VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SYSTEMS

DETEKCE NEŽÁDOUCÍHO PROVOZU V LOKÁLNÍ SÍTI

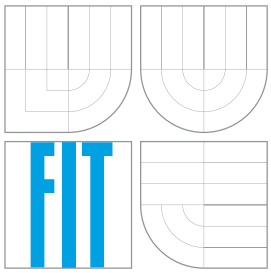
BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

ERIK ŠABÍK

BRNO 2015

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SYSTEMS

DETEKCE NEŽÁDOUCÍHO PROVOZU V LOKÁLNÍ SÍTI

DETECTION OF MALICOUS TRAFFIC IN LOCAL NETWORK

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — ERIK ŠABÍK
AUTHOR

VEDOUCÍ PRÁCE — Ing. MARTIN ŽÁDNIK, Ph.D.
SUPERVISOR

BRNO 2015

Abstrakt

Tato bakalářská práce pojednává o monitorovaní lokálních sítí pomocí sledování síťové komunikace na úrovni IP toků. Popisuje framework Nemea, který slouží na vytváření komplexních systémů pro detekci nežádoucího provozu. Pomocí tohoto frameworku jsou analyzovány tři vzorky dat z rozdílně velkých sítí. Na základě této analýzy je následně vytvořen návrh pro detekci nežádoucího provozu, která v použitém frameworku doposud chybí a je ji možné využít v lokální síti. Způsob detekce spočívá ve vyhledávání komunikace s IP adresami a s URL adresami, které se nacházejí ve veřejně dostupných blacklist seznamech. Výsledky navrženého způsobu detekce jsou vyhodnoceny nad několika vzorky dat.

Abstract

This bachelor's thesis discusses monitoring local networks using IP flows. It describes Nemea framework which is used for building complex systems for detecting malicious traffic. Analysis of data from three different networks was performed by using this framework. Based on this analysis a design for detection of malicious traffic in local network was created. The detection method monitors network traffic for suspicious communication targeting IP or URL addresses that are listed in public blacklists. The detection method is evaluated on various traffic samples and the results show that three analysed samples belong to networks that are well managed and secured since the communication with the blacklisted entities is rare.

Klíčová slova

Nemea,IDS,NetFlow,IPFIX,Blacklist

Keywords

Nemea,IDS,NetFlow,IPFIX,Blacklist

Citace

Erik Šabík: Detekce nežádoucího provozu v lokální síti, bakalářská práce, Brno, FIT VUT v Brně, 2015

Detekce nežádoucího provozu v lokální síti

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Martina Žádnika Ph.D. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Erik Šabík
19. května 2015

Poděkování

Rád by som poďakoval svojmu vedúcemu Ing. Martinovi Žádnikovi Ph.D. za čas venovaný konzultáciam k tejto práci a za prístup k testovacím dátam. Ďalej by som tiež rád poďakoval svojej rodine za ich podporu pri štúdiu.

Obsah

Kapitola 1

Úvod

Už od počiatku vzniku internetu sa objavovali rôzne hrozby v podobe útočníkov, ktorý pomocou neho neoprávnene prenikali do zabezpečených systémov. Útočníci tieto systémy následne buď znefunkčnia, odcudzia z nich citlivé údaje alebo tieto systémy ďalej využívajú pre svoje zámery. Problém takýchto útokov nieje len v kompromitácii daného počítača ale aj v negatívnom dopade na reputáciu počítačovej siete, v ktorej sa tento nakazený počítač nachádza. Správcovia takýchto lokálnych sietí preto musia dbať nato, aby sa v ich sieťach nenachádzali takéto nakazené počítače a tým predišli strate reputácie a umiestnení ich siete na blacklist zoznamy.

Všetka komunikácia zanecháva v počítačovej sieti stopy a podľa týchto stôp môžeme detekovať sieťovú prevádzku, ktorá je pre nás nežiadúca. Jedným zo systémov ktoré sa špecializujú na detekciu takejto nežiadúcej sieťovej prevádzky je framework Nemea, ktorý je vyvíjaný pod združeným CESNET. Na základe detekovaných udalostí sa následne môžeme lepšie zabezpečiť proti aktuálnym hrozbám a taktiež je možne vyvinúť systém, ktorý takéto hrozby bude potláčať.

Následujúca kapitola 2 popisuje čo je to lokálna sieť, ako funguje jej správa a tiež najčastejšie problémy z pohľadu bezpečnosti v takejto sieti. Ďalšia kapitola 3 vysvetľuje čo je to IP tok a ako sa využíva na monitorovanie sietí. Kapitola 4 zmieňuje vyvíjaný framework Nemea, ktorý je tiež použitý v tejto práci. V kapitole 5 sa popisuje analýza vzorov sieťových dát z monitorovaných sietí. Kapitola 6 popisuje návrh spôsobu detekcie nežiadúcej prevádzky. Ďalšia kapitola 7 sa venuje popisu implementácie navrhovaného spôsobu detekcie ako modulu do frameworku Nemea. Naimplementovaný modul je následne otestovaný na všetkých sieťach na ktorých prebehla analýza a taktiež na sieti CESNET2. Výsledky tohto testovania sú zapísané v kapitole 8. Nakoniec kapitola 9 obsahuje zhodnotenie celej práce.

Kapitola 2

Lokálne siete

Lokálna počítačová sieť (angl. *Local Area Network*) je počítačová sieť obmedzená do limitovanej oblasti, napr. budova alebo poschodie. Používajú sa v nej technológie ako Ethernet[1] alebo Token Ring[2]. Je zvyčajne pod správou spoločnosti alebo osoby, ktorá ju vlastní a potrebuje.[15]

2.1 Správa lokálnych sietí

Správu lokálnych sietí zaisťuje správca lokálnej siete. Správca lokálnej siete musí zabezpečiť bezproblémovú prevádzku medzi pripojenými stanicami a prepojením lokálnej siete do chrbtovej siete. Ďalej môže zaisťovať správnu funkčnosť serverov, spravovať účty pre užívateľov siete, prideľovať adresové rozsahy, dohliadať na bezpečnosť v sieti a pod.

Povinnosti správcu siete všeobecne spadajú pod následovné kategórie[1]:

- Návrh a plánovanie siete.
- Vytvorenie a nastavenie siete.
- Udržiavanie siete.
- Rozširovanie siete.

Pre správu lokálnych sietí existuje niekoľko nástrojov. Od typu použitého nástroja sa odvíjajú informácie, ktoré o spravovanej sieti a zariadení v nej získavame. Vymenujem preto niekoľko nástrojov a tiež popíšem aké informácie pomocou nich môžeme získať.

Nástroje pre správu siete:

- Syslog[16] - Jedná sa o štandard pre zaznamenávanie správ o udalostiach, ktoré sa udejú na zariadení. V prípade správy siete môže byť týmto zariadením napr. prepínač alebo smerovač. Pomocou nástroja Syslog môžeme získavať aktuálne informácie o tom, čo sa na zariadení deje. Týmito informáciami môžu byť rôzne systémové správy, chybové správy, stav sieťových rozhraní a ďalšie. Z tohto vyplýva že jeho využitie je na správu zariadení.
- SNMP[17] - Podobne ako pomocou Syslog je možné pomocou SNMP získavať informácie o udalostiach na konkrétnom zariadení. Taktiež typ získaných informácií je

[1] http://www.ieee802.org/3/
[2] http://www.ieee802.org/5/

podobný. Rozdiel medzi Syslog a SNMP napr. v prípade Cisco zariadení môže byť v počte správ, ktoré systém podporuje [10]. Celkovo sa ale znovu jedná o nástroj na správu zariadení a o sieťovej prevádzke nám veľa informácií nezíska.

- NetFlow[3] - Ide o protokol, pomocou ktorého sme schopný získať informácie o sieťovej prevádzke. Výhodou využitia NetFlow oproti predošlým nástrojom, je schopnosť na základe získaných informácií detekovať a následne zamedziť nežiadúcu sieťovú prevádzku. Príklady nežiadúcej sieťovej prevádzky sú uvedené v sekcii 2.1.1. Podrobnejšie informácie o protokole NetFlow sú popísané v sekcii 3.2.1.

2.1.1 Najčastejšie problémy

Na základe viacerých diskusií so správcami sietí, som zostavil následujúci zoznam najčastejšie riešených problémov s pohľadu bezpečnosti:

- Skenovanie portov a hľadanie sieťových služieb pre neskoršie útoky.
- Útoky hrubou silou na vybrané služby. Jedná sa najmä o služby SSH, Telnet, RDP.
- Zamedzenie prístupu k službe (angl. *Denial of Service*).

Kvôli komunikácii rôzneho malwaru môže dôjsť k pridaniu siete na verejné blacklist zoznamy. Týmto klesá reputácia danej siete a následne aj možnosť využitia niektorých služieb. Ako príklad môžem uviesť nakazený počítač, ktorý odosiela SPAM. Sieť v ktorej sa tento počítač nachádza sa dostane na blacklist zoznam a následne nebude možné odosielať emaily organizáciám, ktoré na základe daného blacklist zoznamu filtrujú príchodzie emaily. Preto je nutné takúto komunikáciu detekovať a následne odstrániť objavený problém.

Kapitola 3

Monitorovanie sietí

Monitorovanie sietí je proces na získavanie informácií o množstve a typu prevádzky na meranej sieti. V tejto kapitole vysvetlím čo je to IP tok a popíšem protokoly, ktoré ho využívajú pre monitorovanie sietí. Ďalej sa zameriam na nástroje ktoré slúžia na monitorovanie sietí a popíšem monitorovacie architektúry používané v dnešnej dobe.

Hlavné využitie informácií získaných monitorovaním sietí:

- Optimalizácia sieťovej topológie alebo routovacích pravidiel.
- Analýza sieťových aplikácií alebo užívateľov.
- Na základe počtu prenesených dát môže prebiehať účtovanie, napr. účtovanie zákazníka poskytovateľom internetového pripojenia.
- Ukladanie záznamov o aktivitách na sieti poskytovateľom internetových služieb pre prípadné neskoršie dohladávanie incidentov.
- Bezpečnostná analýza sieťovej prevádzky v reálnom čase, prípadne niekedy v budúcnosti pomocou uložených dát.

3.1 IP tok

Podľa [13] je IP tok definovaný ako jednosmerná sekvencia paketov so spoločnými vlastnosťami, ktorá prejde cez monitorovací bod za určitý časový interval. Keďže sa jedná o jednosmernú sekvenciu tak pre každé spojenie dvoch staníc budú existovať dva IP toky, pre každý smer jeden IP tok. Spoločné vlastnosti IP toku na základe ktorých prebieha agregácia:

- Zdrojová a cieľová IP adresa
- Zdrojový a cieľový port
- Protokol sieťovej vrstvy ISO/OSI modelu

IP tok okrem vyššie uvedených vlastností obsahuje aj ďalšie informácie. Jedná sa napr. o počet paketov, ktoré boli prenesené v rámci jedného IP toku, počet bajtov, časovú značku vzniku IP toku, dĺžku spojenia a ďalšie. Vo všeobecnosti to môžu byť akékoľvek informácie, ktoré sme schopný z daného spojenia extrahovať, pričom ale protokol ktorý používame

na monitorovanie sieti musí podporovať pridávanie užívateľom definované položky. V prípade že toto použitý protokol nepodporuje, sme obmedzení iba na protokolom definovanú množinu položiek.

3.2 Protokoly pracujúce s IP tokmi

Na pracú s IP tokmi môžeme využívať rôzne protokoly, v tejto sekcii popíšem dva najpoužívanejšie z nich.

3.2.1 NetFlow

NetFlow [3] je protokol pre prácu s IP tokmi, vyvinutý spoločnosťou Cisco Systems. NetFlow protokol ma množstvo verzií (v dobe písania tejto práce to boli verzie 1 až 9), ale najpoužívanejšie z nich sú verzie 5 a 9 a preto sa v tejto sekcii budem venovať výhradne týmto verziám.

Verzia 5

NetFlow verzia 5 využíva na odosielanie informácií o IP tokoch protokol UDP. Datagram NetFlow verzie 5 obsahuje hlavičku a záznam o jednom a viac IP tokoch. V hlavičke sa okrem iného nachádza verzia protokolu a počet záznamov v datagrame. Verzia 5 podporuje iba staticky formát IP tokov, to znamená že položky nesúce informácie o IP toku sú pevne dané a nedajú sa zmeniť. Ďalším nedostatkom verzie 5 je absencia podpory IPv6 tokov. Pre úplnosť uvádzam všetky položky ktoré obsahuje verzia 5:

- Zdrojová IP adresa
- Cieľová IP adresa
- Next hop IP adresa - IP adresa ďalšie routru
- Číslo vstupného rozhrania
- Číslo výstupného rozhrania
- Počet bajtov v IP toku
- Počet paketov v IP toku
- Čas príchodu prvého paketu IP toku
- Čas príchodu posledného paketu IP toku
- TCP/UDP zdrojový port
- TCP/UDP cieľový port
- Zjednotenie všetkých TCP príznakov pomocou bitového súčtu (operácia OR)
- Číslo protokolu 3. vrstvy modelu ISO/OSI
- Typ služby
- Číslo zdrojového autonómneho systému

- Číslo cieľového autonómneho systému
- Zdrojová maska podsiete
- Cieľová maska podsiete

Verzia 9

Hlavnou výhodou NetFlow verzie 9 [13] oproti verzii 5 je väčšia flexibilita formátu záznamu o IP tokoch, ktorej sa dosiahlo pomocou šablón. Šablóna popisuje typy položiek, ktoré sa nachádzajú v zázname o IP toku. Datagram NetFlow verzie 9 teda obsahuje hlavičku, jednu a viac šablón a nakoniec jeden a viac záznamov o IP toku. Vďaka tomuto prístupu je možné si upravovať formát záznamu o IP toku podľa aktuálnej potreby. Taktiež je možné spracovávať položky nachádzajúce sa v IPv6 tokoch a iné položky ktoré verzia 5 nepodporovala. Verzia 9 je taktiež nezávislá na použitom transportnom protokole (protokol 4. vrstvy modelu ISO/OSI), teda dovoľuje exportovať záznamy o IP tokoch nielen pomocou UDP ale aj TCP, SCTP a iných protokolov.

3.2.2 IPFIX

IPFIX [14] je rozšírením protokolu NetFlow. Konkrétne sa jedná o NetFlow verziu 10, ktorá bola prehlásená za Štandard IETF. IPFIX oproti NetFlow ma výhodu v tom že dovoľuje užívateľovi špecifikovať vlastné položky, ktoré sa budú exportovať. Toto robí IPFIX oveľa flexibilnejším protokolom ako je NetFlow. V IPFIX protokole je každej položke pridelené identifikačné číslo, na základe ktorého dokážeme rozpoznať o akú položku sa jedná. Každá položka má ešte ďalšie identifikačné číslo, nazývané číslo spoločnosti (angl. *enterprise number*). Toto číslo slúži na identifikovanie položiek, ktoré boli špecifikované nejakou spoločnosťou alebo užívateľom. Základné položky ako napr. IP adresa, číslo protokolu, počet bajtov a pod. majú číslo spoločnosti rovné 0. Jednotlivé položky[2] ako aj prideľovanie čísla spoločnosti[9] spravuje organizácia IANA (Internet Assigned Numbers Authority). IPFIX taktiež ako NetFlow verzie 9 je nezávislý na transportnom protokole.

3.3 Monitorovacia architektúra a jej prvky

V tejto kapitole popíšem čo je to monitorovacia architektúra, exportér, kolektor a aké monitorovacie architektúry sa dnes používajú. Monitorovacia architektúra sa typicky skladá z niekoľkých exportérov a jedného kolektoru. Vzhľadom nato že architektúra pre NetFlow a pre IPFIX je podobná, vysvetlím jej princíp iba pre NetFlow, pričom pre IPFIX je princíp analogický.

3.3.1 Exportér

Exportér je zariadenie, ktoré je pripojené k monitorovanej sieti a ktoré zachytáva pakety prechádzajúce touto sieťou. Na základe informácií z týchto paketov vytvára v pamäti záznamy o IP tokoch. Tieto záznamy sú v pamäti uložené tak dlho dokedy neexpirujú. Po expirácii budú tieto záznamy o IP tokoch odoslané kolektoru pomocou príslušného protokolu (NetFlow alebo IPFIX).

Udalosti vedúce k expirovaniu záznamu o IP toku v pamäti:

- Prekročenie časovej hranice u aktívneho IP toku. To znamená že po určitom čase (typicky 5 minút), prebehne expirácia záznamu aj v prípade, že pre tento záznam stále prichádzajú nové pakety.

- Prekročenie časovej hranice u neaktívneho IP toku. To znamená že po určitom čase (typicky 30 sekúnd), prebehne expirácia záznamu v prípade, že pre daný záznam nepríde žiadny ďalší paket.

- V prípade že sa jedná o TCP spojenie a bol detekovaný paket obsahujúci FIN (koniec spojenia) alebo RST (reset spojenia) príznak.

- V prípade že zaplnenie pamäte pre ukladanie záznamov o IP tokoch je nad určitou hodnotou.

Príkladom exportéru môže byť router podporujúci zbieranie a exportovanie štatistík o IP tokoch alebo samostatne stojaca sonda ako napr. FlowMon[4] od firmy INVEA-TECH. FlowMon sonda je pasívna autonómna sonda (samostatne stojaca sonda, neupravujúca monitorované dáta), ktorá monitoruje sieťovú prevádzku v sieti a vytvára štatistiky o tejto prevádzke vo formáte NetFlow verzie 5, NetFlow verzie 9 alebo IPFIX.

3.3.2 Kolektor

Kolektor je zariadenie, ktoré prijíma dáta od exportéru. Tieto dáta následne ukladá do databáze alebo na disk. V špeciálnych prípadoch môže kolektor prijaté dáta preposlať na ďalší kolektor. Formát v ktorom budú dáta uložené závisí od toho, aké položky chceme sledovať alebo aké nástroje na vizualizáciu dat chceme použiť. Typicky sa používa formát Nfdump[1] pre ukladanie NetFlow záznamov o IP tokoch. Tento formát, rovnako ako NetFlow protokol, neumožňuje ukladať užívateľom definované položky. Preto ak je potreba ukladať aj užívateľom definované položky, (napr. rôzne údaje z aplikačných protokolov) je nutné použiť iný formát pre ukladanie dát. Jednou z možností je použiť FastBit[7] databázu, ktorá toto umožňuje. Kolektor môže taktiež dáta preposlať ďalej na hlbšiu analýzu, napr. do systému Nemea, viď kapitola 4.

Príklad kolektoru môže byť aplikácia vyvíjaná CESNETom s názvom IPFIXcol[8]. Jedná sa o open source implementáciu IPFIX kolektoru v jazyku C/C++ podla špecifikácie v RFC7011[14]. Hlavnou výhodou tohoto riešenia je schopnosť jednoducho pridávať užívateľom implementované zásuvné moduly (angl. *plugin*). Použité zásuvné moduly v IPFIXcole patria do jednej z následujúcich troch hlavných skupín:

- Input: Do tejto skupiny patria zásuvné moduly, ktorých úlohou je počúvať na špecifikovanom sieťovom rozhraní a prijímať z neho správy, ktoré odosiela exportér pomocou podporovaného transportného protokolu. Tieto správy ďalej musia spracovávať podľa toho o aký monitorovací protokol sa jedná (NetFlow alebo IPFIX) a následne pomocou funkcií IPFIXcolu poslať tieto dáta ďalším zásuvným modulom. Príkladom takýchto zásuvných modulov sú UDP input plugin alebo TCP input plugin.

- Intermediate: Zásuvné moduly v tejto skupine prijímajú dáta z input zásuvných modulov a podľa potreby v nich upravujú niektoré položky. Ako príklad môže slúžiť zásuvný modul anonymization, ktorý anonymizuje IP adresy pomocou knižnice CryptoPAn[6]. Po úprave sa dáta následne posielajú output/storage zásuvným modulom.

[1] http://nfdump.sourceforge.net/

- Output/Storage: Jedná sa o zásuvné moduly, ktoré špecifikujú výstupný formát záznamov o IP tokoch a taktiež spôsob výstupu. Môže sa jednať o zápis do súboru na disku, zápis do databáze, preposlanie dát ďalšiemu IPFIXcolu, preposlanie dát do systému Nemea a pod. Príkladom môže byť fastbit storage plugin, ktorý umožňuje ukladať prijaté správy do FastBit databáze.

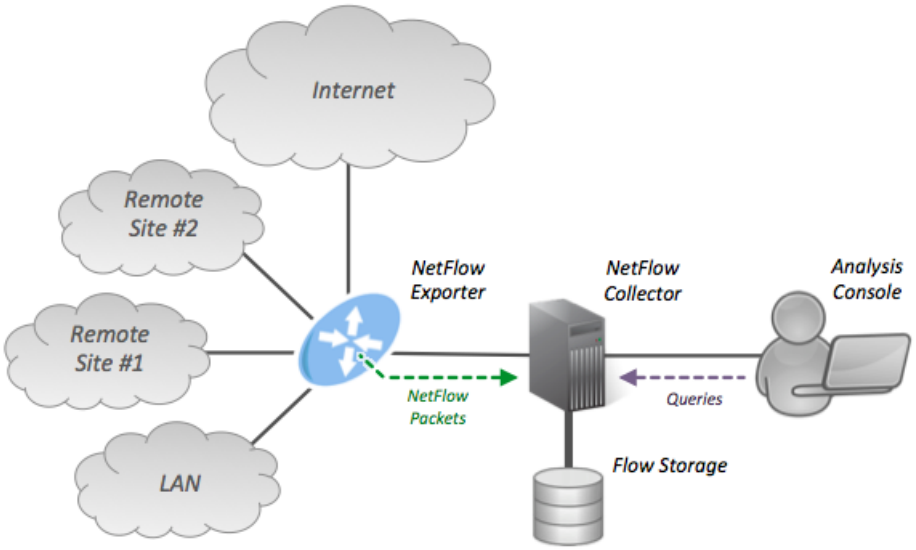
Pomocou týchto zásuvných modulov je možné docieliť úpravu položiek záznamu o IP toku alebo tiež spôsob ich ukladania. Využitie IPFIXcolu je výhodné aj z toho dôvodu, že obsahuje zásuvný modul pre preposielanie záznamov o IP tokoch do systému Nemea a tým dovoľuje využívať systém Nemea na analýzu dát v reálnom čase.

3.3.3 Typy architektúr

V tejto sekcii sa budem venovať dvom hlavným typom architektúr, s ktorými sa môžeme v dnešnej dobe stretnúť. Časť informácií v tejto kapitole som čerpal z online zdroja[5].

Architektúra využívajúca routre

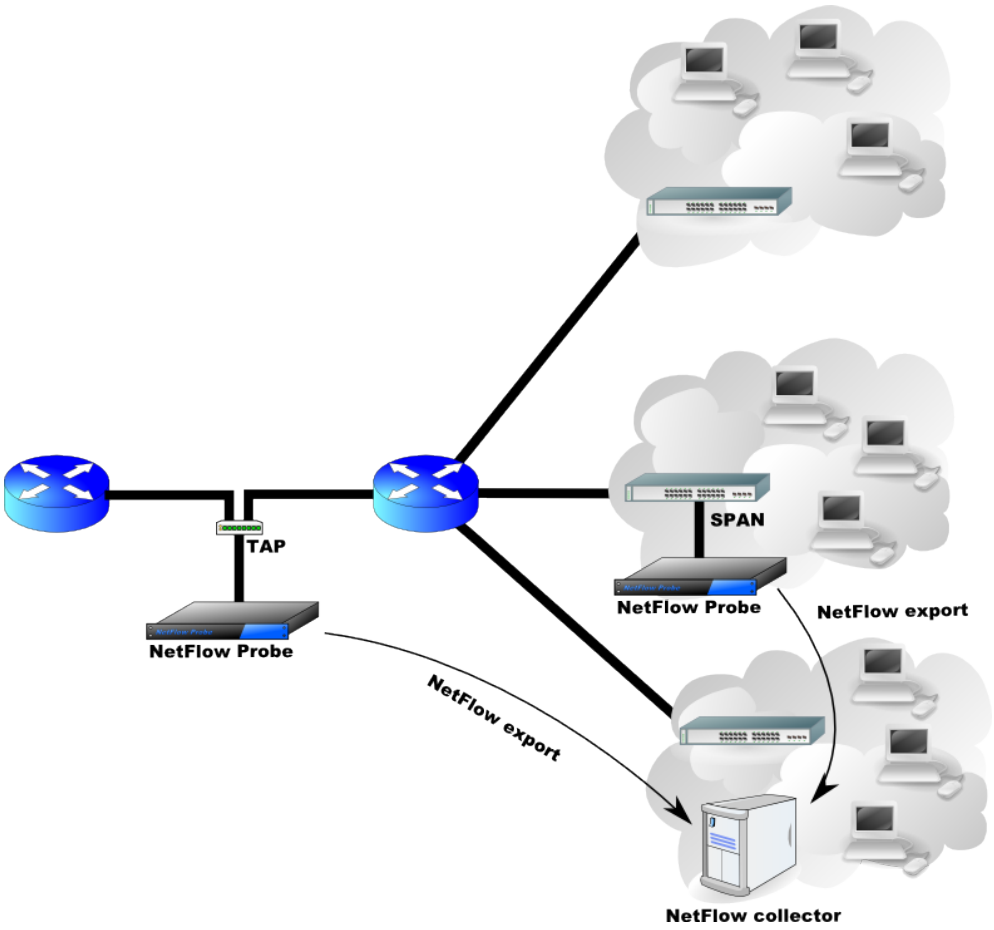
Táto architektúra využíva routre, ktoré sa nachádzajú na hraničných bodoch siete, ako exportéry. Tieto routre musia dokázať analyzovať pakety a vytvárať z nich záznamy o IP tokoch. Toto riešenie má dve hlavné nevýhody. Prvou nevýhodou je cena zariadenia (routru), z čoho vyplýva že toto riešenie nemusí byť vhodné pre malé siete. Samotná analýza paketov a spracovanie štatistiky o IP tokoch taktiež obmedzuje celkový výkon zariadenia, čo u menej výkonných zariadeniach môže viesť až ku vzorkovaniu paketov z ktorých sa vytvárajú IP toky. Druhou nevýhodou môže byť malá množina podporovaných protokolov pre export záznamov o IP tokoch. Ako príklad môžem uviesť Cisco routre podporujúce výhradne protokol NetFlow. Z toho vyplýva že s použitím daného zariadenia nieje možné exportovať užívateľom definované položky, pretože ako bolo popísane vyššie, protokol NetFlow to neumožňuje.



Obrázok 3.1: Ukážka architektúry využívajúcej routre[5]

Architektúra využívajúca sondy

V tejto architektúre sa využívajú pasívne sondy ako exportéry. Jedná sa o špeciálne zariadenie určené výhradne na monitorovanie siete a export štatistík o IP tokoch. Táto architektúra prináša hneď niekoľko výhod. Prvou z nich je cena zariadenia, ktorá oproti routru poskytujúcemu rovnakú funkcionalitu môže byť výrazne nižšia. Ďalšou výhodou je, že toto zariadenie je možné pripojiť transparentne do ľubovolného bodu v sieti pomocou TAP rozhrania. Je treba si ale uvedomiť že v tomto prípade bude monitorovanie prebiehať iba v tomto bode siete. Je tiež možné zapojiť sondu do siete pomocou zrkadlenia portov (angl. *port mirroring*) na routry a teda monitorovať sieťovú prevádzku prechádzajúcu daným routrom. Štatistiky o IP tokoch môže sonda odosielať do kolektoru po odlišnej sieťovej linke akú monitoruje, a tým nezaťažovať monitorovanú sieť. Výhodou použitia sondy namiesto routru taktiež môže byť podpora viacerých monitorovacích protokolov alebo možnosť implementácie a nasadenie vlastného algoritmu pre analýzu sieťovej prevádzky.



Obrázok 3.2: Ukážka architektúry využívajúcej sondy[5]

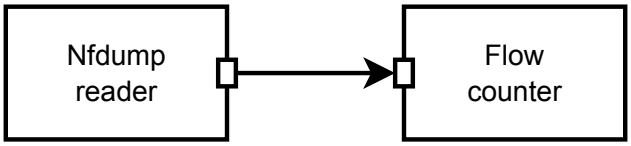
Kapitola 4

Nemea

Nemea (Network Measurements and Analysis) je framework, ktorý umožňuje tvorbu systému pre automatickú analýzu záznamov o IP tokoch v reálnom čase. Tento systém sa skladá z oddeliteľných blokov nazývaných moduly. Jednotlivé moduly sú medzi sebou prepojené rozhraniami a ako celok môžu spracovávať, analyzovať a následne vytvárať správy o rôznych sieťových incidentoch. Väčšinu informácií popísaných v tejto kapitole som čerpal z príslušnej technickej správy[12].

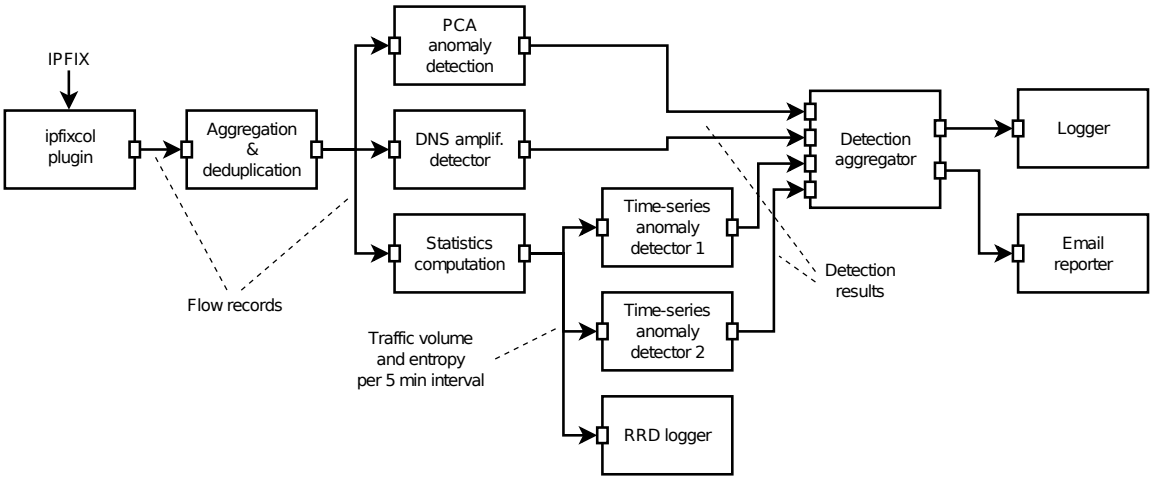
4.1 Moduly

Každý modul je samostatne bežiaca aplikácia, ktorá nejakým spôsobom pracuje so záznamom o IP toku. Vo väčšine prípadov moduly príjmu dáta na vstupnom rozhraní, tieto dáta spracujú a odošlú pomocou výstupného rozhrania. Moduly môžu napr. počítať rôzne štatistiky z prichádzajúcich dat alebo hľadať typické vzory útokov a následne odosielať výsledok tohoto procesu pomocou výstupného rozhrania. Ďalší modul môže tieto výsledky spracovávať, agregovať alebo korelovať s ostatnými výsledkami iných modulov. Takýmto spôsobom je možné vytvoriť komplexný systém pre analýzu sieťovej prevádzky v reálnom čase. Moduly je možné implementovať v jazyku C, C++ alebo Python. Obrázok 4.1 zobrazuje minimálnu konfiguráciu modulov, v ktorej sú dva moduly. Prvý modul načítava záznamy o IP tokoch zo súboru a odosiela ich cez jeho výstupné rozhranie. Druhý modul tieto dáta prijíma a počíta z nich celkový počet IP tokov, paketov a bajtov. Obrázok 4.2 zobrazuje komplexnú konfiguráciu niekoľkých modulov, v ktorej IP toky sú získavané zo siete v reálnom čase pomocou zásuvného modulu pre IPFIX kolektor. IP toky sú v tomto prípade predspracovávané a analyzované niekoľkými algoritmami. Výsledky z tejto analýzy sú agregované a následne z týchto výsledkov je vytvorená správa, ktorá sa odosiela externému systému. Takýmto systémom môže byť napr. systém Warden[1], ktorý je vyvíjaný združením CESNET. Každú úlohu plní samostatný modul.



Obrázok 4.1: Minimálna konfigurácia systému Nemea[12]

[1] https://wardenw.cesnet.cz/



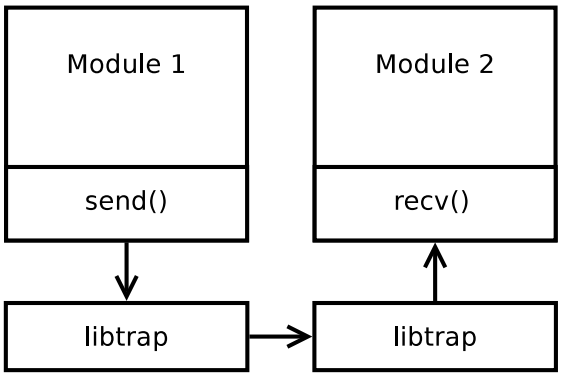
Obrázok 4.2: Komplexná konfigurácia systému Nemea[12]

4.2 Rozhrania

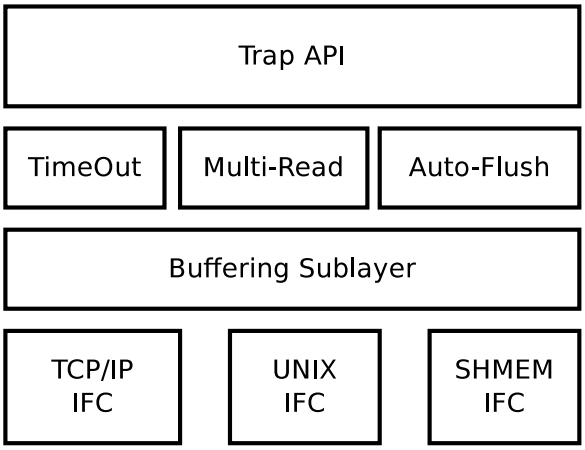
Všetky rozhrania sú jednosmerné a odosielajú dáta vo forme záznamov. Všetky záznamy posielané cez jedno rozhranie musia mať rovnaký formát. To znamená že formát výstupného rozhrania modulu ktorý dáta odosiela a vstupného rozhrania modulu ktorý dáta prijíma musí byť rovnaký. Formát špecifikuje aké položky sú v danom zázname. Formát pre dané rozhranie je špecifikovaný dynamicky v momente kedy sa modul pripája do systému. Dynamická špecifikácia formátu dovoľuje napr. pridať novú položku do tohto formátu bez potreby meniť kód modulov, ktoré pracujú s týmto formátom. Protokol ktorý špecifikuje ako definovať tieto formáty, ako ich vytvárať a používať sa nazýva UniRec, viď sekcia 4.4.

4.3 TRAP

TRAP (Traffic Analysis Platform) je knižnica, ktorá efektívne implementuje rozhrania používané Nemea modulmi. Táto knižnica je linkovaná ku každému Nemea modulu. Obrázok 4.3 zobrazuje koncept komunikácie medzi dvoma modulmi. TRAP abstrahuje modul od aktuálneho rozhrania a jeho špecifík. Odosielací modul odosiela dáta ihneď ako sú dostupné. Operácia odoslania dát môže byť podľa konfigurácie neblokujúca alebo blokujúca. Prijímací modul číta dáta zo vstupu, pričom táto operácia môže byť taktiež neblokujúca alebo blokujúca. TRAP sa taktiež stará o bufferovanie dát a o ich zahadzovanie podľa konfigurácie. Na obrázku 4.4 je zobrazená architektúra knižnice TRAP.



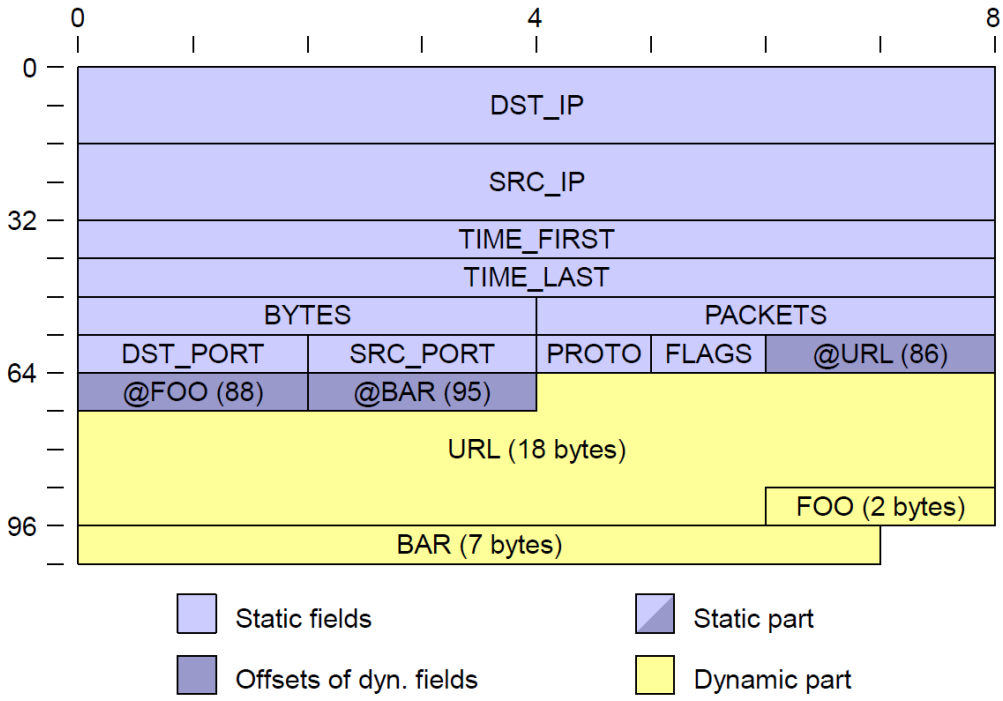
Obrázok 4.3: Typická komunikácia medzi dvoma modulmi v systéme Nemea[12]



Obrázok 4.4: Architektúra libtrap knižnice[12]

4.4 UniRec

UniRec (Unified Record) je špecifický formát správ (záznam), ktoré sa posielajú pomocou TRAP rozhraní. Pozostáva z niekoľkých položiek, pričom každá položka ma svoje meno a dátový typ. Zoznam položiek v zázname sa nazýva šablóna (angl. *template*). Šablóna je špecifikovaná vymenovaním všetkých jej položiek. Pretože TRAP rozhranie používa práve jednu šablónu, tak všetky správy poslané cez toto rozhranie majú rovnaký formát. Akékoľvek dva moduly spojené pomocou TRAP rozhrania musia používať rovnakú šablónu na tomto rozhraní. Šablóna sa pre odosielajúci aj pre prijímací modul špecifikuje pri inicializácii modulu. UniRec záznam obsahuje statickú časť, pre položky ktorých veľkosť sa v čase nemení, a dynamickú časť, pre položky ktorých veľkosť sa môže s časom meniť. Obrázok 4.5 zobrazuje príklad ako môže vyzerať UniRec záznam.



Obrázok 4.5: Príklad UniRec záznamu reprezentujúceho základný IP tok rozšírený o niekoľko dynamických položiek[12]

Na poradí položiek v špecifikácii UniRec šablóny nezáleží. Poradie položiek v samotnom UniRec zázname je ale presne definované podľa nasledujúcich pravidiel:

- Ako prvé sú uložené statické položky, za nimi nasledujú offsety dynamických položiek a na konci sú uložené samotné dynamické položky.
- Položky v statickej časti sú zoradené zostupne podľa ich veľkosti.
- Položky s rovnakou veľkosťou sú zoradené abecedne podľa ich názvu.

4.5 Nemea moduly

V dobe písania tejto práce obsahoval framework Nemea viac ako 20 modulov. Nemea moduly obvykle implementujú základné spracovanie alebo analýzu IP tokov. Vzhľadom na veľký počet modulov popíšem len najpoužívanejšie z nich.

- HostStatsNemea - Detekčný modul, ktorý počíta štatistiky zo sieťovej prevádzky pre každý počítač v sieti. V štatistikách sú následne vyhľadávané typické vzory útokov ako napr. skenovanie siete, útoky hrubou silou alebo DoS útoky. Detekované udalosti následne odosiela pomocou výstupného rozhrania.
- BruteForceDetector - Detekčný modul, ktorý na základe známych vzorov útokov dokáže detekovať útoky hrubou silou na služby SSH, Telnet, RDP. Hlavnou výhodou modulu oproti modulu HostStatsNemea je schopnosť detekovať aj veľmi pomalé útoky.
- Logger - Modul vypisuje všetky prijaté UniRec záznamy na štandardný výstup alebo do súboru v čitateľnom formáte.

- Nfdump Reader - Modul dokáže čítať súbory s IP tokmi zapísanými v Nfdump formáte. Tieto IP toky konvertuje na príslušné UniRec záznamy, ktoré následne odosiela pomocou výstupného rozhrania.
- Anonymizer - Modul anonymizuje IP adresy vo všetkých UniRec záznamoch ktoré prijme na svojom vstupe. Anonymizované UniRec záznamy odosiela pomocou výstupného rozhrania. Anonymizácia využíva knižnicu Crypto-PAn, ktorá zachováva prefix IP adresy.

Kapitola 5

Analýza prevádzky

Analýzu prevádzky som vykonal nad tromi rôzne veľkými sieťami na základe projektu STAAS (Security Tools as a Service), do ktorého boli vlastníci sietí zapojení. Jednalo sa o siete Univerzity Hradec Králové, Akademie věd ČR a Metacentra. Pri všetkých sieťach bola použitá monitorovacia architektúra využívajúca pasívne sondy (viď sekcia 3.3.3). Konkrétne bola použitá pasívna sonda FlowMon a ako kolektor bola použitá aplikácia IPFIXcol. Na detekciu anomálii bol použitý systém Nemea, ktorý prijímal dáta z kolektoru. Všetky IP adresy zobrazené v tejto kapitole boli anonymizované pomocou knižnice Crypto-PAn.

5.1 Univerzita Hradec Králové

V prípade Univerzity Hradec Králové merané obdobie trvalo približne 20 dní. Ide o privátnu sieť nachádzajúcu sa za NATom[19]. Monitorovacia sonda bola umiestnená na rozhranie verejnej a privátnej siete. Pakety sú sondou spracovávané až po preložení adries NATom. V záznamoch o IP tokoch sa toto prejaví ako komunikácia verejných adries s adresami z privátneho rozsahu. Vzhľadom na umiestnenie monitorovacej sondy, nebude možné detekovať nežiadúcu prevádzku, ktorá prichádza z internetu a je zameraná na verejné adresy univerzity. Detekovatelné budú ale udalosti, ktoré pochádzajú z privátnej siete a smerujú do verejnej siete. Na rozhraní medzi privátnou a verejnou sieťou sa tiež nachádza firewall, ktorý filtruje väčšinu nežiadúcej sieťovej prevádzky. Jedná sa hlavne o blokovanie P2P komunikácie, SMTP komunikácie a komunikácie s inými ako povolenými DNS servermi. Kvôli týmto obmedzeniam sa očakáva malé množstvo detekovaných udalosti.

Tabuľka 5.1 zobrazuje štatistiku o prietoku dát v sieti. Jednotkou pre hodnoty v tabuľke je počet IP tokov za sekundu. Z tejto štatistiky sa dá usúdiť že sa jedná o relatívne malú až stredne veľkú sieť.

Minimum	Maximum	Priemer
7	707	186

Tabuľka 5.1: Štatistika o prenesených dátach v sieti v sieti Univerzity Hradec Králové, hodnoty sú v IP tokoch za sekundu

Pomocou systému Nemea sa podarilo detekovať časť nežiadúcej prevádzky. Konkrétne sa jedná o skenovanie portov detekované modulom HostStatsNemea, zobrazené v tabuľke 5.2 a pokus o uhádnutie hesla detekované modulom BruteForceDetector, zobrazené v tabuľke 5.3.

Čas detekcie	Útočiaca adresa	Typ útoku	Intenzita
2015-03-12 13:03:24	164.37.6.173	Horizontálny sken	1692
2015-03-12 13:08:25	164.37.6.173	Horizontálny sken	1135

Tabuľka 5.2: Detekovaná nežiadúca prevádzka modulom HostStatsNemea v sieti Univerzity Hradec Králové

Čas detekcie	Útočiaca adresa	Cieľová služba	Intenzita
2015-03-11 11:40:44.418	164.37.11.33	SSH	30
2015-03-16 11:00:47.874	164.37.8.242	SSH	11

Tabuľka 5.3: Detekovaná nežiadúca prevádzka modulom BruteForceDetector v sieti Univerzity Hradec Králové

5.2 Akademie věd ČR

Merané obdobie v sieti Akademie věd Českej Republiky trvalo približne dva mesiace. Na rozdiel od siete Univerzity Hradec Králové bola monitorovacia sonda umiestnená mimo privátnu sieť. Vďaka tomu v záznamoch o IP tokoch je možné vidieť komunikáciu verejných IP adries a preto je možné očakávať vyšší počet detekcií nežiadúcej prevádzky systémom Nemea ako tomu bolo v prípade predošlej siete.

Tabuľka 5.4 zobrazuje štatistiku o prietoku dát v sieti. Jednotkou pre hodnoty v tabuľke je počet IP tokov za sekundu. V porovnaní so sieťou Univerzity Hradec Králové sa jedná o menšiu sieť.

Minimum	Maximum	Priemer
17	119	50

Tabuľka 5.4: Štatistika o prenesených dátach v sieti Akademie věd ČR, hodnoty sú v IP tokoch za sekundu

Taktiež ako v prípade siete Univerzity Hradec Králové sa aj v sieti Akademie věd ČR podarilo pomocou systému Nemea detekovať nežiadúcu prevádzku. Oproti predošlej sieti sa ale detekovalo omnoho viac udalostí, čo môže byť spôsobené práve tým že, monitorovacia sonda nebola umiestnená do privátnej siete a teda väčšinu nežiadúcej komunikácie nezablokoval NAT alebo firewall. Pretože detekcií bolo veľa, tabuľka 5.5 zobrazuje z každého typu útoku, ktorý bol detekovaný modulom HostStatsNemea, len najsilnejšie útoky (podľa intenzity). V prípade DNS amplifikácie sa nejedná presne o útočiacu adresu, ale skôr o adresu ktorá bola zneužitá na amplifikačný útok. Tabuľka 5.6 zobrazuje taktiež iba jeden najsilnejší útok (podla intenzity) z každého typu útoku (útoky na rôzne protokoly), ktorý bol detekovaný modulom BruteForceDetector.

Čas detekcie	Útočiaca adresa	Typ útoku	Intenzita
2015-03-04 03:51:26.000	228.6.76.53	Horizontálny sken	18356
2015-06-05 01:04:41.000	10.240.21.60	Hádanie hesiel	3065
2015-06-05 01:04:18.000	145.231.171.115	DNS amplifikácia	18205

Tabuľka 5.5: Detekovaná nežiadúca prevádzka modulom HostStatsNemea v sieti Akademie věd ČR

Čas detekcie	Útočiaca adresa	Cieľový port	Intenzita
2015-04-25 14:35:44.249	185.51.26.237	22	1910
2015-04-04 02:41:56.747	41.28.156.39	3389	419

Tabuľka 5.6: Detekovaná nežiadúca prevádzka modulom BruteForceDetector v sieti Akademie věd ČR

5.3 Metacentrum

Doba merania siete Metacentra trvala jeden mesiac. Rovnako ako pri sieti Akademie věd ČR bola monitorovacia sonda zapojená mimo privátnu sieť, takže podobne je v nameraných dátach vidieť komunikáciu verejných adries. Tabuľka 5.7 zobrazuje štatistiku o prietoku dát v sieti. Jednotkou pre hodnoty v tabuľke je počet IP tokov za sekundu. V porovnaní s predošlými dvoma sieťami je tu vidieť niekoľko násobný rozdiel v prietoku dát. Je možné teda predpokladať aj veľký počet detekcií systémom Nemea.

Minimum	Maximum	Priemer
456	1300	565

Tabuľka 5.7: Štatistika o prenesených dátach v sieti Metacentra, hodnoty sú v IP tokoch za sekundu

V následujúcich dvoch tabuľkách sú zobrazené výsledky analýzy pomocou systému Nemea. Skutočne bol detekovaný veľký počet incidentov a preto rovnako ako v prípade Akademie věd ČR sú v tabuľkách zobrazené len najsilnejšie útoky (podla intenzity) z každého detekovaného typu. Tabuľka 5.8 zobrazuje detekcie modulom HostStatsNemea a tabuľka 5.9 zobrazuje detekcie modulom BruteForceDetector.

Čas detekcie	Útočiaca adresa	Typ útoku	Intenzita
2015-06-12 08:04:42.000	145.242.251.199	DNS amplifikácia	22893
2015-04-16 03:01:09.000	33.17.84.241	Hádanie hesiel	12351
2015-06-12 08:04:42.000	95.136.108.26	Horizontálny sken	9636

Tabuľka 5.8: Detekovaná nežiadúca prevádzka modulom HostStatsNemea v sieti Metacentra

Čas detekcie	Útočiaca adresa	Cieľová služba	Intenzita
2015-04-03 19:31:18.745	26.239.143.104	SSH	26972
2015-04-09 22:57:52.924	44.212.251.141	RDP	352

Tabuľka 5.9: Detekovaná nežiadúca prevádzka modulom BruteForceDetector v sieti Metacentra

5.4 Zhodnotenie analýzy

Pomocou systém Nemea sa podarilo detekovať časť nežiadúcej prevádzky. Jednalo sa o časť sieťovej prevádzky, ktorá bola detekovaná na základe typických vzorov útokov. V sieťovej prevádzke sa ale môžu vyskytovať nežiadúca komunikácia, ktorej typické vzory nepoznáme alebo je ich veľmi náročné vytvoriť. Môže sa jednať napr. o komunikáciu s

Command&Control servermi botnetov, sťahovanie malware z webových stránok, SPAM a ďalšie. Takýto typ sieťovej prevádzky stávajúce moduly v systéme Nemea nedetekujú a preto následujúcu časť tejto práce venujem návrhu modulu pre systém Nemea, ktorý bude schopný tento typ nežiadúcej komunikácie detekovať.

Kapitola 6

Návrh spôsobu detekcie

Spôsob detekcie nežiadúcej sieťovej prevádzky bude založený na vyhľadávaní nedôveryhodných účastníkov jednotlivých sieťových spojení. Zoznamy nedôveryhodných účastníkov sa budu získavať z verejne dostupných blacklist zoznamov alebo sa tiež môžu staticky nastaviť užívateľom. Detekcia by mala prebiehať tak, že v jednotlivých záznamoch o IP tokoch sa budú vyhľadávať nedôveryhodné IP adresy. V prípade že ktorákoľvek IP adresa v zázname o IP toku je nájdená na akomkoľvek blacklist zozname (teda je nedôveryhodná), tak je tento záznam o IP toku označený ako podozrivý. Tento spôsob detekcie so sebou prináša niekoľko výhod, ale aj nevýhody oproti stávajúcim modulom na detekciu nežiadúcej prevádzky v systéme Nemea. Rozoberiem preto tie najvýznamnejšie z nich.

Výhody spôsobu detekcie pomocou blacklist zoznamov:

- **Nezáleží na intenzite:** Pri stávajúcich moduloch hrala vždy úlohu intenzita útoku. V menších sieťach je toto problém, pretože vzhľadom na veľkosť siete by sme za útok chceli považovať aj takú sieťovú prevádzku, ktorej intenzita je nižšia ako intenzita podobnej sieťovej prevádzky na väčšej sieti. Tento problém ale nenastáva pri použití navrhovaného spôsobu detekcie, pretože sa zameriavame na každý záznam o IP toku samostatne. V konečnom dôsledku teda záleží iba na tom, či existuje komunikácia s IP adresou ktorú považujeme za nedôveryhodnú alebo nie.

- **Detekcia inak ťažko detekovatelnej nežiadúcej prevádzky:** Pri stávajúcich spôsoboch detekcie sa spolieha na typické vzory útokov. Ak ale nepoznáme vzor útoku, je ho náročné vytvoriť alebo ho vytvoriť vôbec nie je možné, pretože nežiadúca komunikácia vyzerá rovnako ako legitímna komunikácia, tak nie sme schopný takýto útok detekovať alebo je jeho detekcia veľmi náročná. S použitím navrhovaného spôsobu detekcie je ale takúto komunikáciu možné detekovať. Ak máme k dispozícii zoznam IP adries, ktoré boli nahlásené ako potencionálni útočníci v rámci daného typu útoku, môžeme komunikáciu s týmito IP adresami označiť ako nežiadúcu.

Nevýhody spôsobu detekcie pomocou blacklist zoznamov:

- **Dôveryhodnosť zoznamov:** Vzhľadom na použitie cudzích blacklist zoznamov si treba uvedomiť, že presúvame skutočnú detekciu na autorov takýchto zoznamov a využívame len výsledok ich detekcie o ktorej nevieme ako v skutočnosti funguje. Na zozname sa teda môže vyskytnúť IP adresa, ktorá bola falošne označená za iniciátora nežiadúcej komunikácie. Preto je dôležité pred použitím daného zoznamu overiť dôveryhodnosť autorov a spôsob ich detekcie.

- **Falošne pozitívne detekcie:** Môže nastať prípad kedy sa IP adresa dostane na blacklist zoznam pretože skutočne bola iniciátorom nežiadúcej komunikácie, pričom ale táto komunikácia bola jednorázová a už sa nebude opakovať, napr. vytvorená nejakým malwarom. Všetka nasledujúca komunikácia bude označená za nežiadúcu aj keď bol problém už vyriešený, napr. odstránením spomínaného malwaru. Komunikácia bude označovaná za nežiadúcu do tej doby, kým ju autori zoznamu neodstránia z blacklist zoznamu, čo môže trvať aj niekoľko dní. Taktiež môže nastať situácia kedy sa IP adresa dostane na blacklist zoznam, kvôli určitému typu nevyžiadanej komunikácie, napr. SPAM. Keďže sa IP toky vyhodnocujú ako podozrivé iba na základe IP adresy, tak budú ako podozrivé IP toky označované aj tie, ktoré síce obsahujú danú IP adresu ale neodosielajú SPAM.

6.1 Blacklist zoznamy

Nižšie sa nachádzajúce zoznamy sú príklady verejne dostupných blacklist zoznamov. Jedná sa prevažne o blacklist zoznamy IP adries Command&Control serverov botnetov, IP adries odosielajúci SPAM, URL adries obsahujúcich malware alebo URL pokúšajúce sa o phishing. Všetky uvedené zoznamy boli v dobe písania tejto práce verejne dostupné.

- **ZeuS Tracker**[1]: Blacklist zoznam IP adries Command&Control serverov botnetu ZeuS.
- **Palevo Tracker**[2]: Blacklist zoznam IP adries Command&Control serverov botnetu Palevo.
- **Feodo Tracker**[3]: Blacklist zoznam IP adries Command&Control serverov botnetu Feodo
- **Spamhaus**[4]: Blacklist zoznam IP adries odosielateľov SPAMu a služieb odosielacích SPAM.
- **Tor**[5]: Nejedná sa o blacklist v pravom slova zmysle ale ide o zoznam výstupných uzlov zo siete Tor. Komunikácia so sieťou Tor nemusí byť nežiadaná, no vzhľadom na to že sieť Tor poskytuje užívateľovi anonymitu, tak môže byť ľahko zneužitá útočníkmi na ukrytie ich identity. Je teda vhodné túto komunikáciu bližšie sledovať, poprípade korelovať s výsledkami z ostatných analýz.
- **Malware Domains**[6]: Blacklist zoznam URL adries obsahujúcich škodlivý obsah.
- **PhishTank**[7]: Blacklist zoznam URL adries, ktoré boli označené ako phishing.

[1] https://zeustracker.abuse.ch/
[2] https://palevotracker.abuse.ch/
[3] https://feodotracker.abuse.ch/
[4] https://www.spamhaus.org/
[5] https://www.torproject.org/
[6] http://www.malwaredomains.com/
[7] https://www.phishtank.com/

Kapitola 7

Implementácia modulu

V tejto kapitole popíšem implementáciu modulov pre systém Nemea, ktoré budú vykonávať vyššie opísaný spôsob detekcie nežiadúcej sieťovej prevádzky. Pretože sú k dispozícii dva typy blacklist zoznamov (zoznam IP adries a zoznam URL adries), je z hľadiska modularity systému Nemea vhodné implementovať dva moduly, kde jeden sa bude zameriavať na IP adresy a druhý na URL adresy. Moduly zodpovedné za samotnú detekciu sú implementované v jazyku C++. Pre získanie blacklist zoznamov sa využíva samostatná knižnica, ktorá je implementovaná v jazyku C.

7.1 Použitý algoritmus

Algoritmus modulov je z veľkej časti rovnaký. Pretože ale moduly pracujú s rôznym typom dát, tak ich dátové štruktúry a detekčné algoritmy sa mierne líšia. Spoločná časť oboch modulov okrem iného obsahuje získavanie blacklist zoznamov z webových stránok a následne predspracovanie údajov aby s nimi samotné detekčné moduly mohli ľahšie pracovať.

7.1.1 Získanie blacklist zoznamov

Jedná sa o sadu funkcií, ktoré detekčný modul využíva pre jednoduché získanie blacklist zoznamov. Na začiatku behu detekčného modulu sa táto knižnica inicializuje. Pri inicializácii sa špecifikujú nasledovné parametre:

- Názvy blacklist zoznamov, ktoré modul chce získať. Atribúty blacklist zoznamov (identifikačné číslo, názov, webová adresa a pod.) sú uložené v špeciálnom konfiguračnom súbore a detekčné moduly sa na tieto blacklist zoznamy odkazujú iba ich názvom.
- Perióda s akou sa blacklist zoznamy budu aktualizovať. Keďže detekčné moduly môžu bežať aj dlhšiu dobu, je nutné po určitom čase znovu aktualizovať ich blacklist zoznamy a tým predísť falošne pozitívnym detekciám.
- Regulárny vyraz, ktorý špecifikuje časť (IP/URL adresu) blacklist zoznamu, ktorú detekčný modul potrebuje na svoje fungovanie. Toto je nutné pretože blacklist zoznamy, môžu okrem samotného záznamu (IP/URL adresy) obsahovať aj ďalšie informácie, ktoré detekčný modul nepotrebuje a preto je nutné potrebné informácie extrahovať pomocou regulárneho výrazu.
- Režim aktualizácie, pričom je na výber z dvoch režimov.

  - Prvý režim získa aktuálny blacklist zoznam a spraví rozdiel medzi ním a predošlým zoznamom. Výstupom budú 2 zoznamy. Zoznam nových záznamov (záznamy nachádzajúce sa v aktuálnom zozname ale nie v predošlom zozname) a zoznam odobratých záznamov (záznamy nachádzajúce sa v predošlom zozname ale nie v aktuálnom zozname).
  - Druhý režim nerobí rozdiel medzi aktuálnym a predošlým zoznamom ale výstupom je priamo aktuálny zoznam. Výhodou je, že nieje nutné si pamätať predošlí zoznam a teda nároky na pamäť sú nižšie. Naopak nevýhoda oproti predošlému režimu je, že detekčný modul musí pri každej aktualizácii spracovať celý blacklist zoznam, ktorý môže obsahovať aj rádovo tisíce záznamov. Zatiaľ čo pri predošlom riešení stačí spracovať len zmeny, ktoré bývajú rádovo v desiatkach až stovkách.

- Maximálna veľkosť blacklist zoznamu, maximálna dĺžka blacklist záznamu a maximálny počet blacklist záznamov v jednom zozname. Aby knižnicu nespomaľovalo zbytočné alokovanie pamäti za behu, tak je nutné špecifikovať tieto informácie a pamäť sa mohla alokovať hneď pri inicializácii. V prípade zaplnenia tejto pamäte sa všetky následne stiahnuté blacklist záznamy ignorujú.

- Súbor, do ktorého sa zapíšu získané blacklist záznamy spolu s identifikačným číslom blacklist zoznamu, z ktorého tieto záznamy pochádzajú.

Po úspešnej inicializácii knižnica vytvorí nové vlákno, ktoré bude paralelne k behu detekčného modulu získavať a predspracovávať blacklist zoznamy. Následne po predspracovaní a uložení záznamov do súboru, knižnica signalizuje detekčnému modulu že záznamy sú pripravene na prebratie. Knižnica sa následne uvádza do stavu neaktívneho čakania, v ktorom zotrvá po dobu ktorá bola špecifikovaná pri jej inicializácii (perióda aktualizácii).

7.1.2 IP detekčný modul

Detekčný modul zameraný na IP blacklist zoznamy prijíma na svojom vstupe UniRec záznamy obsahujúce IP adresy (zdrojová IP adresa a cieľová IP adresa). Tieto IP adresy porovnáva voči IP adresám uloženým v pamäti, ktoré získal pomocou vyššie popísanej knižnice pre získanie blacklist zoznamov. Ak sa čo i len jedna z IP adries v UniRec zázname zhoduje s ktoroukoľvek IP adresou v pamäti, je celý UniRec záznam označený ako podozrivý a odoslaný pomocou výstupného rozhrania modulu. Pretože aktualizácie blacklist zoznamov sa robia za relatívne dlhý časový interval (typicky každý 5 minút), tak som zvolil na vyhľadávanie IP adries algoritmus s názvom binárne vyhľadávanie[18]. Požiadavkom pre použitie tohto algoritmu je nutnosť mať IP adresy v pamäti zoradené. Tohto sa docieli tým že pri aktualizácii blacklist záznamov sa nové záznamy pridávajú do štruktúry v pamäti medzi stávajúce takým spôsobom, aby záznamy v celej štruktúre boli neustále zoradené. Príkladom štruktúry ktorá umožňuje efektívne vkladať nove záznamy medzi už existujúce je štandardná C++ štruktúra std::vector, ktorá sa preto využíva v detekčnom module.

Vzhľadom na povahu niektorých útokov som zaviedol na výstupe modulu agregáciu detekcií. Táto agregácia spolu spája detekcie, ktoré sa zhodujú v zdrojovej a cieľovej IP adrese a transportnom protokole. Počet agregovaných detekcií je uložený v počítadle, ktoré sa pri odosielaní detekovanej udalosti pridáva k výstupu (viď sekcia 7.2.1 , položka EVENT_SCALE). Odoslanie agregovaných detekcií nastáva pri jednej z následujúcich udalostí:

- Prekročenie časovej hranice pri aktívnej detekcii. To znamená že po určitom čase (typicky 5 minút), prebehne odoslanie agregovanej detekcie aj v prípade že pre túto detekciu sú detekované stále nové IP toky.
- Prekročenie časovej hranice pri neaktívnej detekcii. To znamená že daná detekcia bude odoslaná, ak v rámci časového intervalu (typicky 30 sekúnd) nebude pre túto detekciu detekovaný ďalší záznam o IP toku.
- Pri ukončení modulu sa všetky agregované detekcie odosielajú.

7.1.3 URL detekčný modul

Detekčný modul zameraný na URL blacklist zoznamy prijíma na svojom vstupe UniRec záznamy obsahujúce okrem iného aj URL adresy. Ako bude popísane v sekcii 7.2.2 URL adresy sú rozdelené na dve časti. Na blacklist zoznamoch sa ale URL adresy takto nerozdeľujú. Taktiež sa tam môže vyskytnúť iba názov serveru, teda z pohľadu položiek UniRec formátu sa na blacklist zoznamoch môže vyskytnúť buď samotná položka HTTP_REQUEST_HOST alebo položky HTTP_REQUEST_HOST a HTTP_REQUEST_URL zároveň. Z týchto dôvodov je nutné porovnávať výskyt URL na dvakrát. Prvýkrát sa vyhľadáva iba HTTP_REQUEST_HOST a druhýkrát konkatenácia položiek HTTP_REQUEST_HOST a HTTP_REQUEST_URL, pričom na poradí týchto vyhľadávaní nezáleží. V prípade že aspoň v jednom z týchto vyhľadávaní sa nájde zhoda so záznamom na blacklist zozname, je k tomuto záznamu o IP toku pridané identifikačné číslo blacklist zoznamu na ktorom sa URL adresa nachádza a následne je tento záznam o IP toku odoslaný pomocou výstupného rozhrania. Na vyhľadávanie sa používa štandardná C++ štruktúra std::map, v ktorej kľúčom je URL adresa blacklist záznamu a hodnotou je identifikačné číslo daného blacklist záznamu. Kvôli väčšej kompatibilite prebieha unifikácia URL adries prevzatých z blacklist zoznamov. Unifikácia konvertuje URL na ASCII reťazec pomocou knižnice libidn[11].

7.2 Vstup a výstup modulov

Vstupom a výstupom modulov je myslená UniRec šablóna. Vstupná aj výstupná UniRec šablóna oboch modulov obsahuje základný formát COLLECTOR_FLOW, ktorý obsahuje nasledujúce položky:

- SRC_IP - Zdrojová IP adresa. Môže obsahovať IPv4 alebo IPv6 adresu.
- DST_IP - Cieľová IP adresa. Môže obsahovať IPv4 alebo IPv6 adresu.
- SRC_PORT - Zdrojový port transportnej vrstvy (TCP/UDP).
- DST_PORT - Cieľový port transportnej vrstvy (TCP/UDP).
- PROTOCOL - Číslo protokolu transportnej vrstvy.
- TCP_FLAGS - V prípade že sa jedná o TCP tok, tak položka obsahuje TCP príznaky spojené pomocou bitového súčtu (OR).
- TIME_FIRST - Časová značku vzniku IP toku.

- TIME_LAST - Časová značka konca IP toku, teda jeho exportovanie.
- PACKETS - Počet paketov v rámci celého IP toku.
- BYTES - Počet bajtov v rámci celého IP toku.
- LINK_BIT_FIELD - Bitové pole identifikujúce exportér, ktorý daný záznam o IP toku exportoval. Využíva sa iba v CESNET2 sieti.
- DIR_BIT_FIELD - Položka udávajúca smer IP toku. Využíva sa iba v CESNET2.
- TOS - Položka Type of Contents nachádzajúcu sa v hlavičke IP protokolu.
- TTL - Položka Time To Live nachádzajúcu sa v hlavičke IP protokolu.

7.2.1 IP detekčný modul

Ako bolo z časti zmienené vyššie, detekčný modul pre IP na vstupe obsahuje UniRec šablónu COLLECTOR_FLOW. Keďže pracuje iba s IP adresami, tak mu táto šablóna stačí. Na výstupe IP detekčného modulu sa nachádza taktiež šablóna COLLECTOR_FLOW ale naviac sa tu nachádzajú tri ďalšie položky. Celkovú výstupnú šablónu zobrazuje ukážka 7.1.

```
<COLLECTOR_FLOW> , SRC_BLACKLIST , DST_BLACKLIST , EVENT_SCALE
```

Obrázok 7.1: Celková výstupná šablóna IP detekčného modulu

Sémantika nových položiek:

- SRC_BLACKLIST - V prípade že bola označená zdrojová IP adresa, obsahuje táto položka identifikačné číslo blacklist zoznamu na ktorom sa daná IP adresa nachádza.
- DST_BLACKLIST - V prípade že bola označená cieľová IP adresa, obsahuje táto položka identifikačné číslo blacklist zoznamu na ktorom sa daná IP adresa nachádza.
- EVENT_SCALE - Položka obsahuje intenzitu útoku, ktorá predstavuje počet detekcií za určitý časový interval (viď sekcia 7.1.2).

7.2.2 URL modul

URL detekčný modul obsahuje na vstupe taktiež UniRec šablónu COLLECTOR_FLOW, ale keďže pracuje s URL adresami, potrebuje naviac UniRec šablónu HTTP. Celková vstupná šablóna je zobrazená na ukážke 7.2.

```
<COLLECTOR_FLOW> , <HTTP>
```

Obrázok 7.2: Celková vstupná šablóna URL detekčného modulu

Sémantika položiek nachádzajúcich sa v UniRec šablóne HTTP:

- HTTP_REQUEST_METHOD_ID - Identifikačné číslo HTTP metódy.
- HTTP_REQUEST_HOST - Časť URL obsahujúca názov servera.

- HTTP_REQUEST_URL - Zvyšná časť URL dotazu bez názvu servera.
- HTTP_REQUEST_AGENT_ID - Identifikačné číslo klientskej aplikácie.
- HTTP_REQUEST_AGENT - Názov klientskej aplikácia.
- HTTP_REQUEST_REFERER - URL adresa stránky z ktorej prichádza dotaz.
- HTTP_RESPONSE_STATUS_CODE - Návratová hodnota servera pre daný dotaz.
- HTTP_RESPONSE_CONTENT_TYPE - Typ obsahu ktorý posiela server klientovi.

Výstup URL detekčného modulu obsahuje šablónu COLLECTOR_FLOW a ďalšie 3 položky. Celková výstupná šablóna je zobrazená na ukážke 7.3:

```
<COLLECTOR_FLOW> , URL_BLACKLIST , HTTP_REQUEST_HOST ,
HTTP_REQUEST_URL
```

Obrázok 7.3: Celková výstupná šablóna URL detekčného modulu

Položky HTTP_REQUEST_HOST a HTTP_REQUEST_URL sú popísane vyššie a položka URL_BLACKLIST obsahuje identifikačné číslo blacklist zoznamu na ktorom sa daná URL nachádza.

7.3 Použitie

Použitie oboch modulov je veľmi podobné, preto popíšem použitie iba IP modulu a upozorním na rozdiely medzi IP modulom a URL modulom. Moduly dokážu pracovať v dvoch režimoch, pričom režim modulu určuje ako bude modul získavať a aktualizovať svoje blacklist záznamy. Režimy získavania blacklist zoznamov:

- **Statický režim:** Toto je základný režim v ktorom modul pracuje ak nieje explicitne pri spustení nastavený inak. Tento režim nevyužíva knižnicu na sťahovanie blacklist zoznamov z verejne dostupných zoznamov na internete. Užívateľ si musí sam špecifikovať, ktoré IP adresy (URL adresy v prípade URL modulu) sú nedôveryhodné a komunikácia s nimi má byť označená za podozrivú. Zoznam týchto adries uloží do textového súboru a cestu ku tomuto súboru zadá ako parameter modulu pri spustení. Modul tento zoznam spracuje a uloží si IP/URL adresy do pamäti. V prípade že užívateľ chce upraviť tento zoznam, tak stačí iba upraviť pôvodný súbor s IP/URL adresami a poslať modulu signál SIGUSR1. Modul si po prijatí signálu zoznam aktualizuje podla vstupného súboru.
- **Dynamický režim:** Tento režim využíva knižnicu pre sťahovanie blacklist zoznamov z verejne dostupných zoznamov na internete. Stačí pri spustení modulu špecifikovať o ktoré blacklist zoznamy ma užívateľ záujem a tieto sa následne budú periodicky sťahovať a aktualizovať. Zoznam blacklist zoznamov v tomto prípade ale nieje možné za behu meniť a teda ak užívateľ chce tento zoznam upraviť, musí reštartovať modul.

V prípade IP modulu je tiež možné nastaviť vlastnosti agregácie detekcií (URL modul neimplementuje agregačnú funkciu). Základné nastavenie časovania by malo byť vyhovujúce vo väčšine prípadov. Vzhľadom ale na dostupnú operačnú pamäť, môže užívateľ upraviť veľkosť agregačnej tabuľky a tým znížiť nároky na operačnú pamäť.

Kapitola 8

Testovanie

Testovanie navrhnutých detekčných modulov prebehlo nad dátovou sadou, nad ktorou bola prevedená analýza v kapitole 5 . Testované moduly boli taktiež nasadené do živej prevádzky siete CESNET2. IP adresy, ktoré sa nenachádzajú na blacklist zozname boli anonymizované pomocou knižnice Crypto-PAn. Výhodou použitia knižnice Crypto-PAn na anonymizovanie IP adries je zachovanie prefixu anonymizovanej adresy. To znamená že adresy zdieľajúce rovnaký prefix budú zdielať rovnaký prefix aj po anonymizacii.

8.1 Univerzita Hradec Králové

Počet detekovaných udalostí v rámci meraného obdobia zobrazuje tabuľka 8.1. Jedná sa predovšetkým o anonymizačnú sieť TOR, ktorá nie nutne znamená hrozbu. Druhou najviac detekovanou udalosťou je ZeuS botnet, ktorý už môže predstavovať vážne riziko v podobe počítačov nakazených týmto malwarom. Ďalej nasleduje Spamhaus, ktorý by mal predstavovať posielanie SPAMu. Ako posledná detekcia v tabuľke je botnet Feodo. Je vidieť že tejto komunikácie je málo. Keďže sa jedná ale o malware, bolo by vhodné tieto incidenty ďalej riešiť pomocou antivírových programov. V sieti nebol detekovaný ani jeden incident pomocou URL detekčného modulu.

Blacklist	Počet detekcií	Najväčšia intenzita
Tor	603	10
ZeuS	180	37
Spamhaus	122	1239
Feodo	16	42

Tabuľka 8.1: Počet detekcií v rámci jednotlivých blacklist zoznamov v sieti Univerzity Hradec Králové

Tabuľka 8.2 zobrazuje najväčšie detekované udalosti podľa intenzity. Je z nej vidieť adresy podieľajúce sa na tejto komunikácii a teda je možné sa zamerať na konkrétny počítač v sieti, ktorý danú nežiadúcu prevádzku spôsobil.

Čas výskytu detekcie	Zdrojová adresa	Cieľová Adresa	Intenzita	Blacklist
2015-03-11 15:45:22.121	164.37.9.234	**195.20.141.140**	1239	Spamhaus
2015-03-16 15:54:40.592	164.37.9.150	**195.20.141.140**	1138	Spamhaus
2015-03-23 07:51:28.953	164.37.10.201	**195.20.141.140**	1078	Spamhaus
2015-03-12 21:15:10.213	164.37.8.30	**23.14.92.50**	42	Feodo
2015-03-26 19:06:09.543	164.37.6.55	**23.14.92.50**	41	Feodo
2015-03-10 16:02:27.728	164.37.6.240	**23.14.92.50**	41	Feodo
2015-03-26 19:06:09.544	**23.14.92.50**	164.37.6.55	41	Feodo
2015-03-17 14:49:00.066	164.37.9.245	**37.9.175.4**	37	ZeuS
2015-03-17 14:49:00.066	**37.9.175.4**	164.37.9.245	37	ZeuS
2015-03-16 00:55:09.283	164.37.10.176	**37.9.175.4**	30	ZeuS

Tabuľka 8.2: TOP-10 detekovaných udalostí v sieti Univerzity Hradec Králové, zoradené podľa intenzity

8.2 Akademie věd ČR

V prípade siete Akademie věd ČR boli detekované celkom štyri typy nežiadúcej prevádzky, zobrazené v tabuľke 8.3. Na prvom mieste sa umiestnil Spamhaus, ktorý v počte detekcí a tiež intenzite najväčšej detekcie niekoľko násobne prekonal ostatné typy udalostí. V prípade Spamhausu by sa malo jednať o SPAM. Ako bolo ale spomenuté pri návrhu spôsobu detekcie (viď kapitola 6), porovnáva sa iba na základe IP adresy. Preto aj keď môže ísť o komunikáciu s IP adresami ktoré niekedy odosielali SPAM, nie nutne musí byť aj táto detekovaná sieťová prevádzka považovaná za SPAM. Prinajmenšom môže byť považovaná za podozrivú. Toto taktiež platí o druhej detekcii v poradí, ktorou je komunikácia so sieťou Tor. Na posledných dvoch miestach sa umiestnili dva typy nežiadúcej sieťovej prevádzky. Jedná sa o komunikáciu botnetov ZeuS a Feodo. Z tabuľky vyplýva že tejto komunikácie je málo, čo ale môže byť vysvetlené typom prevádzky. Dá sa totiž predpokladať že komunikácia botnetov s ich C&C servermi bude ich autormi obmedzená na minimum a to z toho dôvodu aby nevzbudzovali pozornosť sieťových administrátorov. URL detekčný modul znova nedetekoval žiadne udalosti.

Blacklist	Počet detekcií	Najväčšia intenzita
Spamhaus	45601	7549
Tor	1491	2224
Feodo	108	29
ZeuS	15	25

Tabuľka 8.3: Počet detekcií v rámci jednotlivých blacklist zoznamov v sieti Akademie věd ČR

Desať najintenzívnejších detekovaných udalostí zobrazuje tabuľka 8.4. Je vidieť že sa jedná výhradne o typ Spamhaus, teda SPAM. Čo pridáva tejto detekcii na podozrivosti je fakt že sa jedná iba o jednu podsieť (43.255.191.0/24). Nie nutne sa musí jednať o SPAM, ale je určite vhodné preskúmať pravú príčinu týchto udalostí správcom siete a určiť či sa jedná skutočne o nežiadúcu komunikáciu alebo nie.

Čas výskytu detekcie	Zdrojová adresa	Cieľová Adresa	Intenzita	Blacklist
2015-04-01 01:29:42.921	145.231.95.183	**43.255.191.153**	7549	Spamhaus
2015-04-01 01:29:25.523	**43.255.191.153**	145.231.95.183	7549	Spamhaus
2015-04-16 21:34:31.377	145.231.95.183	**43.255.191.154**	5632	Spamhaus
2015-04-16 21:34:22.011	**43.255.191.154**	145.231.95.183	5632	Spamhaus
2015-03-29 15:19:21.902	145.231.95.172	**43.255.191.180**	5378	Spamhaus
2015-03-29 15:19:21.891	**43.255.191.180**	145.231.95.172	5378	Spamhaus
2015-04-14 02:14:27.751	145.231.95.172	**43.255.191.145**	5284	Spamhaus
2015-04-14 02:14:17.937	**43.255.191.145**	145.231.95.172	5284	Spamhaus
2015-04-03 11:04:28.580	145.231.95.172	**43.255.191.137**	5261	Spamhaus
2015-04-03 11:04:28.582	**43.255.191.137**	145.231.95.172	5261	Spamhaus

Tabuľka 8.4: TOP-10 detekovaných udalostí v sieti Akademie věd ČR, zoradené podla intenzity

8.3 Metacentrum

Poslednou sieťou, ktorá bola analyzovaná je sieť Metacentra. Ako bolo ukázané v analýze (viď sekcia 5.3), jedná sa o sieť s väčším prietokom sieťových dát ako boli predošlé dve siete. Výsledok tohto väčšieho prietoku dát môžeme vidieť na tabuľke 8.5, kde je vidieť niekoľko násobný nárast detekcií v prípade typu Spamhaus a Tor. Čo sa týka detekcie botnetov tak je to podobné, teda veľmi málo udalostí. Toto bolo ale vysvetlené pri sieti Akademie věd ČR 8.2 a teda nieje nutné sa tým znovu zaoberať. Taktiež ako v predošlých sieťach neboli detekované žiadne udalosti pomocou URL detekčného modulu.

Blacklist	Počet detekcií	Najväčšia intenzita
SPAMHAUS	1332008	6707
TOR	27205	9597
ZEUS	13	48
FEODO	12	4

Tabuľka 8.5: Počet detekcií v rámci jednotlivých blacklist zoznamov v sieti Metacentra

Tabuľka 8.6 zobrazuje najintenzívnejšie detekcie. Oproti predošlým sieťam si môžeme všimnúť že na prvých dvoch miestach je komunikácia so sieťou Tor. Toto môže byť legitímna komunikácia ale taktiež sa môže jednať o nežiadúcu komunikáciu v ktorej sa útočník pokúša skryť svoju skutočnú identitu.

Čas výskytu detekcie	Zdrojová adresa	Cieľová Adresa	Intenzita	Blacklist
2015-04-23 15:03:14.367	**212.24.144.188**	145.242.246.111	9597	Tor
2015-04-23 15:02:11.554	145.242.246.111	**212.24.144.188**	9513	Tor
2015-04-23 17:38:07.251	**43.255.191.146**	145.242.45.108	6707	Spamhaus
2015-04-23 17:42:29.723	**43.255.191.146**	145.242.45.41	6705	Spamhaus
2015-04-23 17:38:41.169	145.242.45.113	**43.255.191.146**	6704	Spamhaus
2015-04-23 17:38:51.163	**43.255.191.146**	145.242.45.115	6699	Spamhaus
2015-04-23 17:38:42.376	145.242.45.115	**43.255.191.146**	6697	Spamhaus
2015-04-23 17:37:51.754	**43.255.191.146**	145.242.45.101	6697	Spamhaus
2015-04-23 17:42:30.079	145.242.45.41	**43.255.191.146**	6693	Spamhaus
2015-04-23 17:37:51.778	145.242.45.101	**43.255.191.146**	6693	Spamhaus

Tabuľka 8.6: TOP-10 detekovaných udalostí v sieti Metacentra, zoradené podla intenzity

8.4 CESNET2

V rámci vysokorýchlostnej siete CESNET2 neprebehla žiadna analýza. Taktiež vzhľadom na veľký objem pretečených dát bolo merané obdobie obmedzené na necelý týždeň. Ako môžeme vidieť v tabuľke 8.7, za toto merané obdobie sa vyskytlo rádovo oveľa viac detekovaných udalostí ako v predošlých sieťach počas dlhšieho obdobia. Zaujímavosťou môže byť absencia detekovaných botnetov, čo môže byť spôsobené krátkym intervalom merania. Ďalšou zaujímavosťou sú udalosti detekované pomocou URL detekčného modulu. V pomere k ostatným detekovaným udalostiam je ale URL detekcií veľmi málo. Vysvetlenie prečo tomu tak môže byť sa nachádza v zhodnotení tejto kapitoly (viď sekcia 8.5).

Blacklist	Počet detekcií	Najväčšia intenzita
Spamhaus	11081757	2816
Tor	823068	310
PhishTank	122	-
Malware domains	53	-

Tabuľka 8.7: Počet detekcií v rámci jednotlivých blacklist zoznamov v sieti CESNET2

Najintenzívnejšie detekované udalosti v rámci siete CESNET2 sú zobrazené v tabuľke 8.8. Jedná sa výhradne o udalosti zo zoznamu Spamhaus, predovšetkým pochádzajúcich z jednej IP adresy. Z tohto dôvodu teda môžme uvažovať že sa skutočne jedná o SPAM, no aj tak je nutná hlbšia analýza problému pre určenie skutočnej príčiny.

Čas výskytu detekcie	Zdrojová adresa	Cieľová Adresa	Intenzita	Blacklist
2015-05-09 13:13:09.646	**192.67.160.132**	242.108.140.125	2942	Spamhaus
2015-05-09 13:13:20.252	**192.67.160.132**	159.35.226.248	2816	Spamhaus
2015-05-09 13:13:00.964	**192.67.160.132**	159.35.254.123	2800	Spamhaus
2015-05-09 13:13:20.713	**192.67.160.132**	159.35.254.122	2780	Spamhaus
2015-05-08 17:44:28.091	148.78.90.129	**176.61.138.138**	2646	Spamhaus
2015-05-09 13:13:21.901	**192.67.160.132**	242.214.39.21	2614	Spamhaus
2015-05-09 13:13:22.744	**192.67.160.132**	242.214.183.173	2613	Spamhaus
2015-05-09 13:13:20.289	**192.67.160.132**	145.235.109.61	2583	Spamhaus
2015-05-09 13:13:20.568	**192.67.160.132**	240.156.28.81	2564	Spamhaus
2015-05-09 13:12:48.928	**192.67.160.132**	159.35.149.151	2553	Spamhaus

Tabuľka 8.8: TOP-10 detekovaných udalostí v sieti CESNET2, zoradené podla intenzity

Tabuľka 8.9 zobrazuje najčastejšie detekované URL adresy. V prípade prvej detekcie bolo nutné skrátiť URL adresu, pretože bola príliš dlhá a namiesto nahradených znakov sa v jej názve nachádzajú tri bodky.

Detekovaná URL adresa	Blacklist	Počet detekcií
rtcnbclient.wmi.amu.edu.pl/.../et.htm	PhishTank	64
www.duba.com	PhishTank	42
www.sugarsync.com	Malware domains	26
ssl.aukro.ua	Malware domains	12
cwbl.eshopcomp.com/search/www.ebay.com	PhishTank	10
directexe.com	Malware domains	8
jl.chura.pl	Malware domains	2
clicksor.com	Malware domains	2
updateyourself.wapka.mobi/index.xhtml	PhishTank	1
xoomer.virgilio.it	Malware domains	1

Tabuľka 8.9: TOP-10 detekovaných udalostí URL detekčným modulom v sieti CESNET2, zoradené podla intenzity

8.5 Zhodnotenie testovania

Pri všetkých analyzovaných sieťach pomocou navrhnutého spôsobu detekcie si môžeme všimnúť isté opakujúce sa vzory v detekovaných udalostiach. Predovšetkým v počte detekovaných typov udalostí. Najviac bolo detekovaných udalostí pomocou zoznamov Spamhaus a Tor. Veľmi málo udalostí bolo detekovaných pomocou zoznamov ZeuS a Feodo. Ako už bolo povedané, takéto rozpoloženie počtu detekcií je vysvetliteľné. V prípade Spamhaus sa jedná o masívne posielanie SPAMu a v prípade anonymizačnej siete Tor sa jedná o prístup na internet, prezeranie webových stránok a pod. Obe tieto aktivity vytvárajú množstvo IP tokov. Na druhej strane komunikácia botnetov s ich C&C servermi bude obmedzená aby sa predišlo ich detekcii a teda budú vytvárať málo IP tokov.

URL detekčný modul oproti IP detekčnému modulu toho detekoval veľmi málo. V troch zo štyroch sietí nedetekoval absolútne nič. Po neskoršej analýze tejto anomálie som prišiel nato že väčšina moderných webových prehliadačov využíva podobné (ak nie rovnaké) blac-

klist zoznamy a prístup na škodlivé URL adresy blokuje. Z tohto dôvodu sa nevytvárajú žiadne IP toky a preto ani neboli detekované skoro žiadne udalosti tohto typu.

Kapitola 9

Záver

V úvode tejto bakalárskej práce bolo popísane čo sú to lokálne počítačové siete, nástroje na ich spravovanie a najčastejšie problémy z pohľadu bezpečnosti, ktoré sa v týchto sieťach vyskytujú. Bolo vysvetlené čo je to IP tok a spolu s ním boli popísané najpoužívanejšie protokoly NetFlow a IPFIX, ktoré pracujú s IP tokmi. Ďalej boli uvedené nástroje ako FlowMon exportér a IPFIXcol kolektor, ktoré sú základom monitorovacej architektúry poskytujúcej rozšírené dáta o IP tokoch. Následne bol objasnený framework Nemea vyvíjaný pod združením CESNET, pomocou ktorého je možné vytvoriť komplexný systém, umožňujúci analýzu IP tokov a detekciu anomálií v reálnom čase. Pomocou frameworku Nemea a tiež manuálnym prechádzaním dát, bola vykonaná analýza troch rozličných sietí. Jednalo sa o sieť Univerzity Hradec Králové, sieť Akademie věd ČR a sieť Metacentra. Na základe tejto analýzy vznikol návrh na spôsob detekcie nežiadúcej sieťovej komunikácie. Princípom tejto detekcie je vyhľadávanie nedôveryhodných účastníkov sieťovej komunikácie podľa IP alebo URL adresy a následne označenie takejto sieťovej prevádzky ako nežiadúcej. Boli tiež uvedené verejne dostupné blacklist zoznamy, obsahujúce potencionálne nebezpečné IP a URL adresy. V oblasti lokálnych počítačových sietí je navrhnutý spôsob výhodný, pretože správci siete môžu ľahko dohľadať nakazený počítač v sieti, ktorý komunikoval s IP/URL adresou na blacklist zozname a následne podniknúť patričné opatrenia.

Aby sa zachovala modularita frameworku Nemea, bol navrhovaný spôsob detekcie rozdelený na dve nezávisle časti, jedna pre detekciu IP adries a druhá pre detekciu URL adries. Ďalej bola popísaná implementácia oboch modulov pre framework Nemea, ktoré vykonávajú navrhovaný spôsob detekcie s použitím predložených verejne dostupných zoznamov. Implementované moduly boli otestované na všetkých sieťach na ktorých prebehla analýza a taktiež na sieti CESNET2. Počas testovania bol IP detekčný modul schopný detekovať nežiadúcu sieťovú prevádzku na každej z otestovaných sietí. Naopak URL detekčný modul v sieťach nad ktorými prebiehala analýza nedokázal detekovať žiadnu nežiadúcu sieťovú prevádzku a jediné detekované udalosti pochádzali zo siete CESNET2. Toto môže byť spôsobené práve tým, že moderné webové prehliadače využívajú podobné blacklist zoznamy a pristúp na tieto URL adresy blokujú.

Z testovania vyplýva že navrhovaný spôsob detekcie dokáže označiť nežiadúcu sieťovú prevádzku, ktorú je inak veľmi ťažké až nemožné detekovať iba na základe typických vzorov. Na druhej strane je tu ale aj šanca falošne pozitívnych detekcií. Celková hodnovernosť detekovaných udalostí teda záleží na použitých zoznamoch, ktoré musia mať určitú mieru dôveryhodnosti inak je použitie takýchto zoznamov kontraproduktívne.

Možné rozšírenie tejto práce, ktoré by zvýšilo mieru pozitívne detekovaných udalostí je odoberanie záznamov ktoré sa ukázali že spôsobujú falošne pozitívne detekcie. Toto by

ale vyžadovalo koreláciu detekčného modulu s iným modulom v systéme Nemea, ktorý by dokázal určiť ktoré detekcie sú falošne pozitívne a ktoré nie.

Literatúra

[1] System Administration Guide, Volume 3 - Sun Microsystems [online]. 2000 [cit. 2015-05-12].
URL http://docs.oracle.com/cd/E19455-01/806-0916/

[2] IP Flow Information Export (IPFIX) Entities - IANA [online]. 2007 [cit. 2015-05-12].
URL http://www.iana.org/assignments/ipfix/ipfix.xhtml

[3] NetFlow Services Solutions Guide - Cisco Systems [online]. 2007 [cit. 2015-05-12].
URL http://www.cisco.com/c/en/us/td/docs/ios/solutions_docs/netflow/nfwhite.html

[4] FlowMon - INVEA-TECH [online]. 2008 [cit. 2015-05-12].
URL https://www.invea.com/cs/products/flowmon

[5] NetFlow - Wikipedia [online]. 2014 [cit. 2015-05-12].
URL http://en.wikipedia.org/wiki/NetFlow

[6] Cryptography-based Prefix-preserving Anonymization [online]. 2015 [cit. 2015-05-12].
URL http://www.cc.gatech.edu/computing/Telecomm/projects/cryptopan/

[7] FastBit: An Efficient Compressed Bitmap Index Technology [online]. 2015 [cit. 2015-05-12].
URL https://sdm.lbl.gov/fastbit

[8] IPFIXcol - CESNET [online]. 2015 [cit. 2015-05-12].
URL https://github.com/CESNET/ipfixcol/

[9] SMI Network Management Private Enterprise Codes - IANA [online]. 2015 [cit. 2015-05-12].
URL
http://www.iana.org/assignments/enterprise-numbers/enterprise-numbers

[10] Building Scalable Syslog Management Solutions - Cisco Systems [online]. December 2002 [cit. 2015-05-12].
URL http://www.cisco.com/c/en/us/products/collateral/services/high-availability/white_paper_c11-557812.html

[11] GNU IDN Library - Libidn [online]. Január 2012 [cit. 2015-05-12].
URL http://www.gnu.org/software/libidn/

[12] Bartoš, V.; Žádnik, M.; Čejka, T.: Nemea: Framework for stream-wise analysis of network traffic. CESNET Technical report 9/2013, December 2013.

[13] Claise, B.: Cisco Systems NetFlow Services Export Version 9. RFC 3954, Október 2004.
URL http://www.ietf.org/rfc/rfc3954.txt

[14] Claise, B.; Trammell, B.; Aitken, P.: Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information. RFC 7011, September 2013.
URL http://www.ietf.org/rfc/rfc7011.txt

[15] Donahue, G. A.: *Network Warrior*. Beijing : O'Reilly, druhé vydání, 2011, ISBN 978-1-449-38786-0, 1-3 s.

[16] Gerhards, R.: The Syslog Protocol. RFC 5424, Marec 2009.
URL http://www.ietf.org/rfc/rfc5424.txt

[17] Harrington, D.; Presuhn, R.; Wijnen, B.: An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks. RFC 3411, December 2002.
URL http://www.ietf.org/rfc/rfc3411.txt

[18] Knuth, D. E.: *The art of computer programming / Vol. 3, Sorting and serching*. Upper Saddle River : Addison-Wesley, třetí vydání, 1998, ISBN 0-201-89685-0, 406-410 s.

[19] Srisuresh, P.; Holdrege, M.: IP Network Address Translator (NAT) Terminology and Considerations. RFC 2663, August 1999.
URL http://www.ietf.org/rfc/rfc2663

Príloha A

Obsah CD

- Zdrojový kód vyvíjaného Nemea modulu.
- Manuál pre sprevádzkovanie modulu.
- Elektronická verzia bakalárskej práce vo formáte PDF.
- Zdrojové texty bakalárskej práce pre systém LaTeX.
- Vzorka dát sieťovej prevádzky s anonymizovanými IP adresami.