Masarykova univerzita
Fakulta informatiky

Vizualizace logů z aktivních síťových prvků

Bakalářská práce

Lukáš Ferda
2017

Prohlášení

Prohlašuji, že tato bakalářská práce je mým původním autorským dílem, které jsem vypracoval samostatně. Všechny zdroje, prameny a literaturu, které jsem při vypracování používal nebo z nich čerpal, v práci řádně cituji s uvedením úplného odkazu na příslušný zdroj.

Lukáš Ferda

Poděkování

Rád bych poděkoval svému vedoucímu, RNDr. Marku Kumpoštovi, Ph.D., za čas a odbornou pomoc, kterou mi poskytl při psaní této bakalářské práce.

Shrnutí

Bakalářská práce se zabývá vybranými vizualizačními technikami pro zobrazení logů síťové komunikace. Obsahuje popis vybraných správných praktik pro jejich tvorbu za účelem efektivní vizualizace. Dále se zabývá správnou kompozicí různých vizualizačních technik do jednoho celku. V praktické části je pak vytvořena vizualizace reálného logu síťové komunikace a obsahuje popis a zhodnocení použitých nástrojů.

Klíčová slova

Vizualizace, Bezpečnostní vizualizace, Metody vizualizace, Vizualizace vícerozměrných dat, Kompozice vizualizací

Obsah

Úvod 1
1 Efektivní graf 2
1.1 Základní terminologie 2
1.2 Proces vizualizace 3
1.3 Volte správné typy grafů 4
1.4 Redukování nedatového inkoustu 4
1.5 Práce s čísly 6
1.6 Práce s barvou 6
1.7 Interakce 7
1.8 Další studium 7
2 Možnosti vizuální reprezentace dat 9
2.1 Základní podoba 9
2.2 Tabulka 10
2.3 Graf 10
2.4 Další studium 19
3 Kompozice vizualizací do jednoho celku 20
3.1 Základní chyby při kompozici vizualizací 21
3.2. Další studium 23
4 Vizualizace reálného logu 24
4.1 Dostupná data 24
4.2 Vizualizace akcí proxy 24
4.3 Použité nástroje v této práci 33
4.4 Shrnutí 37
5. Závěr 38
Slovníček 39
Literatura 40

Úvod

Síťová zařízení jsou zdrojem velkého množství dat v podobě logů síťové komunikace. Objem těchto dat se neustále zvyšuje a s tím vyvstává otázka, jak z těchto dat získat zajímavé informace. V těchto datech můžeme nacházet různé zajímavé souvislosti jako například o stavu sítě, útocích na síť, na skupiny počítačů, jednotlivé počítače, ale také podezřelé aktivity uživatelů. Jedním způsobem, jak analyzovat taková data, je jejich vizualizace. Vizualizací lze zobrazit obrovské množství dat pomocí jednoduššího grafického znázornění. Existuje mnoho vizualizačních technik, některé umožňují vizualizovat pouze jednu proměnnou a jiné jich mohou vizualizovat několik najednou. Dále se také mohou lišit například množstvím zpracovatelných informací do této vizualizace. Pro kvalitní splnění účelu musí vizualizace splňovat jisté podmínky, kterými se také zabývám v této práci.

Cílem teoretické části této práce je představit některé vizualizační techniky, vybrané praktiky jejich správné tvorby a odkázat na další zdroje informací k tématu. V první kapitole jsou uvedeny pojmy použité v této práci a jejich definice a definice typů dat. Dále se kapitola zabývá pojmem efektivní graf. Popisuje správný proces, jak takový graf vytvářet a dále uvádí vybrané správné praktiky tvorby efektivních grafů. Konkrétně popisuje redukování nedatového inkoustu, volbu správných grafů, práci s barvou, čísly a interakci s grafem. Na závěr je uveden odkaz na další zdroje k tématu. Druhá kapitola se věnuje vybraným možnostem vizuální reprezentace dat. Je zde uveden jeden řádek reálného logu, dále se věnuji tabulce a některým grafům, které zde krátce popisuji diskutuji možnost použití a uvádím ukázku. Třetí kapitola se věnuje kompozici více vizualizací do jednoho celku. Kapitola obsahuje výčet správných vlastností kompozice a výčet častých chyb při kompozici vizualizací.

Cílem praktické části je vytvořit vizualizaci reálného logu síťové komunikace. Čtvrtá kapitola obsahuje vizualizaci reálného logu a představuje proces tvorby vizualizace. Konkrétně je zde uvedena vizualizace zamítnuté komunikace. Kapitola obsahuje sloupcové grafy pro přehled celé komunikace a dále detailnější vizualizace pomocí grafů propojených uzlů, tabulky, grafu paralelních souřadnic a hierarchických síťových map. Každý z grafů je doplněn o vysvětlení a případně další návrhy na zlepšení. Dále obsahuje výčet použitých vizualizačních nástrojů v této práci, jejich popis a krátké zhodnocení výhod a nevýhod nástrojů.

V závěru práce jsou shrnuty výsledky práce. Rovněž je zde uvedena úvaha nad dalšími možnostmi rozšíření této práce, slovníček se zdroji překladů použitých názvů a seznam literatury.

1 Efektivní graf

Tato kapitola obsahuje definici základní terminologie, některé vybrané správné principy, které je vhodné dodržovat při tvorbě grafů a poukazuje na potřebné funkce vizualizačních nástrojů. Cílem je představit úvod do tématu tvorby efektivních grafů. Nejprve je ale nutné se zamyslet nad tím, co je to vlastně efektivní graf. Například graf vytvářený pro širokou veřejnost zobrazovaný v nějakém periodiku si pravděpodobně bude klást za cíl zaujmout pozornost čtenáře. Bude se zaměřovat na emoce, estetiku zapamatovatelnost atd. Naším primárním cílem v oblasti bezpečnostní vizualizace by ale mělo být zobrazení dat, které je objektivní a co nejrychleji pochopitelné pro uživatele. Chceme, aby zobrazení podpořilo jeho použití, aby uživatel byl schopný rychle reagovat na nastalé události, mohl lehce popsat události vyššímu managementu atd.

1.1 Základní terminologie

Níže popisuji používané pojmy v dalších částech této práce. Protože se v celé práci většinou odkazuji na zdroje v anglickém jazyce a také většina grafů v této práci je vytvořena v programu bez českého uživatelského rozhraní, používám při výčtu grafů v druhé kapitole anglický i český název. V dalším textu používám pouze české názvy. Slovníček použitých výrazů včetně poznámek k překladu naleznete na konci této práce.

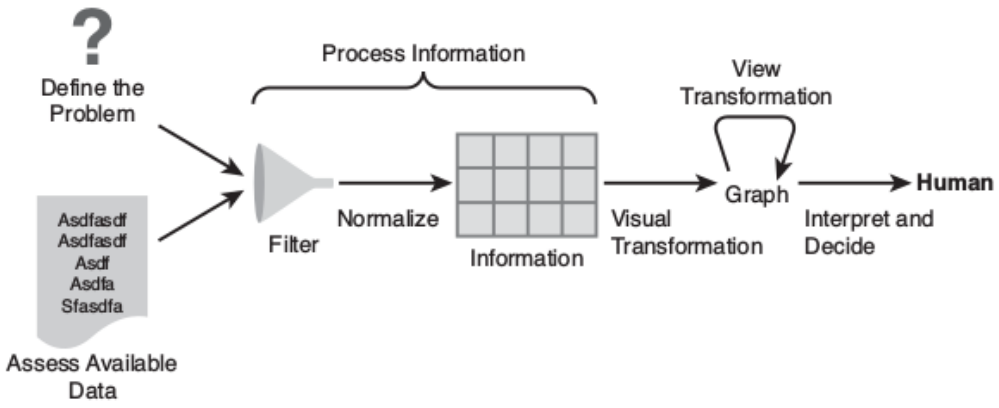
Pod pojmem **vizualizace** můžeme obecně chápat jakékoli zobrazení informace, tak aby byla čitelná pro lidské oko. V této práci je myšleno vždy zobrazení pomocí grafu popřípadě tabulky nebo čísla. **Graf** představuje grafické znázornění pro numerické hodnoty, vztahy, postupy ap. **Uživatel** v kontextu této práce je člověk pracující s programem pro vytváření grafů, tabulek a jejich kompozice do jednoho celku i člověk, který analyzuje logy pomocí vytvořených vizualizací. **Případ použití** představuje záměr za jakým je vizualizace vytvořena. **Interakcí** je myšlena komunikace mezi uživatelem a vizualizačním nástrojem za účelem tvorby a úprav vizualizace. **Bezpečnostní vizualizací** je myšlena vizualizace za účelem odhalení útoků nebo hrozeb pro počítačovou síť.

Abychom mohli lépe popsat různé způsoby vizualizace dat, je nutné nejprve zmínit, že existují různé kategorie dat. Různé metody vizualizace pak jsou vhodné pro různé typy dat. Data můžeme rozdělit do čtyřech základních kategorií podle jejich významu. **Nominální** data můžeme rozdělit do různých kategorií, tedy u dvou hodnot můžeme pouze rozlišovat, zda jsou stejné. Například typ služby nebo číslo portu. **Ordinální** uspořadatelná, lze zavést pořadí tedy u každých dvou hodnot lze uvažovat, zda jsou různé a jestli je jedna větší (kvalitnější) než druhá. Například velikost rizika. **Intervalová** u dvou hodnot můžeme uvažovat, zda jsou stejné nebo různé, o kolik se liší a také lze uvažovat

o rozdílu mezi obměnami těchto hodnot. Například velikost paketů. **Poměrová** zvané také jako **kardinální,** u dvou hodnot lze určit zda jsou různé, stanovit jejich pořadí a přesně změřit jejich rozdíl. Je jasně definovaná nula. Například délka trvání komunikace mezi klientem a serverem.
Další informace o kategorizaci dat lze nalézt v [1] na straně 66 nebo v [2] na straně 2 až 4.

1.2 Proces vizualizace informace

Proces tvorby efektivní vizualizace/grafu se skládá z několika kroků, které lze popsat pomocí následujícího diagramu.



Obrázek 1.1: Schéma vizualizace informace [1]

Nejprve je tedy potřeba začít definováním problému a dohledáním všech relevantních dat k tomuto problému. Které logy a informace jsou dostupné? Které jsou použitelné pro řešení problému? Logy můžou a pravděpodobně budou obsahovat mnohem víc informací a proto následuje další krok filtrování dat. Filtrovaná data mohou být z různých zdrojů (síťová zařízení, operační systém, servery, výstupy programů, atd.) a mohou mít tedy i rozdílný formát. Data se musí sjednotit a uložit například do tabulky nebo jiného formátu pro generování grafu. Dále následuje typicky iterativní proces uprav grafu. Provádí se úpravy jako změny měřítek grafu, přibližování, oddalovaní, různé přeskupování objektů v grafu, obarvení a další úpravy k zvýraznění podstatných informací. U výsledného grafu pak dochází k interpretování a rozhodování.

1.3 Volte správné typy grafů

Volba správného grafu je nejdůležitější pro dosažení efektivní vizualizace dat. V druhé kapitole této práce je popsáno jenom několik možností, ale ve skutečnosti jich je mnohem více a další budou přibývat. Jak tedy zvolit správný typ grafu? Nejprve je potřeba položit správné otázky.

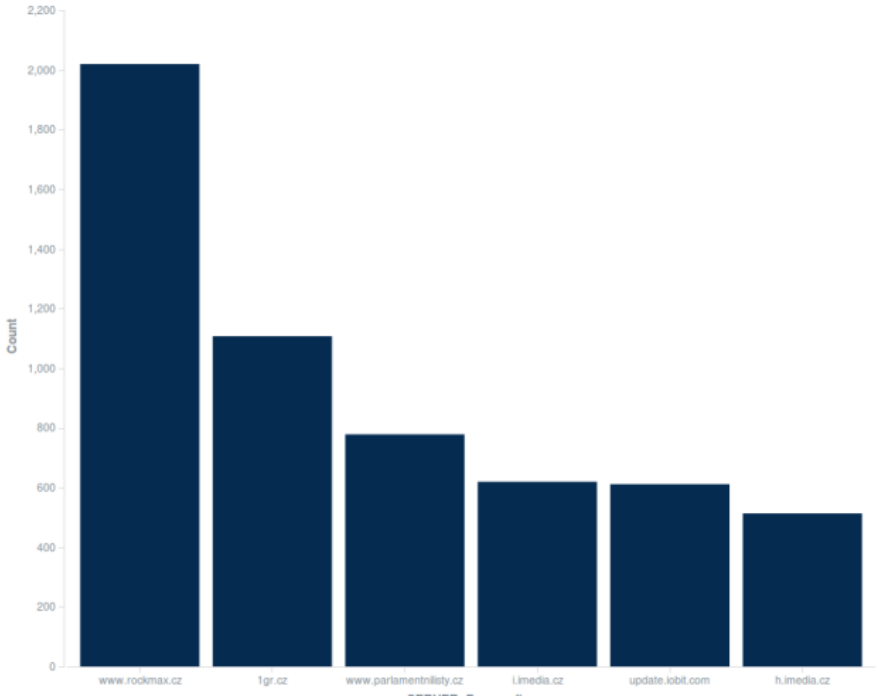
Co chci grafem/kompozicí grafů říct? Jaký má vizualizace účel?
Bude stačit jeden graf? Nebude lepší použít kompozici několika grafů?
Jaké mám data pro splnění tohoto účelu? Jaký je typ těchto dat?
Kolik různých dimenzí dat chci zobrazit?
Kdo bude mojí vizualizaci používat?
Atd.

Není třeba odpovědět na všechny a lze vymyslet i další otázky, ale účelem je si utřídit představy o účelu, datech a možnostech vizualizace dříve než se uživatel pro nějakou vizualizaci rozhodne. V práci [1] je uvedena na straně 110 – 113 přehledná tabulka se srovnáním vybraných grafů i s příklady použití v bezpečnostní vizualizaci.

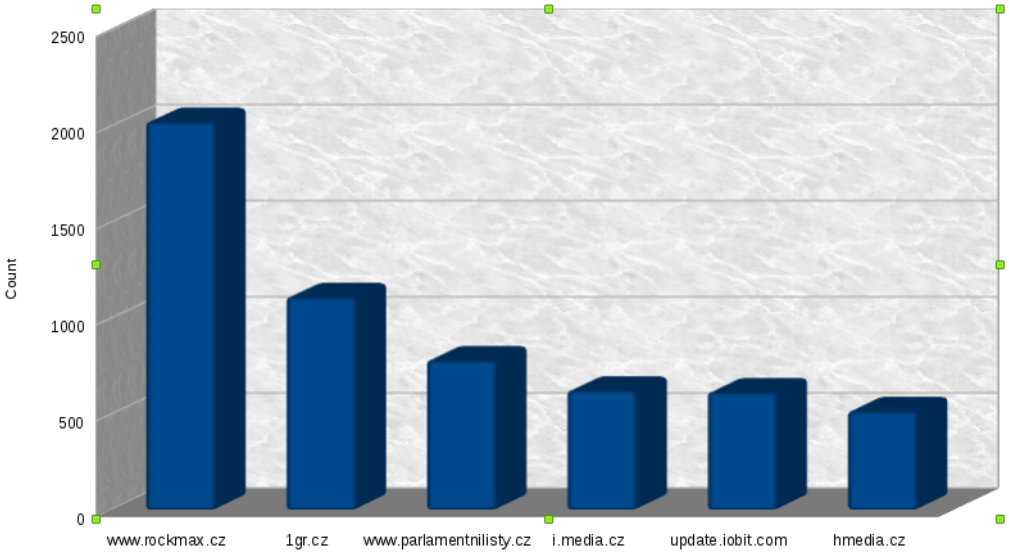
1.4 Redukování nedatového inkoustu

Redukování nedatového inkoustu lze považovat spolu se správnou volbou grafu za úplný základ správné tvorby grafu. Edward Tufte definoval pojem Data-ink Ratio, kde se zabývá poměrem inkoustu použitého pro zobrazení dat a inkoustu použitého pro zobrazení zbytku grafu [3]. V dnešní době lze pojem inkoust ekvivalentně nahradit pixely. Správně vytvořený graf by měl dodržovat princip redukování nedatového inkoustu. Někde též zvané jako „chartjunk“. Grafu nepomůže obrázek na pozadí, nadbytečné informace v grafu, nebo okolo něj ani různé vizuální efekty. Každý takový zbytečný prvek jej činí pouze hůře čitelným a stojí v konečném důsledku uživatele víc času k jeho pochopení.

Dále je uvedeno srovnání dvou grafů vytvořených ze stejných dat a zobrazujících stejnou informaci. Prostá informace o rozložení hodnot je lépe a rychleji komunikována s uživatelem pomocí prvního grafu 1.2.



Obrázek 1.2: Graf se zbytečnými grafickými prvky



Obrázek 1.3: Graf se zbytečnými grafickými prvky

Graf 1.3 je plný přebytečných vizuálních prvků, které pouze odvádějí pozornost od důležité informace.

1.5 Práce s čísly

Při zapisování čísel do tabulek nebo například při vytváření legendy nějakého grafu je vhodné dbát i na správné zapisování čísel. Existují různé normy a postupy, které se tímto problémem zabývají jako například norma ČSN 01 6910.

Takové normy se mohou lišit v různých státech. Například někde se pro oddělení celé a desetinné části čísla používá čárka, ale v anglicky mluvících zemích se čárka používá pro oddělení řádů čísla a pro oddělení celé a desetinné části se používá znak tečky.

Zapisování velkých a malých čísel. Při zapisování velkých čísel přichází v úvahu jejich zápis pomocí vědecké notace. Například číslo 3 720 000 můžeme zapsat jako $3{,}72 \times 10^6$. Pokud se v sloupci tabulky vyskytují velká čísla můžeme do záhlaví sloupce uvést informaci o řádu tím zkrátíme zápis všech čísel ve sloupci [4].

Zapisování čísel do tabulky. Na tabulky máme podobné požadavky jako na grafy. Tedy správná tabulka má být přehledná, úsporná a dobře čitelná. Všechna čísla by se měla uvádět na stejný počet platných číslic. Čísla je pro lepší čitelnost vhodné zarovnat tak, aby číslice stejného řádu byly vždy pod sebou. Uvádět čísla v co nejkratší podobě. Pokud nehrozí ztráta podstatných informací, tak lze uvažovat o vhodném zaokrouhlení čísel [4].

1.6 Práce s barvou

Používejte maximálně 6 – 12 barev. Počet použitých barev by určitě neměl být větší než 12. Důvodem pro omezování počtu barev je, že pokud je graf například složitý a barevné symboly malé může docházet k jejich záměně, nebo graf může ztrácet na přehlednosti. Naopak pokud v černobílém grafu jednu jeho část obarvím například červeně, tak přirozeně bude směrovat pozornost uživatele právě do tohoto obarveného místa [5].

Používejte barvy, které jsou navzájem vizuálně od sebe vzdálené. Tato vzdálenost potom podporuje rozlišování jednotlivých entit a jejich vyhledávání v grafu [5]. Pokud chci použít ve vizualizaci víc barev pro zvýraznění informací je vhodné volit barvy tak, aby byly i dobře pojmenovatelné. Pokud nad vizualizací má diskutovat více uživatelů usnadněte jim diskuzi použitím základních barev. Je jednodušší pojmenovat a mluvit o modré, červené a zelené, než vymýšlet správné pojmenování třech různých odstínů modré barvy.

V západní kultuře červená barva reprezentuje nebezpečí, teplo nebo zastavení. Zelená barva zase představuje pohyb a bezpečí. Modrá reprezentuje zimu atd. Proto by použití červené barvy pro význam běžné hodnoty a zelené barvy pro nějakou nebezpečnou anomálii bylo zbytečně matoucí. Situace je ale ještě komplikovanější uvážíme-li rozdíly ve vnímání barev mezi jednotlivými kulturami. Například v Číně červená barva reprezentuje štěstí a obnovu [5].

1.7 Interakce

Při tvorbě efektivního grafu potřebuje uživatel s grafem interagovat, protože v logu je typicky velmi mnoho informací a množství informací se může velmi lišit v závislosti například na podsíti. Dále pak mohou být velké rozdíly mezi jednotlivými hodnotami, nebo může být logu tisíce záznamů s hodnotou Accepted a pouze jednotky či desítky záznamů Rejected ap. Postupy jako například filtrování obarvení vizualizace aplikované na log z jedné sítě nemusí být použitelné na log ze sítě jiné a výsledné grafy se mohou velmi lišit a některé informace v něm mohou být velmi špatně čitelné nebo vůbec. V tomto případě uživatel potřebuje interagovat s grafem pomocí změn měřítek, přiblížení, oddálení, různé přeskupování, filtrování, obarvení atd. Tyto úpravy potom kromě vizualizačního procesu představují i analýzu samotných dat. Cílem z pohledu interakce je tedy intuitivní rozhraní vizualizačního nástroje s rychlou odezvou a možností okamžitého zobrazení změn.

1.8 Další studium

V [1] jsou v první kapitole uvedeny informace o vizualizaci, jejích výhodách, vlastnostech a některé správné principy její tvorby jako zvýrazňování vyjímek, zobrazení srovnání, popisování dat, zobrazení kauzality a dalších.

V práci [5] se autor zabývá samotným vizuálním vnímáním člověka, jak najít v obraze rychle informaci, jak ji zvýraznit, jak pracovat s barvou a další.

V práci [6] se autoři zabývají mimo jiné i představením a popisem rozhraní pro uživatele, které mu umožňuje nastavovat parametry jako časové okno, počet zobrazených grafů a informací v nich a aplikovat příkaz Grep.

V [7] autor popisuje Gestaltovy principy (jednoduchost, blízkost, dobrého průběhu, uzavření, preference menšího tvaru, ohraničení a symetrie) aplikované v designu.

Pracoval jsem s více než dvaceti zdroji informací a na základě svých zkušeností můžu doporučit pro získání základních znalostí o efektivní vizualizaci logů síťových zařízení, začít knihou **Raffaela Martyho** *Applied security visualization*. Pro získání informací o kompozici vizualizací a práci s nimi doporučuji knihu **Shadana Malika** *Enterprise dashboards: design and best practices for IT.* Dále doporučuji knihu **Colina Warea** *Visual thinking for design*, která se zabývá vizuálním vnímáním člověka a dává rady návrhářům jak se na tyto lidské vlastnosti zaměřit při prezentování informací.

2 Možnosti vizuální reprezentace dat

Počet zařízení připojených do sítě komunikujících nějakým protokolem roste a spolu s tím roste i množství a různorodost těchto informací uložených v záznamech síťové komunikace. Tato kapitola se zabývá vybranými metodami vizuální reprezentace těchto dat. Některé jsou široce používané jako sloupcový nebo koláčový graf a jiné jsou méně známé a používané jako například paralelní souřadnice nebo binary rainfall. Zvolit správnou metodu vizuální reprezentace dat je tedy netriviální otázka, může záviset na typu dat, počtu dimenzí které chceme zobrazit, účelu vizualizace, zda chceme detailní nebo obecný pohled.

2.1 Základní podoba

Logy ve své základní podobě obsahují nejvíce podrobností o komunikaci a poskytují nejdetailnější pohled na jednotlivé položky.

> Jan 4 00:53:00 cw http_proxy[5388.1417.1]: HTTP-888-I PROXY-EVENT PROTOCOL=TCP CLIENT=ad1.tns.int CLIENT-IP=192.168.144.26 CLIENT-PORT=62905 SERVER=crl.microsoft.com SERVER-IP=147.229.255.80 SERVER-PORT=80 SERVER-PORT-NAME='http' USER='AD1$' BYTES-CIN=202 BYTES-COUT=1473 BYTES-SIN=226 BYTES-SOUT=253 DURATION=0.031311 STATUS=ACCEPTED RESULT=OK RULE='default_profile' USER-GROUPS=<NULL> METHOD=GET URI='http://crl.microsoft.com/pki/crl/products/tspca.crl' CONTENT-TYPE=<NULL> STATUS-CODE=304 VIRUS-STATUS=0 PAGE-VIEW=1 BYPASS=0 CATEGORIES='it-services-internet' REFERER=<NULL>

Příklad jednoho řádku z logu síťové komunikace, které byly použity pro zpracování této práce.

Jednou z budoucích výzev bude stanovení jednotného formátu pro logy z různých zdrojů. Při psaní této práce jsem sice používal pouze jeden zdroj logu, ale i tak jsem se setkal s nepříjemnými problémy při parsování. Například jsem ve výstupním souboru pro oddělení hodnot používal znak čárky (CSV formát) a čárka se velmi zřídka, ale přece někdy objevila i v parsované hodnotě a v tomto případě zkazila formátování takového řádku. Velmi nepříjemné je, že problém při parsování může být těžko rozeznatelný a může se projevit až při detailnějších vizualizacích nebo při vizualizaci logu z jiného dne ap. Bude tedy potřeba jasně stanovit formát klíče a hodnoty, jaké znaky se kde mohou a respektive nesmí vyskytovat, jakým znakem jsou hodnoty odděleny atd.

2.2 Tabulka

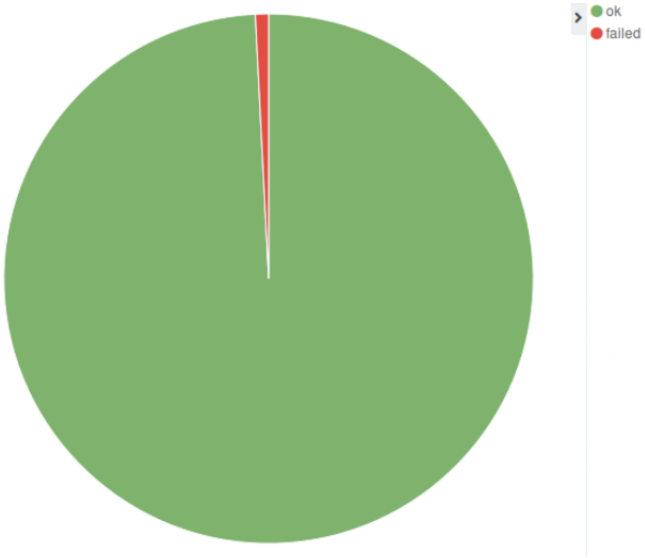
Jedním ze způsobů zobrazení informací z logů síťové komunikace je umístění výsledku(ů) do tabulky. Všeobecně platí, že tabulka se hůře čte než grafické znázornění a proto tam, kde je to možné se číselné údaje nahrazují grafy [8]. Tabulka je tvořena hlavičkou, sloupci a řádky. Názvy sloupců v hlavičce tabulky popisují jejich obsah. Preferovanou variantou je psát názvy vodorovně, zarovnané na střed a začínají velkým písmenem. Šířka sloupců se volí podle nejširšího údaje. Textové položky jsou zarovnávány vlevo. Číselné údaje se zarovnávají na střed sloupce [4]. Úpravou tabulek se zabývá například česká norma ČSN 01 6910.

2.3 Graf

Graf představuje nejlepší a nejjednodušší způsob jak vizualizovat data. Ve dvojrozměrném prostoru papíru či obrazovky zobrazuje informaci z logů síťové komunikace. V tomto dvourozněrném zobrazení můžeme vytvářet i grafy zobrazující několik dimenzí najednou nebo se zaměřit pouze na jednu [1].

Níže je uveden výběr několika typů grafů. Některé jsem vytvářel při zpracování praktické části této práce. Dále je výčet doplněn o další typy, aby byla lépe vidět jejich rozmanitost. Nejedná se zdaleka o všechny typy a není ani cílem uvádět zde všechny. Uvedené grafy lze rozdělit do dvou kategorií a to na základní mezi které patří kruhový, prstencový, sloupcový a spojnicový graf. Tyto grafy jsou dobře známé a velmi rozšířené. Typicky se používají pro zobrazení jedné dimenze z dat. Druhou skupinou jsou grafy složené z více prvků a zobrazující víc dimenzí najednou. Sem patří prstencový graf, složený sloupcový, graf propojených uzlů, hierarchická síťová mapa a graf paralelních souřadnic. Především poslední dvě zmíněné vizualizace nepatří mezi široce rozšířené a při jejich použití musíme uvažovat i o tom zda jejich uživatel v nich dovede číst informace.

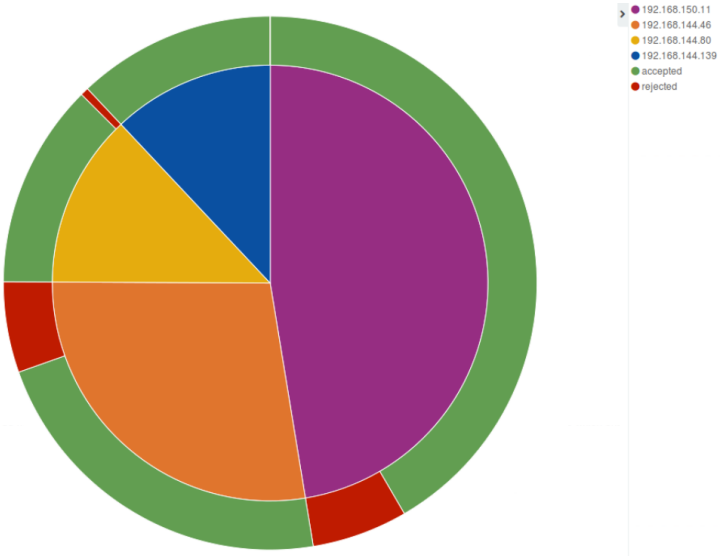
Kruhové grafy *(pie chart)*, někde nazývané výsečové nebo koláčové, se typicky používají pro zobrazení podílu z celku. Je vhodný pro nominální typ dat. Jedná se o základní typ grafu. Tento typ grafu se pro bezpečnostní vizualizaci spíš nedoporučuje a je lepší zvážit použití jiných metod, protože kruhový graf umožňuje znázornit pouze omezený počet hodnot [1]. Pokud se zobrazuje víc hodnot ztrácí graf na přehlednosti. Dále pak nemusí být vhodný pokud obsahuje víc podobných hodnot, které by činily jejich srovnání obtížným.



Plocha výseče grafu 2.1 představuje procentuální podíl komunikace, která skončila s výsledkem Ok/Failed.

Graf 2.1: Kruhový graf celé komunikace

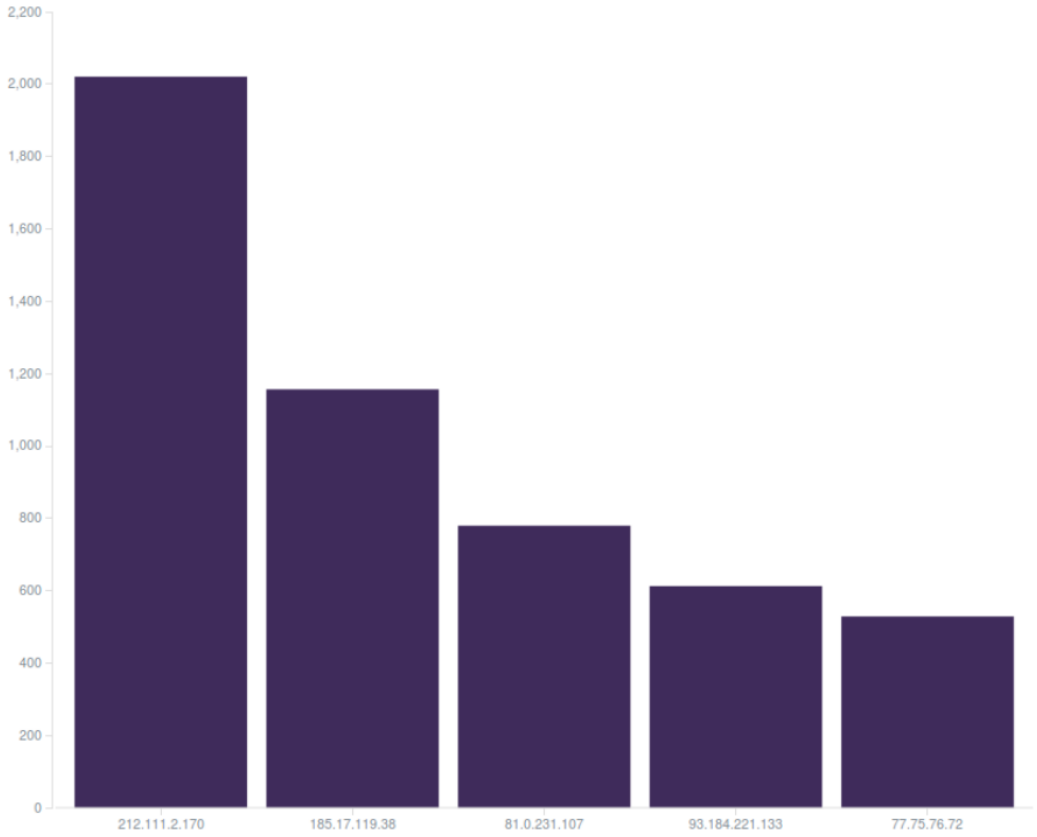
Prstencový graf *(Stacked pie chart)* vychází z výsečového grafu. Například jednotlivé oblasti výsečového grafu můžou být rozděleny na další podoblasti. Je vhodný pro nominální typ dat. Výhodou oproti kruhovému grafu je možnost zobrazení další dimenze dat v grafu. S rostoucím počtem zobrazovaných položek ale klesá jeho přehlednost.



Plocha výsečí v grafu 2.2 představuje procentuální podíl komunikace klientů. Výseče jsou dále rozděleny na dvě části podle výsledku Ok/Failed.

Graf 2.2: Prstencový graf komunikace čtyř klientů

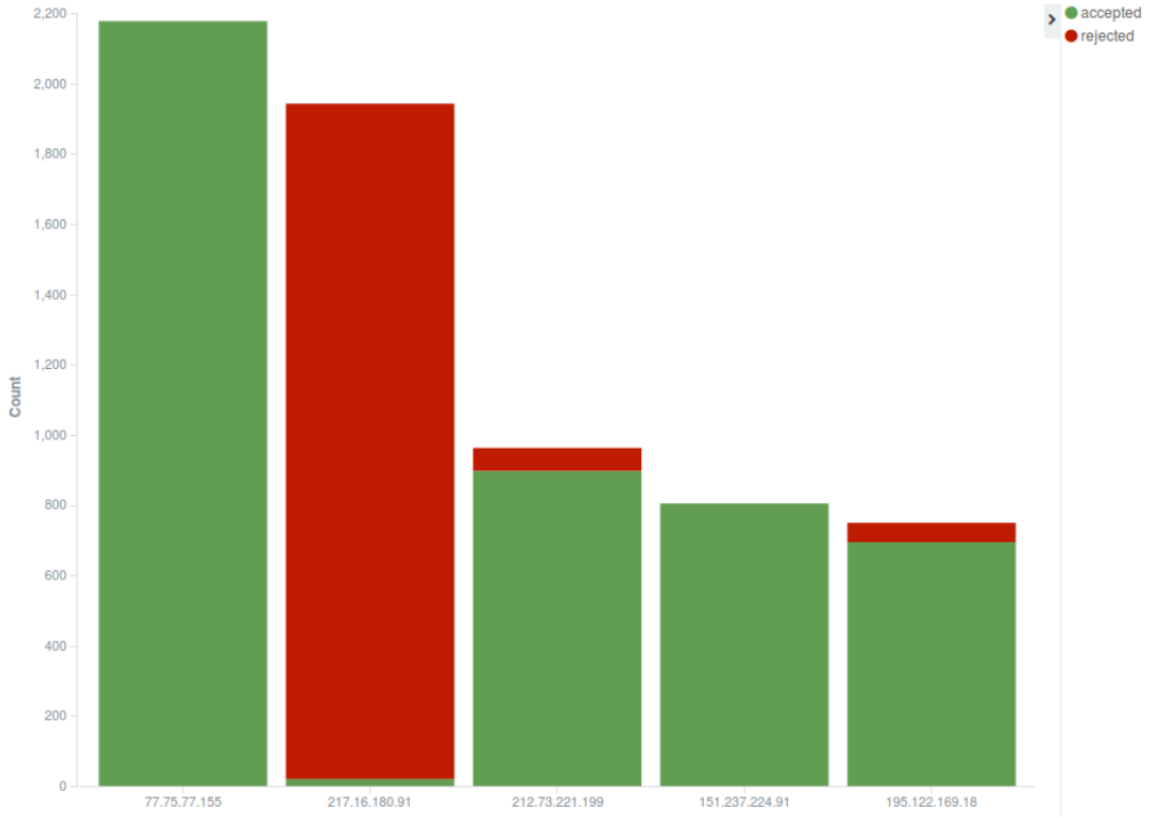
Sloupcové grafy *(Bar chart)* nejčastěji zobrazují počet individuálních hodnot jedné dimenze v datech. Dále jsou velmi intuitivní. Obvykle na vodorovné ose zobrazují jednotlivé kategorie a na svislé ose potom jejich hodnotu. Existují i v prostorových variantách nebo tzv. „Skládané“, kde je každý sloupeček tvořený složením více hodnot. Jsou vhodné pro zobrazení nominálních dat. Umožňují jednoduché intuitivní porovnávání hodnot mezi sebou na základě velikosti sloupců. Na rozdíl od výsečového grafu umožňuje volit různá měřítka na svislé ose (např. lineární, logaritmické). Při zobrazení více hodnot ztrácí graf na přehlednosti.



Graf 2.3: Sloupcový graf pěti nejnavštěvovanějších IP adres

V tomto případě je použita funkce Count, která počítá výskyt hodnot. Dále lze použít i další funkce jako například sumu nebo průměr pro intervalová nebo kardinální data.

Složený sloupcový graf *(Stacked bar chart)* zobrazuje na rozdíl od prostého sloupcového grafu další dimenzi dimenzi v datech pomocí rozdělení sloupců do dalších oblastí.

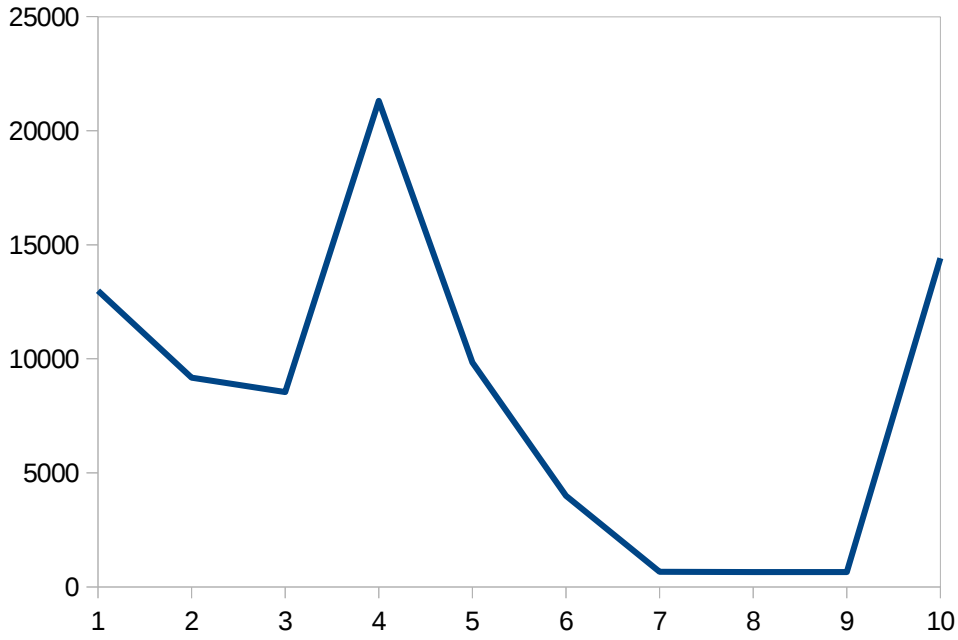


Graf 2.4: Složený sloupcový graf komunikace s pěti servery

V tomto případě je plocha, která zobrazuje množství komunikace s jednotlivými servery rozdělena podle toho, zda byla povolena nebo zamítnuta. V grafu lze vidět anomálii kde komunikace se serverem 217.16.180.91 je z velké většiny zamítnuta. V případě této vizualizace může nastat problém se zobrazením malých počtů zamítnutých spojení, které nemusí být vidět a řešení s logaritmickým měřítkem uvnitř jednotlivých sloupců zatímco na celé sloupce aplikované není by mohlo být pro uživatele matoucí.

Spojnicový graf *(Line chart)* má na vodorovné ose kategorie rovnoměrně rozložené stejně jako jsou rovnoměrně rozložené hodnoty na svislé ose. Spojnicový graf je vhodný především pro zobrazení intervalových dat, kde představuje lepší řešení než graf sloupcový. Dále je to vhodná technika pro vyobrazení trendů jako růst či pokles hodnot v čase a umožňuje zobrazit plynulý průběh hodnoty v čase. Stejně jako u sloupcového grafu můžeme používat různé funkce pro data jako počet, sumu nebo průměr.

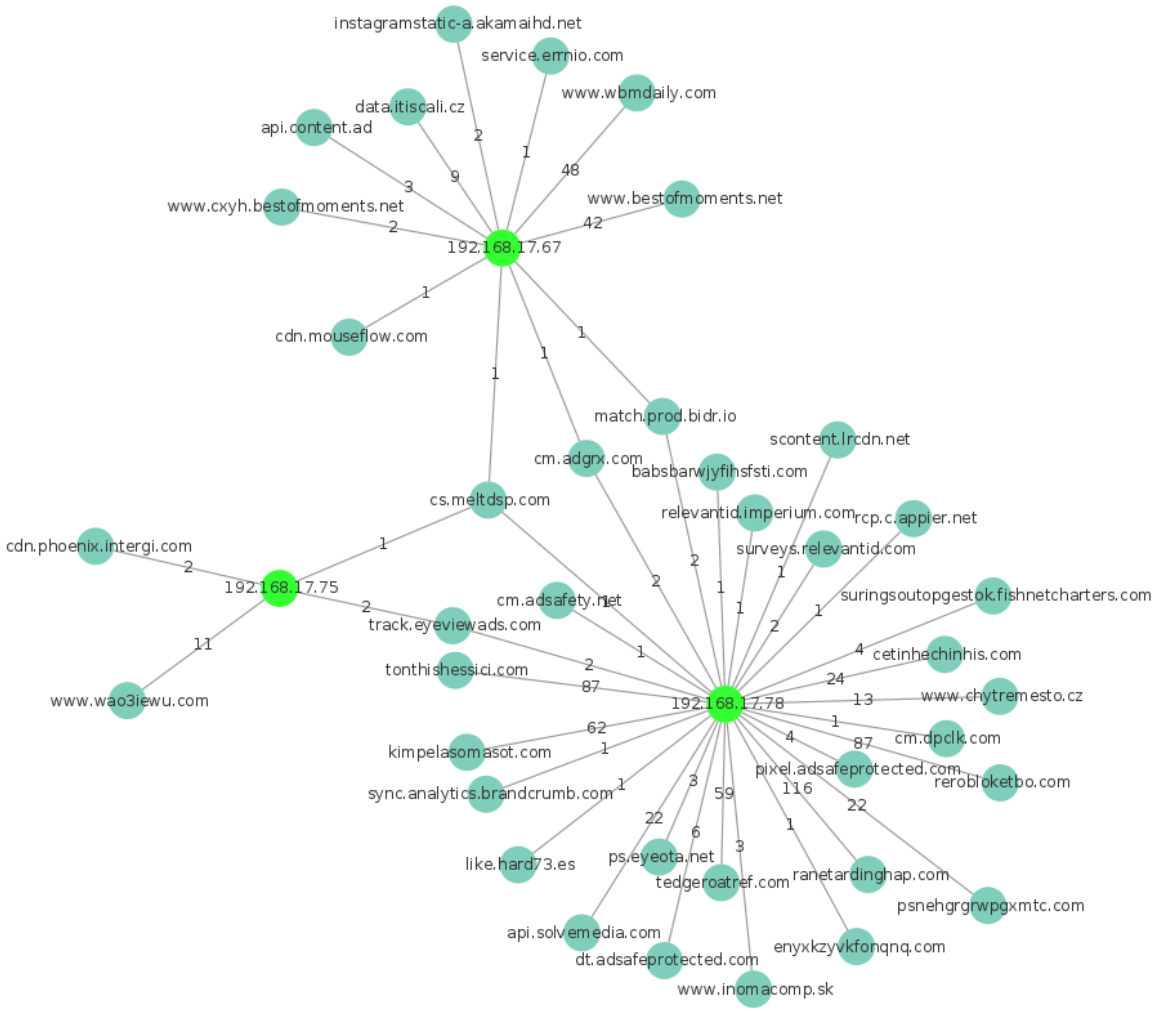
Spojnicový graf má také svojí složenou variantu, kde jsou další dimenze dat zobrazeny pomocí dalších křivek v něm. Příkladem použití by mohl být například vývoj množství komunikace v čase pro některé klienty.



Graf 2.5: Spojnicový graf množství komunikace

Spojnicový graf 2.5 zobrazuje vývoj množství komunikace v deseti dnech. V tomto případě graf také zobrazuje rostoucí a klesající trendy v čase.

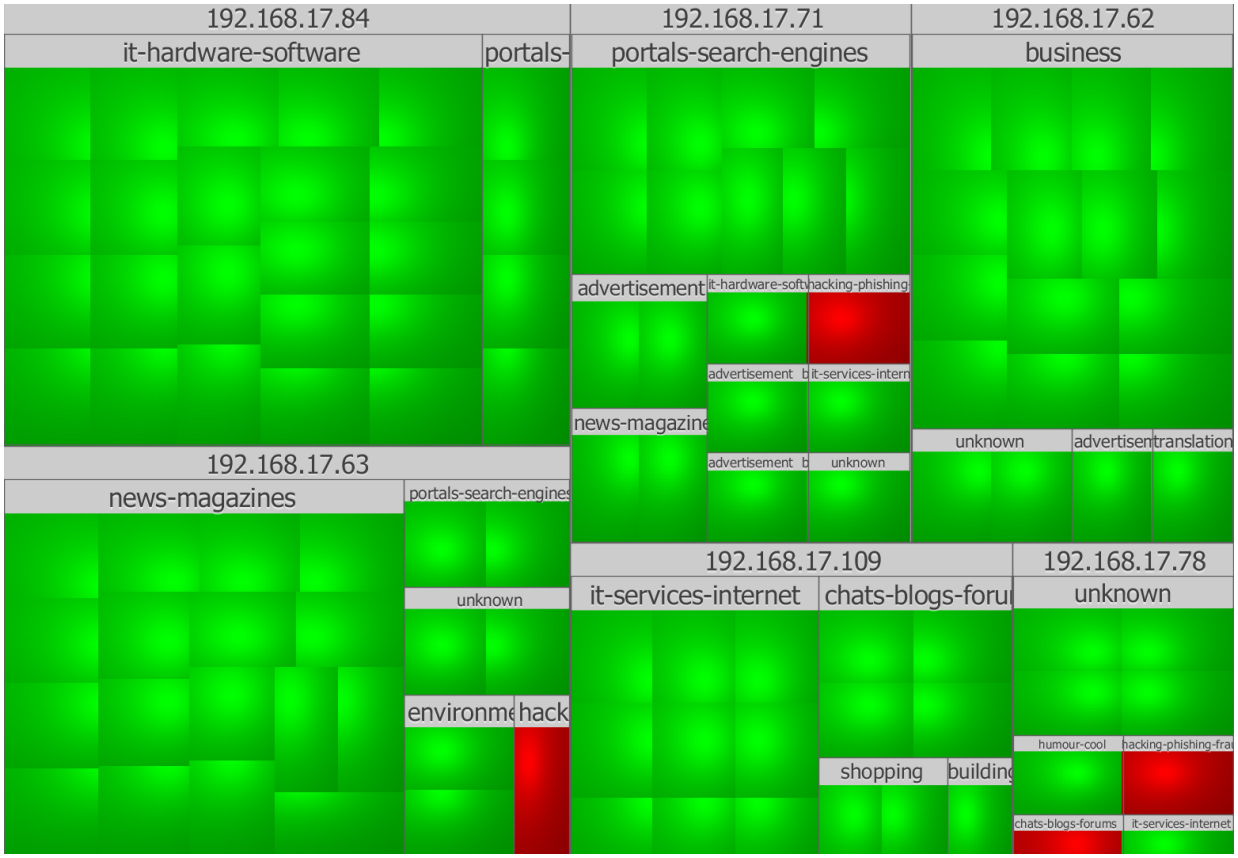
Graf propojených uzlů *(Link graph)* se skládá z uzlů a hran propojujících uzly. Může být orientovaný i neorientovaný. Je vhodný pro všechny druhy dat. Graf propojených uzlů je například vhodný pro zobrazení komunikace mezi uzly, kde naznačuje vztah kdo s kým komunikuje. S rostoucím počtem uzlů a hran vyvstává problém, jak zobrazit graf, aby byl stále přehledný a uzly je třeba filtrovat nebo agregovat do jednoho celku. Další dimenze lze zobrazit například vhodným obarvením grafu nebo ohodnocením jednotlivých hran.



Graf 2.6: Linkový graf komunikace tří uzlů vnitřní sítě

Graf 2.6 zobrazuje komunikaci tří uzlů se servery. Ohodnocení hran udává počet spojení se servery.

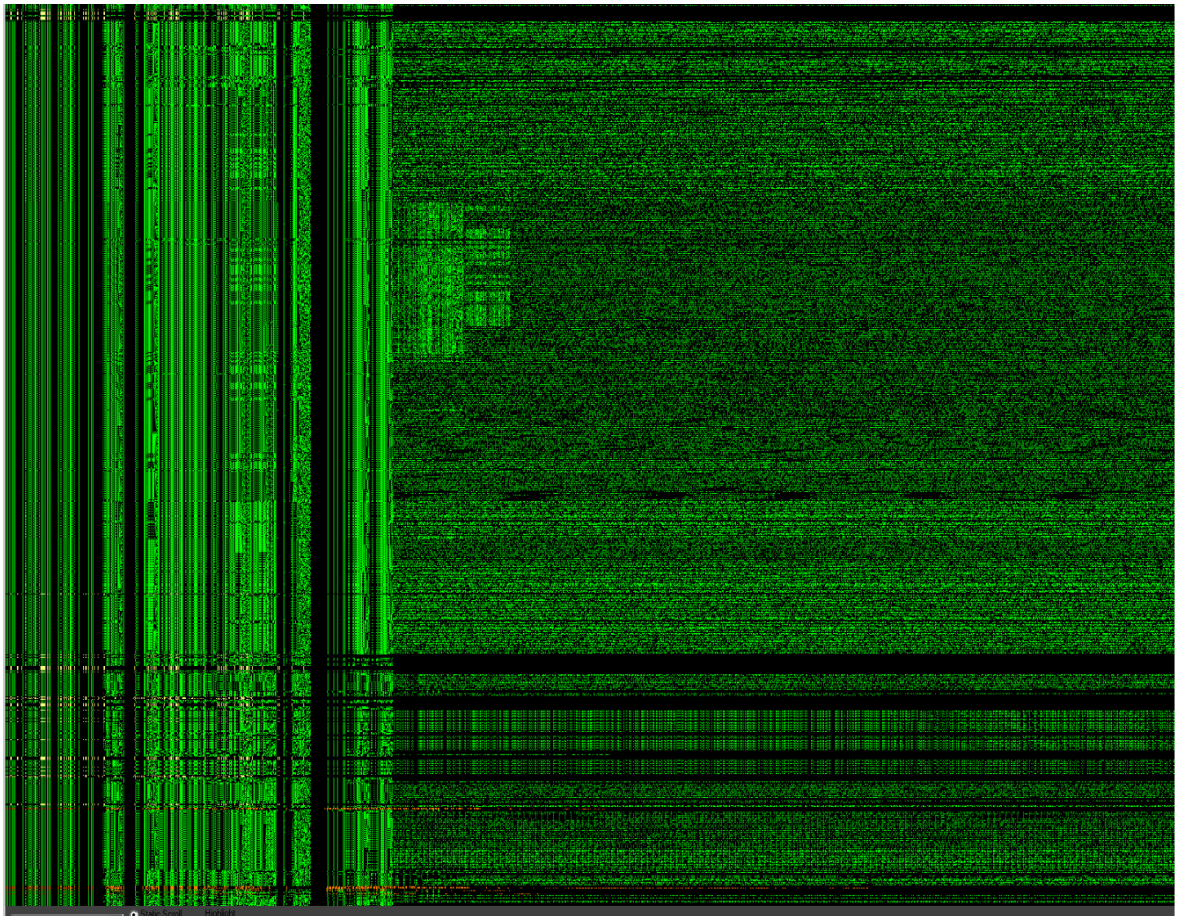
Hierarchická síťová mapa *(Treemap)* jejím principem je zobrazení pomocí čtvercových nebo obdélníkových oblastí. Jejich velikost nebo barva potom slouží k zobrazení informace. Například velikost oblasti může udávat velikost hodnoty a barva může odrážet pozitivní či negativní trend. Pokud nějaká oblast v grafu obsahuje podoblasti, tak je lze seskupit nebo oblast rozdělit na další čtverce atd. Hierarchická síťová mapa nabízí zobrazení jakýkoliv typů dat, je vhodná pro objemná data. Je možné v nich dobře rozpoznávat vzory a umožňují zobrazovat hierarchické nebo stromové struktury datech.



Graf 2.7: Hierarchická síťová mapa komunikace počítačů v síti

Velikost jednotlivých oblastí v grafu 2.7 označuje množství komunikace. Plocha jednotlivých uzlů je dále rozdělena podle kategorií, kde opět velikost odpovídá množství komunikace. Červená barva zvýrazňuje komunikaci s výsledkem Rejected.

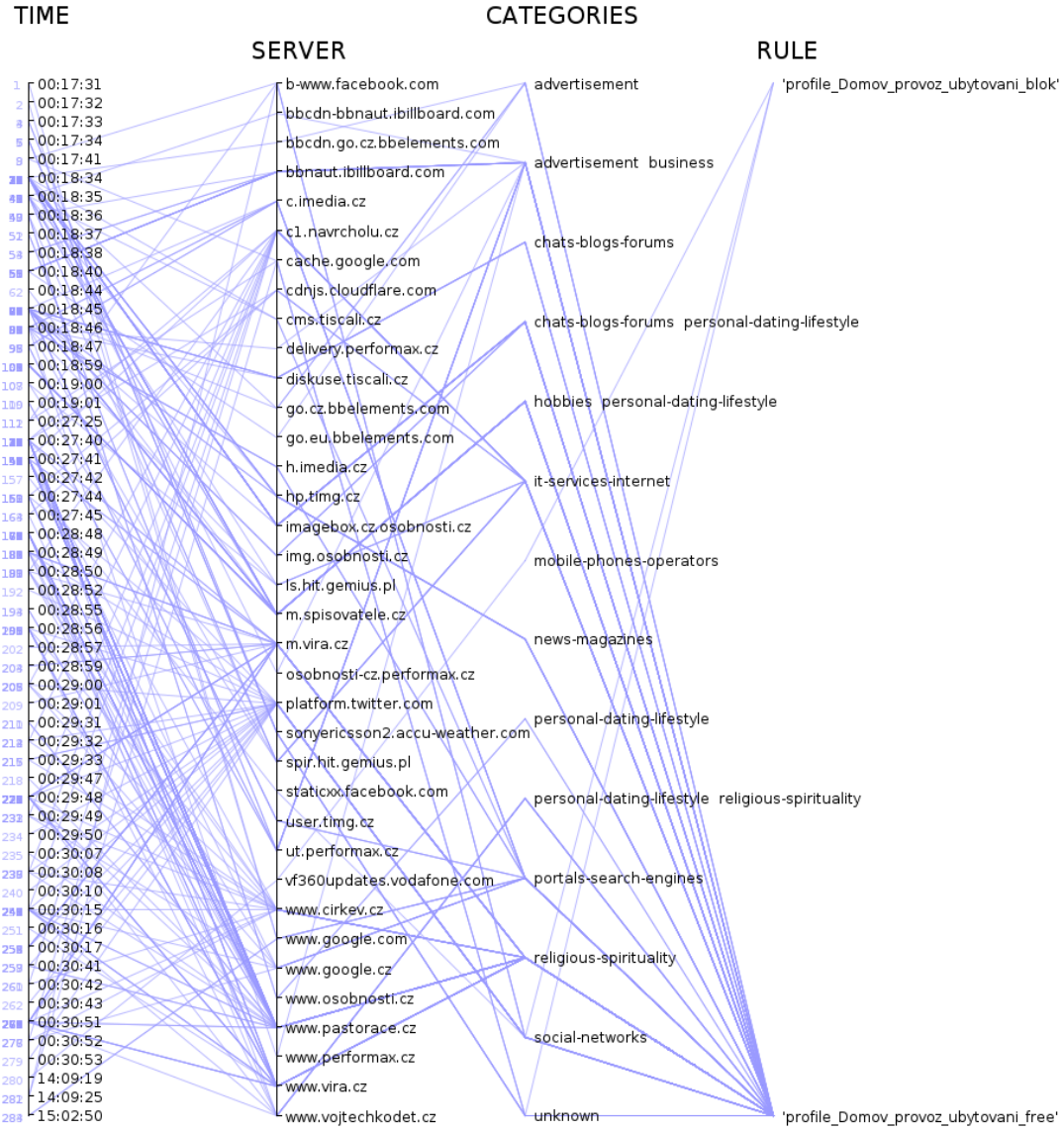
Dalším typem grafu, se kterým se v souvislosti s bezpečnostní vizualizací můžeme setkat, je **binary rainfall**. Lze použít například pro analýzu paketů, kde každý bit paketu představuje obarvení jednoho pixelu. Toto zobrazení lze použít pro rychlé porovnání velkého množství paketů, hledání vzorů, porovnávání délek paketů a identifikování stejných paketů v síti.



Graf 2.8: Graf binary rainfall komunikace jednoho počítače

Graf 2.8 komunikace jednoho počítače při prohlížení různých internetových stránek. Lze vidět obsah stovek paketů najednou a uživatel v něm může rozpoznávat vzory a dále je zkoumat.

Paralelní souřadnice *(Paralel coordinates)* namísto ortogonálního souřadnicového systému jako v jiných grafech se zde využívá paralelní souřadný systém [5]. Tento typ grafu umožňuje zobrazit vícerozměrná data v rovině, kde vertikální osy představují jednotlivé dimenze dat a jejich horizontální spojení představuje jejich vztah. Je vhodný pro zobrazení jakéhokoli typu dat. Změnou měřítka a os lze dosáhnout přiblížení vizualizace a tento typ grafu je vhodný pro interaktivní analýzu dat.



Graf 2.9: Graf paralelních souřadnic komunikace jednoho klienta

Graf 2.9 zobrazuje komunikaci jednoho klienta v síti.

2.4 Další studium

Obsáhlý výčet typů grafů s komentáři přímo pro bezpečnostní vizualizaci lze najít v třetí kapitole práce [1] jsou zde uvedeny i další způsoby vizualizace jako například 3D varianty bodového diagramu, linkového grafu. Dále je zde také uvedena obsáhlejší diskuze o vhodnosti jejich použití.
V češtině lze nalézt výčet některých typů grafů v práci [9] na straně 3 – 22.
Detailnější informace o grafu paralelních souřadnic jsou uvedeny v práci [11].

3 Kompozice vizualizací do jednoho celku

Správná kompozice více vizualizací do jednoho celku vyžaduje menší úsilí ze strany uživatele zpracovávat přijmané informace a v konečném důsledku může i zkrátit dobu potřebnou pro rozpoznání a reakci na událost. Nestačí tedy vytvářet správné grafy, ale je třeba i dbát na to, aby i jejich rozmístění, vzájemná velikost, použité barvy odpovídaly tomu, co chceme uživateli sdělit a aby mu pomáhalo s každodenním používáním. Cílem této kapitoly je uvést některé správné principy jejich kompozice a upozornit na časté chyby.

Správnou kompozici „přístrojové desky“, kde uživatel může sledovat víc vlastností síťového provozu najednou lze charakterizovat pomocí dvou akronymů **SMART** (Synergetic, Monitor key performance indicators, Accurate, Responsive, Timely) a **IMPACT** (Interactive, More data history, Personalized, Analytical, Collaborative, Trackability), které popisují její správné vlastnosti [12].

- **Synergetic**
  Musí být ergonomická a vizuálně efektivní aby umožnila spolu zobrazení víc různých aspektů oproti jejich jednotlivému zobrazení.

- **Monitor key performance indicators**
  Musí zobrazovat klíčové indikátory výkonu pro efektivní rozhodování.

- **Accurate**
  Prezentované informace musí být přesné.

- **Responsive**
  Musí reagovat na předdefinované prahové hodnoty a upozornit uživatele na kritické záležitosti.

- **Timely**
  Musí zobrazovat nejaktuálnější informace.

- **Interactive**
  Měla by být interaktivní a dovolovat uživateli například získání více podrobností.

- **More data history**
  Měla by uživateli umožňovat zobrazit i historický vývoj a hodnoty klíčových ukazatelů v jiné době.

- **Personalized**
  Zobrazení by mělo být specifické pro různé uživatele, jejich pracovní povinnosti, dovoleného přístupu k datům atd.

- **Analytical**
  Měla by uživateli poskytnout rozhraní pro provádění různých analýz.

- **Collaborative**
  Měla by uživatelům poskytovat možnost psát a sdílet svoje poznámky o různých pozorováních.

- **Trackability**
  Měla by uživateli umožňovat přizpůsobení sledovaných parametrů a jejich metrik.

Návrhem kompozice s dodržením těchto vlastností lze docílit výsledku ve kterém uživatel jednodušeji a rychleji rozpozná podstatnou informaci a může rychle reagovat na nastalé situace. Pořadí ve kterém jsou uvedeny tyto vlastnosti nevypovídá nic o jejich vzájemné prioritě.

3.1 Základní chyby při kompozici vizualizací

Stephen Few ve své práci [13] popisuje nejčastější chyby se kterými se setkává při poskytování konzultačních služeb o komunikování kvantitativních informací firmám. Každý problém je v ní popsán a doplněn o názornou ukázku.

- **Rozmístění obsahu v prostoru neodpovídá jeho důležitosti**
  Člověk přirozeně sleduje některé části obrazovky více než jiné. Umístění obsahu na obrazovku by tedy mělo využívat tento fakt a více sledované oblasti by měly obsahovat důležitější informace.

- **Rozmístění obsahu v prostoru nepomáhá jeho použití**
  Pokud některé grafy a informace v nich zobrazované spolu souvisí, tak by měly být zobrazené poblíž sebe.

- **Zobrazení obsahuje prvky které nemají žádný význam**
  Jakékoli dekorace nebo obrázky na pozadí, které nemají žádný význam pro vizualizaci ji pouze znepřehledňují a zapříčiňují, že informace je hůře komunikována. Je otázkou jestli je důležité zobrazovat například firemní logo. Pokud se rozhodnete ho zobrazit, tak jej alespoň umístěte do méně důležitých částí pracovní plochy. Dále je například zbytečné trvale zobrazovat některé popisy a instrukce, které uživatel čte pravděpodobně pouze jednou.

- **Velikost obsahu neodpovídá jeho důležitosti**
  Uživatel přirozeně bude věnovat víc pozornosti většímu grafu.

- **Zvýraznění obsahu neodpovídá jeho důležitosti**
  Nejenom velikost, ale i používání tučného písma nebo barev by mělo odpovídat důležitosti obsahu. Například červená barva může přitahovat víc pozornosti než jiné barvy a proto není vhodné ji používat pro méně podstatné informace.

- **Nadměrné oddělení obsahu**
  Oddělení jednotlivých grafů by mělo být provedeno citlivě a pouze pokud skutečně oddělované informace spolu nesouvisejí. Pracovní plocha by měla být vnímána spíše jako celek.

- **Špatné vizuální spojení souvisejícího obsahu**
  Ne vždy je možné související grafy umístit vedle sebe, aby bylo jasně viditelné srovnání. Potom je důležité tyto grafy alespoň nějak jinak vizuálně spojit. Například pomocí společných barev nebo je seskupovat do logických celků.

- **Zobrazení naznačuje vztah mezi obsahem, který spolu nesouvisí**
  V podstatě se jedná o opačný problém než je zmíněn výše. Například použité barvy mohou naznačovat vztah mezi grafy, které spolu nemají žádnou souvislost.

- **Striktní používání symetrického rozmístění**
  Plocha je striktně rozdělena do oblastí stejné velikosti. Plocha by správně měla být rozdělena na oblasti, tak aby velikost a jejich rozmístění odpovídalo jejich významu.

3.2 Další studium

V práci [12] se autor zabývá obecně pracovní plochou. V první části práce lze najít informace o správných vlastnostech, proces vytváření pracovní plochy, prezentování informace, nebo možnostech varování v pracovní ploše a další. Druhá část se týká různých typů pracovních ploch. Třetí část se zabývá požadavky a hodnocením a čtvrtá část některými reálnými případy užití.

V práci [13] se autor zabývá nejčastějšími chybami, které popisuje na příkladech, vysvětluje proč se jim vyhnout a nakonec uvádí i příklad správné kompozice.

4 Vizualizace reálného logu

Tato kapitola se věnuje vizualizaci reálného logu síťové komunikace. Cílem je přiblížit čtenáři proces vizualizace a bližší použití jednotlivých vizualizačních technik. Pro vizualizaci je použit log ze síťového zařízení Kernun. Log je z jednoho celého dne a obsahuje celkem 108 489 záznamů.

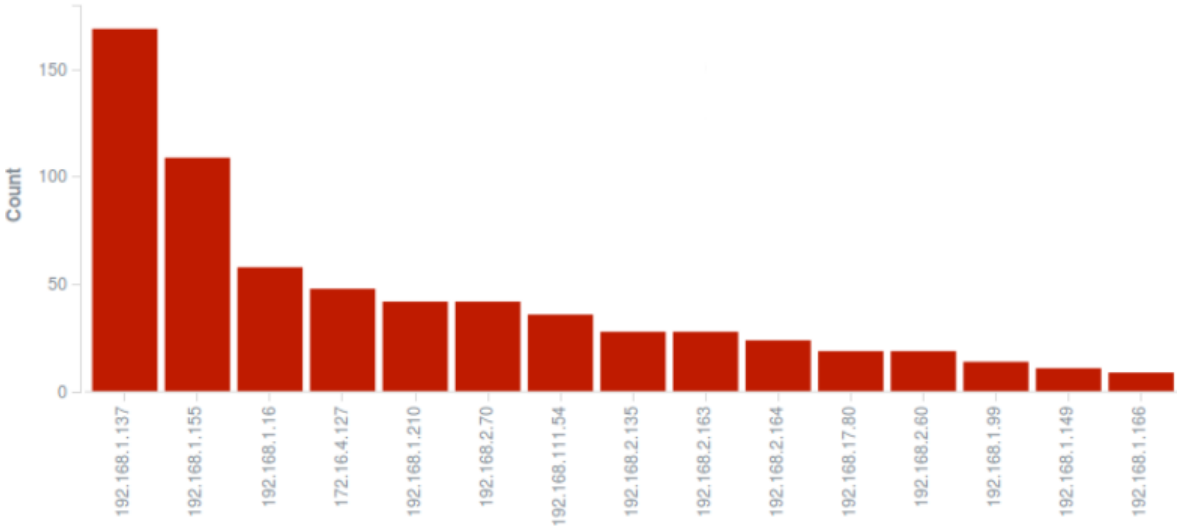
4.1 Dostupná data

Pro následující zobrazení parsuji z logu položky: MONTH, DAY, TIME, PROTOKOL, CLIENT-IP, SERVER, SERVER-IP, SERVER-PORT, DURATION, STATUS, RESULT, RULE, CATEGORIES, VIRUS-STATUS. Dále tento výčet doplňuji položkou COUNTRY a COUNTRY-CODE.

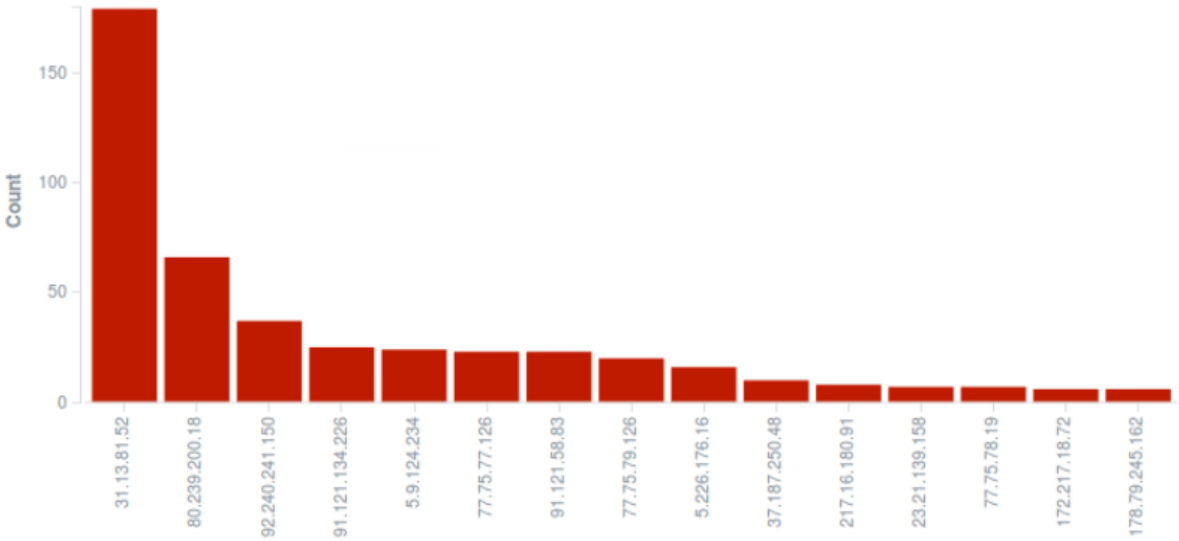
Pro rozparsování používám vlastní jednoduchý program. Parsuji do formátu CSV a SIF. Pro získání státu z jednotlivých Server IP používám výstup příkazu *geoiplookup* v kombinaci se standardními příkazy příkazové řádky *cat*, *cut* a *paste*.

4.2 Vizualizace akcí proxy

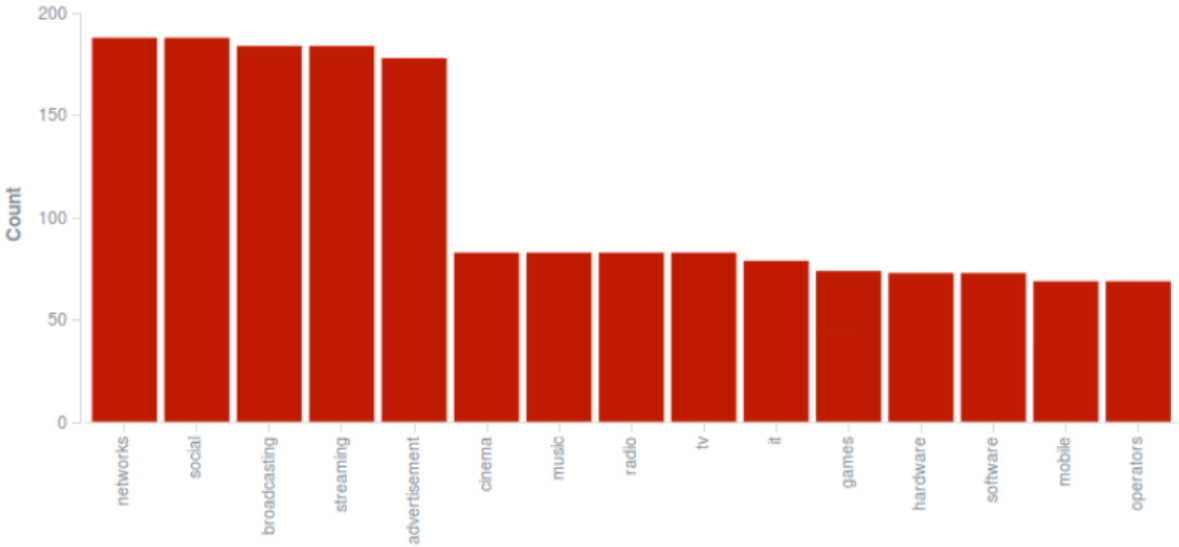
Zaměřuji se na informace o komunikaci se Statusem Rejected. Nejprve vytvořím přehled celé komunikace. Pro tento účel vytvářím sloupcové grafy zobrazující počet záznamů v logu se Status Rejected. Grafy jsou vytvořené pro výběr nepočetnějších Client-Ip dále pak pro Server-Ip, Category, Country-code a Rule.



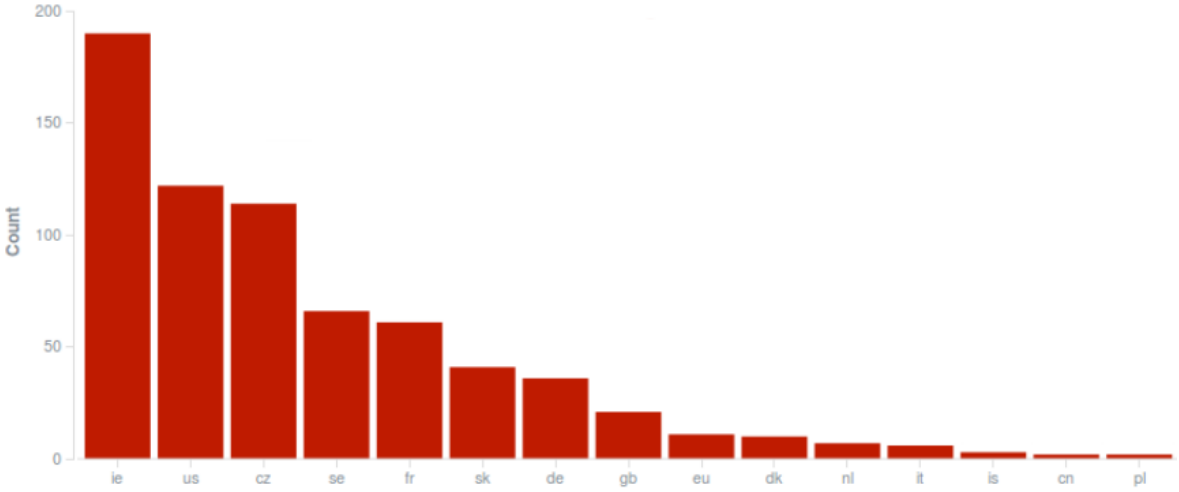
Graf 4.1: Rejected komunikace pro jednotlivé Client Ip



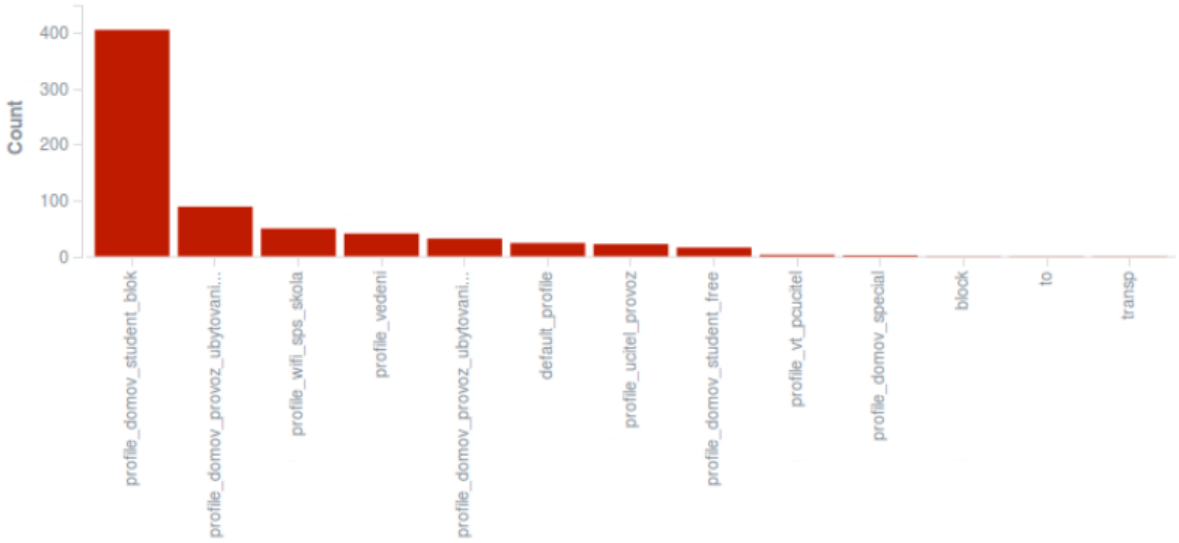
Graf 4.2: Rejected komunikace pro jednotlivé Server Ip



Graf 4.3: Množství Rejected komunikace v jednotlivých kategoriích



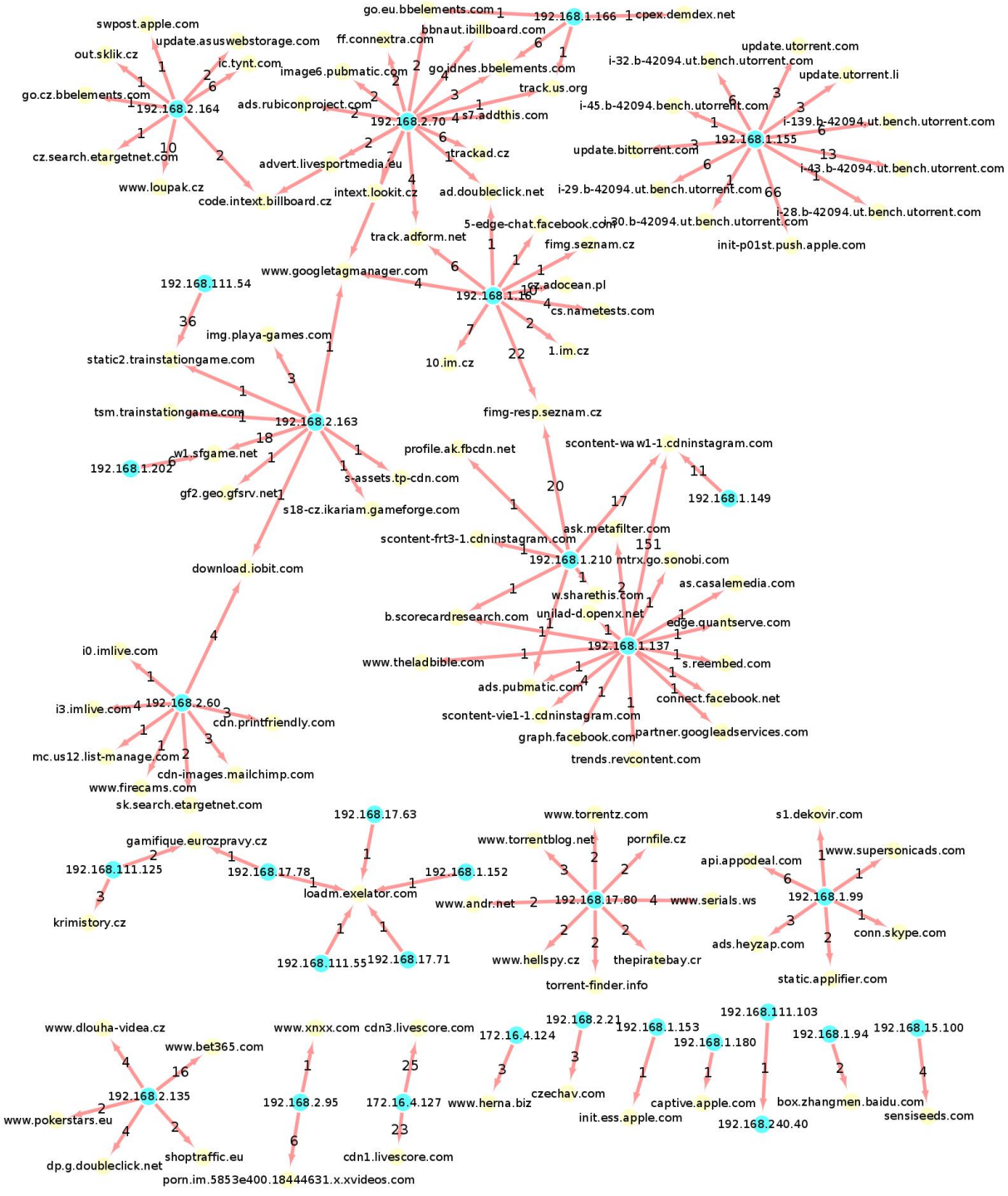
Graf 4.4: Množství Rejected komunikace podle jednotlivých států



Graf 4.5: Množství Rejected komunikace v jednotlivých kategoriích

Kompozicí těchto grafů dávám přehled o množství a rozložení Rejected komunikace. Sloupcové grafy s velkým množstvím položek ztrácejí na přehlednosti a proto uvedené grafy představují výběr patnácti nejpočetnějších položek. Dalším možným řešením velkého počtu položek například pro Client Ip může být kompozice více grafů rozdělených podle podsítí. Například v grafu 4.5 lze hůře rozeznávat hodnoty pro jednotlivé Server Ip a lze zde uvažovat o změně měřítka třeba na logaritmické. Protože se vždy jedná o zobrazení Rejected komunikace je všude použita červená barva. Pro lepší zobrazení dlouhých popisků jednotlivých sloupců v grafu 4.5 by bylo vhodnější zaměnit osy X a Y.

Dále vytvářím detailnější pohled na Rejected komunikaci pomocí grafu propojených uzlů.

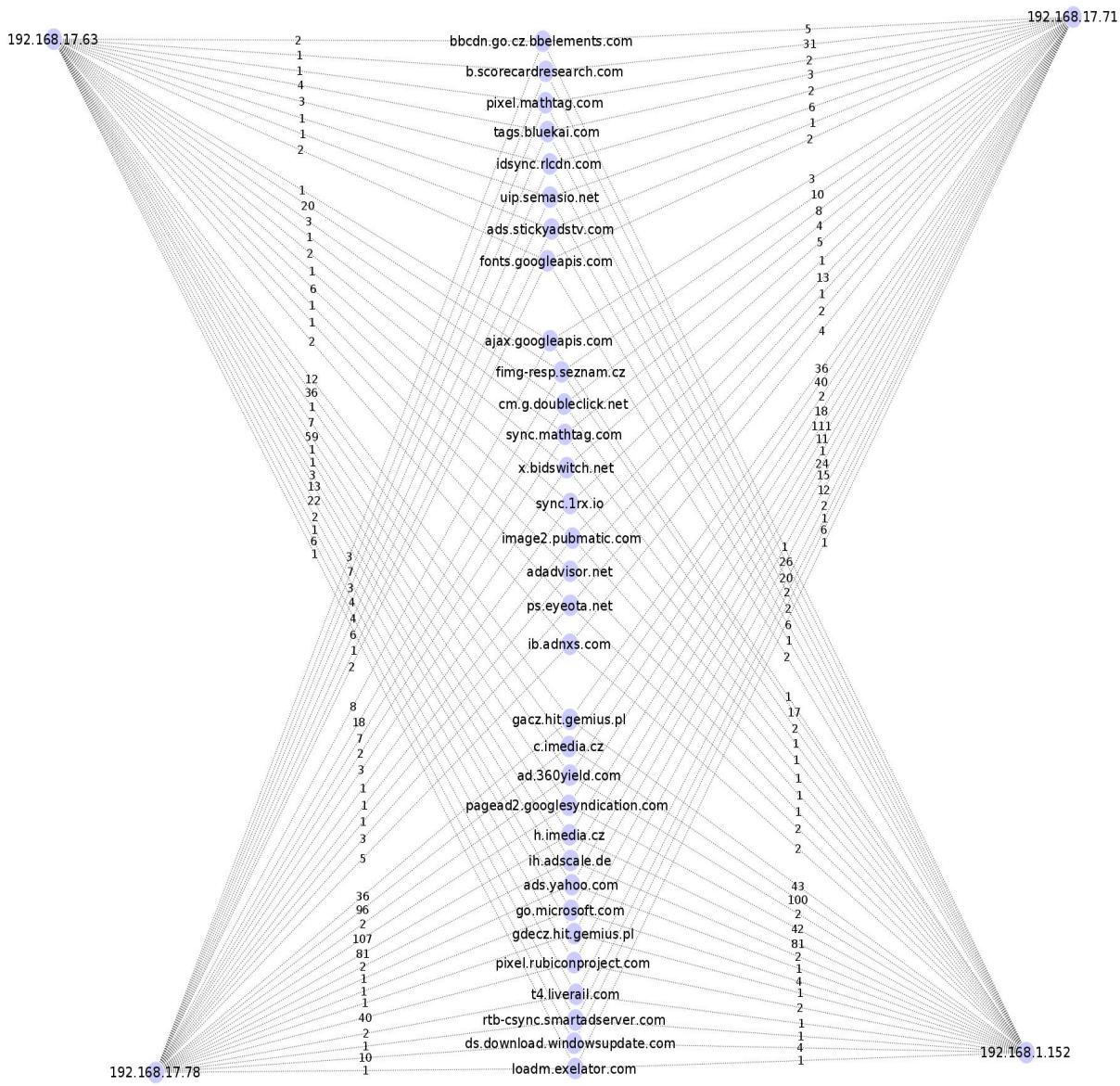


Graf 4.6: Graf propojených uzlů Rejected komunikace

Graf propojených uzlů s ohodnocenými hranami zobrazuje vztah „kdo s kým a jak moc“ komunikuje. Abych dodržel konzistenci mezi grafy jsou hrany grafu opět obarveny červenou barvou, i když je použit světlejší odstín, aby čísla u hran zůstala výrazná. Tyto hrany stále představují Rejected komunikaci jako grafy výše. Pro lepší orientaci v grafu jsou uzly interní sítě obarveny modrou barvou a servery barvou žlutou. Čísla u jednotlivých hran udávají počet položek v logu pro tyto dvojce Client-Ip a Server a vyjadřují tedy počet spojení. Dále by zde mohla být uvedena například hodnota průměru délky spojení.

Po prozkoumání tohoto grafu si můžeme všimnout některých anomálií. V levé spodní části se nachází server *loadm.exelator.com*, který je společný pro pět uzlů naší sítě. Jedná se o uzel s největším počtem společných uzlů sledované počítačové sítě. Podle virusresearch.org se jedná o tzv. hijacker, tedy o zákeřný software napadající internetový prohlížeč. Další anomálie je v pravé spodní části, kde je Rejected komunikace mezi dvěma uzly naší sítě 192.168.111.103 a 192.168.240.40. Déle lze ještě upozornit na uzel *scotent-vie1-1cdninstagram.com* s největším počtem Rejected položek v logu.

Další krok vizualizace je inspirován první anomálií. Zaměřuji se na společnou komunikaci uzlů 192.168.17.63, 192.168.17.71, 192.168.17.78 a 192.168.1.152. Uzel 192.168.111.55 v následujícím zobrazením vynechávám za účelem redukování dat a lepší přehlednosti.



Graf 4.7: Graf propojených uzlů společných pro vybrané Client-IP

Linkový graf veškeré komunikace by byl nečitelný a proto zde uvádím pouze graf společných uzlů. V tomto případě je zvolené jiné obarvení, protože zobrazuji veškerou komunikaci a jedná se tedy o jiné zobrazení než jsou uvedena výše. Hrany jsou zobrazeny pouze pomocí teček pro zlepšení přehlednosti. Modré obarvení uzlů je zvoleno aby byl lépe rozpoznatelný konec hrany. Tento graf je výsledkem postupného filtrování a analýzy,

při které uživatel může provádět různé zvýraznění, odfiltrovat unikátní komunikaci jednoho, dvou uzlů atd. Graf lze ještě vylepšit seřazením serverů. Stejnou informaci, jako je zobrazena v grafu 4.7, lze ekvivalentně zobrazit tabulkou, která je v tomto případě přehlednější.

Počet společné komunikace vybraných Client-IP se Statusem Rejected

Server	192.168.1.152	192.168.17.63	192.168.17.71	192.168.17.78
ad.360yield.com	2	1	2	2
adadvisor.net	1	1	1	1
ads.stickyadstv.com	1	1	1	1
ads.yahoo.com	1	1	1	1
ajax.googleapis.com	1	1	3	8
b.scorecardresearch.com	26	1	31	7
bbcdn.go.cz.bbelements.com	1	2	5	3
c.imedia.cz	100	36	40	96
cm.g.doubleclick.net	2	3	8	7
ds.download.windowsupdate.com	4	6	6	10
fimg-resp.seznam.cz	17	20	10	18
fonts.googleapis.com	2	2	2	2
gacz.hit.gemius.pl	43	12	36	36
gdecz.hit.gemius.pl	1	13	15	1
go.microsoft.com	4	3	24	1
h.imedia.cz	81	59	111	81
ib.adnxs.com	2	2	4	5
idsync.rlcdn.com	2	3	2	4
ih.adscale.de	2	1	11	2
image2.pubmatic.com	1	6	13	1
loadm.exelator.com	1	1	1	1
pagead2.googlesyndication.com	42	7	18	107
pixel.mathtag.com	20	1	2	3
pixel.rubiconproject.com	2	22	12	40
ps.eyeota.net	2	1	2	3
rtb-csync.smartadserver.com	1	1	1	1
sync.1rx.io	1	1	1	1
sync.mathtag.com	1	1	4	2
t4.liverail.com	1	2	2	2
tags.bluekai.com	2	4	3	4
uip.semasio.net	6	1	6	6
x.bidswitch.net	1	2	5	3

Tabulka 4.1: Společná komunikace vybraných klientů

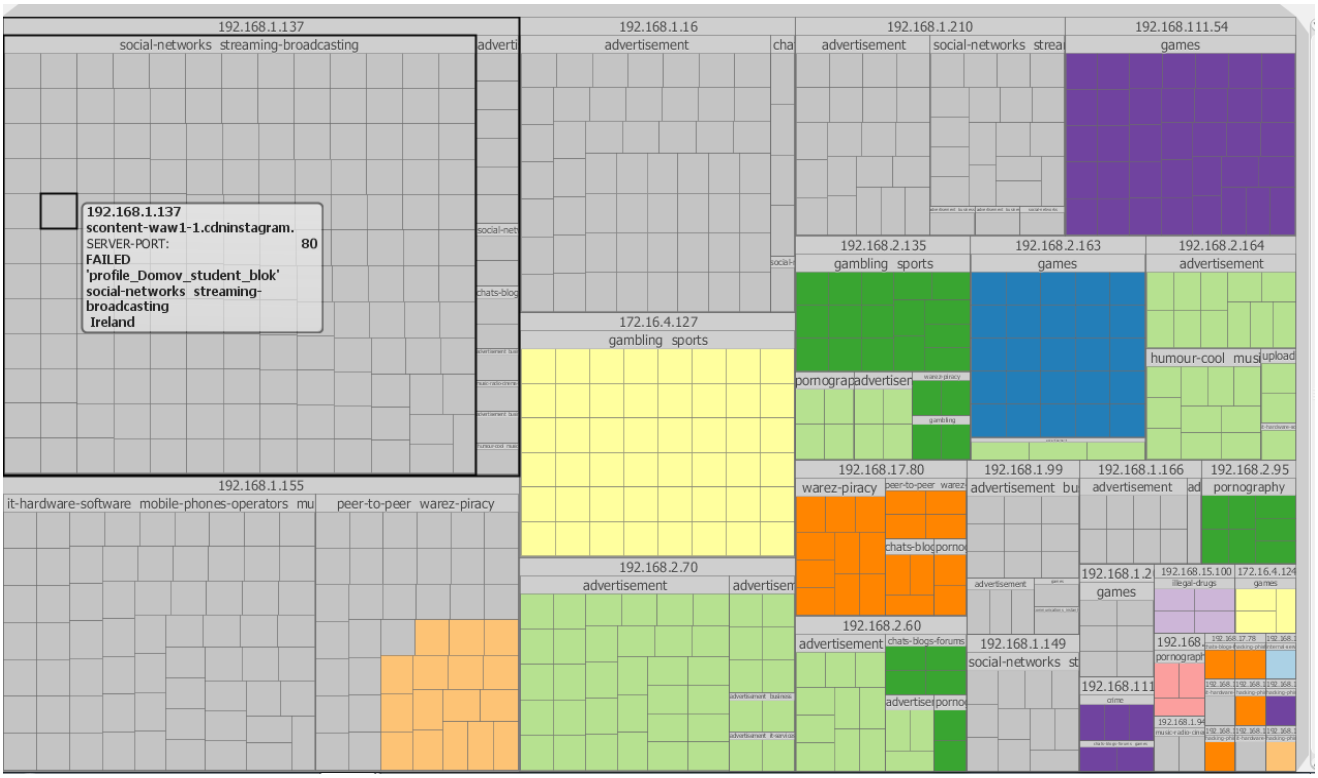
Tabulka 4.1 zobrazuje informaci o množství společné komunikace. Pro lepší přehlednost jsou názvy serverů seřazené podle abecedy a číselné údaje jsou zarovnány podle řádu. Lze uvažovat i o jiných možnostech seřazení například podle řádu domény. Tabulka nebo graf takových informací může sloužit jako podklad pro další analýzu některých serverů.

Další pohled na data lze realizovat pomocí grafu paralelních souřadnic ve kterém můžeme zachytit najednou několik dimenzí.

Graf 4.8: Graf paralelních souřadnic Rejected komunikace

V grafu paralelních souřadnic uživatel může sledovat víc parametrů komunikace v jednom celku. Uživatel tedy může v jedné vizualizaci vidět od jakých klientů,

na které servery vedla Rejected komunikace, které pravidlo ji zablokovalo a do jaké náleží kategorie. Světle červené linie mají nastavenou jistou míru transparentnosti čímž dosahuji sytějšího odstínu barvy v místě s četnějšími záznamy a zvýrazňuji tato místa. Program Xdat dále umožňuje filtrování podle hodnot na jednotlivých osách a uživatel může data analyzovat podle různých hodnot nebo různé kombinace hodnot. Případným vylepšením by mohlo být různé setřídění jednotlivých os. Například setřídění Client-Ip podle podsítě, serveru, domény nebo chronologicky.



Graf 4.9: Hiearchická síťová mapa Rejected komunikace s detailem v tooltipu

Hiearchická síťová mapa zde zobrazuje komunikaci jednotlivých uživatelů jako čtvercové či obdélníkové oblasti, které jsou dále rozděleny na podoblasti podle kategorie. Oblasti jsou obarveny podle položky Rule. Uživatel může získat celkový přehled o rozložení Rejected komunikace mezi klienty, kategorie a pravidel v jednom grafu a díky zobrazovanému tooltipu v programu Treemap může získat další podrobnosti.

Graf 4.10: Hiearchická síťová mapa jednotlivých kategorií rozdělených podle Rule

V grafu 4.10 je zobrazeno rozdělení Rejected komunikace do jednotlivých kategorií, které jsou dále rozděleny dle Rule, podle kterých je graf pro lepší přehlednost i obarven. Uživatel může získat snadno přehled o rozdělení Rejected komunikace do jednotlivých kategorií a zároveň může vidět podle kterých uplatněných pravidel byla v těchto kategoriích komunikace zamítnuta.

4.3 Použité nástroje v této práci

Níže uvádím srovnání použitých nástrojů při tvorbě vizualizací použitých v této práci. Cílem práce není uvést rešerši všech možných možných nástrojů, kterých existuje velké množství použitelných pod různými licencemi. Jednotlivé typy grafů, lze vytvářet různými nástroji. Nejprve je uveden stručný popis nástroje u kterého je vždy v kulatých závorkách uvedena použitá verze programu. Dále jsou uvedeny možnosti vizualizace, které nabízí

a informace o jejich efektivitě a možnostech interakce s programem. Následuje pak stručný seznam výhod a nevýhod nástroje v kontextu vytváření praktické části této práce.

ELK Stack [14] je kombinací třech nástrojů: Elasticsearch (verze 2.2.0), Kibana (verze 4.4.1), Logstash (verze 2.2.2). Elasticsearch je vyhledávací nástroj, Logstash slouží pro import dat do databáze a Kibana je grafické rozhraní pro vytváření vizualizací. Umožňuje vytvářet různé vizualizace a komponovat je do jednoho celku.

Mezi možnosti vizuální reprezentace, které ELK stack nabízí patří plošný graf, tabulka, spojnicový graf, čísla, kruhový graf, mapa a sloupcový graf.

O efektivitě grafů může díky možnosti nastavení počtu položek, barev, velkosti, a filtrace dat rozhodovat sám uživatel. Vizualizace jsou v základu dobře navržené.

ELK stack nabízí široké možnosti interakce. Uživatel může psát výrazy jako například *(CLIENT-IP: 192.168.2.*) AND STATUS: R** které omezí výsledky, z kterých se potom vytváří graf, nebo lze měnit různé parametry grafu a okamžitě je aplikovat.

Výhody:

- Komplexní nástroj, umožňuje vytvářet různé grafy najednou a pracovat s nimi
- Kompozice grafů do pracovní plochy, poskytuje jednoduché rozhraní pro skládání grafů do pracovní plochy
- Široké možnosti interakce s okamžitým zobrazením změn

Nevýhody:

- Ocenil bych především lepší dokumentaci k nástroji
- Absence více možností změn měřítek grafů v základní nabídce

Program **Cytoscape** (verze 3.3.0) [15] je nástroj pro tvorbu linkových grafů. Nabízí pouze jedinou možnost vizuální reprezentace. Uživatel může vytvořit víc linkových grafů a zobrazit je najednou. Cytoscape je rozšiřitelný o pluginy, které může i uživatel sám vytvářet pomocí Cytoscape open API.

Efektivitu grafů může uživatel upravovat pomocí přednastavených zobrazovacích šablon, které může sám dál upravovat. Šablony odpovídají různým použitím a uživatel musí zvolit vhodnou šablonu, aby dosáhl požadovaného efektivního zobrazení.

Interakce uživatele s grafem je realizována pomocí možností výběru a změn šablon. Déle pak lze definovat a aplikovat různé funkce na uzly nebo hrany a může mezi

nimi vyhledávat. Uzly může uživatel přesouvat jednotlivě nebo po skupinách pomocí „táhni a pusť“.

Výhody:

- Velké množství šablon, které může uživatel upravovat
- Možnost zobrazení více linkových grafů najednou
- Kvalitní dokumentace
- Umožňuje hledat mezi uzly a hranami a upravovat je
- Interakce s okamžitým zobrazením změn

Nevýhody:

- Vytváří pouze linkové grafy

X-dimensional Data Analysis Tool (verze 2.0) [16] je nástroj určený pro analýzu multidimenzionálních dat. Umožňuje vytvářet grafy paralelních souřadnic a bodové grafy.

Efektivitu grafů může uživatel ovlivnit množstvím nastavení. Vizualizace jsou v základu dobře navržené, ale v závislosti na množství zobrazovaných informací a použití, lze s grafy dále pracovat za účelem dosažení efektivnějšího grafu.

Interakce je realizována pomocí možností nastavení zobrazení grafů a v případě grafu paralelních souřadnic umožňuje pomocí „táhni a pusť“ přesouvání jednotlivých os, vytváření seskupení (cluster) z hran, přidělení různých barev a nastavení průhlednosti. Dále pak umožňuje výběr hran pomocí omezení intervalů na jednotlivých osách.

Výhody:

- Možnost zobrazení více grafů najednou
- Kvalitní dokumentace
- Možnosti interakce s okamžitým zobrazením změn

Nevýhody:

- Vytváří pouze grafy paralelních souřadnic a bodové grafy
- Program je při velkých objemech dat pomalý
- Graf lze rozšiřovat pouze horizontálně
- Chybí víc možností seřazení na jednotlivých osách

Program **Rumint** (verze 2.14) [17] umožňuje zobrazení probíhající komunikace nebo přehrání zaznamenané komunikace. Umožňuje zobrazit data pomocí grafu paralelních souřadnic, bodového grafu a grafu binary rainfall.

Uživatel nemá mnoho možností jak ovlivnit efektivitu grafů. Grafy se zobrazují pouze černo zelené a nelze nastavit jinou barvu zobrazení. To může působit problém při kompozici více grafů do jednoho celku.

Co se týká interakce, tak v programu Rumint lze pozastavit zobrazení a filtrovat komunikaci podle portů.

Výhody:

- Možnost zobrazení více grafů najednou
- Možnost zobrazení real time komunikace
- Možnost přehrání uložené komunikace

Nevýhody:

- Absence kvalitní dokumentace
- Absence možnosti detailnějšího nastavení zobrazení
- Málo možností interakce

Vizualizační nástroj **TreeMap** (verze 3.8.3) [18] umožňuje vytvářet různé hierarchické síťové mapy a zobrazovat data pomocí treeplotu a tabulky.

Uživatel si může vybrat mezi různými šablonami a spolu s množstvím nastavení tak ovlivnit efektivitu výsledného grafu. Uživatel může nastavovat zobrazované položky, jejich seskupení atd.

Data lze v programu různě filtrovat a uživatel si může nastavit informace zobrazované při pohybu kurzoru nad jednotlivými oblastmi v grafu. Změny nastavení se projevují okamžitě.

Výhody:

- Široké možnosti interakce s okamžitým zobrazením změn
- Zobrazuje tool tip s detailními informacemi o záznamu

Nevýhody:

- Většina šablon cílí na estetiku vizualizace, ale ne na její efektivitu

4.4 Shrnutí

Uvedené grafy výše představují některé z možných řešení vizualizace těchto dat. Grafy byly pro tisk zmenšeny a některé z grafů naleznete v příloze této práce ve větším rozlišení. Kompozice sloupcových grafů, grafu propojených uzlů a grafu paralelních souřadnic je také uvedena v příloze této práce a představuje celkový pohled na Rejected komunikaci, který je vytvořen od souhrnných informací k detailnějším informacím.

Největším problémem, s kterým jsem se při psaní práce setkal, bylo parsování logu. Špatné vytvořený parser může pro některé logy fungovat ale u dalších uživatel ztratí obsaženou informaci a ztráta se může projevit až při některých detailních analýzách nebo se nemusí projevit vůbec. Nalezení tohoto problému může také zneplatnit dříve vytvořené vizualizace, proto je důležité správnému parsování dat věnovat velkou pozornost.

5 Závěr

V této práci jsem se zabýval problematikou vizualizace logů síťové komunikace a vizualizoval jsem log reálné komunikace poskytnutý firmou Trusted Network Solutions, a.s. Cílem práce bylo představit úvod do problematiky tvorby vizualizací, správných technik jejich tvorby za účelem získání nejlepších výsledků a možnostmi vizuální reprezentace logů.

Teoretická část představuje výběr možností vizuální reprezentace, správné praktiky jejich tvorby, správné praktiky jejich kompozice a odkazuje na další informace čímž vytváří úvod do problematiky vizualizace dat. V druhé kapitole jsou také zmíněny možnosti použití.

Při vizualizaci reálného logu se podařila odhalit anomálie, která vznikla působením škodlivého kódu v počítačích. Vizualizace uvedené v této práci obsahují také krátkou diskuzi o jejich výhodách či nevýhodách, použití a s případnými návrhy na zlepšení.

Lze očekávat že vývoj v oblasti bezpečnostní (i jiné) vizualizace dat se nezastaví a budou dále přibývat nové způsoby vizualizace dat. Dalším důležitým krokem bude stanovení jednotného formátu logů, aby se usnadnilo parsovaní logů. Další práce v této oblasti může směřovat k analýze výkonu vizualizačních nástrojů. Například lze srovnávat, jak velké množství dat program dokáže zpracovat v přijatelném čase, rychlost odezvy nebo filtrování dat.

Slovníček

Překlad na základě podobnosti definic pojmů v [1] a [2]

Nominální – Nominal (někde též zvané jako *Categorical* nebo *Discrete*)
Ordinální – Ordinal
Intervalová – Interval
Poměrová – Ratio

Překlad převzatý z [19]

Kruhový graf – Pie chart
Prstencový graf – Stacked pie chart
Sloupcový graf – Bar chart
Spojnicový graf – Line chart
Bodový graf – Scatter plot

Překlad převzatý z [20]

Hierarchická síťová mapa – Tree Map

Překlad převzatý z [10]

Paralelní souřadnice – Paralel coordinates

Literatura

[1] MARTY, R.: *Applied security visualization*. Upper Saddle River, N.J.: Addison-Wesley, c2009. ISBN 978-0-321-51010-5.

[2] CYHELSKÝ, L. a VALENTOVÁ, V.: *Význam základní klasifikace ukazatelů pro korektní interpretaci vzájemných odlišností jejich hodnot: Politická ekonomie* [online]. 2006, vol. 4, s. 542-548, [cit. 1. 7. 2016]
Dostupné z: http://www.vse.cz/polek/download.php?jnl=polek&pdf=573.pdf

[3] TUFTE, E. R.: *The visual display of quantitative information.* 2nd ed. Cheshire: Graphics Press, c2001. ISBN 0-9613921-4-2.

[4] *Úprava písemností psaných strojem nebo zpracovaných textovými editory: ČSN 01 6910.* Praha: Český normalizační institut, 1997. ČSN, 01 6910.

[5] WARE, C.: *Visual thinking for design*. [Repr.]. Burlington, Mass: Morgan Kaufmann, 2008. ISBN 0123708966.

[6] SAMII, A. a KOH, W.: *Interactive Visualization for System Log Analysis* [online]. University of California, Berkeley 2013 [cit. 5. 9. 2016]. Dostupné z: https://people.eecs.berkeley.edu/~kubitron/courses/cs262a-F13/projects/reports/project1_report_ver3.pdf

[7] TUCK, M.: *Gestalt Principles Applied in Design* [online]. 2010 [cit. 29. 11. 2016]. Dostupné z:
http://sixrevisions.com/web_design/gestalt-principles-applied-in- design/

[8] VYMĚTAL, J.: *Průvodce úspěšnou komunikací: efektivní komunikace v praxi.* Praha: Grada, 2008. Manažer. ISBN 978-80-247-2614-4.

[9] KUŽELA, A.: *Vizualizace dat.* [online]. Vysoké učení technické v Brně Fakulta informačních technologií 2008. [cit. 7. 9. 2016]. Dostupné z:
http://www.fit.vutbr.cz/study/courses/VPD/public/0708VPD-Kuzela.pdf

[10] JAMBOR, M.: *Vizualizace a analýza vícerozměrných dat : diplomová práce.* Brno: Mendelova univerzita v Brně, Provozně ekonomická fakulta, 2012. 53 s. Vedoucí diplomové práce Ing. David Procházka, Ph.D.

[11] TRICAUD, S., NANCE K. a SAADE, P.: *Visualizing Network Activity Using Parallel Coordinates* [online]. Proceedings of the 2011 44th Hawaii International Conference on System Sciences , 2011. [cit. 3. 10. 2016]. Dostupné z: https://www.computer.org/csdl/proceedings/hicss/2011/4282/00/04-06-06.pdf

[12] MALIK, S.: *Enterprise dashboards: design and best practices for IT*. Hoboken, N.J.: Wiley, c2005. ISBN 0471738069.

[13] FEW, S.: Formatting and layout definitely matter [online]. Perceptual Edge 2008. [cit. 3. 9. 2016]. Dostupné z: https://www.perceptualedge.com/articles/Whitepapers/Formatting_and_Layout_Matter.pdf

[14] Elasticsearch BV: *The Open Source Elastic Stack* [online]. Elasticsearch: ©2016 [cit. 11. 11. 2016]. Dostupné z: https://www.elastic.co/products

[15] The Cytoscape Consortium: *Cytoscape Network Data Integration, Analysis, and Visualization in a Box* [online]. Cytoscape Consortium ©2001-2016 [cit. 11. 11. 2016]. Dostupné z: http://www.cytoscape.org/

[16] ROCHEFORT, E.: *XDAT - A free parallel coordinates software tool* [online]. Xdat.org: ©2010-2016 [cit. 11. 11. 2016]. Dostupné z: http://www.xdat.org/

[17] CONTI, G. J.: *RUMINT.org - A PVR for Network Traffic and Security Visualization* [online]. [cit. 11. 11. 2016]. Dostupné z: http://www.rumint.org/

[18] Macrofocus GmbH: *TreeMap interactive data visualization software* [online]. Macrofocus © 2000-2016 [cit. 11. 11. 2016]. Dostupné z: http://www.macrofocus.com/

[19] BŘÍZA, V.: *Excel 2007: podrobný průvodce*. Praha: Grada, 2007. ISBN 978-80- 247-1965-8.

[20] KUPČÍK, J.: *Vizualizace multidimenzionálních dat* [online]. Vysoké učení technické v Brně Fakulta informačních technologií 2008. [cit. 10. 9. 2016]. Dostupné z: http://www.fit.vutbr.cz/study/courses/VPD/public/0708VPD-Kupcik.pdf