VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

GRAFICKÁ VIZUALIZACE GEOGRAFICKÝCH DAT SÍŤOVÉHO PROVOZU

GRAPHICAL VISUALIZATION OF NETWORK TRAFFIC GEOGRAPHICAL DATA

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — **JAKUB KACHLÍK**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. JIŘÍ HYNEK, Ph.D.**
SUPERVISOR

BRNO 2020

Vysoké učení technické v Brně
Fakulta informačních technologií

Ústav informačních systémů (UIFS) Akademický rok 2019/2020

Zadání bakalářské práce

Student: **Kachlík Jakub**
Program: Informační technologie
Název: **Grafická vizualizace geografických dat síťového provozu**
Graphical Visualization of Network Traffic Geographical Data
Kategorie: Uživatelská rozhraní

Zadání:

1. Prostudujte nástroje firmy Flowmon pro monitorování síťových toků a hloubkové analýzy paketů v oblasti průmyslových komunikačních sítí. Analyzujte požadavky firmy Flowmon.
2. Prostudujte existující typy vizualizací vhodné pro návrh přehledových obrazovek typu *dashboard.* Prostudujte technologie pro tvorbu webových uživatelských rozhraní a vizualizací dat (React, D3.js, apod.)
3. Navrhněte rozšíření stávajícího systému o pohledy pro vizualizaci dat se zaměřením na geovizalizace.
4. Navržené rozšíření implementujte.
5. Otestujte řešení ve spolupráci s firmou Flowmon.

Literatura:

- Few, S.: *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly, 2006, ISBN: 978-059-6100-162.
- Harris, R. L.: *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, 2000. ISBN: 978-0-1951-3532-9.
- Interní dokumentace firmy Flowmon.

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3

Podrobné závazné pokyny pro vypracování práce viz https://www.fit.vut.cz/study/theses/

Vedoucí práce: **Hynek Jiří, Ing., Ph.D.**
Vedoucí ústavu: Kolář Dušan, doc. Dr. Ing.
Datum zadání: 1. listopadu 2019
Datum odevzdání: 28. května 2020
Datum schválení: 16. října 2019

Abstrakt

Při probíhajících internetových útocích je důležité zjistit co nejvíce informací o útočníkovi a srozumitelně tuto informaci předat síťovému administrátorovi. Dohledový systém Flowmon je aktuálně schopen určit původ zasílaných paketů, shlukovat je do datových toků a zapisovat je do tabulky. Pro tvorbu geografické analýzy útoku jsou však takto vizualizovaná data nepřehledná. Cílem této práce je vytvoření webového informačního dashboardu, který bude zobrazovat geografické vizualizace síťového provozu. Pro uživatele nástrojů tím bude zajištěna detailnější a snáze pochopitelná analýza.

Abstract

During ongoing Internet attacks is important to find out as much information about the attacker as possible and to pass this information clearly to the network administrator. The Flowmon monitoring system is currently able to determine the source destination of sent packets, group them into flows and write them into a table. The data visualized in this way are confusing to create a geographical analysis of the attack. The objective of this work is to create a web information dashboard that will display geographic visualizations of network traffic. It will provide more detailed and easier-to-understand analysis to users of the tools.

Klíčová slova

grafy, geovizualizace, dashboardy, vizuální prezentace, síťový provoz, monitorování sítě.

Keywords

graphs, geovisualization, dashboards, visual presentation, network traffic, network monitoring.

Citace

KACHLÍK, Jakub. *Grafická vizualizace geografických dat síťového provozu.* Brno, 2020. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Jiří Hynek, Ph.D.

Grafická vizualizace geografických dat síťového provozu

Prohlášení

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

.....
Jakub Kachlík
4. června 2020

Poděkování

Rád bych poděkoval vedoucímu mé bakalářské práce Ing. Jřímu Hynkovi, Ph.D. za velmi ochotný přístup, za poskytnutí velkého množství studijních materiálů a především za čas, který mi věnoval při konzultacích. Také bych rád poděkoval členům pracovní skupiny TRACTOR, kteří se průběžně podíleli na testování implementovaného nástroje, a mé přítelkyni, která mi pomáhala při formální kontrole práce.

Obsah

Kapitola 1

Úvod

Obsah komunikace síťového provozu čítá obrovské množství dat, které není správce sítě schopen kontrolovat. Z tohoto důvodu jsou monitorovaná data zpracována pomocí specializovaných nástrojů, které uživatele přehledně informují o situaci ve vybraných oblastech sítě. Jednou z podstatných částí každé sítě je její zabezpečení, o jehož stavu a zjištěných příchozích podezřelých aktivitách musí být správce sítě bezprostředně informován. Protože se však jedná o velké množství informací, je vhodné uživatele informovat takovou formou, díky které co nejrychleji zanalyzuje situaci a pochopí její vývoj.

Za tímto účelem jsou využívány vizualizace, které dokáží ve srovnání s tabulkou jednotlivých záznamů pochopitelnějším způsobem prezentovat vstupní data. Jedním z častých požadavků v oblasti monitorování bezpečnosti je vyhodnocování zdrojů a cílů útoků/anomálií. Pro tento účel není dostačující tabulka, ve které uživatel jen velmi obtížně objeví vztahy mezi záznamy, a tak je nutné použít vhodný diagram, který dokáže srozumitelně vyznačit mimo jiné také geografickou polohu komunikujících zařízení. K těmto účelům slouží různé geovizualizační nástroje.

Požadavek na tento typ vizualizací má také společnosti Flowmon, která je chce využívat ve svých analytických nástrojích. V současné době jsou sice schopni geografické informace síťového provozu získávat, ale nedisponují vhodným vizualizačním nástrojem, který by s nimi byl schopen dále pracovat. Proto se rozhodli požádat o vytvoření tohoto nástroje pracovní skupinu TRACTOR, která vznikla ve spolupráci VUT a MUNI.

Hlavním cílem této práce je navrhnout a vytvořit dashbord[1], který bude zobrazovat geografické vizualizace síťového provozu na základě vstupních dat z interní REST API, ze které budou data získávána pomocí HTTP příkazů. Jelikož jsou současné webové nástroje společnosti Flowmon tvořeny jako samostatné komponenty v knihovně React, budu tímto způsobem implementovat také geovizualizační dashboard.

V kapitole 2 se budu věnovat způsobům monitorování síťového provozu, popisu využití aktuálních nástrojů pro monitorování síťového provozu od společnosti Flowmon a několika dalších konkurenčních společností. V kapitole 3 se zaměřím na práci se vstupními daty, způsoby vizualizace vícedimenzionálních dat, pravidla lidského vnímání obrazců či tvarů. Dále se pak zaměřím na geovizualizace a příklady několika využívaných vizualizací a nástrojů sloužících pro jejich implementaci. Kapitola 4 se zabývá informačními dashboardy, využitím a pravidly pro jejich tvorbu. V kapitole 5 analyzuji práci monitorovacích nástrojů společností Flowmon, SolarWind, NetFort a DataDome s geografickými daty a způsob vizualizace

[1] Dashboard je slovo anglického původu, které není ve slovníku českého jazyka. Pro účely této práce budu dále toto slovo skloňovat podle mužského vzoru *hrad*.

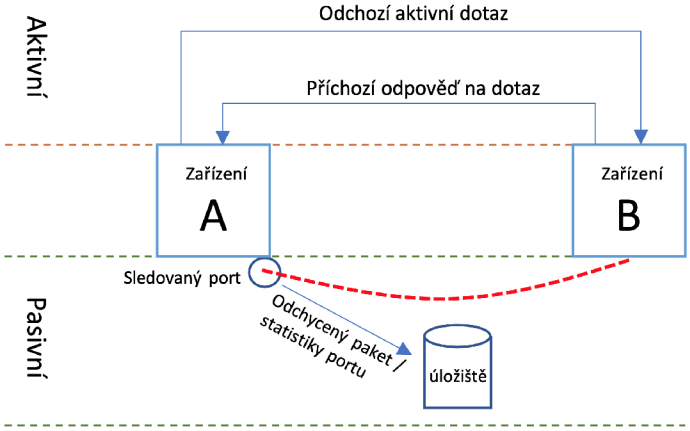
těchto dat se zaměřením na efektivitu vizualizací. Návrh architektury, uživatelského rozhraní a použitých geovizualizací v dashboardu je zaznamenán v kapitole 6. Kapitola 7 popisuje konkrétní postupy a použité nástroje při implementaci dashboardu a možnosti jejich využití. V kapitole 8 popisuji způsob a četnost testování a úprav geovizualizačních vrstev dashboardu v průběhu implementace.

Kapitola 2

Monitorování síťového provozu

Hlavním úkolem monitorování síťového provozu je udržení přehledu o probíhajících událostech a aktivitách v síti. Dle [14] je monitorování síťového provozu definováno jako proces zjišťování a kontrolování stavu či dostupnosti síťových služeb a zařízení. Monitorování sítě neslouží pouze ke sledování stavu sítě v reálném čase, ale také k tvorbě statistik, které pomáhají síťovým analytikům s odhalením bezpečnostních rizik, optimalizačních nedostatků nebo podezřelých aktivit.

Monitorování se dělí do dvou skupin na základě typu komunikace potřebné pro získávání požadovaných informací o síti. Rozdíl mezi aktivním a pasivním monitorováním je naznačen v obrázku 2.1.



Obrázek 2.1: Příklad aktivního a pasivního monitorování síťového provozu

2.1 Aktivní monitorování

Gold [8] popisuje aktivní monitorování, neboli *synthetic monitoring*, jako přístup získávání informací o stavu sítě z odpovědí na zaslané testovací pakety. Pro aktivní monitorování je vyžadováno další aktivní zařízení v síti, které bude zasílat odpovědi na dotazy. Vzhledem k nezbytné vzájemné komunikaci koncových zařízení je vhodné využít protokoly z rodiny TCP/IP. Nevýhodou je zvýšená zátěž sítě způsobená aktivní komunikací mezi koncovými zařízeními. Tento typ monitorování je vhodný pro kontrolu stavu sítě v reálném čase.

Jedním z protokolů využívajících aktivní přístup monitorování sítě je například ICMP[1], pomocí kterého může správce testovat vlastnosti spojené se spolehlivostí a rychlostí sítě jako latenci, kolísání nebo ztrátovost paketů.

2.2 Pasivní monitorování

Pasivní monitorování je založeno na sběru a analýze dat reálného provozu ve specifickém místě v síti. Na rozdíl od aktivního monitorování není vyžadováno další zařízení v síti, kterému by byly zasílány testovací dotazy. Sven Ubik [19] pasivní monitorování označuje jako zpravidla trvalého pozorovatele v síti, který na rozdíl od aktivního monitorování neovlivňuje zatížení sítě, protože pouze odposlouchává síťový provoz.

Jedním z často používaných protokolů je syslog[2], který informace ukládá ve formě logů. Pokud je však vyžadována práce se statistikami *datových toků*[3], je vhodné využít například protokol NetFlow nebo IPFIX. Odchycené pakety jsou zasílány na úložiště, typicky na externí sondu. Zde jsou následně shlukovány do datových toků, které se následně mohou využívat k tvorbě statistik. Pasivní monitorování je vhodné například pro zjišťování QoE[4] nebo řešení problémů se síťovými protokoly.

2.3 Existující řešení společnosti Flowmon

V bakalářské práci se zaměřuji na analýzu existujících nástrojů společnosti Flowmon, která pomáhá firmám spravovat a zabezpečovat jejich síťovou infrastrukturu prostřednictvím monitorování a analýzy chování počítačových sítí. Společnost Flowmon v současné době poskytuje 5 webových pluginů, ve kterých nabízejí služby kombinující přístupy aktivního a pasivního monitorování [7]. Za účelem analýzy jejich nástrojů mi byl poskytnut demo účet, díky kterému jsem měl možnost si všechny nástroje vyzkoušet.

Flowmon Dashboard

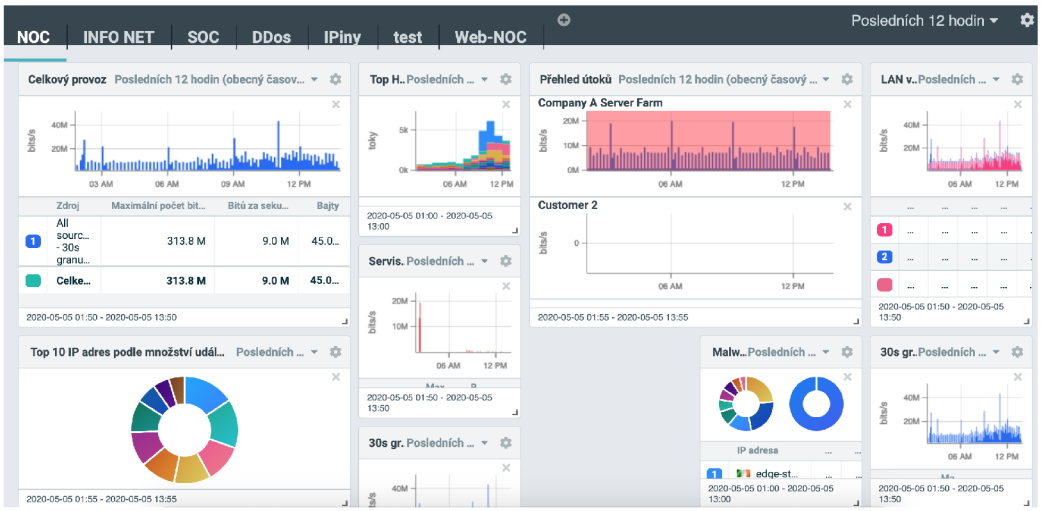
Jedná se o interaktivní plugin pro tvorbu dashboardů. Uživatel má k dispozici pracovní plochu, na které si může rozmisťovat widgety, které jsou k dispozici v seznamech dostupných nástrojů. Widgety jsou zjednodušenou alternativu plnohodnotných nástrojů, které mohou být v dashboardu libovolně přemisťovány a lze upravovat jejich velikost. Uživatelům je umožněno vytváření více dashboardů do oddělených profilů, nastavování časového rozsahu vizualizovaných dat a několik možností pro grafickou úpravu každého widgetu. Na obrázku 2.2 je vidět, že v tomto pluginu jsou využívány jednoduché vizualizace v podobě sloupcových nebo výsečových grafů, které jsou zpravidla doplněny tabulkou se vstupními daty.

[1] ICMP, neboli Internet Control Message Protocol, je síťový protokol rozšiřující IP, který slouží pro zasílání řídících informací o síti[16].

[2] Syslog je zkratka pro System Logging Protocol, což je standartní protokol pro záznam programových zpráv a sběr dat[13].

[3] Datový tok, neboli *flow*, je označován jako jednosměrná sekvence paketů mezi zdrojovým a cílovým koncovým bodem a je identifikován jako kombinace 5 shodných klíčových polí: zdrojová adresa, zdrojový port, cílová adresa, cílový port a číslo protokolu (v novějších verzích může být použit ještě interface routeru/switche a ohodnocení Class of Service, které slouží k přiřazení priority zaslanému paketu).

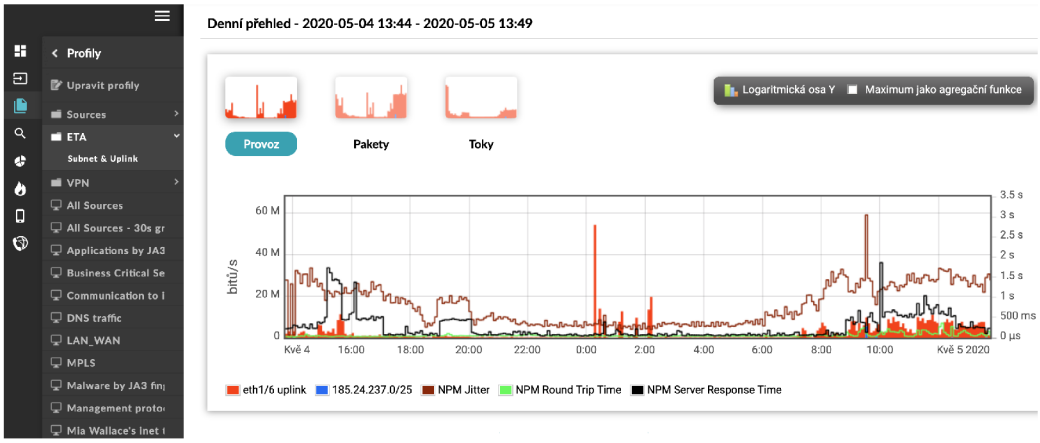
[4] QoE, neboli *Quality of Experience*, značí míru spokojenosti uživatele s vybranou službou [3].



Obrázek 2.2: Ukázka vytvořeného dashboardu, používaných grafů, rozdělení do profilů a výběr zobrazovaného časového úseku

Flowmon Monitoring Center

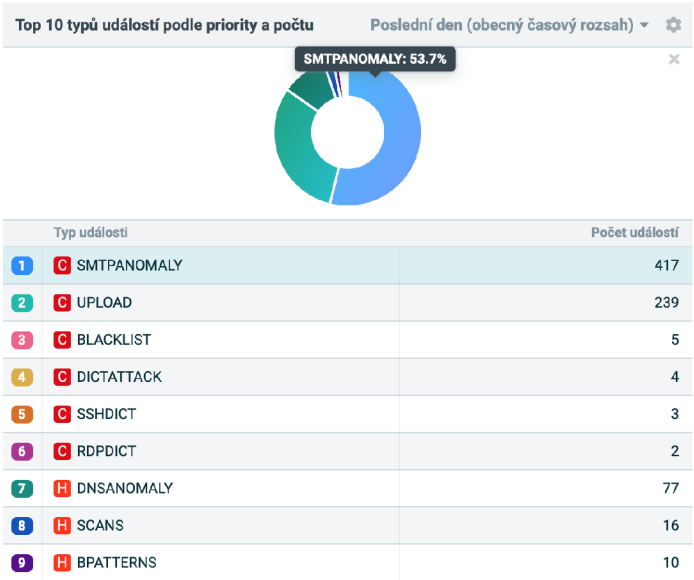
Plugin slouží k analýze provozu v monitorované síti. Přenášené pakety jsou shlukovány do datových toků, ze kterých se pomocí protokolů IPFIX/NetFlow vytváří statistiky, které jsou vizualizovány ve vybraných grafech a tabulkách. Tento plugin disponuje nástrojem pro sdružování jednotlivých částí sítě a uživatel tak může sledovat jak provoz v celé síti, tak i v jejích dílčích částech až na úrovni jednotlivých protokolů, jak je vidět na obrázku 2.3. Funkce *Aktivní zařízení* poskytuje uživateli informace o připojených zařízeních ve formě seznamu nebo grafů. Kromě IP a MAC adresy zpravidla zjistí i informace o výrobci a operačním systému. Pokud společnost využívá technologii internetového volání pomocí VoIP, jsou pomocí služby *VoIP provoz* evidovány informace o ukončených hovorech, použitých protokolech, přenesených paketech a dalších. Pro automatizovanou kontrolu důležitých částí sítě je možné využít nástroj *Alert*, který uživatele upozorní při splnění/nesplnění zadané podmínky.



Obrázek 2.3: Rozdělení sítě do profilů a vizualizace profilu ETA

Flowmon ADS (Anomaly Detection System)

Plugin chrání sítě před kybernetickými hrozbami, které obchází tradiční bezpečností prvky typu firewall. Plugin využívá technologii NBAD[5], která za využití umělé inteligence vyhodnocuje chování připojených zařízení v síti a pomáhá v odhalení neobvyklé, podezřelé a nežádoucí aktivity. Zjištěné anomálie jsou vizualizovány grafem doplněným tabulkou se vstupními daty pro vybranou časovou oblast (viz. obr. 2.4). V sekci *Nastavení* může uživatel aktivovat, deaktivovat a měnit desítky implementovaných detekčních metod, nastavovat zdroje dat, vytvářet a upravovat blacklist seznamy nebezpečných adres nebo tvořit vlastní skripty pro zasílání upozornění na nové anomálie.



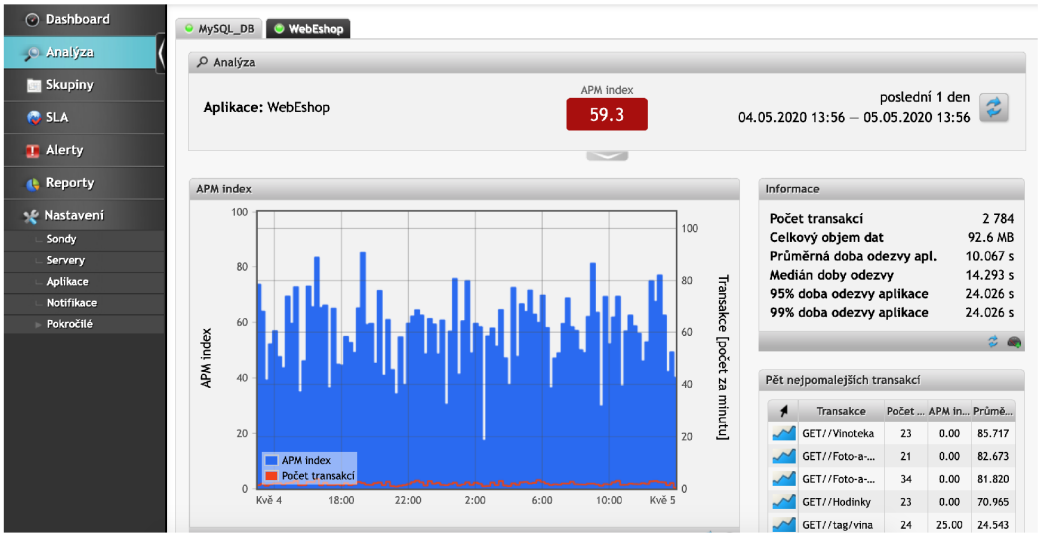
Obrázek 2.4: Vizualizace zaznamenaných anomálií dle typu a počtu výskytů

Flowmon APM (Application Performance Monitoring)

Plugin monitoruje odezvu klientských systémů a odhaluje příčiny chyb aplikací. Při nastavování dojde k logickému rozdělení sítě na vrstvu síťovou, aplikační a databázovou, kdy každá z nich je nepřetržitě monitorována a jsou u ní evidovány informace o počtu transakcí, době odezvy, počtu paralelně připojených uživatelů a chybových hláškách. Na základě těchto informací je zjišťováno plnění podmínek *SLA*[6]. Podle komplexního výsledku SLA dochází k výpočtu APM indexu, který značí spolehlivost, dostupnost a rychlost vybrané části klientského systému (viz. obr. 2.5). Kromě vypočteného indexu APM se v tomto pluginu nachází také vizualizace monitorovaných chyb, díky čemuž uživatel rychleji odhalí problematické sekce i konkrétní chyby (například chybové kódy).

[5]Detekce anomálií chování sítě, neboli *Network Behavior Anomaly Detection*, je způsob nepřetržitého monitorování, který se zaměřuje na neobvyklé nebo podezřelé aktivity v síti.

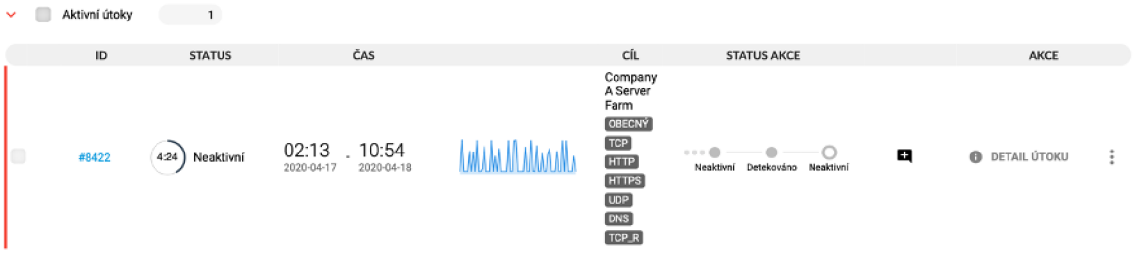
[6]*SLA*, neboli *Service Level Agreement*, je označení pro vzájemně vyjednané smluvní podmínky mezi poskytovatelem a firmou. Obsahem SLA je ujasnění o stupni, rozsahu a kvalitě poskytovaných služeb [9].



Obrázek 2.5: Zobrazení vypočteného APM indexu a časového grafu značícího průběžný vývoj indexu v čase

Flowmon DDoS Defender

Plugin detekuje a s maximální rychlostí zastavuje příchozí volumetrické útoky na základě analýzy statistik síťového provozu. Využíván je interní nástroj využívající umělou inteligenci, který detekuje a v krátkém časovém horizontu zastaví příchozí DDos útok. O těchto událostech je zvolenou formou informován administrátor (syslog, SMS, SNMP . . .). Pro každou událost jsou uživateli k dispozici kromě rychlého přehledu útoku (viz. obr. 2.6) také detailní statistiky proběhlých útoků s informacemi o použitém protokolu, vývoji incidentu v čase, útočících adresách, síle útoku (paketů/sekundu) a dalších. Při chybném označení korektního provozu za probíhající útok je možné incident změnit na falešně pozitivní, čímž dojde k okamžitému uvolnění provozu.

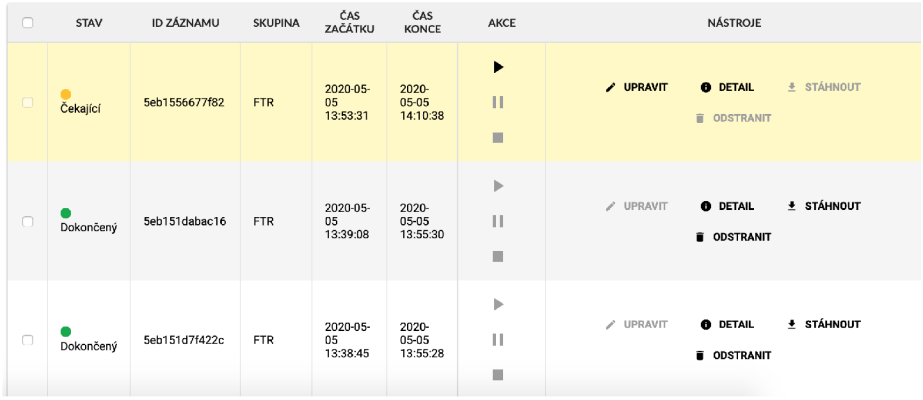


Obrázek 2.6: Přehled zachyceného DDoS útoku s několika základními informacemi ve Flomon DDoS

Flowmon Traffic Recorder

Plugin zaznamenává datový provoz v plném rozsahu na síťovou sondu. Využívá se v případě, kdy uživateli k dostatečné analýze provozu nestačí informace z protokolů IPFIX/NetFlow. Tento způsob monitoringu nabízí na rozdíl od statistik nejen informace o komunikujících zařízeních, ale také detailní obsah komunikace, který může být podstatný pro odhalení

nekorektně reagujících částí sítě. Informace o síťovém provozu jsou zaznamenávány v poměrně tradičním formátu .pcap, který podporuje řada aplikací pro analýzu síťového provozu. Plugin nabízí rychlý přehled aktivních a dokončených záznamů (viz. obr. **??**, kde má uživatel možnost nahlédnout na detaily záznamu nebo si je přímo stáhnout. V sekci *Nastavení* může uživatel měnit pravidla pro zachytávání vybraných paketů dle zvolené IP/MAC adresy nebo síťového portu. Součástí je i správa obsazené paměti, kdy má uživatel k dispozici možnost nastavit pravidelné mazání záznamů po určitém počtu dní, případně při zaplnění procentuální části dostupné paměti.



Obrázek 2.7: Seznam záznamů odposlouchávaného provozu pomocí Flowmon Traffic Recorder

2.4 Konkurenční nástroje

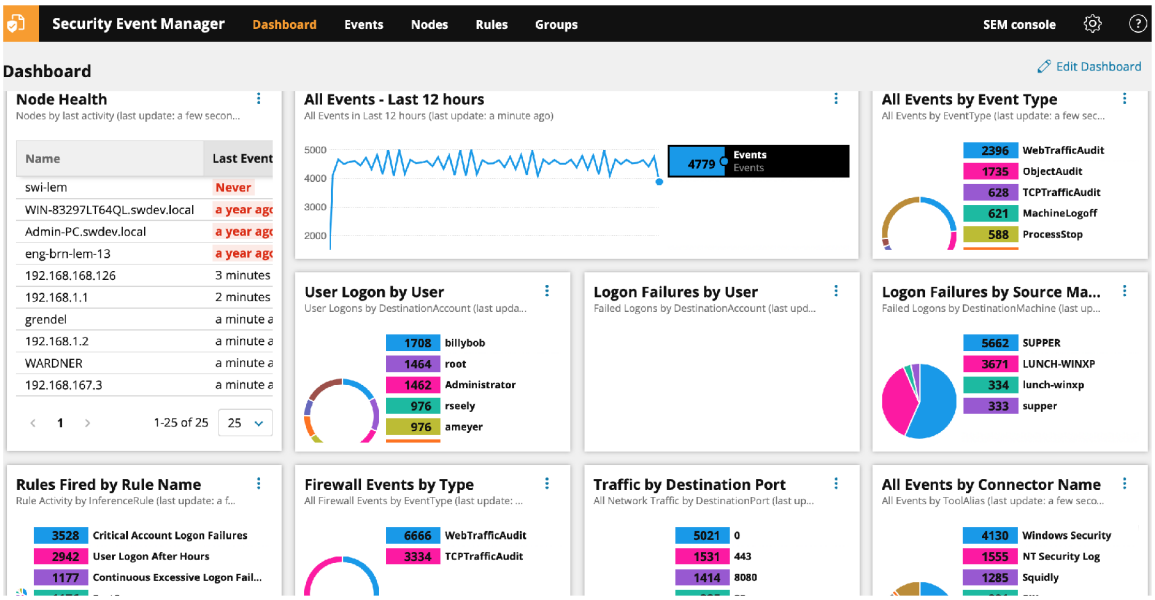
Existuje velké množství nástrojů pro monitorování síťového provozu. Každý z nich nabízí odlišné vizualizace kvůli jiné práci se získanými statistikami. Pro ukázku jsem se rozhodl popsat trojici nástrojů, které přistupují k monitorování sítě podobným způsobem jako společnost Flowmon, ale využívají odlišné vizualizační metody.

2.4.1 SolarWind

Jeden z nástrojů společnosti SolarWind, *Security Event Manager*, slouží k zachytávání a analýze událostí v síti. Jednou z činností tohoto nástroje je průběžná analýza a odhalování potenciálně nebezpečné komunikace. Nebezpečím D-DoS útoku je těžká odhalitelnost, protože nelze jednoduchým způsobem přímo určit, které adresy jsou součástí útočícího botnetu a které korektně komunikují se serverem větším množstvím dotazů. Tento nástroj tedy v případě podezření na útočící adresu zašle kontrolní upozornění, na které bot na rozdíl od skutečného uživatele není schopen korektně zareagovat a následně tak dojde k zablokování komunikace z této IP adresy. Všechny informace o událostech v síti jsou zaznamenávány na společné uložiště *Events*, ve kterém může uživatel pomocí podrobného filtrování analyzovat anomálie a útoky v síti.

Tento nástroj disponuje obazovkou *Dashboard* 2.8, na které si může uživatel rozmisťovat vybrané widgety. Uživatel si tak může zvolit pro něj podstatné typy událostí a vhodným způsobem být informován o změnách a nových záznamech. Ve widgetech jsou využívány jednoduché interaktivní výsečové a sloupcové grafy, tabulky nebo histogramy zobrazující vývoj sledovaného fenoménu v čase. Všechna data zobrazovaná v grafech je možné využít jako rychlý filtr, kdy stačí pouze kliknout na vybranou položku a dojde k otevření tabulky

obsahující prvky zvolené kategorie. Přestože je nástroj schopný zachytit útok a v krátkém časovém intervalu podniknout kroky pro jeho zastavení, neexistuje v současné době implementace samostatné sekce ani podrobnější vizualizace pro analýzu anomálie či útoku.

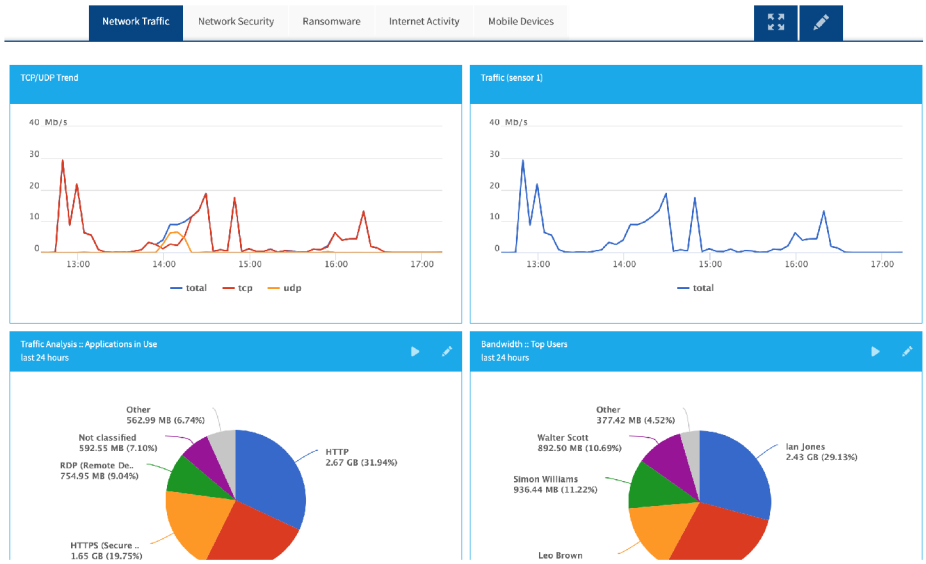


Obrázek 2.8: Ukázka nástroje typu dashboard, uživatelského rozhraní a použitých vizualizací od společnosti SolarWind

2.4.2 Netfort

Jedno z dalších konkurenčních řešení v oblasti monitoringu síťového provozu nabízí společnost Netfort. Za využití protokolu NetFlow webové nástroje odposlouchávají provoz v síti a shromažďují získané informace na uložiště. Uživatel si může vytvořit několik stránek s vizualizacemi typu dashboard, které ho budou informovat o získaných statistikách. Tento nástroj sice neobsahuje rozšíření umělé inteligence pro automatickou detekci útoků nebo anomálií, podle přehledné real-time vizualizace síťového provozu však operátor volumetrický útok snadno odhalí a analytikům jsou k dispozici interaktivní grafy, pomocí kterých jsou schopni prozkoumat podezřelé oblasti až na úroveň IP adres.

Pro vizualizace v dashboardech jsou využity standartní nástroje v podobě výsečového grafu, tabulek a histogramu (viz. obr. 2.9). K vizualizaci konkrétní oblasti sítě si může uživatel zvolit mezi výsečovým a sloupcovým grafem, které však obsahují drobné nedostatky. Není možné upravovat počet kategorií v grafu, a i když je hodnota jedné z nich tak malá, že není ani vizualizována, graf a jeho legenda s ní stále pracují. Nástroje jsou interaktivní a data propojená, díky čemuž je schopen uživatel projít kategoriemi až na seznam jediné IP adresy a záznamy její komunikace.

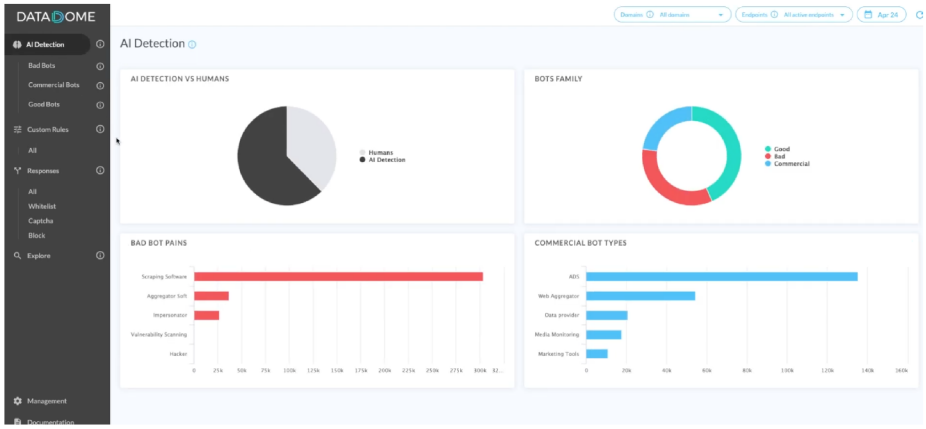


Obrázek 2.9: Ukázka nástroje monitorující síťový provoz v reálném čase, uživatelského rozhraní a použitých vizualizací společnosti NetFort

2.4.3 DataDome

Francouzská společnost Datadome poskytuje platformu zaměřenou ryze na aktivity botů a jejich chování. Jejich jediný nástroj se zaměřuje na detekci botů v síti a důkladnou analýzu jejich aktivit. Boti jsou děleni na 3 kategorie: dobří, špatní a komerčně zaměření. Kategorizace botů je zaznamenávána ve sdílené databázi, díky které jsou uživatelé upozorňování na potenciálně nebezpečné boty ještě před tím, než začnou komunikovat s jejich sítí. Uživateli je umožněna tvorba blacklistů i whitelistů, a pokud dojde k chybnému vyhodnocení aktivit bota, také k označování záznamů za falešně pozitivní.

Informace o síťovém provozu jsou vizualizovány do interaktivního histogramu. Pro zvolenou časovou oblast je pod grafem zobrazena tabulka s informacemi o zaznamenaných botech, jejich aktivitách a detailech o jejich původu. Pro přehled o typech botů připojených do sítě je využíván výsečový graf a pro znázornění jejich aktivity slouží sloupcový graf (viz. obr. 2.10).



Obrázek 2.10: Ukázka uživatelského rozhraní, použitých vizualizací a kategorizace zachycených botů společností DataDome

Kapitola 3

Vizualizace dat

Podle [15] jsou data jakákoli reprezentace skutečnosti schopná přenosu, interpretace či zpracování. Účelem dat je přenášet a dále zpracovávat odraz skutečnosti a jsou to jakékoli zaznamenané poznatky či fakta.

Vizualizace dat je proces generování grafiky na základě dat za účelem snazšího pochopení pro člověka a jeho kognitivní systém [4]. Jejím účelem je pochopení zkoumaných jevů a proniknutí do problému, podobně jako u numerické analýzy, proto bývá někdy označována jako vizuální analýza. Jedná se o velmi názornou formu sdělení informace, což je jeden z důvodů, proč je využívána při tvorbě dashbordů.

3.1 Práce s daty

Před prací s naměřenými daty je nutné provést průzkumovou analýzu, pomocí které zjistíme povahu souboru dat a jeho vlastnosti. Na základě těchto informací můžeme zvolit odpovídající nástroje a postupy pro jejich zpracování. Měřená data dělíme do dvou kategorií: kvantitativní a kvalitativní.

Numerická data, označována také jako kvantitativní, jsou taková data, jejichž hodnota je vyjádřena pomocí číselných znaků, na základě kterých je určena její velikost. Tento typ dat může být spojitý nebo diskrétní. Spojitá data mohou nabývat všech reálných hodnot v rámci stanoveného intervalu, zatímco diskrétní mohou nabývat pouze číselných hodnot izolovaných v intervalu (např. pouze celá čísla). Při monitorování síťového provozu mohou spojitá data vyjadřovat například síťovou latenci a diskrétní data počet připojených uživatelů k určité službě.

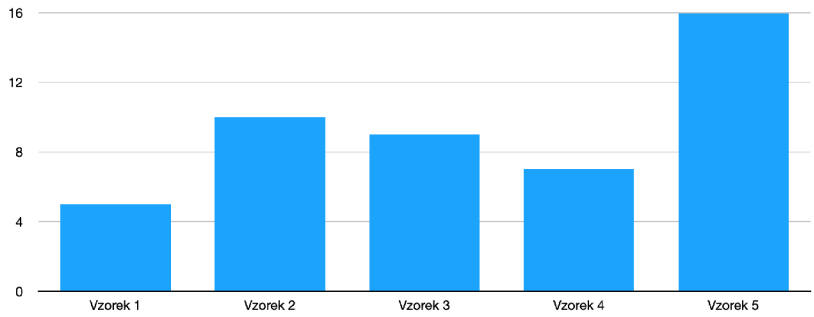
Kvalitativní data nabývají konečného množství diskrétních hodnot a popisují vlastnost jevů. Dělí se na nominální a ordinální, přičemž vztahy mezi ordinálními hodnotami je možné uspořádat, zatímco nominální data řadit nelze. Při monitorování sítě mohou být nominální hodnotou názvy jednotlivých zařízení a ordinální hodnotou může být počet zaznamenaných síťových útoků.

Při správě síťového provozu se zpravidla pracuje s numerickými hodnotami, se kterými je možné provádět řadu matematických operací vhodných pro další analýzu. Data je možné řadit podle velikosti, vypočítat z nich průměrnou hodnotu či rozdělit seřazený soubor dat pomocí mediánu na dvě stejně početné části.

3.2 Dimenze dat

Dimenze dat je množina hodnot určitého typu popisující kvantitativní nebo kvalitativní data. Při práci se souborem dat však může být zaznamenáno více dimenzí a v takovém případě se jedná o multidimenzionální data. Využitím jednotlivých dimenzí můžeme data řadit, kategorizovat, filtrovat, agregovat a další. V oblasti síťového provozu můžeme například pracovat se souborem dat obsahující informace o zdrojové IP adrese, použitém protokolu, cílové IP adrese a přeneseném obsahu dat. Tyto informace můžeme řadit podle kvantitativní hodnoty v podobě objemu dat, filtrovat nebo třídit podle kvalitativních hodnot v podobě IP adres nebo protokolu.

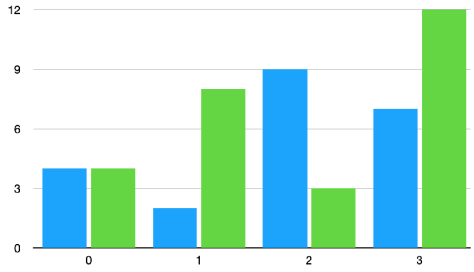
Vizualizací jedné dimenze dat se označuje situace, kdy analyzujeme a zobrazujeme pouze jeden atribut dat. Jedná se o nejjednodušší způsob vizualizace numerických dat a pro tyto účely je vhodné využívat například sloupcové a výsečové grafy, histogramy a další. Při analýze vizualizace se uživatel zaměřuje na jeden zobrazovaný atribut všech vzorků, ze kterého může určit například minimální a maximální hodnotu, překročení stanoveného limitu, průměrnou či nejčastější hodnotu a další (viz. obr. 3.1).



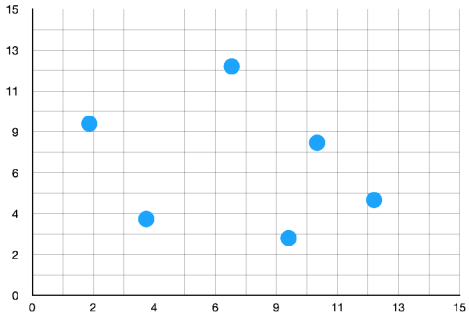
Obrázek 3.1: Ukázka sloupcového grafu vizualizujícího jednodimenzionální data

Multidimenzionální vizualizace bývá chápána jako zobrazování dat v prostoru, což ale nesouvisí přímo s dimenzionalitou dat. Za vizualizaci multivariantních (nebo vícedimenzionálních) dat se označuje taková vizualizace, jejíž výsledek by měl zajistit zobrazení 2 a více atributů datových vzorků. Sarkar [17] uvádí, že díky multivariantní analýze můžeme sledovat nejen distribuci dat, ale zároveň můžeme registrovat vzájemné vztahy, vzorce a korelace (závislosti) mezi jednotlivými atributy.

Existuje množství způsobů a technik, jak sledované atributy zobrazit. Jedna z nejjednodušších možností je skládáním více jednodimenzionálních grafů. Jako příklad lze využít dva sloupcové grafy na obrázku 3.2, kdy každý záznam grafu bude vizualizovat dva atributy. Nejčastěji se však využívá agregační funkce, která vybrané atributy spojí do jednoho a ten následně vizualizuje. Často využívanou technikou je použití dvojice atributů jako koordinát určující umístění vzorku v dvojrozměrném poli (viz. obr. 3.3). Tímto způsobem se často značí vývoj události v čase a sledování vývoje trendů.



Obrázek 3.2: Spojení dvou jednodimenzionálních grafů vytvářející vizualizaci dvoudimenzionálních dat



Obrázek 3.3: Agregace dvojice hodnot do koordinát ve dvoudimenzionální diskrétní vizualizaci

3.3 Geovizualizace

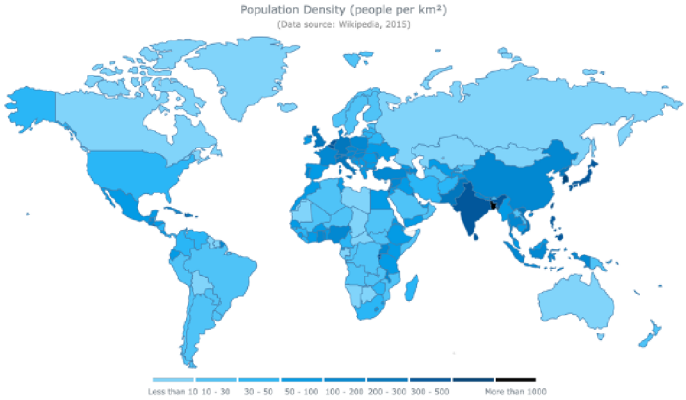
Při monitorování síťového provozu je možné pomocí IP adresy uživatelů odhalit jejich geografickou lokaci minimálně na úrovni státu. Tyto informace je následně vhodné vizualizovat názorným způsobem, díky kterému uživatel ihned pochopí aktuální rozložení hodnot v kontextu celku, kterým může být celá planeta, světadíl a další územní celky. Je tedy vhodné využít náležitý typ vizualizace, který tento náhled a rychlou analýzu situace umožní.

Pojem geovizualizace je zkratka pro geografickou vizualizaci [18]. Jedná se o druh vizualizace dat, jejichž minimálně jeden atribut vyjadřuje geografickou lokalizaci na podkladové mapě. Jejím úkolem je názorně zobrazit vstupní data na konkrétní geografickou oblast na mapě. V následujících podkapitolách zmíním několik geovizualizací, které se liší v postupech a používání geografických dat.

3.3.1 Kartogram

Kartogramem, anglicky *choropleth map*, se označuje typ tematické geografické vizualizace, která jednotlivé vymodelované oblasti stínuje v poměru ke statické proměnné, která představuje souhrn pro zeměpisnou charakteristiku v dané oblasti [2]. Robert L. Harris [10] uvádí, že kartogram bývá označován jako stínovaná, zkřížená nebo texturovaná mapa a předdefinované oblasti, také označované jako jednotky, mohou zastupovat země, státy, regiony, okresy a další samostatné geografické oblasti. Tyto územní celky jsou podle hodnoty zařazeny do třídy intervalů, kdy každá třída je na mapě vizualizována odlišitelnou barvou, odstínem, intenzitou nebo vzorem jako na obrázku 3.4.

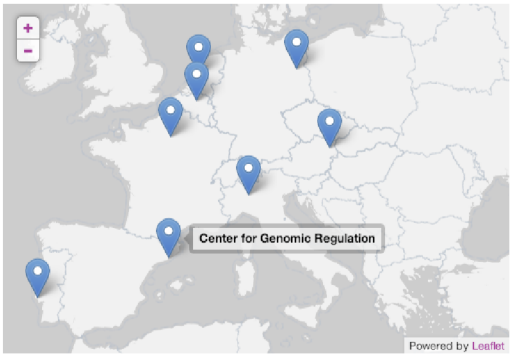
[1] https://www.anychart.com/products/anymap/gallery/Maps_General_Features/World_Choropleth_Map.php



Obrázek 3.4: Kartogram využivající jako třídu intervalů škálu barevné intenzity modré barvy[1]

3.3.2 Body na mapě

Pro vyznačení významných míst na mapě se využívají značky neboli *marker*. Tyto značky jsou odkazem na konkrétní geografickou souřadnici a jejich úkolem je viditelným způsobem vyznačit místo na mapě [10]. Značky bývají zobrazovány nejčastěji ve formě kruhových bodů či speciálních symbolů (viz. obr. 3.5). V některých grafických vizualizacích však mohou být tyto značky využívány k zobrazení statistického souhrnu informací pro určitou oblast a jsou pak vizualizovány v podobě schémat, grafů nebo tabulek.



Obrázek 3.5: Vyznačení několika bodů na mapě Evropy pomocí standartní značky[2]

Pokud dojde k situaci, kdy je nutné na malém prostoru vygenerovat větší množství značek a hrozilo by jejich vzájemné překrývání, lze využít tzv. *shlukování*, tedy vznik speciální značky, která reprezentuje shluk dvou a více standartních značek. V tomto případě musí dojít nejen k vygenerování značky shluku, ale také k přejmutí vlastností a hodnot všech zastoupených značek. Tato funkce se zpravidla využívá u interaktivních map, které umožňují volitelný zoom.

3.3.3 Mapa spojení

Mapa spojení, neboli *connection map*, je typ geografické vizualizace, který zobrazuje spojení mezi několika pozicemi na mapě [11]. Nejjednodušší způsob vizualizace čáry mezi dvěma

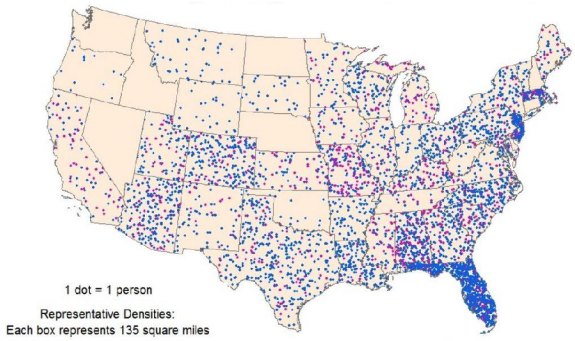
[2] https://www.drupal.org/project/leaflet_label

body je algoritmus nejkratší cesty, který ale na mapě vytvoří běžnou čáru, která nepočítá se zakřivením země, tudíž se nebude jednat o skutečně nejkratší cestu. Proto se často využívá vykreslení pomocí části hlavní kružnice (viz. obr. 3.6), neboli *ortodromy*, která je největší kružnicí tělesa.

Obrázek 3.6
Mapa spojení vizualizující nejkratší cestu mezi body za využití hlavní kružnice[3]

3.3.4 Metoda teček

Metoda teček, označována také jako *dot map*, je druh geovizualizace zobrazující data pomocí teček různé intenzity a velikosti. Tato metoda vychází z vizualizace 3.3.2. Zobrazené body mohou být chápány jakou kvalitativní vizualizace, pokud primárně slouží k určení pozice prvku, nebo jako kvantitativní v případě, kdy umístění všech bodů na mapě značí intenzitu trendu v dané oblasti jako na obrázku 3.7.

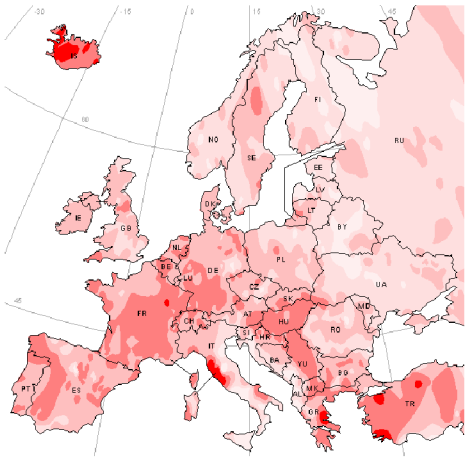


Obrázek 3.7: Mapa teček zaznamenávající počet úmrtí způsobených bleskem v USA mezi lety 2007 až 2017[4]

3.3.5 Isopleth mapa

Isopleth je druh geografické vizualizace barevně zvýrazňující hodnotu zvolené oblasti vzhledem ke statické proměnné. Tato geovizualizace vychází z kartogramu, ale na rozdíl od něj není hodnota mapována do předdefinovaných polygonů, nýbrž do oblastí, jejichž souřadnice jsou přímo součástí vstupních dat (viz. obr. 3.8).

[3] https://www.r-graph-gallery.com/connection-map.html
[4] https://twitter.com/gijn/status/1139504012628299777/photo/1
[5] https://setis.ec.europa.eu/example-of-isopleth-map



Obrázek 3.8: Isopleth mapa zobrazující rozšíření termálních pramenů v Evropě[5]

3.3.6 Současné nástroje

Existuje množství nástrojů, pomocí kterých je možné vizualizovat geografická data. Úkolem této práce je však vytvořit geovizualizační dashboard pro webový prohlížeč, pro což je vhodné použít některou z knihoven JavaScript.

Jednou z možností vytváření geografické vizualizace je vykreslovat jednotlivé polygony. Jedná se o postup práce na nízké úrovni, protože uživatel nemá k dispozici žádné předpřipravené geovizualizační nástroje a pomocí pravidel popisuje vlastnosti a tvar jednotlivých elementů. Existujícím nástrojem, který tvoří vizualizace tímto způsobem je například knihovna D3.js, která je založená na práci s dokumenty zaměřenými na data. Pracuje se standartními webovými jazyky HTML, CSS a bývá využívána pro zjednodušené generování SVG elementů popisující dvojrozměrnou vektorovou grafiku.

Protože se při tvorbě geovizualizací vytváří 2D vizualizace z původního 3D objektu, musí docházet k projekci původního modelu do výsledné vizualizace. V případě geografických vstupních dat, kdy jsou oblasti popsány pomocí dvojice hodnot (zeměpisné šířky a délky), tedy musí dojít k výpočtu nových souřadnic dle zvolené geografické projekce. Je možné využít knihovny d3-geo a d3-geo-projection, jejichž součástí jsou desítky projekcí, které se řídí rozdílnými přístupy pro vytvoření 2D mapy z původního 3D povrchu Země.

Tento způsob je vhodné využívat pro základní vizualizace, protože pomocí tohoto přístupu dojde k vizualizaci mapy relativně obtížným způsobem. Jelikož tento přístup pracuje na velmi nízké úrovni, pokud bude uživatel chtít přidat nějaké vylepšení mapy, implementace a propojení s mapou bude opět velmi náročné.

Geovizualizace lze vytvářet i opačným přístupem než generováním každého jednotlivého polygonu, a to výběrem geovizualizačních šablon s nastavitelnými vlastnostmi. Do této šablony uživatel pouze nahraje vstupní data a upraví dostupné vlastnosti mapy. Množstvím vizualizačních schémat disponuje například sbírka vizualizací NivoRocks, která pro práci s grafy kombinuje knihovny D3 a ReactJS. Zaměření vizualizací je však poměrně široké a NivoRocks nabízí pouze jednu geografickou vizualizaci. Jedná se o kartogram, který uživateli umožní namapování vstupních dat na konkrétní státy za pomoci JSON kolekce. Uživateli je následně umožněno upravovat množství vlastností mapy, jako je styl hranice mezi státy, zoom, zobrazovanou oblast, typ projekce a mnohá další.

Přestože se jedná o přístup s relativně jednoduchou implementací, bývá využíván zpravidla pro základní vizualizace, protože možnosti úprav a nastavení jsou pro každé schéma omezené a neumožňují jednoduché rozšíření.

Další metodou je využití nástrojů specializovaných na geografické vizualizace, které umožňují snazší implementaci a podporují množství funkcí vhodných pro geovizualizace (zoom, posun na mapě, shlukování značek...) a jednoduché možnosti rozšíření. Jednou z těchto knihoven je například Leaflet, který je prezentován především jako vhodný nástroj pro uživatele se zařízením ovládající se dotykovou obrazovkou. Nabízí však nejširší možnosti úprav z trojice zmíněných knihoven a má k dispozici nástroje pro snadnou implementaci geografických vizualizací typu kartogram a body na mapě, které je možné dále upravovat a kombinovat s dalšími knihovnami. Díky aktivní komunitě disponuje tato knihovna velkým množstvím pluginů, které jsou volně k využití, a nabízí široké možnosti úprav mapy od předpřipravených uživatelských menu a popupových oken až po funkce shlukování překrývajících se značek či získání přesných souřadnic kurzoru myši.

3.4 Pravidla kognitivního vnímání

Jak zmiňuje Few [6], zrak je naším dominantním smyslem a je úzce spjatý s myšlením. Abychom tedy mohli efektivně předat vizuální informaci uživateli, musíme pochopit, jak vnímá objekty lidský mozek. Při návrhu dashboardu je nutné zvolit takové techniky vizualizace, které budou pro uživatele jednoznačné a srozumitelné.

Gestaltismus neboli tvarová psychologie je jedna z disciplín, která studuje lidské vnímání obrazu a podvědomé reakce.

Na konci 19. a začátku 20. století v oblastech Rakouska a Německa začíná psychologie studovat vnímání tvarů, podobností a struktur. V roce 1912 vydal Wertheimer [20], která byla počátkem tvarové psychologie. Postupně vzniklo několik zákonů, které popisují zajímavosti podvědomého vnímání jako například:

- **Zákon blízkosti**: Jedná se o pravidlo, které popisuje tendenci vnímat objekty rozmístěné blízko sebe jako součást jednoho celku nebo série.
- **Zákon podobnosti**: Popis tendence mozku shlukovat objekty podobných barev, tvarů, velikosti a orientace do jedné skupiny.
- **Zákon pokračování**: Snaha lidského mozku o vnímání na sebe plynule navazujících částí jako celku.

Při mapování kvantitativních dat by měla být použita správná barevná škála, která umožní uživateli snadno odlišit jednotlivé intervaly hodnot. V knize [1] autoři detailněji popisují následující přístupy pro tvorbu barevných škál:

- *Jednobarevný postup (single-hue)*: Jedná se o metodu využití jedné barvy, která se v jednotlivých intervalech liší intenzitou.
- *Hodnotový přístup (value)*: Jednobarevný přístup, který může být využit pro jakoukoli barvu. V průběhu barevné škály se zvolená barva kombinuje s odpovídající intenzitou černé nebo bílé barvy.
- *Dvoubarevný přístup (bi-polar)*: Využívají se dvě, zpravidla opačné, barvy, které svojí nízkou/vysokou intenzitou dosahují ve střední hodnotě škály stejné barvy (zpravidla

bílý střed škály). V první polovině škály tedy svoji intenzitu mění jedna barva a v druhé polovině barva druhá.

- *Smíchaný barevný přístup (blended hue)*: Tento přístup využívá kombinaci dvou příbuzných barev, které na sebe postupnou kombinací dokážou plynule navázat.
- *Částečně spektrální přístup (partial spectral)*: Dochází ke spojení dvou odstínů sousedící se stejnou barvou, pro oba by byl s touto barvou tedy vhodný přístup blended hue. Zajištěno je tak tedy rozšíření barevného spektra ve zvolené škále.
- *Plně spektrální přístup (full spectrall)*: Spojuje se celé spektrum barev od modré po červenou, což zajistí nejširší možnou barevnou škálu z pohledu viditelného spektra. Tento přístup je doporučeno využívat jen ve speciálních případech, protože široká paleta barev v jedné mapě může často uživatele mást.
- *Kvalitativní přístup (qualitative)*: Účelem kvalitativní vizualizace není zobrazení intenzity hodnoty, ale vyznačení oblastí, které tuto podmínku splňují. Není tedy vhodné používat vzájemně si podobné barvy, ale takové, které od sebe budou snadno odlišitelné.

Důležitá je i práce s periferním viděním[6] uživatele. Na rozdíl od zaostřeného místa se totiž rozlišení vnímané oblasti periferního vnímání značně snižuje. Periferní vidění však lze využít například pro zaregistrování Pop-up[7] notifikace. Je však doporučeno tento způsob upozornění používat jen zřídka.

[6] Periferní vidění je vnímání okrajovou částí sítnice. Člověk dokáže pomocí periferního vidění vnímat i objekty mimo hlavní zornou oblast.

[7] Pop-up je vizuální element uživatelského rozhraní, známý také jako vyskakovací okno.

Kapitola 4

Dashboard

Dashboard je vizualizační nástroj předávající informace uživateli. K těmto účelům využívá množství grafických doplňků v podobě grafů, tabulek, symbolů nebo map. Tyto elementy musí být vhodně zvoleny a vizualizovány, protože jednou z nejdůležitějších vlastností dashboardu je rychlost a snadná pochopitelnost předávaných informací.

Stephen Few [6] definuje dashboard jako vizuální zobrazení nejdůležitějších informací pro dosažení jednoho nebo více cílů rozmístěné na obrazovce tak, že jsou informace viditelné na první pohled. Ve slovníku dictionary.com je definován jako uživatelské rozhraní nebo webová stránka, která poskytuje aktuální informace týkající se pokroku či výkonu, obvykle v grafické a snadno čitelné podobě. Nejprve byla tímto pojmem nazývána palubní deska automobilu, přičemž vzhledem ke společným vlastnostem rychlého a snadno pochopitelného zobrazení nejdůležitějších elementů začal být tímto slovem označován i informační dashboard. Pro oba nástroje je společné, že předávají cílovému uživateli informace o situaci za pomoci několika grafických nástrojů bez nutné znalosti aktivit probíhajících na pozadí.

Dashboard může pracovat s více zdroji dat, které s využitím procesů agregace a zjednodušování musí zobrazovat validní informace potřebné pro dosáhnutí stanovených cílů. Zobrazované informace musí být přehledné a snadno pochopitelné, aby dokázal uživatel co nejrychleji zareagovat na vzniklý problém, s čímž souvisí i skutečnost, že při návrhu dashboardu jsou upřednostňovány grafické vizualizace před textovými [6]. Pokud se nejedná o ryze analytické nástroje, není nutné zobrazovat mnoho informací s velkým množstvím detailů, protože správně navržený dashboard by měl poskytovat dostatek informací na jedné obrazovce bez nutnosti pohybu na stránce či přepínání mezi obrazovkami. Jedním ze vhodných ukázek informačního dashboardu je obrázek 4.1 od Stephena Fewa, který se při jeho návrhu řídil pravidly jednoduchého a srozumitelného zobrazení bez zbytečně široké škály barev a rušivých elementů.

Obrázek 4.1: Ukázka informačního dashboardu navrženého podle pravidel Stephena Fewa[1]

4.1 Využití

Dashbboardy mohou zobrazovat různorodá data a díky tomu mohou být užívány v mnoha odvětvích. Využití v konkrétních oblastech je velmi individuální vzhledem k ojedinělým představám a podmínkám jednotlivých společností. Dashboardy se mohou lišit detailností vizualizovaných dat, obnovovací frekvencí nebo typem požadovaných vizualizací. Manažerské dashboardy kontrolují výkonnost zaměstnanců, plnění norem, plánování a dodržování firemních cílů, zjednodušují firemní komunikaci. Zatímco některé organizace nejlépe využijí rychlou efektivní kontrolu jednoho primárního elementu, prioritou jiné může být udržení dostupnosti spojené jak s monitoringem služeb, tak s analytickým nástrojem pro detailní průzkum vzniklých problémů.

4.2 Vizualizace v dashboardech

Pomocí vizualizačního média jsou graficky nebo textově reprezentována data. Vizualizačním nástrojem může být například tabulka, která dokáže zobrazit n-dimenziální data, ale vzhledem k množství zachycených informací zabírá hodně místa a obtížně se v ní odvozují vazby mezi hodnotami.

Naopak diagramy neboli grafy kladou důraz na grafickou reprezentaci dat, která umožňuje snazší modelování vztahů mezi hodnotami, ale je pomocí nich možné zobrazit pouze omezený počet dimenzí. Grafy jsou při návrhu dashboardu nejvyužívanější vizualizační nástroj. Jejich velkou výhodou je relativně snadná pochopitelnost pro uživatele, protože se se zobrazováním

[1] http://startuplifeblog.com/tag/stephen-few/

dat do grafů setkávají poměrně často, a schopnost zobrazit velké množství dat na poměrně malém prostoru.

Pro využívání diagramů však existuje několik doporučení, kterých je vhodné se držet:

- Srozumitelnost grafu by měla zůstat podobná i při černobílém náhledu, protože téměř 10 % populace trpí nějakou poruchou rozlišování barev.
- Je důležité využívat takové grafy, ze kterých je uživatel schopen jednoznačně vyčíst podstatné informace. Pokud jsou tedy rozdíly mezi jednotlivými hodnotami nepatrné, je vhodné vyhnout se například kruhovým objektům a využít takové grafy, které jsou schopné odlišit i nepatrné rozdíly.
- Není vhodné využívat zbytečně komplikované grafy, které sice dokáží zobrazit velké množství dimenzí, ale uživatel bude muset nad pochopením pravidel grafu trávit delší čas a zároveň se snižuje přehlednost.
- Graf by neměl být doplňován o zbytečné vizuální doplňky jako fotky, loga, kresby...

4.3 Dělení dashboardů

Dashboardy se mohou dělit podle množství kritérií. Výběr konkrétního typu dashboard závisí na konkrétních požadavcích uživatele nebo organizace. Zatímco některé organizace potřebují pouze vizuální doplněk pro týdenní kontroly splněných norem, jiné vyžadují nepřetržité monitorování kritických sekcí s maximální obnovovací frekvencí. Few [6] zmiňuje rozdělení dashboardů podle mnoha kritérií, z nichž některé jsou zobrazeny v tabulce 4.1.

Kategorie dělení	**Druhy**
Role	strategický, analytický, operační
Typ dat	kvantitativní, nekvantitativní
Rozpětí dat	celofiremní, pro oddělení, individuální
Obnovovací frekvence	měsíční, týdenní, denní, hodinový, real-time
Interaktivita	statické zobrazení, interaktivní zobrazení (filtrování)

Tabulka 4.1: Ukázka několika kategorií dashboardů

4.3.1 Klasifikace dashboardů dle rolí

Přestože se dashboardy dělí podle mnoha kritérií, v mojí bakalářské práci se zaměřuji na využití při monitorování síťového provozu, což vhodně vymezuje dělení dashboardů dle Eckersona [5]. Ten je dělí podle primárních rolí na 3 typy: strategické, analytické a operační. Není nutné, aby výsledný dashboard striktně plnil pouze jednu roli, finální vizualizace jich může částečně kombinovat několik, ale je důležité pochopit rozdíly mezi nimi.

Strategický dashboard

Poskytuje přehled informací o dlouhodobém stavu organizace. Jedná se o typ dashboardu určený primárně pro nejvyšší management, jehož primárním úkolem je plánování a dohled nad dodržováním plánu organizace [12]. Zobrazuje data z delšího časového období, protože na základě dlouhodobého monitorování lze snáze odhalit prostor ke zlepšení či změnám.

V oblasti monitorováním síťového provozu vlastník sítě deklaruje podmínky dostupnosti, rychlosti a spolehlivosti svých služeb, na základě kterých je vytvořen dashboard, který plnění těchto cílů sleduje. Ke kontrole těchto hodnot dochází po delších časových úsecích, aby byl vzorek dostatečně reprezentativní.

Analytický dashboard

Nabízí uživatelské rozhraní pro tvorbu analytických operací. Využívá se v oblastech, které detailněji zkoumají vzniklé problémy. Tento dashboard pracuje s velkým množstvím dat, protože musí poskytovat detailní informace pro univerzální analytické nástroje. Měla by být dostupná možnost porovnání s jinými záznamy v čase pro tvorbu základní analýzy, aby uživatel pochopil trendy dané problematiky.

V případě vzniku anomálie v monitorované síti je kromě zjištění samotného výskytu události nutná i podrobná analýza situace za účelem odhalení původce anomálie a zavedení preventivních opatření. Dashboard by měl umožňovat interaktivní komunikaci, díky které může analytik na existující událost nahlížet z více perspektiv a rychleji tak pochopit jádro dané problematiky.

Operační dashboard

Jedná se o nejpoužívanější typ dashboardu, protože je založen na předávání rychlých zpráv s aktuálními údaji a udržení si přehledu o dané oblasti, což je v dnešní digitální době žádoucí. Jeho primárním úkolem je správa a kontrola každodenních rutin. Je určen pro uživatele, kteří musí disponovat informacemi o aktuálním stavu situace, a proto pracuje primárně s daty s vysokou obnovovací frekvencí, často dokonce přímo v reálném čase.

Při nepřetržitém monitorování sítě upozorňuje operátory na překročení průměrné hodnoty či stanovených prahů. Je kladen důraz na jednoduchost a zvýraznění kritických hodnot, aby mohl uživatel co nejrychleji rozpoznat problém a pohotově na tuto situaci zareagovat.

Kapitola 5

Analýza

V této části práce se zaměřím na problém vizualizace geografických dat. Provedu analýzu nástrojů společnosti Flowmon, které mají přístup ke geografickým datům, a porovnám zobrazení těchto dat s řešením některých konkurenčních společností. Za účelem analýzy jsem využil možnosti interaktivních demo účtů, které většina společností v tomto oboru nabízí, a v případě nástrojů společnosti Flowmon jsem se zúčastnil i schůzek se zkušenějšími uživateli, kteří vysvětlovali využití nástrojů z pohledu konkrétních person.

Geografická data, stejně jako další druhy dat, je možné vizualizovat mnoha způsoby. Jedním z možných způsobů je textová reprezentace, která však v tomto případě uživateli zkomplikuje a prodlouží čas strávený u analýzy dat. Ani vizualizace v podobě statistických grafů tento problém neřeší. Pokud má uživatel co nejrychleji pochopit podstatu řešené problematiky, je nutné, aby na první pohled porozuměl grafu, a pokud se snažíme předat uživateli informace geografického charakteru, vhodný způsob zobrazení je danou informaci ukázat na mapě.

5.1 Geografická data v nástrojích Flowmon

Monitorovací nástroje společnosti Flowmon získávají informace o datových tocích pomocí protokolů Netflow a IPFIX. Pro každý datový tok je k dispozici dvojice komunikujících IP adres, ze kterých lze, například pomocí nástroje IP-API[1] nebo dalších jemu podobných, získat informace o geolokaci IP adresy.

V nástroji Monitoring Center, popsaném v kapitole 2.3, je však v sekci *Analýza* z IP adresy dohledáno geografické umístění a je tedy zřejmé, že tuto funkcionalitu mají ve společnosti Flowmon již implementovanou. Informace o zemi původu IP adresy je však vizualizována pouze jako ikona vlajky daného státu (obrázek 5.1) a záznamy jsou umístěny v tabulce, ze které je uživatel jen těžko schopen vyvodit rychlé závěry a udělat si přehled o situaci. Navíc s informací o lokalizované zemi nelze při tvorbě statistiky pracovat jako s filtrovacím parametrem, takže není možné ani shlukovat adresy náležící pod stejnou zemi. Přestože je nástroj pro geografickou lokalizaci implementován a použit v jednoho pluginu, žádný další jej při práci nevyužívá. Geografické informace by našly využití v mnoha dalších nástrojích. Velmi užitečné by mohly být například v nástrojích *ADS* a *DDoS*, kde by díky těmto informacím mohlo být možné blokovat zahraniční provoz pouze ve vybraných zemích.

[1] https://ip-api.com

Obrázek 5.1: Současný způsob vizualizace geografických dat pomocí symbolu vlajky konkrétního státu v nástroji Flomon Monitoring Center

Pro lepší porozumění rutinních aktivit je nutné se na užívání nástrojů ADS a DDoS Defender podívat z pohledu konkrétních uživatelů. Na základě několika uskutečněných schůzek se zaměstnanci Flowmonu byli definováni dva hlavní uživatelé jejich nástrojů: operátor a analytik.

Operátor sleduje aktuální dění v síti a snaží se udržet si přehled o aktivitách a anomáliích v monitorované síti. Na případné nové události potřebuje být jednoznačně vizuálně upozorněn. Ať jsou útoky nebo anomálie odchozí nebo příchozí, nastává situace, která vybočuje ze standartu a musí být neprodleně řešena. V co možná největší míře pracuje s real-time daty a nestará se o řešení incidentu, protože v případě vzniklého problému pouze deleguje událost dále na analytika.

Druhou skupinou jsou analytici, kteří zkoumají záznamy incidentů. Na rozdíl od operátorů se zajímají o detaily vedoucí ke vzniku situace a informace o průběhu incidentu a potřebují tedy mít k dispozici co nejpodrobnější informace. Lze předpokládat, že mají o problematice lepší znalosti než operátor, protože se starají o řešení a o případnou prevenci.

5.2 Zpracování geografických dat konkurencí

Všechny analyzované nástroje pro monitoring síťového provozu pracují minimálně s daty v podobě IP adres, ze kterých je pomocí některého z mnoha API nástrojů možné snadno získat informace o geografickém umístění. Provedu tedy analýzu konkurenčních produktů za účelem průzkumu, jak a zda pracují s geografickými daty ve svých nástrojích. Pro analýzu jsem využil interaktivní demo účty, které společnosti poskytují pro vyzkoušení jejich služeb. Zaměřím se na to, zda geografická data vůbec zpracovávají a pokud ano, tak i na způsob jejich vizualizace.

SolarWind

Tato společnost disponuje desítkami nástrojů a pro analýzu jsem měl k dispozici několik interaktivních simulací síťového provozu. Informace o probíhající komunikaci v síti jsou k dispozici v nástroji *Network* v sekci *My Dashboards*. Nástroj shromažďuje informace a statistiky provozu a agregované výsledky následně vizualizuje podle vybraných kategorií, přičemž jednou z možných kategorií je i stát. Ani v tomto případě však nejsou geografická data použita pro geovizualizaci. V demo ukázce je uživateli zobrazen widget pro top 5 zemí komunikující prostřednictvím IPv4[2]. Informace jsou vizualizovány prostřednictvím výsečového grafu, který při najetí na jednu z částí zobrazí název státu a hodnoty provozu. Vizualizované statistiky jsou zobrazeny a seřazeny také v tabulce pod grafem (viz. obr. 5.2), kde je u názvu státu použita vlajka dané země.

[2]IP protokol ve verzi 4 je datově orientovaný síťový protokol, který slouží k zasílání paketů na internetu.

COUNTRY	INGRESS BYTES	EGRESS BYTES	INGRESS PACKETS	EGRESS PACKETS	PERCENT
Private Address	50.5 Gbytes	40.2 Gbytes	39.84 M	32.51 M	59.8%
North Korea	20.7 Gbytes	20.7 Gbytes	2.51 M	2.51 M	27.23%
United States	11.0 Gbytes	6.2 Gbytes	12.72 M	8.45 M	11.32%
Sweden	640.2 Mbytes	88.9 Mbytes	237.44 k	136.25 k	0.48%
Netherlands	451.3 Mbytes	269.3 Mbytes	1.57 M	1.21 M	0.47%

Obrázek 5.2: Vizualizace geografické informace formou vlajky do tabulky států, které nejvíce komunikují prostřednictvím protokolu IPv4

Geografické vizualizace, které jsem v demo nástroji objevil, byly mapa aktuálního počasí a úderů blesku, obě typu isopleth, které však uživateli nepředají žádné informace o síťovém provozu, a ještě dvě mapy světa, na kterých je umístěno několik bodů symbolizující geografické rozmístění firemních technologií (viz. obr. 5.3). Obě mapy předávaly informaci stejným způsobem, ale jedna informovala o výpadku a druhá o jiných chybách. V případě výpadku nebo jiné chyby změní dané body barvu a uživatel je tak informován o chybě na konkrétním zařízení a jeho geografické poloze, což využijí hlavně společnosti, které nemají všechny technologie v jedné oblasti.

Obrázek 5.3: Geovizualizace zobrazující body na mapě, která vyznačuje oblasti, ve kterých má uživatel uložené technologie. Barva značek se mění v závislosti na stavu zařízení

Byl jsem překvapen, že jsem v dostupných nástrojích tak velké společnosti jako SolarWind nenalezl žádnou plnohodnotnou geovizualizaci. Pro ověření, jestli jsem žádnou geovizualizaci nebo práci s geografickým daty neopomenul, jsem kontaktoval oddělení zákaznické podpory společnosti, od které mi bylo oznámeno, že se skutečně v současné době jedná o jediné nástroje i využití geografických dat v jejich produktech.

NetFort

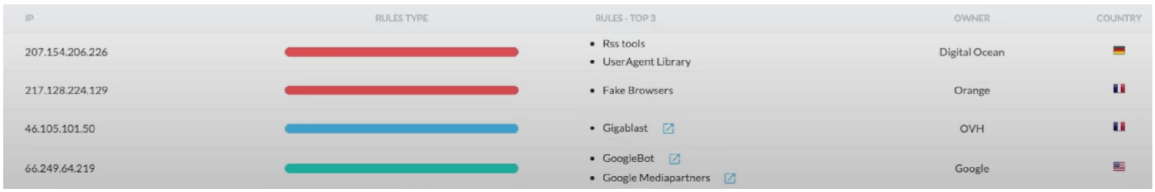
Nástroj společnosti Netfort je nazýván LANGuardian. V tomto pluginu jsem nenašel žádný typ geovizualizace. Většina jejich grafů je založená na zobrazení výsečového nebo sloupcového grafu a tabulky s IP adresami, výjimečně kategoriemi. Kdykoli, kdy je v tabulce zobrazen seznam IP adres, je každá adresa doplněna o vlajku státu (viz. obr. 5.4), ke kterému náleží, případně symbol privátní sítě. Přestože je tento nástroj schopen získávat geografické informace z IP adres, je tento typ dat využíván pouze jako grafický doplněk do tabulky záznamů. Geografická příslušnost k vybrané zemi není systémem využita ke kategorizaci a nelze tedy pracovat se sítěmi daného státu jako s celkem. Tento nástroj není v žádné oblasti zaměřen na geografická data.

PROTOCOL	SOURCE IP	DESTINATION IP	SERVER PORT
TCP	10.2.2.100	10.1.1.97 (DELYNAS)	445 (microsoft-ds)
TCP	10.2.2.100	185.102.218.80 (a3.aliez.me)	8080 (http-proxy)
UDP	10.2.2.100	213.127.233.218	51413
UDP	10.2.2.100	185.113.128.99	51413

Obrázek 5.4: Tabulka adres obsahující informace o použitých síťových protokolech, IP adresách s vlajkou státu a číslech komunikujících portů

DataDome

Aktivity botů, které tento nástroj monitoruje, jsou zaznamenávány na základě IP adresy, pomocí které se serverem komunikují. Při tvorbě záznamu se k informaci o adrese připojuje i zjištěná geografická lokalizace (viz. obr. 5.5), která je v přehledu záznamů zobrazována. V nástroji nejsou využity žádné geografické vizualizace pro znázornění původu komunikujících botů, nebo četnost botů v jednotlivých zemích. Příslušnost k danému státu je však brána jako samostatná kategorie, podle které je možné filtrovat výsledky a nastavovat další pravidla. Přestože tedy v současné době není v dashboardu žádná geovizualizace implementována, nástroj je schopen geografická data z IP adresy získat, třídit a následně s nimi pracovat jako s kategorickým celkem.



Obrázek 5.5: Vizualizace statistik o botech v síti včetně míry nebezpečí, majitele, zaměření a státní příslušnosti

5.3 Shrnutí analýzy

V současné době společnost Flowmon nedisponuje ve svých pluginech žádnými geovizualizacemi a jen v ojedinělých případech pracuje s geografickými daty, které však využívá pouze jako grafický doplněk v podobě vlajky k záznamům v tabulce. S informacemi o příslušnosti IP adresy k odpovídajícímu státu není možné nijak pracovat ani podle nich filtrovat záznamy. Přestože společnost k analýze nenabízí data o geografické poloze, detailnějším informacím o operačních systémech a názvech výrobců zařízení je věnována samostatná sekce *Aktivní zařízení*.

Konkurence v současné době sice také pracuje s geografickými informacemi velmi střídmě, ale umožňuje alespoň práci s tímto typem dat jako s jakoukoli jinou kategorií. V řešení společnosti SolarWind je geografickým datům věnován samostatný oddíl *Country*, kde jsou uživateli k dispozici statistiky o zemích, které s jeho zařízením komunikují. V jednom z nástrojů je využita geografická vizualizace v podobě bodů na mapě, které symbolizují pozici nastavených zařízení uživatele. Jedná se o interaktivní nástroj, který měnící se barvou značek upozorňuje uživatele na změny stavu zařízení. V žádném z analyzovaných nástrojů se však nenachází geovizualizace, která by pracovala s informacemi o síťovém provozu.

Obecně by se mělo zapracovat na využití geografických dat, aby uživatel získal přehled o situaci komunikujících zařízení z pohledu jejich geografického umístění. Pro vizualizaci tohoto typu dat je vhodné použít geovizualizace, které uživateli náležitým způsobem vyznačí problematické oblasti v mapě celého světa.

Kapitola 6

Návrh řešení

Cílem mé práce je vytvořit dashboard geografických vizualizací pro společnost Flowmon. Existující nástroje nabízí několik pohledů na danou problematiku a mým úkolem je vytvořit nástroj, který nabídne uživateli nový pohled z hlediska geografického zaměření vzniklých síťových anomálií a usnadní mu tak analýzu aktuální situace. Pomocí tohoto nástroje budou zobrazeny kritické oblasti a uživatel může rychleji zareagovat na vzniklý problém. Jelikož jsou současné webové nástroje společnosti Flowmon navrženy v JavaScript knihovně React, bude moje práce tvořena ve stejném programovacím jazyce. Grafické návrhy vizualizací jsem vypracoval pomocí nástroje Figma.com[1]

6.1 Architektura

Data budou získávána z interní REST API konkrétního nástroje společnosti Flowmon a budou přijímána ve formátu JSON. Jelikož prozatím neznáme formát a názvy dimenzí výsledku, bude navržený nástroj připraven pracovat s generickými daty a implementované funkce tedy nebudou závislé na typu ani tvaru vstupních dat. Aby byla umožněna práce s generickými daty, bude si sám uživatel mapovat jednotlivé dimenze dat na konkrétní dimenze grafu.

Informace o souřadnicích polygonů vymodelovaných států nebudou součástí vstupních dat, obsaženy budou pouze zkratky dle normy ISO 3166-1 alpha-3, jejichž kompletní seznam je sepsán například v uživatelské příručce nástroje Flowmon DDoS Defender[2]. Příklad vstupního záznamu může být následující:

```
{
    "state":"mitigated",
    "from": "BEL",
    "to": "AGO",
    "value": 334330
}
```

Uživatel si pomocí menu vybraného nástroje bude moci zvolit dimenzi, která obsahuje zkratku státu, do kterého bude zvolená hodnota vizualizována. Dále zvolí dimenzi obsahující sledovanou hodnotu a operaci, kterou bude chtít s agregovanými záznamy daného státu

[1] Zdroj použité podkladové mapy v návrzích: https://cs.m.wikipedia.org/wiki/Soubor:Eastern-Europe-map.svg

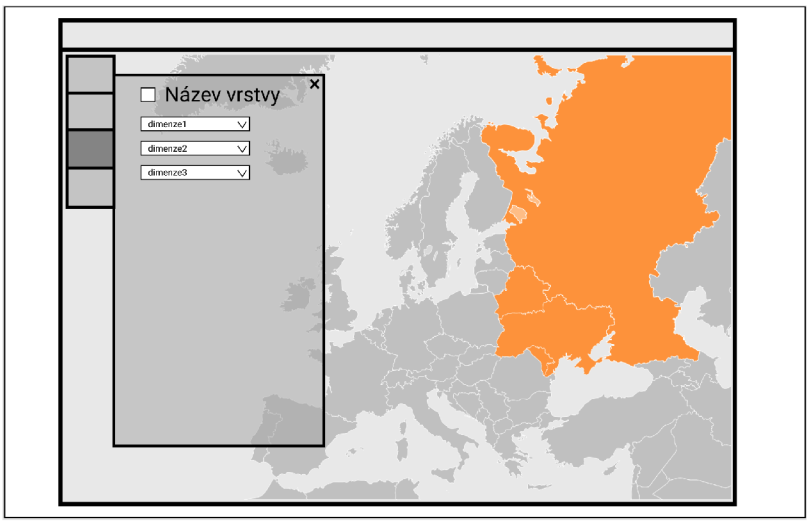
[2] https://demo.flowmon.com/doc/iad_userguide_en.pdf

provádět jako počet výskytů, nebo součet (sčítání hodnot, pokud se bude jednat pouze o číselné hodnoty).

Vymodelované státy tedy budou sepsány v samostatném souboru a popis jejich tvaru a umístění bude ve standartním formátu geoJSON, který slouží k reprezentaci jednoduchých geografických prostorových objektů. Geovizualizace budou kvantitativní a podle charakteru dat ve zvolené dimenzi bude vizualizovaná hodnota získána buď počtem výskytů v souboru, nebo, pokud se bude jednat pouze o číselné hodnoty, součtem těchto hodnot.

6.2 Návrh UI

V případě geovizualizací je vhodné zobrazit na celé obrazovce pouze mapu, případně doplněnou o menu pro drobná nastavení. Proto bude v dashboardu umístěn pouze jeden widget v podobě interaktivní mapy s postranním uživatelským menu. Jednotlivé geovizualizace budou zobrazovány ve formě vrstev, které se budou vzájemně překrývat. Pro zajištění přehlednosti budou všechny vrstvy skrývatelné, například pomocí check boxu, aby bylo uživateli umožněno zobrazovat a kombinovat požadované vrstvy. Tato možnost zobrazování/skrývání vrstev bude součástí uživatelského menu. Menu bude obsahovat sekce pro každou implementovanou vizualizaci, ve kterých budou k dispozici všechny možnosti nastavení dané vrstvy, jak je znázorněno na obrázku 6.1. Jelikož bude dashboard navržen pro práci s generickými daty, budou v menu k dispozici uživatelské nástroje, pomocí kterých bude moci uživatel mapovat vybrané dimenze vstupních dat na dimenze georafických vizualizací.



Obrázek 6.1: Návrh vzhledu a velikosti widgetu v dashboardu a rozmístění uživatelského menu a ovládacích prvků

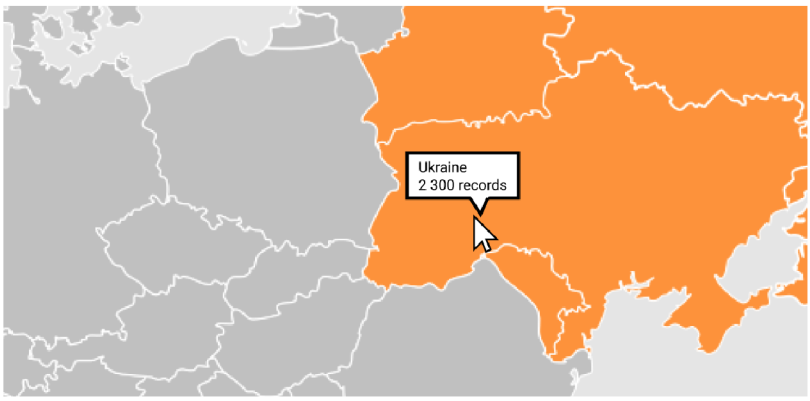
6.2.1 Kartogram

Využita bude geovizualizace kartogram, která bude zobrazovat intenzitu problematiky v jednotlivých státech. Intenzita bude zobrazována pomocí barevné škály *smíchaný barevný přístup* 3.4, tedy složením dvou příbuzných barev, které jsou ve středním intervalu zkombinovány, na jednom okraji je využita jedna barva a na opačném okraji druhá. Tento přístup jsem zvolil, protože je pomocí něj snadné zdůraznit výrazněji vychýlené hodnoty.

Kromě barevného odlišení vizualizovaných hodnot bude mít uživatel v každém polygonu státu k dispozici i informační tooltip[3], který ho bude informovat o názvu lokace a přesné hodnotě sledovaného prvku v dané oblasti jako na obrázku 6.2.

Uživatel bude v menu pomocí tří select boxů mapovat dimenze dat do dimenzí grafu. První select box bude určovat, ve které dimenzi se nachází zkratka daného státu. Tento údaj bude zapsán v souboru pomocí zkratky dle ISO 3166-1 alpha-3. Druhý select box bude určovat dimenzi udržující informaci o hodnotě spojenou s vybraným státem a třetí bude sloužit k výběru operace s touto hodnotou. Dostupnými operacemi bude součet hodnot a počítání záznamů obsahujících zvolenou vlastnost z druhého select boxu.

Vizualizovány budou všechny vymodelované státy. Pokud ve vstupních datech nebude některý ze států definován, bude vyplněn neutrální barvou a v jeho tooltipu bude zaznamenán pouze jeho název. Když bude některá z uživatelem zvolených dimenzí chybně zvolena, nedojde ke korektnímu namapování hodnot na geografické oblasti a všechny státy budou standardně zobrazeny jako nedefinované.



Obrázek 6.2: Návrh zvýraznění namapovaných dat a zobrazení tooltipu při najetí na konkrétní stát pro vizualizace kartogram

6.2.2 Body na mapě

Druhou použitou vizualizací budou body na mapě. Značky budou umístěny uprostřed každého státu a budou nést informaci o uživatelem namapované hodnotě. Tyto značky, označovány jako centroidy, budou, stejně jako polygony států, samostatným souborem, protože ani tyto informace nebudou součástí vstupních dat z interní API.

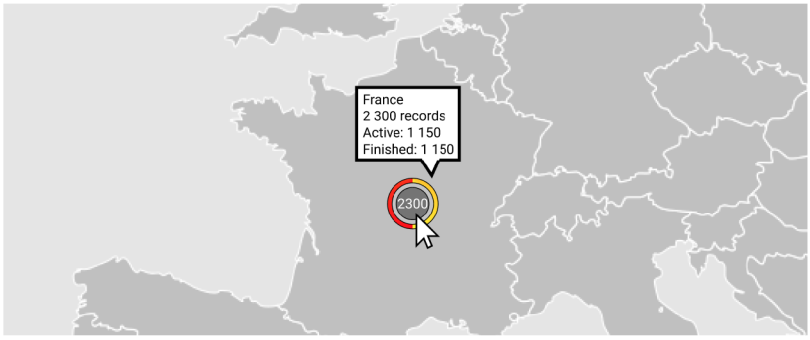
Protože bude mapa interaktivní a umožní uživateli funkce přiblížení/oddálení, musí tato vizualizace disponovat možností shlukování značek. Díky této funkci bude zamezeno překrývání blízkých značek a bude tak zajištěna vyšší přehlednost vizualizovaných dat. Značky budou ve tvaru výsečového grafu typu donut s přidanou dimenzí uvnitř kružnice.

V uživatelské menu se budou nacházet 4 select boxy pro výběr dimenzí dat. Tři z nich budou zastupovat dimenze pro určení státu, hodnoty a agregační operace. Čtvrtý bude sloužit k určení dimenze, která rozdělí zvolenou hodnotu do kategorií. Výsledná vizualizace tak bude uživateli zobrazovat dvě informace, číselnou hodnotu pro danou oblast a poměr jednotlivých kategorií, ze kterých je složena. Díky tomu bude možné například zobrazit

[3]Tooltip je datová položka jazyka HTML, která zobrazuje uživateli nápovědu při najetí myší na konkrétní element.

celkový počet útoků v konkrétním státě a zároveň graficky znázornit, které z útoků jsou ukončené, plně aktivní a mitigované[4].

Stejně jako v případě kartogramu bude i tato vizualizace disponovat pro každou zobrazenou značku tooltipem, který bude uživateli zobrazovat název oblasti, výčet kategorií a hodnoty, které jsou k nim přiřazeny jako na obrázku 6.3.



Obrázek 6.3: Výsečový graf vizuálně zachycující poměr velikosti zaznamenaných kategorií vůči celku a zobrazení celkové hodnoty namapovaných dat doplněný o tooltip s výčtem hodnoty a jednotlivých kategorií

[4]Za mitigované jsou považovány stále aktivní útoky, jejichž provoz je již filtrován.

Kapitola 7

Implementace

V této kapitole bude sepsána implementace geografického dashboardu, použitých technologií a práce s knihovnou LeafletJS.

7.1 Použité nástroje

Při práci s daty je využívána knihovna ReactJS. Aplikace je rozdělená do několika spolupracujících komponent. Pro skládání komponent je využíváno rozšíření syntaxe JavaScriptu JSX, který je využíván k popisu vkládání jednotlivých komponent do dokumentu a k předávání speciálního vstupu jazyka React označované jako `props` neboli properties (vlastnosti).

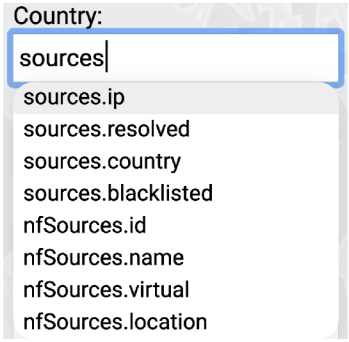
Pro tvorbu geovizualizací je použita knihovna LeafletJS. Jako jediná z analyzovaných knihoven totiž nabízí široké možnosti úprav a přidávání vlastních rozšíření, což zajistí do budoucna možnost případných vylepšení a rozšíření nástroje například o plánovanou geovizualizaci mapy spojení. Navíc disponuje geovizualizacemi typu kartogram a body na mapě. Pro vizuální úpravy elementů dashboardu je používán jazyk pro tvorbu stylů CSS. Aplikace využívá další podpůrné balíčky o jejichž instalaci a spouštění se stará správce balíčků NPM.

7.2 Zpracování vstupních dat

Jelikož nebylo doposud stanoveno, se kterými daty z REST API bude vhodné pracovat, současná implementace geografického dashbordu pracuje s daty z externího souboru ve formátu JSON a implementace je připravena pro práci s generickými daty.

Při načtení souboru do skriptu dojde k uložení obsahu souboru do zvolené proměnné. Pro zvolení mapované dimenze dat je vytvořen seznam dostupných dimenzí, který je uživateli k dispozici v menu konkrétní vrstvy geovizualizace. Za tímto účelem byla vytvořena pomocná třída *GeoData*, která zkopíruje obsah souboru a provede jeho analýzu. Protože vstupní data mohou být libovolně zanořená, jsou obsahy záznamů procházeny rekurzivně až na úroveň struktury nebo pole. Pro každou platnou cestu je otestováno, zda obsahuje hodnotu, a následně je uložena i se svým obsahem do instanční proměnné `metaPaths`, která udržuje informace o dostupných cestách a hodnotách vstupního JSONu. V komponentách pro kartogram a body na mapě je poté využívána třídní funkce `getMetaPaths()`, která vrací seznam dostupných cest. Po konzultaci s pracovním týmem jsem se rozhodl místo select boxů využít sofistikovanější řešení v podobě našeptávače. Po analýze několika nástrojů jsem využil React komponentu `react-autocomplete` (viz. obr. 7.1), která nabízí možnosti úpravy

vzhledu pomocí CSS a nastavení libovolné akce při zvolení nabízené hodnoty. Navíc díky správci balíčků NPM a komponentě jazyka React je snadné i její zprovoznění.



Obrázek 7.1: Použitá React komponenta `react-autocomplete` filtrující seznam parsovaných cest vstupního JSON souboru na základě uživatelského vstupu

7.3 Konzistence databází

Jako seznam dostupných vymodelovaných států jsem použil volně dostupné *countries.geo.json*[1] a pro středové hodnoty států *un-country-centroids.json*[2]. REST API společnosti Flowmon disponuje, dle manuálu k nástroji DDoS Defender, databází 249 států. Na základě normy ISO 3166-1, která se zabývá reprezentací názvů a zkratek států a jejich subdivizí, je v současné době uznáváno 249 států, ze kterých je 193 suverénních a členy Organizace spojených národů.

Pro zajištění konzistence všech používaných souborů včetně databáze států v REST API jsem vytvořil skript v jazyce Node.js, který porovnává názvy a tříznakové zkratky států s oficiální normou. Skript otestoval názvy a použité zkratky všech dostupných záznamů v souborech *countries.geo.json*, *un-country-centroids.json* a *FlowmonAPI.json* vůči aktuální normě ISO 3166-1. Pouze databáze *FlowmonAPI.json* byla v souladu s touto normou. V případě odlišného názvu doplnil skript automaticky do souboru oficiální název dle normy ISO 3166-1. Několik zbývajících států, které nebyly vymodelovány formou polygonů nebo zaznamenány mezi centroidy, bylo vypsáno na standartní výstup. Jednalo se zpravidla o menší ostrovy v oblasti Karibského moře. Tyto chybějící záznamy jsem následně vymodeloval pomocí nástroje geojson.io a doplnil je do souborů.

7.4 Vrstvy dashboardu

Dashboard se skládá z jediného widgetu, který je umístěn přes celou obrazovku. Ten je doplněn o uživatelské menu a ve spodní části jsou zmíněny použité technologie. Aby bylo možné pracovat s více geovizualizacemi, je nutné aplikovat možnost skládání vrstev jednotlivých vizualizací a vyřešit přepínání mezi nimi. Z toho důvodu jsem vytvořil nadřazenou komponentu `CombinedMap`, která spravuje nastavení a zobrazení implementovaných vrstev. Tato vrstva slouží k vytvoření základní prázdné Leaflet mapy a jejích vlastností a postranního uživatelského menu, které bude spolupracovat s vrstvami. Všechny vrstvy využívají stejně pojmenované sdílené funkce, díky čemuž může nadřazená komponenta pracovat s jakoukoli

[1] https://github.com/johan/world.geo.json/blob/master/countries.geo.json
[2] https://gist.github.com/rosszurowski/415eb3175249fff3714b14c1c51582cc

vrstvou stejným způsobem pro inicializaci, zobrazení/skrytí vizualizační komponenty, tvorbu položek, nastavení atributů a dalších.

Jelikož použité vizualizace nezobrazují geografickou mapu světa, ale pouze vybraným způsobem vizualizují mapované geograficky zaměřené hodnoty, které jsou obsaženy ve vstupních datech, bylo nutné použít podkladovou mapu, která bude sloužit ke geografické orientaci uživatele. V práci jsem využil mapu OpenStreetMap[3] s licencí Open Database License, která zajišťuje volné využití a sdílení za podmínky poskytnutí stejné možnosti dalším uživatelům (viz. obr. 7.2). Mapa je umístěna pod všemi implementovanými vrstvami a vizualizace tedy budou umístěny na podkladovou vrstvu. Tato vrstva nedisponuje žádnými interaktivními funkcemi a slouží pouze k zpřehlednění obsahu dashboardu.

Obrázek 7.2: Implementace otevíratelného uživatelského menu, podkladové mapy OpenStreetMap a lišty použitých technologií

7.5 Kartogram

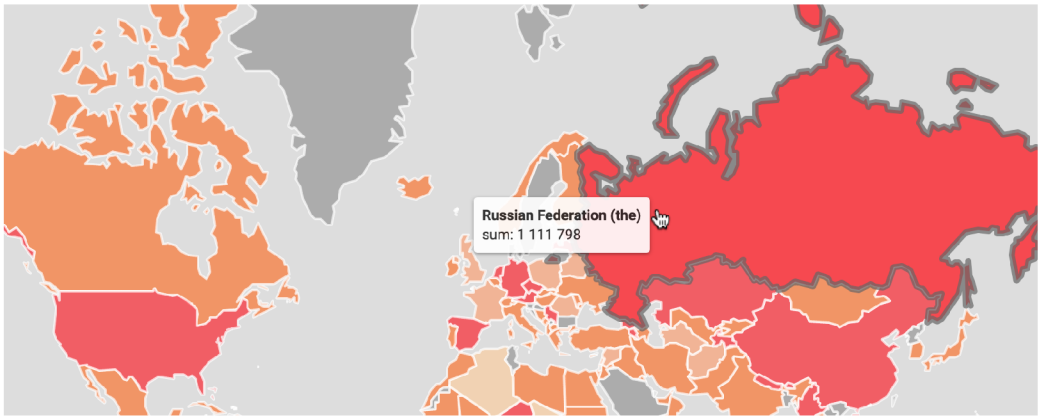
Tato komponenta je inicializována ve vyšší vrstvě `CombinedMap`, která jí pomocí *props* předává informace o vymodelovaných státech, jejich zkratkách, odkaz na vytvořenou Leaflet mapu a vstupní data. V současné implementaci jsou k dispozici 3 přednastavená barevná schémata obsahující každá 8 odstínů barev, jejichž prahy jsou nastaveny manuálně a po výsledcích testování a upřesnění požadavků je možné tento rozptyl prahů změnit. Informace o vymodelovaných státech jsou zkopírovány do třídní proměnné, do které bude následně ukládána i celková hodnota pro zařazení do odpovídajícího barevného intervalu. Pokud není dvojice hodnot zvolena nebo neposkytují platné hodnoty, jsou polygony států vyplněny výchozí nastavenou barvou.

Data jsou uživatelem mapována pomocí stažené React komponenty `AutoComplete`, která při zvolení hodnoty změní obsah vybrané třídní proměnné a zavolá funkci pro update vykreslené vizualizace. Při zvolení vstupních dat dojde k cyklickému čtení dvojice namapovaných dimenzí obsahující zkratku státu a sledovanou hodnotu. Pokud je pro získanou zkratku nalezen odpovídající výsledek v seznamu vymodelovaných států, pokusí se získat zapsanou hodnotu `properties` tohoto státu, případně vytvoří novou s nulovou hodnotou a na základě zvolené matematické operace hodnotu přičte nebo inkrementuje obsah této proměnné.

[3] https://www.openstreetmap.org

Tímto způsobem cyklus doplní seznam států o novou dimenzi dat vyjadřující vizualizovanou hodnotu a následně pro každý záznam zkontroluje, zda původní barva polygonu hodnotou stále odpovídá rozdělení barev dle prahů.

Výsledná vizualizace je doplněná o interaktivní prvky zvýraznění a zobrazení přesné hodnoty. Při najetí kurzorem myši na konkrétní stát se sníží průhlednost výplně polygonu a změní se barva a šířka jeho obrysu jako na obrázku 7.3. Navíc je uživateli pomocí *sticky*[4] tooltipu zobrazena informace o názvu státu a přesné vizualizované hodnotě.



Obrázek 7.3: Vizualizace polygonů států na základě hodnot ze vstupních dat, ukázka zvýraznění jednoho státu a použití tooltipu

7.6 Body na mapě

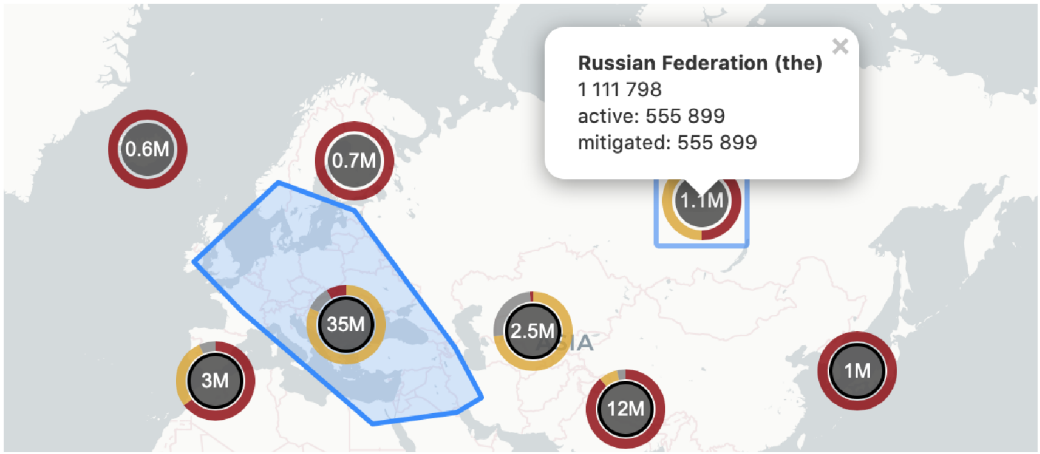
Stejně jako vrstva kartogram, je i komponenta prezentující body na mapě inicializována ve vyšší vrstvě. Na rozdíl od kartogramu jsou jí předány pouze vstupní data a záznamy o centroidech států. V této komponentě je využita knihovna D3, která slouží k vizualizaci značek. Tyto značky, na rozdíl od původních z knihovny Leaflet, nabízí vizualizaci více dimenzí dat. V tomto případě se ve středu vizualizace nachází celková hodnota z namapovaných záznamů pro danou oblast a obrys vizualizace je tvořen výsečovým donut grafem, který ukazuje poměr kategorií další dimenze.

Na základě zvolených dimenzí v uživatelském menu jsou informace o státech, celkové hodnotě a typech kategorií ukládány do společného pole. Jakmile jsou všechna data zpracována, dojde k vytvoření značek pro každý záznam v poli. Vzhled značek se řídí pravidly pro grafickou komponentu knihovny D3. Vnější výsečový donut graf pracuje s počtem a poměrem hodnot kategorií a vnitřní vizualizace číselně vyjadřuje celkovou hodnotu pro konkrétní stát.

Funkce knihovny Leaflet pro tvorbu značek `markerClusterGroup` disponuje funkcí pro shlukování značek. Značky se shlukují na základě vzájemné vzdálenosti. Tato funkce se využívá při přibližování/oddalování mapy a vzhled nových značek zastupující 2 a více původních značek zůstává stejný, pouze přejímá a sčítá hodnoty zastoupených států a při najetí kurzorem myši barevně zabarví oblast, kterou zastupuje (viz. obr. 7.4).

Kromě vizualizace informací prostřednictvím grafů je součástí každého grafu také popup, který uživatele informuje o názvu zvolené oblasti, celkové hodnotě a rozložení hodnot v rámci kategorií.

[4]Vlastnost sticky značí změnu pozice elementu podle pohybu kurzoru uživatele.



Obrázek 7.4: Body na mapě vizualizující počet útoků v jednotlivých oblastech, ukázka zobrazení konkrétních informací prostřednictvím popupového okna a vizualizace zastoupené oblasti prostřednictvím barevného polygonu.

Kapitola 8

Testování

V této kapitole se budu věnovat způsobům, kterými byl v průběhu implementace testován dashboard, a problémům, které byly díky testování objeveny a vyřešeny.

8.1 Testování v týmu

Na projektu frontendových vizualizací v projektu TRACTOR[1] pro společnost Flowmon pracujeme v 7členném týmu, ve kterém byly na začátku projektu členům rozděleny požadované vizualizace, na kterých samostatně pracujeme. Využíváme repositář na Gitlabu Masarykovy univerzity, na který průběžně ukládáme provedené změny a úpravy. Každých 14 dní se pořádá schůze týmu, na které prezentujeme náš dvoutýdenní pokrok. Při dokončení úseku práce si s dalšími členy týmu vzájemně testujeme vytvořené a upravené komponenty a předáváme si zpětnou vazbu a nápady na změny. V geovizualizačním dashboardu jsou uživateli k dispozici dva předpřipravené soubory. První z nich mi byl poskytnut jako jeden z možných výstupů API a z něj jsem následně vytvořil druhý soubor, u kterého jsem upravil záznamy tak, aby vizualizace demonstrovaly všechny implementované vlastnosti grafů.

Jeden z elementů, který jsem na základě zpětné vazby zbytku týmu změnil, bylo uživatelské menu, konkrétně nástroj pro výběr mapování dimenzí. Původní návrh počítal se select boxem, ve kterém by byly uživateli zobrazeny všechny dostupné dimenze, ze kterých by si mohl jednu zvolit. Nástroj však musí podporovat práci s generickými daty, ve kterých se může nacházet klidně několik desítek dimenzí a pro uživatele by mohlo být zbytečně zdlouhavé požadovanou dimenzi najít (viz. obr. 8.1). Z toho důvodu jsem ve finální implementaci využil našeptávač, který dle uživatelského vstupu filtruje odpovídající dimenze.

Závažná chyba z pohledu nekonzistence databází byla odhalena při testování vstupních dat. Volně dostupný soubor vymodelovaných států ve formátů geoJSON i centroidů totiž neobsahoval všechny státy dle normy ISO 3166-1, ale pouze částečně rozšířený seznam členů Organizace spojených národů. Při kontrole dat bylo navíc odhaleno nejen množství chybějících, ale i chybně pojmenovaných států.

Díky testování jsem upravil i některé elementy za účelem zvýšení přehlednosti. U zobrazovaných hodnot v popupových oknech a tooltipech jsem upravil formát zápisu čísel oddělením násobků tisíce. Dále jsem použil jednu z vlastností Leaflet mapy pro nastavení ohraničení oblasti, ve které se může uživatel pohybovat, díky čemuž v interaktivní mapě nemůže změnit koordináty mimo oblast vizualizací.

[1] TRACTOR je zkratkou TRaffic Analysis and seCuriTy OpeRations for ICS/SCADA.

Obrázek 8.1: Srovnání přehlednosti dvojice přístupů pro výběr požadované dimenze.

8.2 Porada se zaměstnanci Flowmonu

V rámci hlášení postupu práce došlo ke schůzce se zaměstnanci Flowmonu. Aby byly prezentované vizualizace reprezentativní, ale zároveň byly schopné ukázat snadnou odlišitelnost podle hodnoty, rozhodl jsem se používat v dashboardu dva vstupní soubory, mezi kterými je možné přepínat a je přidána také možnost nahrát vlastní soubor ve formátu .json.

V písemném shrnutí odvedené práce zaměstnanec UX týmu Radovan Dvorský konstatoval: „*Prototyp obsahuje všetky základné interakcie, ktoré užívateľ očakáva pri práci s mapou z iných produktov pracujúcich s mapovými podkladmi. Konkrétne ide o priblíženie a oddialenie mapy, ktoré je možné dosiahnuť prostredníctvom samostatného tlačidla na priblíženie/oddialenia, prostredníctvom myši, ale aj pomocou touchpad a gestami na ňom. Na priblíženie reagujú všetky vrstvy očakávaným spôsob, čiže zvýšením resp. znížením množstva informácií na danej vrstve podľa dostupnej plochy. Posun v mape je možný ťahaním a je nazačený zmenou kurzoru myši. Podobne aj reakcia podkladu na kurzor dáva užívateľovi odovzu o možnej akcii na prvku. Z pohľadu interakcií v mape boli požiadavky splnené.*“

Ve finálním zprávě hodnotící dosavadní práci zaměstnanci Flowmonu uvedli, že na základě provedeného testování prototypu považují všechny stanovené cíle projektu za splněné v požadovaném rozsahu.

Jako další postup práce byla navržena integrace grafické úpravy vizualizací na základě interní styleguide. Na základě této úpravy bude Flowmon schopen zahájit první vlnu testování za účelem zjištění srozumitelnosti vizualizací na reprezentativním vzorku uživatelů, kterým je nástroj určen.

Kapitola 9

Závěr

Cílem mojí práce bylo analyzovat současné nástroje společnosti Flowmon a zaměřit se na možnosti využití geografických vizualizací. Pro vytvoření dashboardu bylo nutné nastudovat způsoby monitorování sítě, vizualizace dat a pravidla kognitivního vnímání. Jelikož bylo klíčovou částí mé práce vytvoření návrhu a implementace geografického dashboardu, bylo nutné nastudovat pravidla pro správné používání vizualizací v dashbordech. Z původně jediné geograficky zaměřené vizualizace formou vlajek států v informativní tabulce vznikl samostatný analytický nástroj zaměřený pouze na geovizualizace, který dokáže namapované dimenze dat vizualizovat formou kartogramu, nebo informativních bodů na mapě.

Přestože vytvořené vizualizace zajišťují uživateli možnost analýzy dat z geografického pohledu, je možné využít i další geograficky zaměřené vizualizace, které nabídnou uživateli jiný náhled na danou problematiku. Jedním z možných rozšíření může být například plánovaná mapa spojení, která by umožnila spojovat dvojice komunikujících států a sledovat geografické trendy komunikace se sítí uživatele.

V současné době pracujeme na dokončení úprav vzhledem k diagramům užití pro konkrétní nástroje, které budou využívat geovizualizační dashboard. Výsledky pravidelně konzultujeme se zaměstnanci společnosti Flowmon. Implementovaný dashboard se může v průběhu testování a využívání měnit vzhledem k novým nebo odlišným požadavkům ze strany společnosti nebo uživatelů.

Literatura

[1] Arthur H. Robinson, P. C. M. A. J. K. S. C. G. *Elements of Cartography*. 6. vyd. Wiley, 1995. ISBN 978-0471555797.

[2] Borden Dent, T. H. *Cartography: Thematic Map Design*. McGraw-Hill Education, 2008. ISBN 978-0-072-94382-5.

[3] Bouzian, M. *Modeling and optimization of the quality of customer experience (QoE) of data services on the mobile network. Application to video streaming*. Nice, FR, 2017. PhD thesis. Université Côte d'Azur. Dostupné z: https://webcache.googleusercontent.com/search?q=cache:FhoungGNYa0J:https://www.theses.fr/2017AZUR4061.pdf.

[4] Donolo, R. M. *Contributions to geovisualization for territorial intelligence*. Rome, IT, 2017. PhD thesis. the University of Rome Tor Vergata. Dostupné z: https://tel.archives-ouvertes.fr/tel-01371535/document.

[5] Eckerson, W. W. *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. 2. vyd. Wiley, 2010. ISBN 978-0-470-58983-0.

[6] Few, S. *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, 2006. ISBN 978-0596100162.

[7] Flowmon, a. s. *Přehled řešení*. [cit. 2020-04-20]. Dostupné z: https://www.flowmon.com/cs/prehled-reseni.

[8] Gold, K. *The role of active and passive monitoring in virtual networks* [online]. Únor 2019 [cit. 2020-04-20]. Dostupné z: https://www.exfo.com/en/resources/blog/active-passive-network-monitoring/.

[9] GROUP, A. *SLA* [online]. 2016 [cit. 2020-04-20]. Dostupné z: https://www.sprava-site.eu/sla/.

[10] Harris, R. L. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, 2000. ISBN 978-01-951-3532-9.

[11] Holtz, Y. *Connection map* [online]. 2018 [cit. 2020-04-20]. Dostupné z: https://www.r-graph-gallery.com/connection-map.html.

[12] Hynek, J. *Impact of subjective visual perception on automatic evaluation of dashboards design* [online]. 2019 [cit. 2020-04-20]. Dostupné z: https://www.fit.vut.cz/study/phd-thesis-file/906/906.pdf.

[13] Lonvick, C. *The BSD syslog Protocol* [online]. 2001 [cit. 2020-04-20]. Dostupné z: https://tools.ietf.org/html/rfc3164#page-2.

[14] Matoušek, P. *Síťové aplikace a jejich architektura.* VUTIUM, 2014. ISBN 978-80-214-3766-1.

[15] Ondřej Přibyl, J. P. *Úvod do analýzy dat.* [cit. 2020-04-20]. Dostupné z: https://zolotarev.fd.cvut.cz/static/mamy/mamy-2016-03-slides.pdf.

[16] Postel, J. *INTERNET CONTROL MESSAGE PROTOCOL* [online]. 1981 [cit. 2020-04-20]. Dostupné z: https://tools.ietf.org/html/rfc792.

[17] Sarkar, D. *The Art of Effective Visualization of Multi-dimensional Data* [online]. Leden 2018 [cit. 2020-04-20]. Dostupné z: https://towardsdatascience.com/the-art-of-effective-visualization-of-multi-dimensional-data-6c7202990c57.

[18] Shroder, J. F. *Treatise On Geomorphology.* Academic Press, 2013. ISBN 978-0-08-088522-3.

[19] Ubik, S. *Trendy v monitorování vysokorychlostních počítačových sítí* [online]. Červen 2006 [cit. 2020-04-20]. Dostupné z: https://www.ist-lobster.org/publications/articles/sdel_tech.pdf.

[20] Wertheimer, M. *Experimentelle Studien über das Sehen von Bewegung.* 1912.

Příloha A

Obsah přiloženého paměťového média

- `/`
  - `README.md`
  - `text.pdf` ..... text práce
  - `text_print.pdf` ..... text práce ve verzi pro tisk
  - `text_source/` ..... zdrojové soubory pro sestavení práce
  - `script-data/`
    - `script.js` ..... skript pro kontrolu dodržení ISO 3116-1
    - `data/` ..... vstupní soubory
  - `ReactApp/`
    - `package.json` ..... balíky potřebné pro spuštění aplikace
    - `webpack.config.js` ..... webové konfigurace
    - `src/`
      - `index.html`
      - `index.js`
      - `static/`
        - `world_countries.json` ..... vymodelované polygony
        - `un-country-centroids.json` ..... centroidy pro markery
        - `data/` ..... připravená vstupní data
      - `components/`
        - `Geo2.js` ..... načtení a zpracování dat
        - `CombinedMap.js` ..... společné rozhraní pro vrstvy
        - `common.scss` ..... styly pro elementy leafletu
        - `util/`
          - `TabDOMUtil.js` ..... tvorba DOMu
          - `GeoData.js` ..... parsování dat
        - `layers/`
          - `PointClustersLayer.js` ..... vrstva markery
          - `ChoroplethLayer.js` ..... vrstva choropleth
          - `MapLayer.js` ..... vzor pro vrstvy
          - `Autocomplete.js` ..... komponenta našeptávače
          - `TileLayer.js` ..... vrstva OpenStreetMap