenia detekcie. Ako príklad môžem uviesť mnou implementované riešenie dvoch detektorov. Prvý detektor sa snaží odhaliť útok DNS Amplification. Druhý detektor zachytáva v sieti IP spoofing. Koreláciou výsledkov týchto dvoch detektorov sa potvrdí existencia anomálie. V tomto prípade oba detektory pracujú s tokovými dátami.

Ukážka vpravo predstavuje spoluprácu na rozdielnej úrovni zdrojových dát. Jednoduchý detektor avizuje druhému detektoru udalosť, na základe ktorej druhý detektor extrahuje a využije informácie z aplikačnej vrstvy. Konkrétny príklad znova uvediem z vlastnej práce. Prvý, jednoduchý detektor monitoruje a zaznamenáva priebeh SMTP prevádzky. Na jej základe druhý detektor zachytávajúci DNS dáta obohatené o položky aplikačnej vrstvy vo formáte IPFIX dohľadá v prevádzke prípadnú existenciu reverzného dotazovania sa na zdroj SMTP prevádzky a výsledok tohto dotazovania. V prípade negatívnej odpovede je možné zdroj pokladať za škodlivý, kvôli distibúcií nevyžiadanej pošty vo forme spamu.



Obrázok 1: Ukážka spolupráce niekoľkých detektorov.

Koreláciu dát je nutné vykonať z rôznych pohľadov - dáta z rôznych zdrojov, dáta rôznych typov či úrovne. Zaujímavá môže byť aj korelácia na rôznych časovo merateľných intervaloch a na základe rôznych množín. Získané poznatky z korelačných experimentov sú dôležité z pohľadu súvislostí jednotlivých skupín dát, a rovnako aj z pohľadu vhodnosti použitia určitej detekčnej metódy. Na základe dôkladnej analýzy je potom potrebné určiť konkrétne položky dát, ktoré sú pre detekciu daného incidentu nevyhnutné alebo prospešné. Tento krok vedie k návrhu optimalizovaných detekčných metód.

Výsledky jednotlivých detektorov alebo ich častí bude potrebné vhodne kombinovať. Je preto nevyhnutné navrhnúť hierarchiu jednotlivých ukazovateľov a ich podiely na výslednej detekcii. Výsledky niektorých detektorov by napríklad mali byť zohľadnené pri rozhodovaní iných.

5 Záver

Hlavnou úlohou tohto príspevku bolo predstaviť ciele mojej dizertačnej práce a načrtnúť spôsoby ich dosiahnutia. Venoval som sa problematike vhodnosti zdrojových dát a dospel som k záveru, že najlepším riešením je využívanie IPFIX s tokovými dátami obohatenými o položky aplikačnej vrstvy a kombinovanie viacerých typov dát. Ďalej som popísal vlastný prístup k vybraným anomáliám, vybral zaujímavé metódy ich detekcie a priblížil získané poznatky. V kapitole o dizertačnej práci som potom poskytol návrh riešenia spolupráce viacerých detektorov, respektíve detekcie pomocou viacerých typov dát. Momentálne sa venujem optimalizácií a zlepšovaniu algoritmov detekcie, pričom sa snažím nachádzať súvislosti a vyvodiť návaznosti jednotlivých typov dát a výsledkov detektorov.

Poďakovanie

Táto práca bola podporená projektom IT4Innovations Centre of Excellence CZ.1.05/1.1.00/02.0070.

Reference

[1] *Detecting Reflection Attacks in DNS Flows*, ročník 19, University of Twente, 2013.

[2] Bartoš, V.; Žádník, M.; Čejka, T.: Nemea: Framework for stream-wise analysis of network traffic. URL <http://www.cesnet.cz/wp-content/uploads/2014/02/trapnemea.pdf>

[3] Bilge, L.; Kirda, E.; Kruegel, C.; aj.: EXPOSURE : Finding malicious domains using passive DNS analysis. In *NDSS 2011, 18th Annual Network and Distributed System Security Symposium, 6-9 February 2011, San Diego, CA, USA*, 2011.

[4] Karasaridis, A.; Meier-Hellstern, K.; Hoeflin, D.: NIS04-2: Detection of DNS Anomalies using Flow Data Analysis. In *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, 2006, ISSN 1930-529X, s. 1–6.

[5] Kováčik, M.: Liberouter: DNS plugin [online]. [cit. 2014-06-24]. URL <https://www.liberouter.org/technologies/dns-plugin/>

[6] Kováčik, M.: *Detekce síťových anomálií a bezpečnostních incidentů s využitím DNS dat*. Pojednání k tématu disertační práce, Fakulta informačních technologií VUT v Brně, Brno, CZ, 2014.

[7] Kováčik, M.; Kajan, M.; Žádník, M.: Detecting IP-spoofing by modelling history of IP address entry points. In *Emerging Management Mechanisms for the Future Internet*, *Lecture Notes in Computer Science 7943*, ročník 2013, Springer Verlag, 2013, ISBN 978-3-642-38997-9, ISSN 0302-9743, s. 73–83.

[8] Qi, C.; Chen, X.; Xu, C.; aj.: A Bigram based Real Time DNS Tunnel Detection Approach. *Procedia Computer Science*, ročník 17, 2013: s. 852 – 860, ISSN 1877-0509.

[9] Roolvink, S.: Detecting attacks involving DNS servers : A netflow data based approach. 2008. URL <http://essay.utwente.nl/58497/>

[10] Villamarin-Salomon, R.; Brustoloni, J. C.: Identifying Botnets Using Anomaly Detection Techniques Applied to DNS Traffic. In *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*, 2008, s. 476–481.