Sem vložte zadání Vaší práce.

České vysoké učení technické v Praze
Fakulta informačních technologií
Katedra počítačových systémů

Diplomová práce

Analýza síťového provozu s pomocí komunikačních map

Bc. Tomáš Vicher

Vedoucí práce: Ing. Tomáš Čejka

17. března 2016

Poděkování

Rád bych poděkoval vedoucímu práce Ing. Tomáši Čejkovi za rady a čas, které mi věnoval. Dále bych rád poděkoval rodině a přítelkyni za podoporu při práci.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 46 odst. 6 tohoto zákona tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou, a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla, a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené. Každá osoba, která využije výše uvedenou licenci, se však zavazuje udělit ke každému dílu, které vznikne (byť jen zčásti) na základě Díla, úpravou Díla, spojením Díla s jiným dílem, zařazením Díla do díla souborného či zpracováním Díla (včetně překladu), licenci alespoň ve výše uvedeném rozsahu a zároveň zpřístupnit zdrojový kód takového díla alespoň srovnatelným způsobem a ve srovnatelném rozsahu, jako je zpřístupněn zdrojový kód Díla.

V Praze dne 17. března 2016 .....

Odkaz na tuto práci

Vicher, Tomáš. *Analýza síťového provozu s pomocí komunikačních map.* Diplomová práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2016.

Abstrakt

Podstatou a cílem diplomové práce je navrhnout metodu zpracování informací o síťových tocích k vygenerování a udržování grafu komunikace mezi jednotlivými uzly sítě, popřípadě agregovaných uzlů (podsítí). Navrhnout využití pro vzniklé struktury ke sledování vývoje provozu, případně k detekci neobvyklých jevů na síti. Výsledkem práce je detekční metoda implementovaná jako modul do systému NEMEA. Součástí práce je také rešerše možností monitoringu sítě, útoků na síť a statistických metod pro analýzu síťových toků.

Klíčová slova síťová analýza, síťový tok, Holt-Wintersova metoda, python, monitorování, graf, simulace

Abstract

The purpose and goal of the master thesis is to design method of processing information about network flow to generate and preserve graph of communication between network nodes or agregated nodes (subnetworks). The purpose is to design and implement usage for the formed structures to monitor network traffic and detect anomaly. The result of master thesis is the detection method implemented as module to the NEMEA system.

Keywords network analysis, network flow, Holt-Winters method, python, monitoring, graph, simulation

Obsah

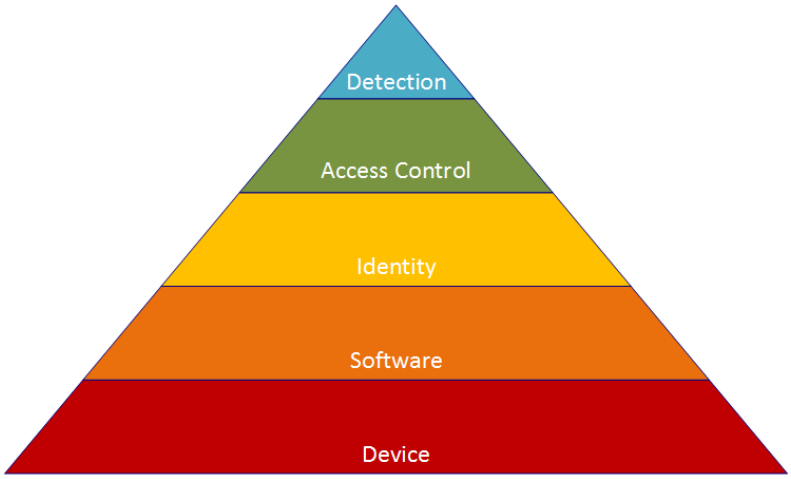
Seznam obrázků

Úvod

Motivace

Počet počítačů a dalších zařízeních spojených sítí se stále zvyšuje jak v domácnostech, tak ve firmách spolu s celosvětovým trendem zlepšující se dostupnosti internetového připojení [5]. Reakcí na tyto změny je potřeba vývoje a zlepšování síťového monitoringu. Motivací pro vývoj monitoringu sítí je mnoho, zmínit mohu například zvýšení bezpečnosti pro citlivá a důležitá data přenášená po síti i uchovávaná v zařízeních připojených k síti. Kvůli stále rostoucímu trendu kyberkriminality je důležité nebrat monitoring jako okrajovou záležitost v IT.[6] Síťový monitoringu lze využít i z hlediska lepšího využití hardwaru a sítí.

Pro zařazení monitoringu do komplexnějšího přehledu bezpečnosti v IT je uveden obrázek s hierarchií zabezpečení.1. Politika zabezpečení v IT začíná od správy zařízení a inventarizace, pokračuje správou a údržbou software. Další vrstvou je kontrola nad identitou uživatelů, zaměstnanců,... Na předchozí vrstvy zabezpečení navazuje řízení přístupu ke zdrojům a informacím. Na vrcholu této bezpečnostní pyramidy se nachází detekční metody - téma na které je zaměřena tato diplomová práce.[1]



Obrázek 1: Hierarchie zabezpečení v IT. Zdroj: [1]

Struktura práce

Tato práce se zabývá monitorováním sítí. Je zaměřena především na zpracování síťových toků, na kterém je založena návrhová a implementační část práce. V první části je vypracována rešerše zpracování síťového provozu, anomálií a statistických metod. Ve druhé části je s informacemi získanými rešerší navrženo řešení pro monitoring síťového provozu s využitím síťových toků. Třetí část popisuje použité nástroje a implementaci navrženého řešení jako modul do systému Nemea[1]. Ve čtvrté části jsou uvedeny testy s výsledky testování implementovaného modulu. V závěrečné části je uvedena diskuze, zhodnocení splněných cílů práce a možnosti budoucího rozšíření.

[1]Nemea je systém fungující na základě modulů umožňující real-time monitoringu sítě.

Kapitola 1

Specifikace cílů

Kapitola popisuje rozbor zadání, tedy rozbor cílů práce.

1.1 Rozbor zadání

Nastudujte způsob monitorování síťového provozu na základě sledování síťových toků

Úkolem je vyhledat a nastudovat informace o síťových tocích. Nastudovat možnosti monitoringu sítě na základě sledování síťových toků. S tímto úkolem souvisí zpracování informací o dostupných technologiích pro práci s toky, způsobů síťových útoků a statistických metod pro zpracování a vyhodnocování dat.

Navrhněte metodu zpracování informací o síťových tocích k vygenerování a udržování grafu komunikace mezi jednotlivými uzly sítě popř. agregovaných uzlů (podsítí)

Na základě nastudovaných informací z předchozího bodu je úkolem navrhnout metodu, jakou lze zpracovat a udržovat informace přijaté ze síťových toků v podobě grafu.

Navrhněte využití vzniklých grafových struktur ke sledování vývoje provozu na síti, případně k detekci neobvyklých jevů na síti

S využitím informací z prvního bodu zadání je úkolem navrhnout způsob využití dat připravených v grafové struktuře za účelem vyhodnocování provozu v síti. Při sledování provozu detekovat neobvyklý stav ukazující na anomálii v provozu.

Navrženou detekční metodu implementujte jako modul do systému Nemea

Z návrhu vzniklého v předchozích bodech je úkolem implementovat aplikaci, která bude odpovídat specifikacím systému Nemea pro použití jako modul a integraci do Nemea systému.

Otestujte tento mechanismus za pomoci simulace příp. s použitím reálného provozu z počítačové sítě

Úkolem je implementovanou aplikaci otestovat na datech z reálného nebo simulovaného síťového provozu a otestovat implementované monitorovací a detekční metody.

Kapitola 2

Rešerše

V této kapitole jsou uvedeny způsoby monitoringu a detekce změn v síti, útoků na síť a statistických metod. V první sekci kapitoly jsou uvedeny možnosti monitoringu změny v síti. V dalších sekcích rešerše jsou popsány metody monitoringu síťového provozu na úrovni paketů a síťových toků. V souvislosti s monitoringem toků je v rešerši zahrnuta sekce věnující se systému Nemea pro který byl v rámci práce navrhnut a implementován modul. Dále se rešerše zaměřuje na útoky na síť, které lze monitorovat a detekovat s využitím analýzy síťových toků. V závěrečné části uvádím rešerši statistických metod pro použití k detekci anomálií v síťovém provozu.

2.1 Zpracování síťového provozu

Monitorování síťového provozu lze podle způsobu zpracování síťových dat rozdělit na dvě skupiny. Analýza paketů, při které je vyhodnocování anomálií prováděno pomocí signatur a analýzu chování sítě, která probíhá na základě zpracování síťových toků.

2.1.1 Analýza paketů pomocí signatur

Jednou z možností rozpoznání anomálií v síťovém provozu je detekce na základě signatur na úrovní jednotlivých paketů. Tento způsob využívají IDS systémy[2]. Signatury jsou vzory, které jsou vyhledávány v síťovém provozu. Mohou být vytvořeny podle návrhu systému ve kterém jsou použity. Záleží na konkrétním IDS jakou možnost tvorby nebo modifikace signatur uživatelům poskytuje, ale základním předpokladem pro úspěšnou detekci pomocí signatur je výběr IDS nebo vytvoření správných signatur pro problém, na který se má detekční systém zaměřovat. Níže je uvedeno příkladem několik vybraných signatur s popisem situace, kterou s nimi lze monitorovat. [7]

[2]Intrusion detection system

- Detekce připojení IP adresy, která je rezervovaná pro jiné zařízení. Lze odhalit kontrolou parametru zdrojové adresy v hlavičce paketu.
- Detekce paketů se špatně nastavenými flagy. Lze odhalit porovnáním s dobrými/špatnými vzorovými kombinacemi.
- Identifikace nežádoucího doručeného emailu. Lze rozpoznat porovnáním obsahu předmětu, nebo jiných částí zprávy se vzorem.
- Pokus o DNS buffer overflow může být odhalen rozparsováním DNS dotazu a kontrolou délky obsahu.
- DOS útok který je realizován pomocí velkého množství stejných dotazů nebo obsahujících charakteristickou část lze odhalit pomocí jejich počítání.

Z principu používání signatur vychází nevýhoda v nízké účinnosti odhalování zero-day útoků[3]. V bezpečnostních systémech založených na principu signatur v té době ještě neexistují signatury, se kterými by je mohl systém odhalit. V softwaru na který je útok cílen obvykle existuje chyba, která nebyla ještě opravena.[8] Silnou stránkou systémů založených na signaturách je detekce útoků, které byly již v minulosti zaznamenány. Proto tyto systémy vyžadují pravidelné aktualizace signatur.

2.1.2 Analýza toků a chování sítě

Analýza chování sítě (behaviorální analýza) je technika monitorování sítě využívající sledování a vyhodnocování síťového provozu pomocí toků. Pro definování síťového toku lze v literatuře najít různé výklady. V tomto článku je použita definice podle IPFIX od IETF organizace.[9] Síťový tok je definován jako sada IP paketů procházejících pozorovaným místem v síti v určeném časovém intervalu. Všechny pakety patřící do jednoho toku se vyznačují množinou shodných parametrů. Každý z těchto parametrů toku je vyhodnocen z následujících parametrů paketu:

- jedno nebo více políček z hlavičky paketu (např. cílová IP adresa), políčka z transportní části hlavičky paketu(např. číslo cílového portu), nebo políčka z aplikační části hlavičky (např. RTP políčka [10])
- jedna nebo více charakteristik samotného paketu
- jedno nebo více políček odvozených ze zpracovaných paketů (např. IP adresa dalšího hopu)

V IPFIX terminologii jsou pak používané parametry nazývány klíči. Formát toku složený z pěti klíčů může vypadat například takto:

[3]Zero-day je typ útoků, které jsou realizovány v den, kdy byly vytvořeny.

```
(ip_src,ip_dst,port_src,port_dst,protocol)
```

Monitoringem toků je možné sledovat a detekovat různé způsoby využití sítě a anomálie v provozu. Sledovat lze síťové spoje mezi počítači, síťovými počítačovými periferiemi, například síťovými tiskárnami, mezi routery a dalšími připojenými zařízeními. Monitorovat lze různé obecné statistiky provozu (množství přenesených dat, počet přenesených paketů, počet přenesených toků,...). Ze sledování toků lze s využitím vhodné techniky a správně vybraných parametrů monitorovat i specifičtější případy. Příkladem monitoringu sítě zaměřeného na konkrétní způsob užívání sítě je monitoring struktury a provozu v síti z pohledu emailové komunikace a detekce změny struktury spojení ve firemní emailové síti.

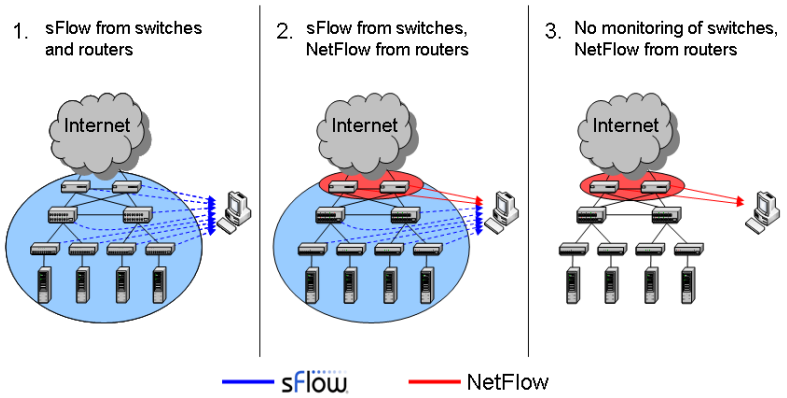
Analýza chování sítě doplňuje metodu analýzy sítě založené na signaturách. Oproti signaturové analýze lze pomocí analýzy chování sítě lépe odhalit zero-day útoky, na které ještě nebyly aktualizovány signatury. Nevýhodou zpracování toků jsou horší výsledky v detekci již známých útoků zanesených do signatur. Zpracování toků nabízí oproti paketové analýze vyšší rychlost, které je dosaženo tím, že zpracování paketů vyžaduje operace s větším množstvím dat, než zpracování toků. Do toku může být sdruženo více paketů a obsažena jen část informace z paketů, které sdružuje. Do toku se nezahrnuje například datová a další části, které nejsou ve formátu toku obsaženy. IDS pracující s toky může být navrhnut tak aby zpracovával toky obsahující pouze informace, které jsou potřebné k jeho funkčnosti. Tím lze eliminovat množství zpracovaných dat a zvýšit rychlost.

Díky odlišným vlastnostem analýzy toků a paketů lze využít oba mechanizmy tak, aby se vzájemně doplnily jako dvoufázová ochrana. Rychlejší zpracování toků lze využít jako první fázi na ochranu celé sítě (například celé infrastruktury ve firmě) a pomalejší zpracování paketů lze nasadit jako druhou fázi na místech, kde byla v první fázi zachycena podezřelá aktivita, nebo na místech, která jsou z pohledu infrastruktury nebo důležitosti zpracovávaných a uchovávaných dat kriticky důležitá.

2.1.3 Technologie monitoringu sítí

V praxi jsou pro monitoring sítí využívány různé technologie (NetFlow, sFlow, JFlow, NetStream, IPFIX, atd.). Zásadní odlišnost se nachází mezi sFlow a ostatními jmenovanými. SFlow operuje na druhé L2 (linkové) vrstvě ISO/OSI síťového modelu. NetFlow a další jmenované technologie fungují na třetí L3 a čtvrté L4 vrstvě ISO/OSI.[2] V obrázku 2je vidět typické nasazení a případně vzájemné doplnění sFlow a NetFlow Ve třech scénářích je shrnuto typické využití.[2]

1. Všechny switche a routery podporují sFlow. Data ze switchů a routerů sbírá a vyhodnocuje centrální zařízení, které poskytuje přehled nad všemi zapojenými prvky.



Obrázek 2: Nasazení NetFlow a sFlow. Zdroj: [2]

2. Situace nastává v prostředí, kdy switche podporují sFlow a routery NetFlow. Stává se tak typicky v situacích kdy jsou switche od jiného výrbce než routery. Síťová infrastruktura zůstává celá monitorována.

3. Switche nepodporují monitoring provozu, routery podporují NetFlow. Monitoring zařízení a provozu je omezen pouze na routery.

2.1.3.1 Netflow/IPFIX

V práci jsou uvedeny informace o NetFlow/IPFIX architektuře, která je podporována kolektorem v systému Nemea. IPFIX je protokol vycházející z NetFlow. NetFlow je protokol vytvořený firmou Cisco Networks, jeho hlavním účelem je monitoring sítě s pomocí toků. IPFIX lze popsat jako kopii NetFlow, která původní NetFlow rozšiřuje o některé důležité parametry.[11] Dvě zásadní rozšíření, která přináší IPFIX oproti NetFlow jsou:

- IPFIX umožňuje zaznamenávat vendor ID. Tím může výrobce síťového hardwaru přidat své informace do toku. Tento postup umožňuje přenášet v tocích informace, které obvykle zpracovávají monitorovací systémy pomocí SNMP, nebo jsou ukládány do syslogu.

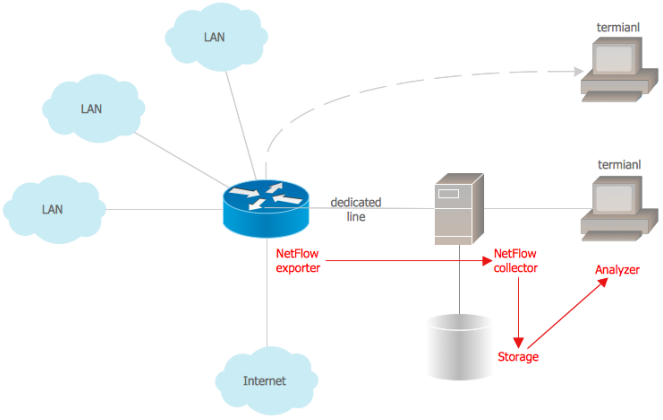
- IPFIX oproti NetFlow povoluje proměnnou délku políček

V obrázku 3 je zakresleno základní schéma NetFlow architektury, obsahující exportér, kolektor a analyzér.

Exportér je podle IETF popsán jako zařízení (například router) podporující NetFlow služby. Exportér monitoruje pakety vstupující do pozorovaného bodu a z těchto paketů vytváří toky. Informace exportuje v podobě záznamu toku do kolektoru.

NetFlow kolektor je popsán jako zařízení, které přijímá záznamy toků z jednoho nebo více exportérů. Obdržené záznamy zpracovává a ukládá informace o záznamu toku. Ukládané záznamy mohou být před uložením ještě

agergovány.[12] Data z kolektoru zpracovává třetí prvek zakreslený v nákresu architektury, analyzer. Analyzer je aplikace, která vyhodnocuje výsledná data, případně vytváří reprezentaci dat pro uživatele.[3]



Obrázek 3: Architectura NetFlow. Zdroj: [3]

2.2 Anomálie v síťovém provozu

Detekce anomálií v síťovém provozu je problémem nalezení neobvyklých vzorů chování v síti, které zasahují mimo hranice síťového provozu, který je určen jako normální.[13] Pro detekování anomálií je potřeba určit referenční stav sítě. Při porovnání s referenčním stavem lze detekovat anomálie. Při vyhodnocování anomálií mohou vznikat falešně pozitivní a falešně negativní stavy. V případě, kdy je nahlášena anomálie, která ve skutečnosti neexistuje, jedná se o falešně pozitivní stav. V případě nenahlášení existující anomálie se jedná o falešně negativní stav. Detekovat anomálie lze s určením konfidenčního intervalu. Konfidenční interval (jinak interval spolehlivosti) lze definovat takto takto:

Nechť Θ_1 a Θ_2 jsou dvě výběrové statistiky takové, že:

$$P(\Theta \in (\Theta_1, \Theta_2)) = 1 - \alpha$$

kde $\alpha \in (0, 1)$. Pak interval (Θ_1, Θ_2) nazveme intervalem spolehlivosti pro parametr Θ s parametrem spolehlivosti $1 - \alpha$. Používá se též termínu $100(1 - \alpha)\%$ interval spolehlivosti nebo konfidenční interval.

Konfidenční interval je tedy oblast, kde se s určitou pravděpodobností nachází pozorovaný parametr.[14] S tímto mechanizmem lze určovat anomálie s určitou spolehlivostí, na základě toho, jestli se pozorovaný parametr nachází

uvnitř nebo mimo konfidenční interval. Toto vyhodnocování vede na statistickou disciplínu testování hypotéz.[15] Určení konfidenčního intervalu ovlivní citlivost detekce, kdy v případě většího konfidenčního intervalu může docházet k vyššímu množství falešně negativních stavů. Zvolení malého konfidenčního intervalu může naopak vyvolat větší množství falešně pozitivních detekcí.

Ovlivnit spolehlivost detekce lze zvoleným postupem pro detekci anomálie. Jednodušší způsob detekce anomálie je detekovat vždy, když je naměřená hodnota mimo konfidenční interval. Takový způsob detekce může vést na velké množství falešně pozitivních hlášení. Robustnější systém pro detekci je s využitím klouzavého okénka fixní velikosti s řadou výsledků pozorování. Pokud je množství naměřených hodnot překračující konfidenční interval vyšší než předem zvolená hranice, je zaznamenána anomálie.[16] Nevýhodou takového řešení je zpoždění v detekci, které odpovídá velikosti klouzavého okénka.

Vyhodnocováním různých parametrů sítě lze sledovat změny ve struktuře, v množství síťového provozu, změny rozložení provozu v síti, změny zastoupení různých složek v provozu a dalších vlastností.

2.2.0.2 Vyhodnocování změn

Jednou z možností pro sledování a vyhodnocování změn v síti je využití některého generativního síťového modelu, ve kterém je definováno pravděpodobnostní rozdělení pro zkoumané parametry sítě. Jako model může být zvolen některý z generativních modelů pro náhodné grafy(GHRG, stochastický blokový model, hierarchický grafový model, Kroneckerův graf).[19][20][21]

Tímto způsobem je možné vyhodnocovat změny ve struktuře sítě. Podle vyhodnocování různých parametrů sítě lze pozorovat například:

- Rozdělování jedné komunity v síti do dvou.
- Spojování dvou komunit do jedné.
- Formování komunity, když jedna skupina vytvoří hrany, kterými se spojí s druhou.
- Fragmentace, když jedna ze skupin ztratí všechny hrany.

Další možností je změny v síti sledovat na základě vyhodnocování po sobě jdoucích obrazů sítě, rozdělených po časových oknech, kde se jednotlivé obrazy sítě vyhodnocují jako řady skalárních hodnot vyčtených ze sledovaných parametrů sítě. Takto připravené parametry lze zpracovat některou z metod vhodných pro vyhodnocení časových řad.[22] Statistické metody jsou podrobněji popsány v kapitole Statistické metody2.4. Tímto způsobem je možné monitorovat změny provozu v síti, které mohou ukazovat na probíhající síťové útoky. Níže jsou uvedeny některé druhy síťových útoků a jejich typické projevy jako síťové anomálie.2.2.12.2.22.2.32.2.4

2.2.1 Odmítnutí služby (DOS)

Mezi DOS útoky patří více konkrétních druhů útoků. Různé druhy jsou více či méně dobře rozpoznatelné monitorováním toků. Mezi lépe rozpoznatelné patří brute force DOS V tomto typu útoku je podstatou přetížení zdroje nebo sítě, které způsobí nedostupnost služby.

Typy DOS útoky obtížně rozeznatelné z analýzy toků jsou založeny na principu sémantického útoku. V těchto případech nedochází k odmítnutí služby na základě zahlcení, ale kvůli obsahu. Příkladem je starší způsob, dnes již známého útkou s názvem Ping of Death. Útočník odesílá záměrně poškozené ping pakety, které způsobí pád systému. Nemožnost identifikace tohoto útoku pomocí toků je způsobena tím, že je při tomto útoku použit pouze jeden ICMP[4] tok.

...doplnit

2.2.2 Skenování

2.2.3 Viry, červy

2.2.4 Botnety

2.3 Nemea

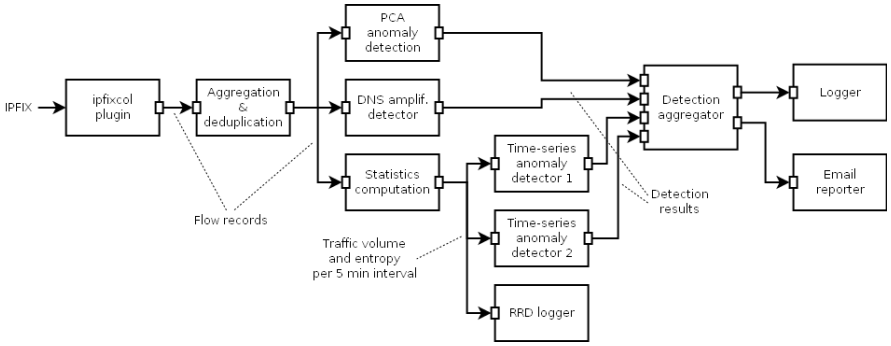
2.3.1 Struktura systému

Nemea je framework, který umožňuje sestavení monitorovacího systému z jednotlivých modulů. Moduly komunikují přes vstupní a výstupní rozhraní. Data mezi moduly jsou přenášena pomocí TRAP platformy ve formátu, který specifikuje protokol UniRec. Zpracování dat v Nemea probíhá v reálném čase. Data se posílají jako proudy z výstupních do vstupních rozhraní. Načítání dat je možné provádět v reálném čase z vlastního IPFIX kolektoru nebo číst předem připravená uložená data. V obrázku 4 je zobrazen Nemea systém v konfiguraci pro komplexní monitoring sítě. Aktivní jsou moduly pro sběr dat, zpracování dat, vyhodnocení a detekci anomálií, zpracování a report výsledků. Rozhraní v Nemea jsou implementována jako jednostranná, pro obousměrnou komunikaci tedy musí mít modul definováno minimálně jedno vstupní a jedno výstupní. Formát zasílaných dat je mezi moduly dohodnut při připojení modulu do systému. Modul tak může přijímat a odebírat pouze potřebná data.

2.3.2 Moduly

Každý modul funguje v systému jako samostatný proces. Architektura modulů jako samostatných procesů přináší více možností pro návrh, implementaci a používání modulů, než zpracování systému do jedné zkompilované aplikace.

[4]Protokol používaný pro oznamování chyb nebo nedostupnosti služby.



Obrázek 4: Příklad Nemea sestavení systému. Zdroj: [4]

- Moduly mohou být spuštěny a zastaveny v různém čase nezávisle na ostatních.
- Moduly lze spravovat ze strany operačního systému nezávisle na dalších částech Nemea systému.
- K implementaci modulů je možné použít různé programovací jazyky.
- Zdroje přidělované operačním systémem jsou odděleny pro každý modul (proces).
- V případě pádu aplikace se nevyřadí z provozu celý Nemea systém
- Moduly lze samostatně aktualizovat.

Moduly zachovávají jednotnost ve formátu dat a v použití rozhraní(TRAP,UniRec). Tím si zachovávají kompatibilitu a možnost zapojení do systému. Moduly mohou mít více vstupních i výstupních rozhraní a mohou být spojeny s více moduly. V případě ztráty spojení se pokouší o jeho obnovení.

2.4 Statistické metody

V této sekci jsou popsány metody jednoduchého a exponenciálního klouzavého průměru pro odhadování trendu posloupnosti časových řad, které mohou být využity při vyhodnocování síťového provozu. Dále jsou popsány metody využívající exponenciálního klouzavého průměru pro vícenásobné vyhlazování a predikci hodnot počítajících s trendem a sezónností. S popisem metod je uveden i způsob detekce anomálií.[17]

2.4.1 Jednoduchý klouzavý průměr

Metoda jednoduchého klouzavého průměru je založena na výpočtu aritmetického průměru z předchozích období zpracovávané časové řady. Výhodou jed-

noduchého klouzavého průměru je například oproti exponenciálnímu klouzavému průměru zobrazení skutečného aritmetického průměru z vyhodnocovaných dat.[18] Vzorec pro výpočet je:

$$SMA = \frac{p1 + p2 + \ldots + pn}{n}$$

P1 až pn jsou hodnoty z časové řady, n je počet hodnot.

2.4.2 Exponenciální vyhlazování

Exponenciální vyhlazování, jinak exponenciální vyrovnávání, či exponenciální klouzavý průměr, je obdobně jako jednoduchý klouzavý průměr založeno na výpočtu z předchozích pozorovaných hodnot - z časové řady. Oproti jednoduchému klouzavému průměru přináší výhodu v rychlejší reakci na změny při analýze časových řad.[18]

Z časové řady je u exponenciálního vyhlazování sestavován trend, který lze vyjádřit jako polynom k-tého stupně vzorcem:

$$\beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_k t^k$$

S využitím vzorce trendu lze získat vzorec pro výpočet konkrétní pozorované hodnoty v čase jako:

$$y_{t-j} = T_{t-j}^{(k)} + \epsilon_{t-j} = \beta_0 - \beta_1 j + \beta_2 j^2 - \ldots + (-1)^k \beta_k j^k + \epsilon_{t-j}$$

kde
y_{t-j} je pozorovaná hodnota,
$t = 1, 2, \ldots, n$ index aktuální pozice v pozorování,
$j = 0, 1, \ldots, t-1$ index stáří pozorování vztažený k indexu v okamžiku pozorování,
ϵ náhodná složka.
Záporné hodnoty ve vzorci jsou z důvodu, že je vzorec odvozován od času t, od kterého se další složky odvíjí do minulosti$t - j$. U exponenciálního vyhlazování jsou přiřazeny k jednotlivým hodnotám ve zpracovávané časové řadě váhy, které směrem do vzdálenější minulosti exponenciálně klesají. Odhad parametrů lze vyjádřit váženou metodou nejmenších čtverců:

$$\min_{\beta_0,\ldots,\beta k} \sum_{j=0}^{t-1} a^j (y_{t-j} - T_{t-j}^{(k)})^2$$

kde $0 < \alpha < 1$, tudíž α^j s rostoucím j klesá. Z tohoto vzorce jsou odvozeny parametry $\beta_0, \ldots, \beta_k$. Tímto postupem jsou parametry odvozovány znova z aktuálních dát při každém opakování pozorování. Pro snížení časové náročnosti mohou být využity vyhlazovací statistiky, sestavené z počátečních odhadů parametrů $\beta_0, \ldots, \beta_k$, které jsou dále rekurentně aktualizovány.

2.4.2.1 Jednoduché exponenciální vyhlazování

V jednoduchém exponenciálním vyhlazování je trend považován za konstantní:

$$T_{t-j} = \beta_0, j = 0, 1, \ldots, t-1$$

Hodnoty vyhlazování v čase t jsou vypočteny ze vzorce:

$$\hat{y}_t = (1-\alpha)y_t + \alpha\hat{y}_{t-1}, t = 1, \ldots, n$$

Počáteční hodnota $\hat{y}_0$ je určena jako aritmetický průměr:

$$\hat{y}_0 = \frac{1}{n}\sum_{t=1}^{n} y_t$$

2.4.2.2 Dvojité exponenciální vyhlazování

V metodě dvojitého exponenciální vyhlazování se uvažuje lineární trend:

$$T_{t-j} = \beta_0 - \beta_1 j, j = 0, 1, \ldots, t-1$$

Při výpočtu s využitím vyhlazovacích statistik popsaných v kapitole jednoduchého exponenciálního vyhlazování jsou použity pro výpočet vyhlazovacích statistik tyto vzorce:

$$S_t = (1-\alpha)y_t + \alpha S_{t-1}$$
$$S_t^{[2]} = (1-\alpha)S_t + \alpha S_{t-1}^{[2]}$$

Počáteční hodnoty S_0 a $S_0^{[2]}$ jsou získány z odhadnutých parametrů $\beta_0(0)$ a $\beta_1(0)$:

$$S_0 = \hat{\beta}_0(0) - \frac{\alpha}{1-\alpha}\hat{\beta}_1(0)$$
$$S_0^{[2]} = \hat{\beta}_0(0) - \frac{2\alpha}{1-\alpha}\hat{\beta}_1(0)$$

kde $\beta_0(0)$ a $\beta_0^{[2]}(0)$ jsou spočítány metodou nejmenších čtverců.

Vyrovnaná hodnota v čase t je spočítána vzorcem:

$$\hat{y}_t = 2S_t - S_t^{[2]}$$

2.4.2.3 Trojité exponenciální vyhlazování

Trojité exponenciální vyhlazování pracuje s kvadratickým trendem:

$$T_{t-j} = \beta_0 - \beta_1 j + \beta_2 j^2, j = 0, 1, \ldots, t-1$$

Vyrovnávací statistiky lze zapsat rovnicemi:

$$S_t = (1-\alpha)y_t + \alpha S_{t-1}$$

$$S_t^2 = (1-\alpha)S_t + \alpha S_{t-1}^2$$
$$S_t^3 = (1-\alpha)S_t^2 + \alpha S_{t-1}^3$$

Počáteční hodnoty pro vyhlazovací statistiky jsou vypočítány rovnicemi:

$$S_0 = \hat{\beta}_0(0) - \frac{\alpha}{1-\alpha}\hat{\beta}_1(0) + \frac{\alpha(1+\alpha)}{2(1-\alpha)^2}\hat{\beta}_2(0)$$

$$S_0^{[2]} = \hat{\beta}_0(0) - \frac{2\alpha}{1-\alpha}\hat{\beta}_1(0) + \frac{2\alpha(1+2\alpha)}{2(1-\alpha)^2}\hat{\beta}_2(0)$$

$$S_0^{[3]} = \hat{\beta}_0(0) - \frac{3\alpha}{1-\alpha}\hat{\beta}_1(0) + \frac{3\alpha(1+3\alpha)}{2(1-\alpha)^2}\hat{\beta}_2(0)$$

β_0,β_1,β_2 jsou získány metodou nejmenších čtverců.
Vyhlazené hodnoty v čase t jsou vypočítány vzorcem:

$$\hat{y}_t = 3S_t - 3S_t^{[2]} + 3S_t^{[3]}$$

2.4.3 Holtova metoda s lineárním trendem

Holtova metoda exponenciálního vyhlazování trendů (zpracována 1957) pracuje se dvěma vyhlazovacími konstantami α a β. Konstanta α slouží k adaptivnímu odhadu úrovně β_0 v čase t a β pro adaptivní odhad směrnice lineárního trendu β_1 v čase t

Odhad parametrů β_0 a β_1 v čase t je:

$$\hat{\beta}_{0,t} = \alpha y_t + (1-\alpha)(\hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1})$$

$$\hat{\beta}_{1,t} = \beta(\hat{\beta}_{0,t} - \hat{\beta}_{0,t-1}) + (1-\beta)\hat{\beta}_{1,t-1}$$

kde
$\hat{\beta}_{0,t}$ je odhad úrovně lineárního trendu v čase t,
$\hat{\beta}_{1,t}$ je odhad směrnice lineárního trendu v čase t,
$\hat{\beta}_{0,t-1}$ je odhad úrovně lineárního trendu v čase $t-1$,
$\hat{\beta}_{1,t-1}$ je odhad směrnice lineárního trendu v $t-1$,
$\alpha \in < 0, 1 >$ je vyhlazovací konstanta úrovně,
$\beta \in < 0, 1 >$ je vyhlazovací konstanta směrnice. Bodová předpověď s horizontem období $h > 0$ v čase t je definována:

$$\hat{y}_t(h) = \hat{\beta}_{0,t} + h\hat{\beta}_{1,t}$$

Pro $h = 1$ v čase $t-1$ jsou bodová předpověď a chyba předpovědi:

$$\hat{y}_{t-1}(1) = \hat{\beta}_{0,t-1} + \hat{\beta}_{1,t-1}$$

$$\hat{a}_t = y_t - \hat{\beta}_{0,t-1} - \hat{\beta}_{1,t-1}$$

2.4.4 Holt Wintersova metoda se sezónními trendy

Winters, Holtův žák, rozšířil Holtovu metodu exponenciálního vyhlazování trendů o aditivní a multiplikativní sezónnost. Metoda využívá třech vyhlazovacích konstant pro hladinu, trend a sezónnost

...doplnit

2.5 Klasifikace vytížení sítě

Kapitola 3

Návrh

3.1 Struktura navrhnutého modulu

Kapitola 4

Implementace

4.1 Použité nástroje

4.2 Součásti modulu

Kapitola 5

Použití modulu

5.1 Nasazení v systému NEMEA

5.2 Spouštění modulu

Kapitola 6

Testy

Závěr

Literatura

[1] The Hierarchy of Cyber Needs [online]. Dostupné z: https://blogs.technet.microsoft.com/askpfeplat/2016/01/25/the-hierarchy-of-cyber-needs/

[2] sFlow and Netflow [online]. [cit. 2016-02-07]. Dostupné z: http://blog.sflow.com/2009/05/sflow-and-netflow.html

[3] Netflow architecture. Computer and Network Examples [online]. [cit. 2016-02-07]. Dostupné z: http://www.conceptdraw.com/How-To-Guide/netflow-architecture

[4] Nemea [online]. [cit. 2016-02-07]. Dostupné z: https://www.liberouter.org/technologies/nemea/

[5] Gartner Says 6.4 Billion Connected Things Will Be in Use in 2016, Up 30 Percent From 2015 [online]. [cit. 2016-02-07]. Dostupné z: http://www.gartner.com/newsroom/id/3165317

[6] Cybercrime [online]. [cit. 2016-02-07]. Dostupné z: http://www.interpol.int/Crime-areas/Cybercrime/Cybercrime

[7] Network Intrusion Detection Signatures [online]. [cit. 2016-02-07]. Dostupné z: http://www.symantec.com/connect/articles/network-intrusion-detection-signatures-part-one

[8] What is Zero Day Exploit? [online]. [cit. 2016-02-07]. Dostupné z: http://www.kaspersky.com/internet-security-center/definitions/zero-day-exploit

[9] Requirements for IP Flow Information Export (IPFIX) [online]. RFC. Dostupné z: https://tools.ietf.org/html/rfc3917

[10] RTP: A Transport Protocol for Real-Time Applications [online]. Dostupné z: https://tools.ietf.org/html/rfc3550

[11] NetFlow vs. IPFIX[online]. [cit. 2016-02-07]. Dostupné z: https://www.whatisipfix.com/

[12] Group, N. W.: Cisco Systems NetFlow Services Export Version 9[online]. [cit. 2016-03-05]. Dostupné z: https://www.ietf.org/rfc/rfc3954.txt

[13] Monowar H. Bhuyan, D. K. B.; Kalita, J. K.: Network Anomaly Detection: Methods, Systems and Tools. *IEEE COMMUNICATIONS SURVEYS AND TUTORIALS*, 2014.

[14] Limpouch, J.: ANALÝZA EKONOMICKÝCH ČASOVÝCH ŘAD S PŘÍKLADY [online]. [cit. 2016-02-07]. Dostupné z: http://www-troja.fjfi.cvut.cz/~limpouch/sigdat/statodn/node12.html

[15] Ing. Josef Bednář, P.: Testování statistických hypotéz [online]. [cit. 2016-02-07]. Dostupné z: http://mathonline.fme.vutbr.cz/download.aspx?id_file=479

[16] Brutlag, J. D.: Aberrant Behavior Detection in Time Series for Network Monitoring [online]. [cit. 2016-02-07]. Dostupné z: http://usenix.org/legacy/publications/library/proceedings/lisa2000/full_papers/brutlag/brutlag_html/index.html

[17] ANALÝZA EKONOMICKÝCH ČASOVÝCH ŘAD S PŘÍKLADY [online]. [cit. 2016-02-07]. Dostupné z: http://nb.vse.cz/~arltova/vyuka/crsbir02.pdf

[18] Moving Averages - Simple and Exponential [online]. [cit. 2016-02-07]. Dostupné z: http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:moving_averages

[19] Predikce vítěze sportovního utkání využitím PageRanku [online]. Bakalářská práce. Dostupné z: https://otik.uk.zcu.cz/bitstream/handle/11025/13540/BP_%20sudap_A11B0612P.pdf

[20] C. SESHADHRI, A. P.; KOLDA, T. G.: An In-Depth Analysis of Stochastic Kronecker Graphs. *Sandia National Laboratories*, 2013.

[21] Peel, L.; Clauset, A.: Detecting change points in the large-scale structure of evolving networks. *Department of Computer Science, University of Colorado, Boulder, CO 80309*, Nov 2014.

[22] Detecting Change in Longitudinal Social Networks [online]. [cit. 2016-02-07]. Dostupné z: https://www.cmu.edu/joss/content/articles/volume12/McCullohCarley.pdf

Příloha A

Seznam použitých zkratek

GHRG Generalized hierarchical random graph

Nemea Network measurements analysis

IDS Intrusion detection system

IPFIX IP Flow Information Export

TRAP Traffic analysis platform

UniRec Unified record

SNMP Simple network management protocol

ICMP Internet control message protocol

Příloha B

Instalační příručka

Příloha C

Uživatelská příručka

Příloha D

Obsah přiloženého CD

- `readme.txt` ..... stručný popis obsahu CD
- `exe` ..... adresář se spustitelnou formou implementace
- `src`
  - `impl` ..... zdrojové kódy implementace
  - `thesis` ..... zdrojová forma práce ve formátu LaTeX
- `text` ..... text práce
  - `thesis.pdf` ..... text práce ve formátu PDF
  - `thesis.ps` ..... text práce ve formátu PS