CESNET Technical Report 3/2014

Distribuovaný kolektor záznamů o IP tocích: návrh a první experiment

Martin Žádník, Pavel Krobot, Lukáš Kekely, Viktor Puš, Jan Kořenek

Received 20. 12. 2014

Abstract

Monitoring síťového provozu je dnes již nezbytnou součástí moderní síťové infrastruktury. V rámci monitorovacího systému dochází ke sběru a analýze dat na kolektoru. Současné kolektory ve většině případů nepodporují distribuované zpracování sbíraných dat a není tak možné škálovat jejich výkonnost přidáváním výpočetních uzlů. V této technické zprávě je nastíněn koncept distribuce úloh kolektoru a provedeny první experimenty se zpracováním sbíraných dat o síťovém provozu v paradigmatu MapReduce.

Keywords: distribution, cloud, collector, big data

1 Úvod

Monitoring počítačového provozu poskytuje cenné informace pro správu počítačové sítě. Ve většině případů jsou sbírána metadata o síťovém provozu například SNMP, NetFlow či IPFIX. Monitorovací systém se typicky sestává z měřicích sond a sběrného místa - kolektoru. Zatímco úlohou sond je přesně změřit veškerý provoz, úlohou kolektoru je změřená data sbírat, uchovat, analyzovat a sbíraná data dále zpřístupnit správci dané síťové infrastruktury, který provádí jejich expertní analýzu, například dohledání nahlášeného incidentu. Kolektor tedy představuje centrální místo celého systému a tedy i potenciální úzké hrdlo celého systému. Strop výkonnosti současných kolektorů se projeví především ve velkých sítích s mnoha sondami a potřebou důkladně analyzovat stav sítě, vyhledávat významné události a dohledávat nahlášené incidenty.

Při detailní analýze jednotlivých činností kolektoru jsou především problémy s:

- nedostatečnou propusností diskového subsystému a
- zpožděná analýza průběžně přijímaných dat.

Při analýze uložených dat je úlohou kolektoru tato data co nejrychleji získat z permanentního úložiště a následně vykonat dotaz nad daty. Většina současných řešení ale uvažuje pouze s centrálním úložným prostorem a neřeší distribuci dat na více úložných uzlů. Výkonnost je tedy limitována možnostmi centrálního diskového pole a propojením tohoto pole se zbytkem systému. Druhým problémem je průběžná analýza přijímaných dat, kdy dochází nejprve k uložení dat na permanentní úložiště a až následně dochází k analýze dat. Tím vzniká zbytečná prodleva mezi přijetím dat a jejich analýzou a tedy klesá možnost rychle reagovat na zjištěné události.

Nutnost zvyšování výkonnosti kolektoru je dále podpořena několika trendy, které lze pozorovat v Internetu. Významným trendem je neustále se zvyšující množství

zařízení připojených do Internetu, zvyšující se objem provozu a s tím rostoucí množství dat, která je potřeba sbírat, uchovávat a analyzovat. Dalším trendem je centralizace služeb do datových center či cloudů a vznik nových služeb. Centralizace a zavedení nových služeb vede opět na zvýšené množství dat, která se přenáší přes síťovou infrastrukturu. Bezpečnost, spolehlivost a kvalita poskytované služby jsou klíčovými faktory pro poskytovatele. Výpadek služby může generovat nejen obrovské ekonomické ztráty, ale i ohrožovat či narušovat některé základní služby státu s dopady na zdraví a majetek. Dalším trendem je připojování nových zařízení do Internetu (tj. připojování jakýchkoliv zařízení do Internetu nazvané jako Internet of Things nebo také Internet of Everything). Tento trend bude hrát významnou roli v dalším nárůstu provozu a jeho diverzifikaci. Navíc připojovaná zařízení ze své podstaty disponují pouze omezeným výkonem, který je dedikován aplikaci, nikoliv její ochraně. Tento stav podtrhuje nutnost včasného odhalování hrozeb a napadení. Významným trendem je rovněž kyberkriminalita, projevující se v podobě velkého množství síťových útoků, které využívají vlastnosti či zranitelnosti aplikací dostupných prostřednictvím sítě. Útoky a hrozby se stávají mohutnějšími (na začátku roku byly reportovány DDoS s rychlostí dat převyšující 300 Gb/s) nebo sofistikovanějšími, které svou intenzitou nebo vlastnostmi zůstávají mimo detekční schopnosti současných metod. Posledním trendem, který podtrhuje vhodnost distribuce úlohy kolektoru, je budování datových center s velkým množstvím levných výpočetně-úložných uzlů v podobě běžných serverů. Distribuovaný kolektor by tak mohl využít úložné a výpočetní kapacity takové architektury. Shrneme-li předchozí trendy, existuje zvyšující se potřeba monitorovat provoz v síti za účelem poskytnoutí vyšší míry spolehlivosti a bezpečnosti infrastruktury, služeb a jejich uživatelů a zároveň snižovat náklady na tento monitoring.

V reakci na problémy současných řešení, trendy ve vývoji počítačového provozu a ve zpracování velkého množství dat, navrhujeme upravit architekturu kolektoru s cílem maximálně podpořit distribuci a online zpracování dat. Tento návrh počítá s využitím paradigmatu MapReduce, tedy distribuce činnosti k datům, a dále s využitím distribuovaného proudového zpracování dat, které bude probíhat paralelně k uložení dat do permanentní úložiště.

Samozřejmě je možné pozorovat obdobné projekty, které se snaží distribuovat úložiště pomocí frameworku Apache Hadoop [3], např. [2]. Naším cílem je ovšem vytvořit kompletně distribuovaný kolektor, včetně příjmu dat, jejich okamžitého zpracování a v neposlední řadě uložení. Z tohoto důvodu plánujeme ověřit některé technologie pro zpracování velkého množství dat, i ty které již byly publikovány, neboť idealizované prostředí již publikovaných experimentů často ovlivní výsledky. Tato technická zpráva proto nejprve popisuje navrhovaný distribuovaný kolektor a následně prezentuje experimenty s první platformou pro distribuci úložiště.

2 Návrh

Z pohledu problému rychlého získání dat z úložiště se jeví jako logické využití paradigmatu MapReduce. V tomto paradigmatu budou sbíraná data ukládána do distribuovaného úložiště. Samotný kolektor bude tvořen běžnými servery, které poskytují jak výpočetní tak i samotný úložný prostor. Dolování dat na této ar-

chitektuře bude využívat extremní celkovou propustnost diskových subsystémů v jednotlivých uzlech. Je tedy nutné distribuovat velkou část dotazu/výpočtu přímo na uzly s lokálními daty a získané mezivýsledky spojit do výsledku tak, aby uživatel kolektoru získal informaci shodnou s informací na běžném kolektoru, nicméně v daleko kratším čase.

Z pohledu problému zpožděná analýza průběžně přijímaných dat se jeví jako vhodná proudová analýza dat. Vzhledem k množství přijímaných dat a náročnosti samotné analýzy je nutné rovněž distribuovat výpočet na více uzlů. Z toho výplývá nutnost připravit vhodně data pro proudové zpracování, přizpůsobit současné typy výpočtů na proudové zpracování a vhodně upravit samotně výpočty, aby je bylo možné efektivně paralelizovat.

Obrázek Obrázek 1 uvádí zjednodušené blokové schéma kolektoru SecurityCloud, ve kterém je kolektor rozdělen do několika logických částí, které spolu komunikují.

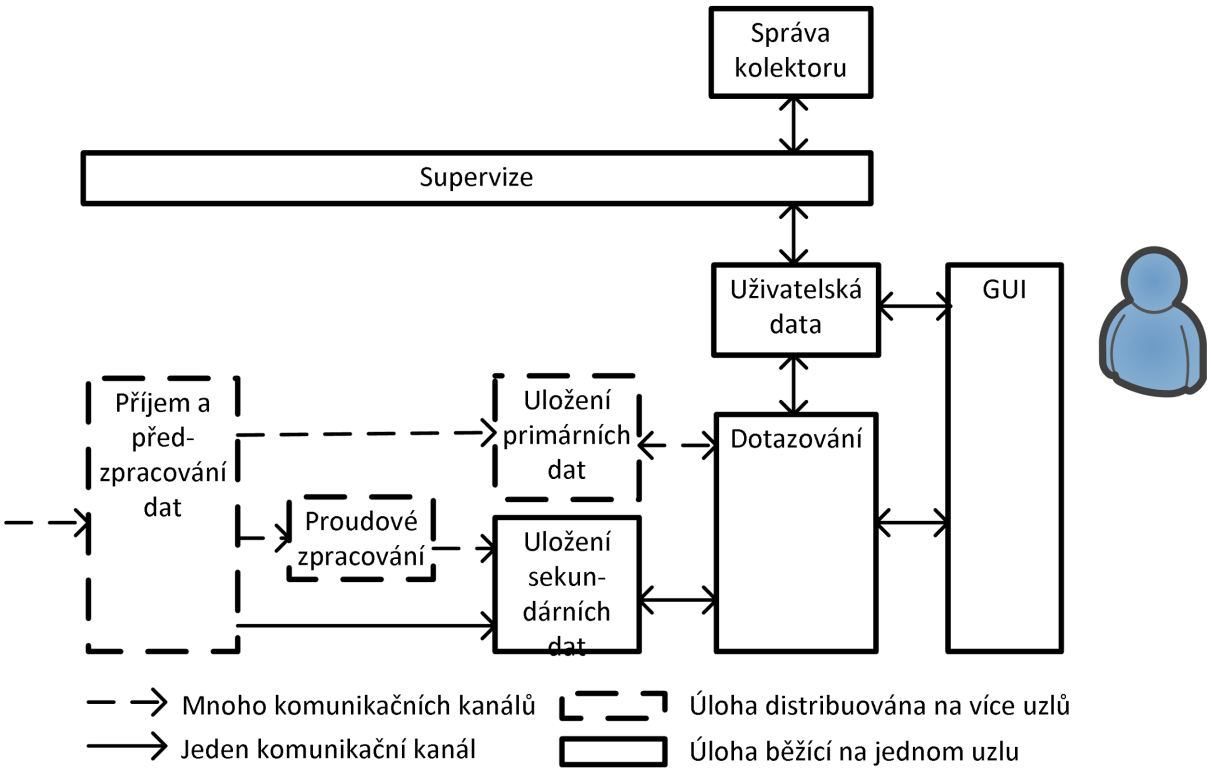

Figure 1. Architektura kolektoru.

Hlavní vstup dat do kolektoru tvoří záznamy o IP tocích (primární data), které jsou přijímány z jednotek až stovek zařízení schopných tato data exportovat. Jeden kolektor může být sdílen více organizacemi, kdy každá organizace vidí pouze data, která ji logicky přísluší. Při příjmu dat je zpracován samotný protokol pro přenos dat o IP tocích a následně jsou data předzpracována. Předzpracování dat je odpovědné především za obohacení data a za asociaci dat ke zdroji, organizacím a profilům, a případně k výpočtu některých hodnot časových řad. Primární obohacená data jsou uložena do distribuovaného trvalého úložiště. Duplikát primárních dat je poslán k proudovému zpracování a data vypočtená během předzpracování

jsou uložena do úložiště sekundárních dat. Proudové zpracování má za úkol distribuovaně vypočítat pohledy na data a dále provést behaviorální analýzu dat. Výsledky proudového zpracování jsou uloženy do úložiště sekundárních dat. Nad úložištěm primárních dat i nad úložištěm sekundárních dat je vykonáváno dotazování. Dotazovací vrstva odděluje primární a sekundární úložiště od GUI. Za účelem oddělení dat jednotlivých organizací je vedena uživatelská databáze. Veškeré procesy v úložišti jsou monitorovány skrze supervizi, která sděluje provozovateli kolektoru aktuální stav kolektoru pro dané organizace a dále poskytuje data a konfigurační rozhraní pro správu celého kolektoru.

2.1 Příjem a předzpracování dat

Na vstupu je potřeba zpracovat různé protokoly, ve kterých jsou záznamy o tocích přenášeny. Jedná se o protokoly rodiny NetFlow a IPFIX. Tyto protokoly budou přenášeny v otevřené či šifrované podobě (TLS) přes TCP a v otevřené podobě přes UDP. Kolektor naslouchá jednotlivým protokolům na předem definovaných portech. Zpracování vstupu bude probíhat distribuovaně na několika uzlech vyhrazených pro tuto činnost. Každý zdroj zasílá data o IP tocích právě na jeden vstupní uzel. K distribuci zátěže mezi uzly dochází tak, že zdroje jsou rozděleny do disjunktních podmnožin a každý vstupní uzel bude přijímat a zpracovávat data právě z jedné podmnožiny. Nový zdroj dat se připojí na nejméně vytížený uzel nebo je mu alokován nový uzel. Záznamy o tocích mohou být rozšířeny o:

- Data o přiřazení k organizacím a profilům. U záznamu bude doplněna informace, kterým organizacím přísluší a v rámci organizací, kterým profilům odpovídá.
- Data o lokalitě IP adres. Za účelem obohacení dat bude využita databáze o geolokaci IP adresy a čísla AS.
- Data o identitě uživatelů. Za účelem obohacení dat o identitu uživatele jsou získávána z DHCP a podobných mechanismů v prostředí IPv6. Protokolem pro získávání těchto dat je syslog, který bude do kolektoru proudit skrze vedlejší vstupní rozhraní.

Hlavními výstupy příjmu a předzpracovaní dat jsou výstupní rozhraní obsahující záznamy o tocích. Organizace IANA definuje množinu ustálených položek záznamu . Tato množina položek není konečná a úložiště i proudové zpracování musí počítat s rozšířením množiny položek. Tato data budou ukládána do úložiště a přenášena do proudového zpracování. Dalším výstupním rozhraním je rozhraní obsahující sekundární data (statistiky o časových řadách). Výstupní rozhraní do úložiště dat bude přímo zapisovat data do distribuovaného úložiště primárních dat. K parametrizaci zápisu bude využívat data asociovaná k záznamům během jejich příjmu, normalizace a obohacení. Konkrétní typ rozhraní a formát dat bude reflektovat nativní vstupní rozhraní úložiště (např. zápis do souboru ve formátu JSON). Výstupní rozhraní do proudového zpracování bude přímo přenášet data do distribuovaného proudového zpracování dat. Konkrétní typ rozhraní a formát dat bude reflektovat nativní vstupní rozhraní úložiště. Sekundární data budou ukládána do úložiště sekundárních dat v interním formátu tohoto úložiště. Sekundární data tvoří hodnoty statistické ukazatele spočítané pro N sekundové in-

tervaly. Administrativní rozhraní bude sloužit k předávání některých parametrů, ovládání načtení konfigurace a reportování vytíženosti vstupní části či chybových stavů.

Příjem a zpracování dat bude implementováno s využitím software IPFIXcol. Tento software poskytuje základní funkcionalitu pro příjem a zpracování protokolů pro export záznamů o IP tocích a bude dále rozšiřován o funkcionalitu potřebnou pro splnění požadavků na SecurityCloud kolektor. Výsledný software pak bude tvořit jeden ucelený proces a tento proces poběží v 1 až N instancích v rámci SecurityCloud kolektoru na vstupních uzlech. IPFIXcol pracuje následovně. Na vstupu dochází ke zpracování transportního protokolu a protokolu pro export záznamů o tocích ve vstupních pluginech. Přijímané pakety se záznamy jsou převáděny do interní struktury odpovídající IPFIX paketu. Tento paket následně prochází mediačními pluginy. Mediační pluginy upravují či rozšiřují samotný paket nebo upravují či rozšiřují metadata asociovaná k paketu. Následně je paket zpracováván výstupními pluginy, které paket zapisují či přenáší do výstupních rozhraní ve formátu daného rozhraní. IPFIXcol software je nutné rozšířit o funkcionalitu popsanou v následujícím textu a schematicky zobrazenou na obrázku Obrázek 2.

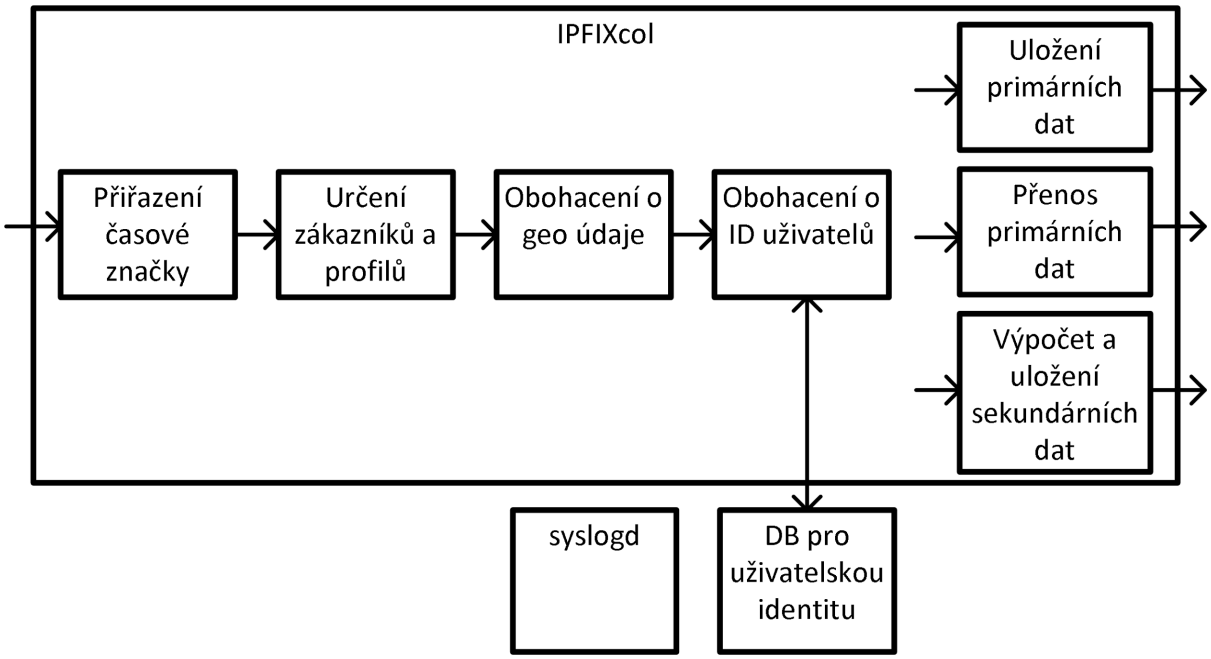

Figure 2. Schéma rozšíření IPFIXcol software.

Ke každému internímu IPFIX paketu budou asociovány informace, které dále rozšiřují informace o příchozím paketu a informace o záznamech o tocích. Informace vztahující se k celému záznamu budou rozšířeny o časovou značku příchodu paketu na kolektor ve vstupním modulu.

Informace vztahující se k jednotlivým záznamům budou rozšířeny o pole struktur, kde každá struktura se vztahuje právě k jednomu záznamu a bude obsahovat:

— odkaz na záznam v IPFIX paketu,
— identifikátor zdrojové AS,
— identifikátor cílového AS,

— identifikátor zdrojové země,
— identifikátor cílové země,
— zdrojová identita,
— cílová identita,
— odkaz na seznam strukturovaných informací k organizacím a profilům.

Strukturovaná informace k organizaci a jejím profilům se sestává z:

— identifikátor organizace (číslo),
— identifikátor pravidla zdroje (číslo),
— odkazu na zdrojovou identitu,
— odkazu na cílovou identitu,
— odkazu do seznamu profilů.

Tato struktura bude vyplňována v jednotlivých modulech. Na vstupu dojde k přiřazení časové značky kolektoru k příchozímu paketu. Následně další modul určí příslušnost záznamu k organizaci na základě zdroje dat, rozsahu IP adres záznamu, MPLS či VLAN tagů. Záznam může náležet více organizacím. Každá organizace si může definovat vlastní profily (filtrační podmínky), proto modul určí všechny profily, kterým záznam odpovídá. V následujícím modulu je záznam obohacen o geolokační údaje a následně o údaje vztahující se k identitě uživatelů. Takto obohacený záznam je zpracován několika výstupními moduly.

Výstupní moduly zapisují/přenáší záznamy. Tento zápis bere v úvahu metadata vztahující se k záznamu. Díky tomu je například záznam při ukládání do primárního úložiště replikován, pokud náleží více organizacím či profilům. Stejně tak při přenosu primárních dat k proudovém zpracování dochází pouze u záznamů určitého proffilu a jen pro vybrané organizace. Metadat rovněž využívá výstupní modul počítající sekundární data (statistiky za daný časový interval, např. počet toků, paketů, bytů per TCP, UDP, ICMP či další protokoly).

2.2 Uložení primárních dat

Vstupní rozhraní poskytuje zápisové rozhraní pro přímý zápis záznamů do úložiště bez dalších mezivrstev. Zápis záznamů skrze zápisové rozhraní je řešen již ve výstupním rozhraní příjmu a předzpracování dat, aby byla minimalizována režie spojená s kopírováním či přenosem dat. Zápisové rozhraní bude dovolovat zapsat odděleně data od různých organizací a oddělit data v rámci profilů.

Výstupní rozhraní poskytuje rozhraní pro tvorbu, spuštění, ukončení dotazovacích požadavků z dotazovací komponenty.

Administrativní rozhraní poskytuje funkce pro sledování uložených dat, sledování zaplnění diskového prostoru celkově i per organizaci, a dále funkce pro management dat jako jsou rotace či mazání dat, agregace a komprese dat.

Vzhledem k tomu, že platforma pro primární úložiště bude zvolena až na základě experimentů, je nutné uvažovat o ukládání dat jak přímo do souborového systému tak i do databáze. Z tohoto důvodu zde definujeme funkcionalitu na nejnižší úrovni, tedy na úrovni souborového systému, a pokud bude využita databáze, pak databáze musí zajistit stejnou funkcionalitu, nicméně implementace této funkcionality bude využívat prostředky samotné databáze. Data jsou při uložení distribuována mezi uzly ve svazku. Uložiště bude podporovat replikaci dat, koeficient replikace

lze nastavit genericky pro každý soubor. Přijatá data jsou do úložiště ukládána v logické adresářové struktuře, která začíná na úrovni organizace. Každý adresář dané organizace dále obsahuje složky, odpovídající jednotlivým profilům této organizace. V další struktuře jsou data rozdělena podle roku, měsíce a dne. V adresářích odpovídajících jednotlivým dnům jsou pak samotná data o tocích uložena podle nastavitelné granularity (např. po 5 minutách, po 1 hodině apod.). Pokud některý z profilů obsahuje další reálné podprofily, jsou data náležící těmto podprofilům uložena opět do adresáře tohoto podprofilu ve stejné struktuře (rok/měsíc/den...). Adresář tohoto podprofilu je pak uložen ve složce profilu nadřazeného.

Indexace bude realizována v rámci celých souborů. Indexování záznamů či hodnot v záznamech bude řešeno pomocí externích indexů. To znamená, že indexy budou ukládány do dedikovaných indexových souborů. K jednomu souboru se záznamy tak může vzniknout několik souborů s různými indexy. Takový přístup umožní případnou modifikaci způsobu indexování, přidání či odebrání indexů apod. Na žačátku souboru bude hlavička souboru udávající typ a verzi indexu. Název indexového souboru bude odpovídat názvu souboru se záznamy o tocích a určuje tak, ke kterému souboru se index vztahuje, přípona souboru bude značit, kterou položku soubor indexuje. V případě více indexů nad jednou položkou bude přípona obsahovat suffix s názvem indexu.

Příklad je uveden na návrhu indexace v rámci zdrojové a cílové ip adresy. Pro indexování IP adres budou použity Bloomovy filtry. Každý index je pak reprezentován vektorem binárních hodnot, jehož délka bude záviset na počtu unikátních adres dané sady dat (týkajících se dané organizace). Indexový soubor bude obsahovat hlavičku a následovat bude délka vektoru binárních hodnot (8 B) a samotný vektor binárních hodnot.

2.3 Uložení sekundárních dat

Vstupní rozhraní poskytuje rozhraní pro zápis sekundárních dat do permanentního úložiště. Zápis dat skrze rozhraní je řešen ve výstupním rozhraní příjmu a předzpracování dat (pro data z předzpracování) a ve výstupním rozhraní proudového zpracování (pro výstupy z proudového zpracování). Zápisové rozhraní musí umožnit oddělit data, vztahující se k různým organizacím.

Výstupní rozhraní poskytuje rozhraní pro tvorbu, spouštění, ukončení dotazovacích požadavků z dotazovací komponenty.

Administrativní rozhraní poskytuje funkce pro správu uložených dat, především managementu dat jako je například nastavení podmínek pro uchování pouze posledních N hodnot, mazání dat a další.

Uložení sekundárních dat bude realizováno prostřednictvím databázového systému, neboť tato část již není kritická z pohledu propustnosti. Sekundární data sestávají z hlášení o událostech a reportů, které jsou výstupem proudového zpracování a ze statistických ukazatelů z fáze předzpracování (pravidelným zápisem vznikají časové řady). Pro tato sekundární data bude v databázovém systému existovat způsob jednoznačného rozlišení dat jednotlivých organizací. Rozlišení dat bude probíhat na základě identifikátoru organizace, který je přiřazen výstupním datům již v části příjmu a předzpracování dat. Díky tomuto identifikátoru bude možné data dané organizace zapsat do oddělených prostorů nebo ponechat identi-

fikátor součástí dat a rozlišit tak mezi hlášeními dané organizace. Hlášení nebudou mít pevně definovanou strukturu, nicméně množina základních ukládaných údajů je následující:

— čas začátku události
— čas konce události
— adresa útočníka
— adresa cíle
— typ události
— počet toků/paketů/bytů odeslaných směrem od/k útočníkovi
— poznámka
— ...

Správa či modifikace sekundárních dat bude prováděna pravidelně skrze administrativní rozhraní za účelem udržení definovaného objemu dat. Dále bude existovat omezená možnost modifikace dat skrze dotazovací rozhraní v případě, že si uživatel přeje vyznačit důležitost události, její uchování či přiřadit poznámku k události.

2.4 Uložení uživatelských dat

Vstupní rozhraní poskytuje zápisové rozhraní pro zápis a modifikaci uživatelských dat do permanentního úložiště. Zápis dat je realizován přes DB rozhraní. Zápisové rozhraní musí zajistit přístup pouze k datům, která patří danému uživateli.

Vstupní rozhraní poskytuje zápisové rozhraní pro zápis a modifikaci uživatelských dat do permanentního úložiště. Zápis dat je realizován přes DB rozhraní. Zápisové rozhraní musí zajistit přístup pouze k datům, která patří danému uživateli. Čtecí rozhraní

Výstupní rozhraní poskytuje rozhraní pro tvorbu, spuštění, ukončení dotazovacích požadavků z dotazovací komponenty. Čtecí rozhraní musí zajistit přístup pouze k datům, která patří danému uživateli.

Uživatelská data v kolektoru popisují uživatele, jejich role, práva, nastavení. Uživatelská data nevznikají z primárních ani sekundárních dat. Databáze uživatelských dat bude obsahovat údaje pro:

— správu organizací, tj. definice pravidel, která přiradí zdroj dat dané organizaci, alokace výpočetních a diskových prostředků, nastavení obohacování dat o geolokační údaje a o identitu uživatelů,
— správu zákaznických profilů, tj. které profily se mají nad daty dané organizace uplatnit,
— správu uživatelů a jejich nastavení (např. nastavení reportů),
— správu skupin a jejich přístupů k datům a hlášením.

Obecně budou uživatelská data využívána pro třídění příchozích dat a pro podmínění přístupu k datům.

2.5 Dotazování

Uživatelské rozhraní řeší komunikaci z GUI komponentou kolektoru. Rozhraní poskytuje uživateli možnost zadávat prostřednictvím GUI dotazy jednak nad úložištěm primárních dat, jednak nad databází dat sekundárních. Smyslem tohoto rozhraní je odstínění konkrétní platformy, použité k uložení dat a poskytnutí jednotného způsobu zadávání dotazů. Dále musí toto rozhraní poskytovat možnost spouštění připravených dotazů, a to jak na popředí tak i na pozadí.

Dotazovací vrstva využívá všechny 3 úložiště v systému:

- úložiště primárních dat,
- úložiště sekundárních dat,
- úložiště uživatelských dat.

Přístup k datům je řešen nativním rozhraním daného úložiště.

Z pohledu návrhu bude dotazování rozděleno do tří částí - dotazování nad primárními daty, dotazování nad sekundárními daty a dotazování nad uživatelskými daty. Z pohledu uživatele se však bude jednat o jednotné rozhraní. Dotazovací vrstva určí, na které úložiště požadavek směrovat, na základě kontextu předaného z GUI komponenty. Dotazovací vrstva tedy obdrží požadavek na primární, sekundární či uživatelská data.

Před vykonáním uživatelských požadavků nad úložišti musí dojít k ověření přístupových práv daného uživatele. Za tímto účelem je proveden dotaz na úložiště uživatelských dat vedoucí k ověření, zda uživatel má oprávnění ke spuštění dotazu nad danými daty, a k získání identifikátoru dat pro daného uživatele. Získaný identifikátor identifikuje data daného uživatele v úložišti a slouží pro vytvoření samotného dotazu.

Dotazy na primární data budou z GUI předávána ve formátu shodném pro NfDump. V případě potřeby budou tyto dotazy přeloženy do nativního jazyku primárního úložiště. V případě, že není využíváno nativního rozhraní primárního úložiště (například z důvodu výkonnosti či chybějící podpory některých typů dotazů), pak je toto dotazování realizováno v rámci dotazovací komponenty. Dotaz se skládá z kombinace následujících operací:

- filtrace,
- agregace a
- seřazení.

Pořadí operací typicky odpovídá pozici v přechozím seznamu. Nejprve dochází k filtraci záznamů, následně jejich agregaci a seřazení. Speciálním typem agregace je spojení záznamů o tocích náležících jednomu obousměrnému spojení do tzv. biflow. Filtrovací podmínka na primární data může být zadána v GUI přímo uživatelem. Z tohoto důvodu je nutné nejprve dotaz rozebrat a ověřit jeho platnost z pohledu syntaxe dotazu. Výsledkem tohoto ověření bude chybové hlášení s určením syntaktické chyby, v případě selhání syntaktického rozboru. Validovaný dotaz bude dle potřeby přeložen pro nativní dotazovací jazyk daného úložiště anebo vykonán přímo za pomoci vlastní dotazovací knihovny. Data, nad kterými má být dotaz proveden, jsou určena na základě příslušnosti uživatele k organizaci a příslušným profilem, se kterým uživatel v GUI pracuje. Při přístupu k datům je nutné zachovat hierarchii profilů a dodržet postupné aplikování profilů na zdro-

jová data. V případě reálných profilů se jedná o přístup k datům v příslušné lokaci struktury daného uložiště (např. v příslušném adresáři). Pokud se jedná o profily virtuální (stínové), musí být pravidla tohoto profilu načtena a přidána k dotazu. Výsledný dotaz je proveden nad daty nadřazeného reálného profilu. Před provedením dotazu učiní dotazovací vrstva odhad doby vykonání dotazu na základě odhadu rozsahu objemu přistupovaných dat spolu s odhadem aktuální dotazovací výkonností (výkonnost se mění v závislosti na velikosti svazku a v závislosti na vytížení). Vykonávané dotazy mohou být uživatelem přesunuty na pozadí explicitně. Dlouho trvající dotazy jsou po uživatelem definovaného intervalu přesunuty na pozadí automaticky. Uživatel může běžící dotaz předčasně ukončit.

Dotazy směrované na sekundární úložiště budou v GUI zadávána prostřednictvím předdefinovaných možností. Z tohoto důvodu není nutné provádět kontrolu validity dotazu. Před spuštěním dotazu jsou ověřena přístupová práva a získán identifikátor k datům sekundárního úložiště. Vzhledem k tomu, že sekundární úložiště bude uchovávat o několik řádů méně dat než primární úložiště, budou dotazy vykonávány na popředí. Z pohledu Dotazovací vrstva rovněž bude zprostředkovávat přístup do uživatelských dat za účelem:

— zjištění objemu uložených dat a alokované kapacity,
— správy profilů (dle oprávnění),
— správy zdrojů (dle oprávnění),
— správy uživatelů (dle oprávnění).

3 Experimenty porovnávající výkonnost dotazování

Před samotným vývojem distribuovaného kolektoru plánujeme experimenty s rozšířenými platformami pro distribuované zpracování uložených dat, například Hadoop a jeho nádstavby, ElasticSearch [4]. Stejně tak budou provedeny experimenty z distribuovanými platformami pro proudové zpracování, například Kafka [5], Storm [6], StreamMine3G [7]. V této technické zprávě uvádíme prvotní experimenty s platformou Hadoop a porovnáváme výkonnost vůči tradičnímu, vysoce optimalizovanému kolektoru, NfDump [1].

Apache Hadoop je volně dostupný framework, který realizuje spolehlivé distribuované výpočty na velkých datech s využitím počítačového clusteru v paradigmatu MapReduce. Pro práci s tímto frameworkem byly vybrány 3 přístupy. První z nich je vlastní implementace dotazů, zvolených pro experimenty. Tato vlastní implementace je pak provedena nad daty ve formátu CSV a nad binárními daty. Dále byly použity dvě nástavby pro Hadoop, poskytující rozhraní pro dotazování nad uloženými daty. Jedná se o nástavby Hive a Pig. Měření doby provedení dotazů pomocí těchto přístupů bylo porovnáváno vůči výsledkům, získaným pomocí nástroje Nfdump, který byl spuštěn nad stejnými daty v komprimované či nekomprimované formě.

Pro účely experimentů s distribuovaným zpracováním dat byly navrženy 4 dotazy. Tyto dotazy byly následně spouštěny nad anonymizovanými daty z jednoho dne reálného provozu, která byla rozdělena na soubory s přírůstkem dat odpovídajícím jedné hodině provozu (tj. první soubor neobsahuje žádná data, druhý soubor obsahuje data z jedné hodiny, třetí soubor data ze dvou hodin atd.). Celkově tato

data obsahovala 1.1 mld. toků, 33.8 mld. paketů a 28.5 terabytů přenesených počítačovou sítí v tomto jednom dni. Pro každý blok dat byl každý ze 4 dotazů spuštěn třikrát. Výsledná doba trvání dotazu se pak spočítala jako průměr těchto hodnot. Dotazy určené pro Hadoop byly spuštěny v rámci clusteru o 6 uzlech, na kterých byl systém Hadoop provozován. Stejné dotazy byly provedeny pomocí NfDump na jednom serveru.

První z dotazů počítá ze všech záznamů o tocích sumu počtu paketů a sumu počtu bytů. Obrázek Obrázek 3 zobrazuje výsledky provedení tohoto dotazu. V grafu na tomto obrázku (a třech následujících) je na ose x velikost dat, vyjádřena v počtu hodin uložených dat. Na ose y je pak průměrná doba vykonání dotazu v sekundách

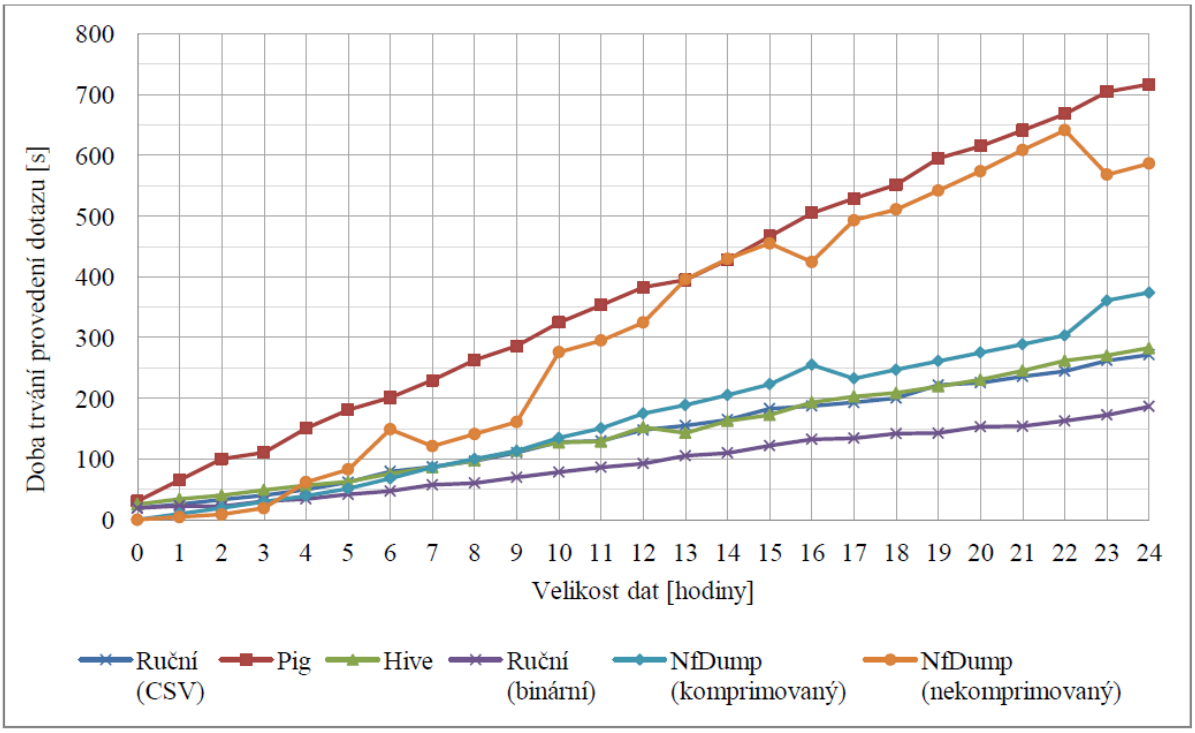

Figure 3. Dotaz na sumu počtu paketů a bytů.

Další dotaz získává celkový počet všech záznamů s cílovým portem 53, tj. zaměřuje se na dotazy protokolu DNS. Obrázek Obrázek 4 zobrazuje časy provedení dotazu pro jednotlivá měření.

Třetí dotaz vybírá jen zvolená pole (časová značka příchodu záznamu, protokol, zdrojová a cílová IP adresa, zdrojový a cílový port, počet paketů a počet bytů) pro záznamy o IP tocích přenášené spolehlivým protokolem TCP na portu 53. Obrázek Obrázek 5 ukazuje výsledky provedení tohoto dotazu.

Poslední ze sady dotazů pro každou zdrojovou adresu počítá sumu paketů, bytů a celkový počet záznamů přenesených z této adresy pomocí protokolu TCP. Obrázek Obrázek 6 zobrazuje doby odezvy dotazu pro jednotlivá data.

Z výše uvedených experimentů dosáhla nejlepších výsledků vlastní implementace dotazů, realizovaná nad binárními daty. Dobrých výsledků dosahovala tato vlastní implementace i nad daty ve formátu CSV (tj. textový soubor). Stejně tak

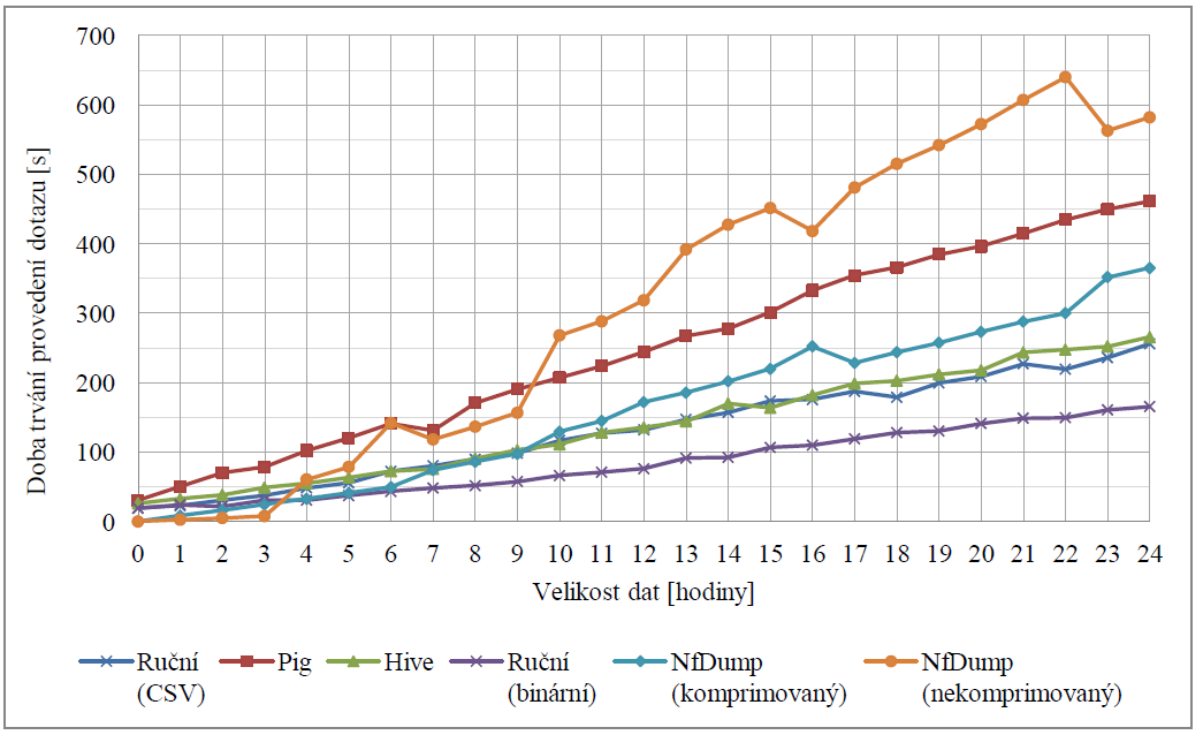

Figure 4. Dotaz na počet záznamů s portem 53.

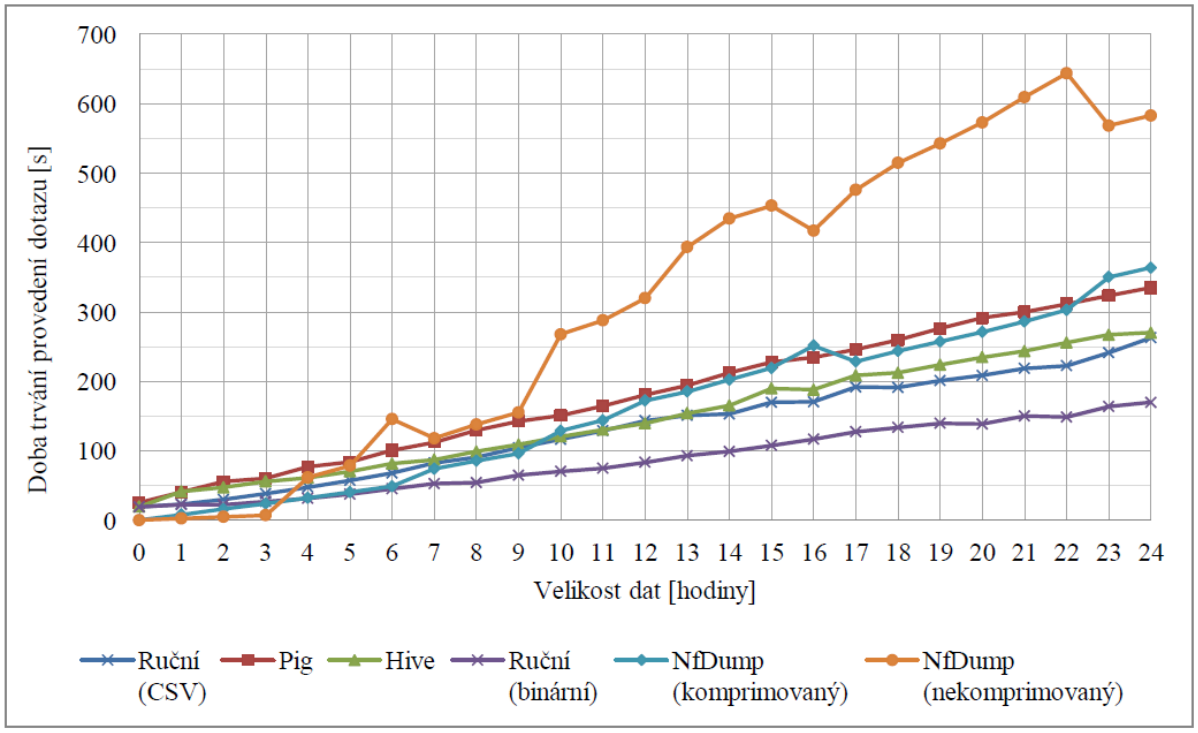

Figure 5. Dotaz vypisující vybrané položky záznamu s portem 53 a protokolem TCP.

za zmínku stojí i nástavba Hive, která oproti implementaci konkrétního dotazu poskytuje rozhraní pro zadávání obecně libovolných dotazů. Nejhůře ze všech textovaných možností dopadla nástavba Pig. Další informací, kterou je možné z

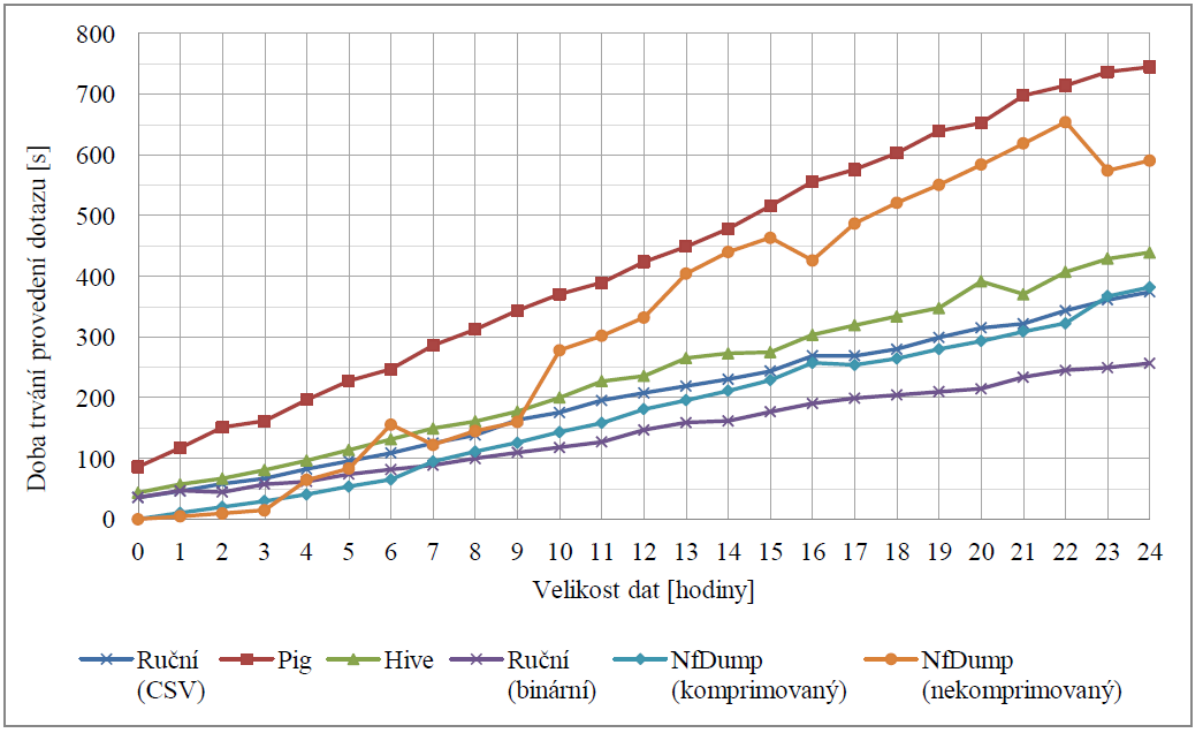

Figure 6. Dotaz na výpočet sumy paketů, bytů a počet záznamů pro každou zdrojovou IP adresu.

výsledků vyčíst je stálost doby trvání dotazů prováděných nástrojem Nfdump. Ta je způsobena tím, že Nfdump při vykonávání těchto dotazů musí vždy sekvenčně projít všechna data, což je z celého dotazu časově nejnáročnější část. Samotný výpočet pak již způsoboval odchylky pouze v rámci vteřin. Důležitým jevem, na který se bude nutné v následujícím výzkumu zaměřit, je počáteční doba latence dotazů spouštěných nad systémem Hadoop. Ta se i pro dotazy nad velmi malým množstvím dat pohybuje okolo 20 vteřin, což není pro navrhovaný kolektor vhodné.

4 Závěr

Koncept distribuovaného zpracování velkého množství dat na levných výpočetních uzlech se zdá být vhodný pro potřeby kolektoru síťových dat. V rámci této technické zprávy jsme identifikovali základní komponenty tohoto kolektoru a specifikovali jejich funkcionalitu. Za účelem řešení dvou nejpalčivějších problémů současných kolektorů jsme navrhli využití paradigmatu MapReduce a proudového zpracování dat.

První experimenty se známým frameworkem Apache Hadoop ukázaly, že je možné zvyšovat propustnost přístup k datům na permanentním uložišti. Nicméně distribuované zpracování ve frameworku Apache Hadoop má svá úskalí v podobě vysoké režie a latence. Na druhou stranu Hadoop obsahuje velké množství konfiguračních možností. V naší budoucí práci se proto budeme snažit nalézt vhodnou konfiguraci, která by snížila režii i latenci při zpracování dotazů.

5 Poděkování

Tento dokument vznikl v rámci projektu TA04010062 - Technologie pro zpracování a analýzu síťových dat velkého rozsahu (SecurityCloud) programu ALFA(4) TAČR.

References

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

[2] Yeonhee Lee, Youngseok Lee. Toward scalable internet traffic measurement and analysis with Hadoop, ACM SIGCOMM, 2013.

[3] Apache Hadoop. *http://hadoop.apache.org/*

[4] ElasticSearch. *http://www.elasticsearch.org/*

[5] Apache Kafka. *http://kafka.apache.org/*

[6] Apache Storm. *https://storm.apache.org/*

[7] StreamMine3G. *https://streammine3g.inf.tu-dresden.de/trac*